Robust Energy-Efficient Design for MISO Non-Orthogonal Multiple Access Systems

Faezeh Alavi, Kanapathippillai Cumanan, Milad Fozooni, Zhiguo Ding, Sangarapillai Lambotharan and Octavia A. Dobre

***Abstract*—Non-orthogonal multiple access (NOMA) has been envisioned as a promising multiple access technique for 5G and beyond wireless networks due to its significant enhancement of spectral efficiency. In this paper, we investigate a robust energy efficiency design for multi-user multiple-input single-output (MISO) NOMA systems where imperfect channel state information is available at the base station. A clustering algorithm is applied to group the users into different clusters, and then NOMA technique is employed to share the available resources fairly among the users in each cluster. To remove the interference between clusters, two different types of zero-forcing (ZF) designs, namely, hybrid-ZF and full-ZF are employed at the BS. The full-ZF scheme completely removes the interference leakage at the cost of more number of antennas and the hybrid-ZF scheme partially mitigates the interference leakage. To solve the problem, the Dinkelbach's algorithm is employed to convert the non-linear fractional programming problem into a simple subtractive form. Finally, simulation results reveal that hybrid-ZF outperforms the full-ZF scheme with a few clusters, while full-ZF shows a better performance with the higher number of clusters. The numerical results confirm that our proposed robust scheme outperforms the non-robust scheme in terms of the rate-satisfaction ratio at each user.**

***Index Terms*—Convex optimization, Multiple-input single-output (MISO), Non-orthogonal multiple access (NOMA), Robust energy efficiency (EE), Worst-case performance optimization, Zero-forcing (ZF).**

I. Introduction

In recent years, mobile communication technologies have been facing various key challenges, such as increasing demand for high data rate services, massive connectivity requirements and scarcity of radio resources, which need to be addressed in the next generation of wireless networks [1]–[6]. On the other hand, this explosive growth of data traffic has triggered a rapid increase in energy consumption. The statistics show that the information and communication technology infrastructures consume more than 3% of the world-wide energy consumption [7]. Hence, an appropriate performance metric is required to strike a good balance between the achievable data rate and power consumption. To this end, energy efficiency (EE), defined as the number of bits that can be reliably transmitted per Joule of energy consumption, has been recently considered as one of the key performance metrics to evaluate the performance of communication networks [8], [9].

To accommodate a large number of connected devices with higher data rates, non-orthogonal multiple access (NOMA) has been recently advocated as a prospective candidate for multiple access technique in the fifth generation (5G) and beyond wireless networks [10]–[16]. In NOMA, multiple users can share the same wireless resources, i.e., time, frequency and code domains by applying superposition coding (SC) and power domain multiplexing at the transmitter. More specifically, NOMA allocates higher transmit power to the users with poor channel conditions, while the users with better channel conditions are served with less transmit power. Then, successive interference cancellation (SIC) technique is employed at the receiver for multi-user detection. In other words, NOMA mitigates the interference through a non-orthogonal approach to significantly increase the system throughput while introducing an affordable additional complexity at the receiver [12]. As a result, more mobile terminals can be served simultaneously with higher spectral efficiency (SE). Hence, NOMA has recently attracted a considerable amount of research interests from both industry and academia, thanks to its great potential capabilities in future wireless networks.

A. Literature

Most of the existing works on NOMA in the literature mainly focuses on improving the overall SE of communication systems [17]–[22]. However, there is a dearth of literature considering the EE which has been identified as one of the key performance metrics in future wireless networks. The EE of NOMA systems was investigated in [23] for a given statistical channel state information (CSI) at the transmitter. A crucial step forward was followed in [24] to maximize the EE of downlink NOMA systems by recalling a non-linear fractional programming method. In addition, the authors in [25] proposed a power allocation and subchannel assignment to maximize the EE in NOMA networks by assigning only two users per subchannel. The joint user scheduling and power allocation in this context was further explored in [26], [27] under the assumption of imperfect CSI. In [26], it was assumed that only two users can be multiplexed on each subchannel whereas a general case with more number of users on same

The work of K. Cumanan and Z. Ding and was supported by the UK EPSRC under grant number EP/P009719/2 and by H2020-MSCA-RISE-2015 under grant number 690750.

F. Alavi and K. Cumanan are with the Department of Electronic Engineering, University of York, York, YO10 5DD, U.K. (e-mail: {sa1280, kanapathippillai.cumanan}@york.ac.uk). M. Fozooni was with the Institute of Electronics, Communications and Information Technology, Queen's University Belfast, Belfast BT3 9DT, U.K. at the time of this work. He is currently with Ericsson AB, 41756 Gothenburg, Sweden (e-mail: milad.fozooni@ericsson.com). Z. Ding is with the School of Electrical and Electronic Engineering, The University of Manchester, Manchester, M13 9PL, U.K. (e-mail: zhiguo.ding@manchester.ac.uk). S. Lambotharan is with the School of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough, U.K. (e-mail: S.Lambotharan@lboro.ac.uk). O. A. Dobre is with the Faculty of Engineering and Applied Science, Memorial University, Canada (e-mail: odobre@mun.ca).

subchannel was developed in [27]. These results confirmed that the NOMA system can achieve a better performance in terms of sum rate and EE compared to the conventional orthogonal multiple access (OMA) systems, for example orthogonal frequency-division multiple access (OFDMA). An energy-efficient power and bandwidth allocations were derived in [28] for a NOMA system which has multiple subchannels with unequal bandwidth. Some other related works can be found in [29], [30]. In [29], the authors proposed two user scheduling schemes combined with a power allocation scheme to enhance the EE in the multiple-input multiple-output (MIMO) NOMA system. As such, another optimal power allocation strategy has been proposed in [30] to solve the EE maximization problem for a multi-cluster multi-user MIMO-NOMA system.

Most existing research works on NOMA scheme have assumed that perfect CSI is available at the base stations (BSs) [31]–[34] which is not a realistic assumption in practice due to estimation and quantization errors, or inevitable delays in feedback links. Furthermore, the channel uncertainties can deteriorate the performance of SIC-based receivers where the users are sorted with respect to their channel gains [35]. Hence, it is of paramount importance to incorporate CSI uncertainties into problem formulations of NOMA-based networks to guarantee the required quality-of-service (QoS) at different users. To this end, robust design is a standard approach to tackle the channel uncertainties [36]–[38] and it can be categorized primarily into two groups: I) worst-case design with norm-bounded channel uncertainties, where CSI errors are bounded within a known region [39], [40]; II) outage probability-based design by assuming that the channel errors are random variables with a known probability density function which is available at the transmitter [41], [42]. In [43], [44], robust designs for the multiple-input single-output (MISO) NOMA systems have been developed to maximize the sum rate and minimize the total transmit power, under the assumption of bounded channel uncertainties. An outage probability-based design has been proposed in [45] to minimize the total required transmit power in MISO NOMA systems. Motivated by the above discussion, we focus on robust resource allocation schemes to appropriately address the impact of channel uncertainties on EE of a MISO NOMA system. In [46], a worst-case rate maximization problem is investigated in downlink MIMO NOMA networks which is solved by using cutting-set method with alternating optimization and pessimization steps.

B. Contributions

In this paper, we consider a downlink transmission of NOMA wireless network where a BS equipped with multiple antennas serves a set of single-antenna users that are uniformly distributed within a cell. By employing a clustering algorithm, the users are grouped into several clusters with two users per cluster. We consider a bounded channel uncertainty model to define the CSI errors, and design the beamfomers to optimize the worst-case EE problem. To the best of the authors' knowledge, the resource allocation problem that maximizes the robust EE has not been studied in the literature for MISO NOMA systems. The main contributions of this work are summarized as follows:

1) Having defined the system EE as the ratio between total sum rate and total power consumption, we focus on the robust EE maximization problem for a downlink MISO system, relying on NOMA principles in each cluster. The QoS requirement of each user is also included and guaranteed by an individual minimum data rate.
2) To incorporate practical scenarios, we assume that only the imperfect CSI is available at the BS and the channel uncertainties are bounded by predefined ellipsoids. Then, we consider the worst-case EE to ensure providing a required QoS at each user regardless of the channel uncertainties.
3) To effectively mitigate mutual interferences among different clusters, we present two different zero-forcing (ZF) schemes for the beamforming design, namely, I) hybrid-ZF and II) full-ZF. Although the full-ZF scheme can completely remove the interference between different clusters, it requires more number of transmit antennas at BS than that of the hybrid-ZF scheme to serve the same number of users. By increasing the number of clusters, the residual interference increases, and hence, the full-ZF approach can achieve a better performance in terms of EE as the residual interference can be completely cancelled.
4) To solve the power allocation problem, we cast the original problem in hand by considering the lower bound of SINR to present the constraints in a more tractable form. Then, an iterative algorithm is developed to transform the non-convex problem into sequential convex problems, which can be tackled by means of the standard power allocation techniques in each iteration. In particular, the Dinkelbach's algorithm is employed in each iteration to convert the non-linear fractional programming problem into a simple subtractive form.

C. Paper Organization

The rest of the paper is structured as follows. In Section II, we describe the system model and the hybrid-ZF scheme for beamforming design, while the robust EE design under the channel uncertainties is delineated in Section III. The full-ZF scheme is motivated and developed in Section IV. Finally, numerical results to validate the effectiveness of the proposed schemes are provided in Section V, before concluding the paper in Section VI.

D. Notation

Throughout this paper, we use lowercase boldface letters for vectors and uppercase boldface letters for matrices. The conjugate transpose and inverse of a matrix are denoted by $(\cdot)^H$ and $(\cdot)^{-1}$, respectively. The symbol $\mathbb{C}^n$ shows the n-dimensional complex space, and $\mathbb{R}_+$ represents the non-negative real numbers. The Euclidean norm of a vector is denoted by $\|\cdot\|$, and $|\cdot|$ represents the absolute value of a complex number. The notation $(x)^+$ stands for $\max(0, x)$, while $\mathcal{N}$ and $\mathcal{CN}$ denote a real and complex Gaussian random variable, respectively.

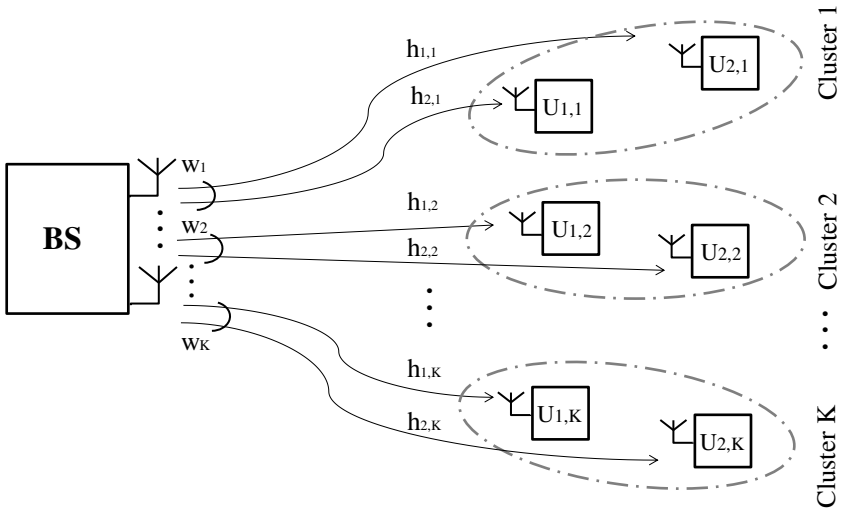


Fig. 1. A MISO NOMA system with K clusters and two users per cluster.

II. System Model

We consider a MISO NOMA downlink transmission where a BS equipped with N antennas intends to communicate with $2K$ single antenna users. All users are grouped into K clusters ($K \leq N$) with two users per cluster by employing the clustering algorithm [47], [48]. Note that the number of users in a cluster can be more than two; however, we assume only two users in each cluster for the sake of brevity. The l^{th} user in the k^{th} cluster is denoted by $U_{l,k}$, for all $k \in \{1, \ldots, K\}$ and $l = 1, 2$. Let $\mathbf{h}_{l,k} \in \mathbb{C}^{N\times 1}$ represent the channel vector from the BS to $U_{l,k}$, which can be modeled as $\chi\sqrt{d_{l,k}^{-\alpha}}$ [49], where χ denotes the Rayleigh fading channel gain, $d_{l,k}$ is the distance between the BS and $U_{l,k}$, and α represents the path loss exponent. For user pairing, we apply the clustering algorithm in [47] which is based on the channel correlation, $\frac{|\mathbf{h}_i^T \mathbf{h}_j|}{\|\mathbf{h}_i\| \, \|\mathbf{h}_j\|}$, and gain difference, $\big|\, \|\mathbf{h}_i\| - \|\mathbf{h}_j\| \,\big|$, between two users i and j. This algorithm selects two users that have a high correlation and a large channel gain difference in each cluster.

Among two users in a cluster, we consider $U_{2,k}$ has a higher channel gain than $U_{1,k}$, so that $\|\mathbf{h}_{1,k}\| \leq \|\mathbf{h}_{2,k}\|, \ \forall k$. The users in each cluster are supported by a NOMA beamforming vector to share the same time-frequency block but with different power levels through power domain multiplexing. Motivated by realistic scenarios in practice, we assume that the perfect CSI is not available at the transmitter due to quantization, channel estimation errors and feedback delays. Hence, we model the actual channel by the worst-case model [44], [50], [51], and incorporate the norm-bounded channel uncertainties in our analysis such that

$$\mathbf{h}_{l,k} = \hat{\mathbf{h}}_{l,k} + \Delta\hat{\mathbf{h}}_{l,k}, \tag{1}$$

where $\hat{\mathbf{h}}_{l,k}$ is the estimated channel, and $\Delta\hat{\mathbf{h}}_{l,k}$ is the corresponding channel uncertainty. In this model, it is assumed that $\Delta\hat{\mathbf{h}}_{l,k}$ is confined in a certain region, i.e., $\|\Delta\hat{\mathbf{h}}_{l,k}\| \leq \varepsilon$.

Let $\mathbf{w}_k$ and $p_{l,k}$ denote the beamforming vector steering towards the k^{th} cluster and the transmit power allocated to user $U_{l,k}$, respectively. From the NOMA protocol, the BS broadcasts the superposition coded users' signals as

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{w}_k(\sqrt{p_{1,k}}\ s_{1,k} + \sqrt{p_{2,k}}\ s_{2,k}), \tag{2}$$

where $s_{1,k}$ and $s_{2,k}$ are the unit power information symbols for the weak and strong users, respectively. Thus, the received signals at the weak user $U_{1,k}$ and the strong user $U_{2,k}$ are given by

$$y_{1,k} = \mathbf{h}_{1,k}^H \mathbf{x} + n_{1,k}, \tag{3}$$

$$y_{2,k} = \mathbf{h}_{2,k}^H \mathbf{x} + n_{2,k}, \tag{4}$$

where $n_{l,k} \sim \mathcal{CN}(0, \sigma^2)$ for $l = 1, 2$ is zero-mean additive white Gaussian noise with variance σ^2. By utilizing the SIC at the receivers, $U_{2,k}$ decodes and removes the data of $U_{1,k}$ from the aggregated received signal $y_{2,k}$, and then, decodes its own data.

Next, we utilize the ZF beamformer at the BS to eliminate the interference between clusters by deploying $N \geq K$ antennas at the BS. To this end, the beamforming vector is designed based on the user's channel, $\hat{\mathbf{h}}_{l,m}$, and fulfills the following conditions:

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

Note that when there are $K \leq N < 2K - 1$ antennas at the BS, it is not possible to simultaneously satisfy (5) for both channel vectors $\hat{\mathbf{h}}_{i,m}$ and $\hat{\mathbf{h}}_{2,m}$. Therefore, if it is assumed that the channel $\hat{\mathbf{h}}_{l,m}$ is aligned with one of these users' channels, while the other user will suffer from the interference caused by transmission of signals to other clusters. Consequently, this residual interference can severely degrade the performance of SIC at the strong user to decode the weaker user's signal [47]. Therefore, to efficiently implement SIC, beamforming vectors

are generated based on the channels of the stronger users $\hat{\mathbf{h}}_{2,m}$, to satisfy the condition in (5) such that

$$\hat{\mathbf{h}}_{2,m}^H \mathbf{w}_k = 0, \quad \forall m \neq k. \tag{6}$$

However, note that $\hat{\mathbf{h}}_{1,m}^H \mathbf{w}_k \neq 0$, for any $m \neq k$, which is the source of residual interference. Since there is residual interference for the the weak user, we refer this scheme as a hybrid-ZF scheme. By defining $\mathbf{H} = [\hat{\mathbf{h}}_{2,1} \ \cdots \ \hat{\mathbf{h}}_{2,K}]$, the beamforming vector can be obtained as

$$\mathbf{W} = [\mathbf{w}_1 \ \cdots \ \mathbf{w}_K] = \mathbf{H}^\dagger = \mathbf{H}(\mathbf{H}^H\mathbf{H})^{-1}, \tag{7}$$

where $\mathbf{H}^\dagger$ denotes the pseudo-inverse of the matrix $\mathbf{H}$, and $\mathbf{w}_k$ is the beamforming vector for the k^{th} cluster. Therefore, the received signal at $U_{2,k}$ can be written as

$$\begin{aligned} y_{2,k} =& \mathbf{h}_{2,k}^H \mathbf{w}_k (\sqrt{p_{1,k}} \ s_{1,k} + \sqrt{p_{2,k}} \ s_{2,k}) \\ &+ \Delta\hat{\mathbf{h}}_{2,k}^H \sum_{j \neq k} \mathbf{w}_j (\sqrt{p_{1,j}} \ s_{1,j} + \sqrt{p_{2,j}} \ s_{2,j}) + n_{2,k}, \end{aligned} \tag{8}$$

where the second term in (8) refers to the residual interference which cannot be completely removed during the ZF process due to imperfect CSI [43]. Overall, the signal-to-interference-and-noise ratio (SINR) at the strong user to decode the weak user's signal is given by

$$\text{SINR}_{2,k}^{(1)} = \frac{p_{1,k}|\mathbf{h}_{2,k}^H \mathbf{w}_k|^2}{\underbrace{p_{2,k}|\mathbf{h}_{2,k}^H \mathbf{w}_k|^2}_{\text{intra-cluster interference}} + \underbrace{\sum_{j \neq k} |\Delta\hat{\mathbf{h}}_{2,k}^H \mathbf{w}_j|^2 (p_{1,j} + p_{2,j})}_{\text{residual interference due to imperfect CSI}} + \sigma^2}, \tag{9}$$

and after removing the weak user's signal via SIC technique, the strong user achieves the SINR in (10). The first term of the denominator in (10) is considered due to the fact that the stronger user cannot completely remove the detected weaker user's signal during the SIC process. At the other end, the SINR of weak user to decode its own signal is given by

$$\text{SINR}_{1,k}^{(1)} = \frac{p_{1,k}|\mathbf{h}_{1,k}^H \mathbf{w}_k|^2}{\underbrace{p_{2,k}|\mathbf{h}_{1,k}^H \mathbf{w}_k|^2}_{\text{intra-cluster interference}} + \underbrace{\sum_{j \neq k} |\mathbf{h}_{1,k}^H \mathbf{w}_j|^2 (p_{1,j} + p_{2,j})}_{\text{residual interference}} + \sigma^2}. \tag{11}$$

Thus, the achievable rate at $U_{1,k}$ and $U_{2,k}$ can be respectively defined as follows [17]:

$$R_{1,k} = \log_2 \big(1 + \min\{ \inf_{\Delta\hat{\mathbf{h}}_{1,k}} \text{SINR}_{1,k}^{(1)}, \inf_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(1)} \}\big), \tag{12}$$

$$R_{2,k} = \log_2 (1 + \inf_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(2)}). \tag{13}$$

III. Robust Energy Efficiency Maximization

In this section, we develop a robust energy-efficient power allocation scheme for a MISO NOMA system by incorporating the inevitable channel uncertainties. First, we define the EE formulation and then use it to model the worst-case power optimization problem. After applying a set of appropriate lemmas to transform the non-convex problem into a convex one, we solve the obtained problem by employing the Dinkelbach's algorithm.

A. Problem Formulation

To design an energy-efficient system, we consider a global EE which is defined as the ratio of the achievable sum rate of the system (bits/s/Hz) and the total power consumption (Watt). The overall EE of the NOMA system with the worst-case performance design can be mathematically expressed in (14), where P_c is the power dissipated in circuit blocks. Accordingly, the optimization problem can be formulated to determine the transmit power allocation that maximizes the worst-case EE under limited power budget and the QoS constraint for each user as follows:

$$\max_{p_{1,k}, p_{2,k}} \quad EE, \tag{15a}$$

$$s.t. \quad \sum_{k=1}^{K} (p_{1,k} + p_{2,k}) \ \leq \ P^{\max}, \tag{15b}$$

$$R_{1,k} \ \geq \ R^{\min}, \quad R_{2,k} \ \geq \ R^{\min}, \quad \forall k, \tag{15c}$$

where $P^{\max}$ is the maximum transmit power available at the BS and $R^{\min}$ is the minimum required data rate for each user.

This optimization problem is a non-convex and non-linear fractional programming problem. To solve this EE maximization problem, we present an iterative approach, where the Dinkelbach's algorithm is employed to optimize an approximated convex problem.

B. Power Allocation Design

In this subsection, we propose a power allocation scheme that maximizes the robust EE through an iterative algorithm. First, we introduce variables $\{\gamma_{1,k}, \gamma_{2,k}\} \in \mathbb{R}_+$ to further simplify the optimization problem in (15) as follows:

$$\max_{\gamma_{1,k}, \gamma_{2,k}, p_{1,k}, p_{2,k}} \quad \frac{\sum_{k=1}^{K} \big(\log_2(1 + \gamma_{1,k}) + \log_2(1 + \gamma_{2,k})\big)}{\sum_{k=1}^{K} (p_{1,k} + p_{2,k}) + P_c}, \tag{16a}$$

$$s.t. \ \sum_{k=1}^{K} (p_{1,k} + p_{2,k}) \leq P^{\max}, \tag{16b}$$

$$\gamma^{\min} \leq \gamma_{1,k} \leq \min\{ \inf_{\Delta\hat{\mathbf{h}}_{1,k}} \text{SINR}_{1,k}^{(1)}, \inf_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(1)} \}, \ \forall k, \tag{16c}$$

$$\gamma^{\min} \leq \gamma_{2,k} \leq \inf_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(2)}, \ \forall k, \tag{16d}$$

where $\gamma^{\min} = 2^{R^{\min}} - 1$ is the minimum required SINR at each user. The equivalent problem in (16) is still non-convex

$$\text{SINR}_{2,k}^{(2)} = \frac{p_{2,k}|\mathbf{h}_{2,k}^H \mathbf{w}_k|^2}{\underbrace{p_{1,k}|\Delta\hat{\mathbf{h}}_{2,k}^H \mathbf{w}_k|^2}_{\text{intra-cluster interference due to imperfect CSI}} + \underbrace{\sum_{j\neq k} |\Delta\hat{\mathbf{h}}_{2,k}^H \mathbf{w}_j|^2 (p_{1,j}+p_{2,j})}_{\text{residual interference due to imperfect CSI}} + \sigma^2}. \tag{10}$$

$$EE = \frac{\sum_{k=1}^K (R_{1,k}+R_{2,k})}{\sum_{k=1}^K (p_{1,k}+p_{2,k})+P_c} = \frac{\sum_{k=1}^K \Big(\log_2\big(1+\min\{\inf_{\Delta\hat{\mathbf{h}}_{1,k}} \text{SINR}_{1,k}^{(1)}, \inf_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(1)}\}\big) + \log_2(1+\inf_{\Delta\hat{\mathbf{h}}_{2,k}} \text{SINR}_{2,k}^{(2)})\Big)}{\sum_{k=1}^K (p_{1,k}+p_{2,k})+P_c}. \tag{14}$$

and NP-hard. As there is a common parameter $\Delta\hat{\mathbf{h}}_{l,k}$ in both numerator and denominator of the SINR expression, the constraints in (16c) and (16d) are intractable. To circumvent this issue, we consider their lower bounds through the following lemma:

Lemma 1: Consider

$$\text{SINR}_{i,k}^{(j)} = \frac{p_j|(\hat{\mathbf{h}}_{i,k}+\Delta\hat{\mathbf{h}}_{i,k})^H \mathbf{w}_k|^2}{\sum_n p_n|(\hat{\mathbf{h}}_{i,k}+\Delta\hat{\mathbf{h}}_{i,k})^H \mathbf{w}_n|^2 + \sum_m p_m|\Delta\hat{\mathbf{h}}_{i,k}^H \mathbf{w}_m|^2 + \sigma^2}$$

which represents the SINR at the i^{th} user in the k^{th} cluster to decode the j^{th} user's signal. A lower bound of $\inf_{\Delta\hat{\mathbf{h}}_{i,k}}(\text{SINR}_{i,k}^{(j)})$ can be expressed as

$$\varphi_{i,k} = \frac{p_j \ f_{i,k}^k}{\sum_n p_n \ g_{i,k}^n + \sum_m p_m \ \overline{g}_{i,k}^m + \sigma^2}, \tag{17}$$

where

$$f_{i,k}^k = \left| \left(|\hat{\mathbf{h}}_{i,k}^H \mathbf{w}_k| - \varepsilon\|\mathbf{w}_k\| \right)^+ \right|^2, \tag{18}$$

$$g_{i,k}^n = \left| |\hat{\mathbf{h}}_{i,k}^H \mathbf{w}_n| + \varepsilon\|\mathbf{w}_n\| \right|^2, \tag{19}$$

$$\overline{g}_{i,k}^m = \big(\varepsilon\|\mathbf{w}_m\|\big)^2. \tag{20}$$

Proof: Please refer to Appendix A.

By applying the lower bound function $\varphi_{i,k}$ in (17), to the main problem (16) the following optimization problem can be formulated:

$$\max_{\gamma_{1,k},\gamma_{2,k},p_{1,k},p_{2,k}} \ \frac{\sum_{k=1}^K \big(\log_2(1+\gamma_{1,k})+\log_2(1+\gamma_{2,k})\big)}{\sum_{k=1}^K (p_{1,k}+p_{2,k})+P_c}, \tag{21a}$$

$$s.t. \ \sum_{k=1}^K (p_{1,k}+p_{2,k}) \le P^{\max}, \tag{21b}$$

$$\gamma_{1,k} \le \frac{p_{1,k} \ f_{1,k}^k}{p_{2,k} g_{1,k}^k + \sum_{m\neq k}(p_{1,m}+p_{2,m}) g_{1,k}^m + \sigma^2}, \forall k, \tag{21c}$$

$$\gamma_{1,k} \le \frac{p_{1,k} \ f_{2,k}^k}{p_{2,k} g_{2,k}^k + \sum_{m\neq k}(p_{1,m}+p_{2,m}) \overline{g}_{2,k}^m + \sigma^2}, \forall k, \tag{21d}$$

$$\gamma_{2,k} \le \frac{p_{2,k} \ f_{2,k}^k}{p_{1,k} \overline{g}_{2,k}^k + \sum_{m\neq k}(p_{1,m}+p_{2,m}) \overline{g}_{2,k}^m + \sigma^2}, \forall k, \tag{21e}$$

$$\gamma^{\min} \le \gamma_{1,k}, \quad \gamma^{\min} \le \gamma_{2,k}, \forall k. \tag{21f}$$

Although all the constraints in (21) can be rearranged as standard posynomials, this problem cannot be formulated as a geometric program (GP) as the objective function cannot be written as a posynomial function. To solve this fractional programming problem, we employ the Dinkelbach's algorithm which converts a non-linear fractional optimization problem into an equivalent and a tractable problem. For more details, please refer to Appendix B.

TABLE I
DINKELBACH'S ALGORITHM

Algorithm 1 Dinkelbach's Algorithm

1. Initialization: Set $\epsilon > 0, n = 0, \lambda_n = 0$,
2. repeat
3. $\quad \mathbf{x}_n^* = \arg\max \{f(\mathbf{x}_n) - \lambda_n g(\mathbf{x}_n)\}$,
4. $\quad F(\lambda_n) = f(\mathbf{x}_n^*) - \lambda_n g(\mathbf{x}_n^*)$,
5. $\quad \lambda_{n+1} = \dfrac{f(\mathbf{x}_n^*)}{g(\mathbf{x}_n^*)}$,
6. $\quad n = n+1$,
7. until $F(\lambda_n) < \epsilon$.

According to the requirement of Dinkelbach's algorithm, we have to reformulate the problem in (21) in a concave-convex fractional problem (CCFP) form to apply this algorithm. To deal with the non-convex nature of constraints in (21c)-(21e),

we introduce new variables $\vartheta_{1,k}, \vartheta_{2,k}$ and ϑ_k and redefine the corresponding constraints in the following inequalities:

$$(21c)\Rightarrow\begin{cases}\gamma_{1,k}\vartheta_{1,k}\leq p_{1,k}f_{1,k}^k,\\ p_{2,k}g_{1,k}^k+\sum_{m\neq k}(p_{1,m}+p_{2,m})g_{1,k}^m+\sigma^2\leq\vartheta_{1,k},\end{cases}\forall k, \tag{22}$$

$$(21d)\Rightarrow\begin{cases}\gamma_{1,k}\vartheta_{2,k}\leq p_{1,k}f_{2,k}^k,\\ p_{2,k}g_{2,k}^k+\sum_{m\neq k}(p_{1,m}+p_{2,m})\overline{g}_{2,k}^m+\sigma^2\leq\vartheta_{2,k},\end{cases}\forall k, \tag{23}$$

$$(21e)\Rightarrow\begin{cases}\gamma_{2,k}\ \vartheta_k\leq p_{2,k}f_{2,k}^k,\\ p_{1,k}\overline{g}_{2,k}^k+\sum_{m\neq k}(p_{1,m}+p_{2,m})\overline{g}_{2,k}^m+\sigma^2\leq\vartheta_k,\end{cases}\forall k. \tag{24}$$

Next, to deal with the product of optimization variables in (22)-(24), we utilize the following expression:

$$\gamma_{i,k}\ \vartheta_{j,k}=\frac{1}{4}\big[(\gamma_{i,k}+\vartheta_{j,k})^2-(\gamma_{i,k}-\vartheta_{j,k})^2\big]. \tag{25}$$

Then, the second quadratic term can be approximated by the first order Taylor series around $\gamma_{i,k}^{(t)}$ and $\vartheta_{j,k}^{(t)}$. As such, the product of two variables can be transformed into a convex term as

$$\begin{aligned}\gamma_{i,k}\ \vartheta_{j,k}&\approx\frac{1}{4}(\gamma_{i,k}+\vartheta_{j,k})^2-\frac{1}{4}[(\gamma_{i,k}^{(t)}-\vartheta_{j,k}^{(t)})^2\\&\quad+2(\gamma_{i,k}^{(t)}-\vartheta_{j,k}^{(t)})(\gamma_{i,k}-\gamma_{i,k}^{(t)}-\vartheta_{j,k}+\vartheta_{j,k}^{(t)})]\\&\triangleq G(\gamma_{i,k}\,\vartheta_{j,k},\ \gamma_{i,k}^{(t)}\,\vartheta_{j,k}^{(t)}).\end{aligned} \tag{26}$$

By recalling the above approximation and applying the Dinkelbach's algorithm, we should treat the following optimization problem in the t^{th} iteration:

$$\max_{p_{1,k},p_{2,k},\mathbf{A}}\ \sum_{k=1}^{K}\big(\log_2(1+\gamma_{1,k})+\log_2(1+\gamma_{2,k})\big)-\lambda_n\left(\sum_{k=1}^{K}(p_{1,k}+p_{2,k})+P_c\right), \tag{27a}$$

$$s.t.\ \sum_{k=1}^{K}(p_{1,k}+p_{2,k})\leq P^{\max}, \tag{27b}$$

$$G(\gamma_{1,k}\,\vartheta_{1,k},\,\gamma_{1,k}^{(t)}\,\vartheta_{1,k}^{(t)})\leq p_{1,k}f_{1,k}^k,\qquad\forall k, \tag{27c}$$

$$p_{2,k}g_{1,k}^k+\sum_{m\neq k}(p_{1,m}+p_{2,m})g_{1,k}^m+\sigma^2\leq\vartheta_{1,k},\forall k, \tag{27d}$$

$$G(\gamma_{1,k}\,\vartheta_{2,k},\,\gamma_{1,k}^{(t)}\,\vartheta_{2,k}^{(t)})\leq p_{1,k}f_{2,k}^k,\qquad\forall k, \tag{27e}$$

$$p_{2,k}g_{2,k}^k+\sum_{m\neq k}(p_{1,m}+p_{2,m})\overline{g}_{2,k}^m+\sigma^2\leq\vartheta_{2,k},\forall k, \tag{27f}$$

$$G(\gamma_{2,k}\,\vartheta_k,\,\gamma_{2,k}^{(t)}\,\vartheta_k^{(t)})\leq p_{2,k}f_{2,k}^k,\qquad\forall k, \tag{27g}$$

$$p_{1,k}\overline{g}_{2,k}^k+\sum_{m\neq k}(p_{1,m}+p_{2,m})\overline{g}_{2,k}^m+\sigma^2\leq\vartheta_k,\forall k, \tag{27h}$$

$$\gamma^{\min}\leq\gamma_{1,k},\quad\gamma^{\min}\leq\gamma_{2,k},\qquad\forall k, \tag{27i}$$

TABLE II
ENERGY EFFICIENCY MAXIMIZATION ALGORITHM

Algorithm 2 Energy Efficiency Maximization Algorithm

1. Initialize $\mathbf{\Lambda^{(0)}}$ to a feasible value of (21), and set $t=0$,
2. repeat
   Solve (27a) by using Dinkelbach's algorithm,
   Set $\mathbf{\Lambda^{(t+1)}}=\mathbf{A}^*$,
   Update $t=t+1$,
3. until required accuracy or maximum number of iterations.

where $\mathbf{A}\triangleq\{\gamma_{1,k},\gamma_{2,k},\vartheta_{1,k},\vartheta_{2,k},\vartheta_k\}$. For notational simplicity, all the variables that are used in the approximations of the product of two variables in t^{th} iteration are defined as

$$\mathbf{\Lambda^{(t)}}\triangleq\{\gamma_{1,k}^{(t)},\ \gamma_{2,k}^{(t)},\ \vartheta_{1,k}^{(t)},\ \vartheta_{2,k}^{(t)},\ \vartheta_k^{(t)}\}. \tag{28}$$

Since the problem in (27a) approximates the problem in (21) around $\mathbf{\Lambda}^{(t)}$, we should iteratively solve the problem in (27a) for different values of $\mathbf{\Lambda}^{(t)}$ and update the approximations to obtain the best local solution. Towards this end, if the solution of problem (27a) in the t^{th} iteration is $\mathbf{A}^*\triangleq\{\gamma_{1,k}^*,\gamma_{2,k}^*,\vartheta_{1,k}^*,\vartheta_{2,k}^*,\vartheta_k^*\}$, it is considered as the initial point of the next iteration, i.e., $\mathbf{\Lambda}^{(t+1)}$, until the algorithm converges. The pseudo-code of the proposed iterative algorithm is summarized in Table II. Furthermore, the minimum threshold to terminate the algorithm is chosen as the difference between two successive values of achieved EE or the number of iterations is reached to a predefined maximum value.

C. Feasibility of Problem (15)

It is worth mentioning that before solving the problem in (15), it is important to check the feasibility of the problem. Note that the minimum data rate constraints in (15c) might be unattainable at all users if the available total power is not sufficient at the BS. Hence, there exists a minimum required transmit power $P^{\min}$ which satisfies minimum data rate requirement for each user and makes the problem in (15) feasible only under the condition $P^{\max}\geq P^{\min}$. Thus, it is important to determine a feasible range of $P^{\max}$ that should be able to provide the data rate requirements at each user. To obtain $P^{\min}$, we formulate an auxiliary optimization problem that determines the minimum required transmit power to satisfy the minimum data rate requirement for all users as

$$P^{\min}=\min_{p_{1,k},p_{2,k}}\ \sum_{k=1}^{K}(p_{1,k}+p_{2,k}), \tag{29a}$$

$$s.t.\ \ R_{1,k}\ \geq\ R^{\min},\ \ R_{2,k}\ \geq\ R^{\min},\ \forall k. \tag{29b}$$

This optimization problem can be converted into a linear programming problem by invoking the same technique discussed for solving the main problem in (15). By obtaining the $P^{\min}$ from the problem (29), the feasibility of problem in (15) can be determined. With $P^{\max}\geq P^{\min}$, the problem in (15) is feasible and the power allocation can be determined to maximize the EE of the system while satisfying all the constraints.

IV. Full-ZF beamforming scheme

In this section, we present the full-ZF beamforming scheme to completely mitigate the interference between clusters. In particular, it is assumed that the number of antennas employed at the BS is $N \geq 2K-1$, which provides sufficient degrees of freedom for the ZF beamformer to completely remove the residual interference [52]:

$$\hat{\mathbf{h}}_{l,j}^H \mathbf{w}_k = 0, \quad \forall j \neq k, \ l = 1, 2. \tag{30}$$

To design the beamforming vector by satisfying the conditions in (30), we define

$$\mathbf{H}_k = [\hat{\mathbf{H}}_1 \cdots \hat{\mathbf{H}}_{k-1} \ \hat{\mathbf{H}}_{k+1} \cdots \hat{\mathbf{H}}_K], \tag{31}$$

where $\hat{\mathbf{H}}_k = [\hat{\mathbf{h}}_{1,k} \ \hat{\mathbf{h}}_{2,k}]$. Then, the null space of the matrix $\mathbf{H}_k$ can be utilized for the beamforming vector $\mathbf{w}_k$ which results in $\mathbf{H}_k^H \mathbf{w}_k = \mathbf{0}$. By exploiting this condition, referred to as full-ZF beamformer, the aggregated received signal at $U_{l,k}$ is given by

$$\begin{aligned} y_{l,k} = \mathbf{h}_{l,k}^H \mathbf{w}_k (\sqrt{p_{1,k}} \ s_{1,k} + \sqrt{p_{2,k}} \ s_{2,k}) \\ + \Delta\hat{\mathbf{h}}_{l,k}^H \sum_{j \neq k} \mathbf{w}_j (\sqrt{p_{1,j}} s_{1,j} + \sqrt{p_{2,j}} s_{2,j}) + n_{l,k}, \ l = 1, 2, \end{aligned} \tag{32}$$

where the second term in (32) shows the impact of imperfect CSI on ZF design. Hence, the SINR at the weak user to decode its own signal can be defined as

$$\widetilde{\text{SINR}}_{1,k}^{(1)} = \frac{p_{1,k} |\mathbf{h}_{1,k}^H \mathbf{w}_k|^2}{\underbrace{p_{2,k} |\mathbf{h}_{1,k}^H \mathbf{w}_k|^2}_{\text{intra-cluster interference}} + \underbrace{\sum_{j \neq k} |\Delta\hat{\mathbf{h}}_{l,k}^H \mathbf{w}_j|^2 (p_{1,j} + p_{2,j})}_{\text{residual interference due to imperfect CSI}} + \sigma^2}. \tag{33}$$

Similarly, the SINR at the strong user to decode the weak user's signal is given by

$$\widetilde{\text{SINR}}_{2,k}^{(1)} = \frac{p_{1,k} |\mathbf{h}_{2,k}^H \mathbf{w}_k|^2}{\underbrace{p_{2,k} |\mathbf{h}_{2,k}^H \mathbf{w}_k|^2}_{\text{intra-cluster interference}} + \underbrace{\sum_{j \neq k} |\Delta\hat{\mathbf{h}}_{2,k}^H \mathbf{w}_j|^2 (p_{1,j} + p_{2,j})}_{\text{residual interference due to imperfect CSI}} + \sigma^2}, \tag{34}$$

and the strong user achieves the following SINR to decode its own message after performing SIC in (35). Based on these definitions of SINRs at both users, the worst-case EE of the full ZF scheme can be expressed in (36).

Accordingly, we solve the following optimization problem to determine the best power allocation that maximizes the worst-case EE:

$$\max_{p_{1,k}, p_{2,k}} \ EE^{\text{full-ZF}}, \tag{37a}$$

$$s.t. \ \sum_{k=1}^{K} (p_{1,k} + p_{2,k}) \ \leq \ P^{\max}, \tag{37b}$$

$$\begin{cases} \log_2 \left(1 + \min \left\{ \inf\limits_{\Delta\hat{\mathbf{h}}_{1,k}} \widetilde{\text{SINR}}_{1,k}^{(1)}, \inf\limits_{\Delta\hat{\mathbf{h}}_{2,k}} \widetilde{\text{SINR}}_{2,k}^{(1)} \right\} \right) \geq R^{\min}, \forall k, \\ \log_2 (1 + \inf\limits_{\Delta\hat{\mathbf{h}}_{2,k}} \widetilde{\text{SINR}}_{2,k}^{(2)}) \geq R^{\min}, \ \forall k. \end{cases} \tag{37c}$$

To solve the fractional programming problem in (37), we apply the same procedure as in Section III.B. Towards this end, we equivalently reformulate the problem in (37) by introducing variables $\widetilde{\gamma}_{1,k}$ and $\widetilde{\gamma}_{2,k}$ as follows:

$$\max_{\widetilde{\gamma}_{1,k}, \widetilde{\gamma}_{2,k}, p_{1,k}, p_{2,k}} \frac{\sum_{k=1}^{K} \left(\log_2(1 + \widetilde{\gamma}_{1,k}) + \log_2(1 + \widetilde{\gamma}_{2,k}) \right)}{\sum_{k=1}^{K} (p_{1,k} + p_{2,k}) + P_c}, \tag{38a}$$

$$s.t. \sum_{k=1}^{K} (p_{1,k} + p_{2,k}) \leq P^{\max}, \tag{38b}$$

$$\gamma^{\min} \leq \widetilde{\gamma}_{1,k} \leq \min\{ \inf_{\Delta\hat{\mathbf{h}}_{1,k}} \widetilde{\text{SINR}}_{1,k}^{(1)}, \inf_{\Delta\hat{\mathbf{h}}_{2,k}} \widetilde{\text{SINR}}_{2,k}^{(1)} \}, \ \forall k, \tag{38c}$$

$$\gamma^{\min} \leq \widetilde{\gamma}_{2,k} \leq \inf_{\Delta\hat{\mathbf{h}}_{2,k}} \widetilde{\text{SINR}}_{2,k}^{(2)}, \ \forall k. \tag{38d}$$

By invoking Lemma 1, we have

$$\max_{\widetilde{\gamma}_{1,k}, \widetilde{\gamma}_{2,k}, p_{1,k}, p_{2,k}} \frac{\sum_{k=1}^{K} \left(\log_2(1 + \widetilde{\gamma}_{1,k}) + \log_2(1 + \widetilde{\gamma}_{2,k}) \right)}{\sum_{k=1}^{K} (p_{1,k} + p_{2,k}) + P_c}, \tag{39a}$$

$$s.t. \sum_{k=1}^{K} (p_{1,k} + p_{2,k}) \leq P^{\max}, \tag{39b}$$

$$\widetilde{\gamma}_{1,k} \leq \frac{p_{1,k} \ \widetilde{f}_{i,k}^k}{p_{2,k} \widetilde{g}_{i,k}^k + \sum_{m \neq k} (p_{1,m} + p_{2,m}) \widetilde{\widetilde{g}}_{i,k}^m + \sigma^2}, \quad \forall k, \ i = 1, 2, \tag{39c}$$

$$\widetilde{\gamma}_{2,k} \leq \frac{p_{2,k} \ \widetilde{f}_{2,k}^k}{p_{1,k} \widetilde{\widetilde{g}}_{2,k}^k + \sum_{m \neq k} (p_{1,m} + p_{2,m}) \widetilde{\widetilde{g}}_{2,k}^m + \sigma^2}, \quad \forall k, \tag{39d}$$

$$\gamma^{\min} \leq \widetilde{\gamma}_{1,k}, \quad \gamma^{\min} \leq \widetilde{\gamma}_{2,k}, \qquad \forall k, \tag{39e}$$

where

$$\widetilde{f}_{i,k}^k = \left| \left(\left| \hat{\mathbf{h}}_{i,k}^H \mathbf{w}_k \right| - \varepsilon \|\mathbf{w}_k\| \right)^+ \right|^2, \tag{40}$$

$$\widetilde{g}_{i,k}^k = \left| \left| \hat{\mathbf{h}}_{i,k}^H \mathbf{w}_k \right| + \varepsilon \|\mathbf{w}_k\| \right|^2, \tag{41}$$

$$\widetilde{\widetilde{g}}_{i,k}^m = \left(\varepsilon \|\mathbf{w}_m\| \right)^2. \tag{42}$$

Finally, the fractional programming problem in (39) can be solved by leveraging Dinkelbach's algorithm which converts

$$\widetilde{\text{SINR}}_{2,k}^{(2)} = \frac{p_{2,k}|\mathbf{h}_{2,k}^H \mathbf{w}_k|^2}{\underbrace{p_{1,k}|\Delta\hat{\mathbf{h}}_{2,k}^H \mathbf{w}_k|^2}_{\text{intra-cluster interference due to imperfect CSI}} + \underbrace{\sum_{j\neq k}|\Delta\hat{\mathbf{h}}_{2,k}^H \mathbf{w}_j|^2(p_{1,j}+p_{2,j})}_{\text{residual interference due to imperfect CSI}} + \sigma^2}. \tag{35}$$

$$EE^{\text{full-ZF}} = \frac{\sum_{k=1}^{K}\Big(\log_2\left(1+\min\{\inf_{\Delta\hat{\mathbf{h}}_{1,k}} \widetilde{\text{SINR}}_{1,k}^{(1)}, \inf_{\Delta\hat{\mathbf{h}}_{2,k}} \widetilde{\text{SINR}}_{2,k}^{(1)}\}\right) + \log_2(1+\inf_{\Delta\hat{\mathbf{h}}_{2,k}} \widetilde{\text{SINR}}_{2,k}^{(2)})\Big)}{\sum_{k=1}^{K}(p_{1,k}+p_{2,k})+P_c}. \tag{36}$$

a non-linear fractional optimization problem to an equivalent but more tractable problem. For more details, please refer to Appendix B. According to the condition in Dinkelbach's algorithm, we have to reformulate the problem in a CCFP form to apply this algorithm. To deal with the non-convex nature of constraints in (39c) and (39d), we introduce new variables $\widetilde{\vartheta}_{1,k}, \widetilde{\vartheta}_{2,k}$ and $\widetilde{\vartheta}_k$ and redefine the corresponding constraints in the following inequalities:

$$(39c) \Rightarrow \begin{cases} \widetilde{\gamma}_{1,k}\ \widetilde{\vartheta}_{i,k} \leq p_{1,k}\ \widetilde{f}_{i,k}^{k}, \\ p_{2,k}\widetilde{g}_{i,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m})\widetilde{\overline{g}}_{i,k}^{m} + \sigma^2 \leq \widetilde{\vartheta}_{i,k}, \end{cases} \forall k,\ i=1,2, \tag{43}$$

and

$$(39d) \Rightarrow \begin{cases} \widetilde{\gamma}_{2,k}\ \widetilde{\vartheta}_k \leq p_{2,k}\ \widetilde{f}_{2,k}^{k}, \\ p_{1,k}\widetilde{\overline{g}}_{2,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m})\widetilde{\overline{g}}_{2,k}^{m} + \sigma^2 \leq \widetilde{\vartheta}_k, \end{cases} \forall k. \tag{44}$$

In order to deal with the product of optimization variables in (43) and (44), we utilize the expression in (25). Similar to the previous section, the quadratic term can be approximated by the first order Taylor series in (45) around $\widetilde{\gamma}_{i,k}^{(t)}$ and $\widetilde{\vartheta}_{j,k}^{(t)}$, to transform it into a convex term. As such, the product of two variables can be transformed into a convex term as

$$\begin{aligned}\widetilde{\gamma}_{i,k}\ \widetilde{\vartheta}_{j,k} &\approx \frac{1}{4}(\widetilde{\gamma}_{i,k}+\widetilde{\vartheta}_{j,k})^2 - \frac{1}{4}[(\widetilde{\gamma}_{i,k}^{(t)}-\widetilde{\vartheta}_{j,k}^{(t)})^2 \\ &\quad + 2(\widetilde{\gamma}_{i,k}^{(t)}-\widetilde{\vartheta}_{j,k}^{(t)})(\widetilde{\gamma}_{i,k}-\widetilde{\gamma}_{i,k}^{(t)}-\widetilde{\vartheta}_{j,k}+\widetilde{\vartheta}_{j,k}^{(t)})] \\ &\triangleq \widetilde{G}(\widetilde{\gamma}_{i,k}\,\widetilde{\vartheta}_{j,k},\ \widetilde{\gamma}_{i,k}^{(t)}\,\widetilde{\vartheta}_{j,k}^{(t)}).\end{aligned} \tag{45}$$

By recalling the above approximation and applying the Dinkelbach's algorithm, we should treat the following optimization problem in the t^{th} iteration

$$\max_{p_{1,k},p_{2,k},\widetilde{\mathbf{A}}} \quad \sum_{k=1}^{K}\left(\log_2(1+\widetilde{\gamma}_{1,k})+\log_2(1+\widetilde{\gamma}_{2,k})\right) - \lambda_n\left(\sum_{k=1}^{K}(p_{1,k}+p_{2,k})+P_c\right), \tag{46a}$$

$$s.t.\ \sum_{k=1}^{K}(p_{1,k}+p_{2,k}) \leq P^{\max}, \tag{46b}$$

$$\widetilde{G}(\widetilde{\gamma}_{1,k}\,\widetilde{\vartheta}_{i,k},\ \widetilde{\gamma}_{1,k}^{(t)}\,\widetilde{\vartheta}_{i,k}^{(t)}) \leq p_{1,k}\widetilde{f}_{i,k}^{k},\quad \forall k,\ i=1,2, \tag{46c}$$

$$p_{2,k}\widetilde{g}_{i,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m})\widetilde{\overline{g}}_{i,k}^{m} + \sigma^2 \leq \vartheta_{i,k}, \forall k,\ i=1,2, \tag{46d}$$

$$\widetilde{G}(\widetilde{\gamma}_{2,k}\,\widetilde{\vartheta}_k,\ \widetilde{\gamma}_{2,k}^{(t)}\,\widetilde{\vartheta}_k^{(t)}) \leq p_{2,k}\widetilde{f}_{2,k}^{k}, \qquad \forall k, \tag{46e}$$

$$p_{1,k}\widetilde{\overline{g}}_{2,k}^{k} + \sum_{m\neq k}(p_{1,m}+p_{2,m})\widetilde{\overline{g}}_{2,k}^{m} + \sigma^2 \leq \widetilde{\vartheta}_k,\ \forall k, \tag{46f}$$

$$\gamma^{\min} \leq \widetilde{\gamma}_{1,k},\quad \widetilde{\gamma}^{\min} \leq \widetilde{\gamma}_{2,k}, \qquad \forall k, \tag{46g}$$

where $\widetilde{\mathbf{A}} \triangleq \{\widetilde{\gamma}_{1,k}, \widetilde{\gamma}_{2,k}, \widetilde{\vartheta}_{1,k}, \widetilde{\vartheta}_{2,k}, \widetilde{\vartheta}_k\}$. For notational simplicity, all variables that are used in the approximations of the product of two variables in the t^{th} iteration are defined as

$$\widetilde{\mathbf{\Lambda}}^{(t)} \triangleq \{\widetilde{\gamma}_{1,k}^{(t)},\ \widetilde{\gamma}_{2,k}^{(t)},\ \widetilde{\vartheta}_{1,k}^{(t)},\ \widetilde{\vartheta}_{2,k}^{(t)},\ \widetilde{\vartheta}_k^{(t)}\}. \tag{47}$$

Finally, we iteratively solve the approximated problem in (46a) for different values of $\widetilde{\mathbf{\Lambda}}^{(t)}$ and update the approximations to obtain the best local solution similar to the proposed iterative algorithm in Table II.

V. Computational Complexity Analysis

We analyze the computational complexity of the proposed algorithm by quantifying the required number of arithmetic operations in the worst-case at each iteration, along with the required number of iterations to achieve the solutions with a certain accuracy [53], [54]. We define the computational complexity for the algorithm as presented in the floowing:

In each iteration of Algorithm 2, a fractional program defined in (27a) and (46a) is solved via the Dinkelbach's algorithm in Algorithm 1. In particular, the Dinkelbach's algorithm solves a fractional program by solving a series of auxiliary problems. Hence, the main contributions to the computational complexity of the proposed algorithm come from the complexities introduced by solving problems defined in (27a)

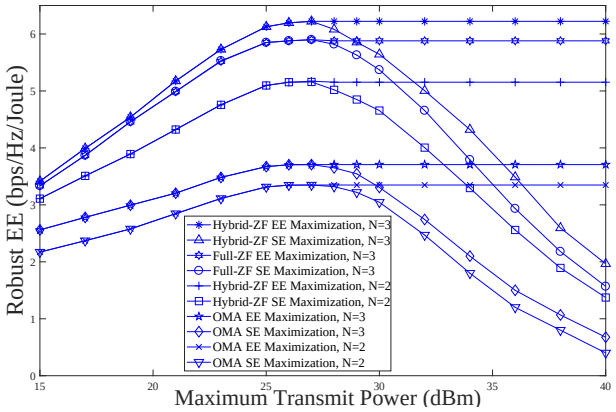


Fig. 2. Robust EE performance versus the maximum available power at the BS in hybrid-ZF, full-ZF and OMA schemes. System parameters are $K = 2$ clusters, $R^{\min} = 1$ and error bound $\varepsilon = 0.001$.

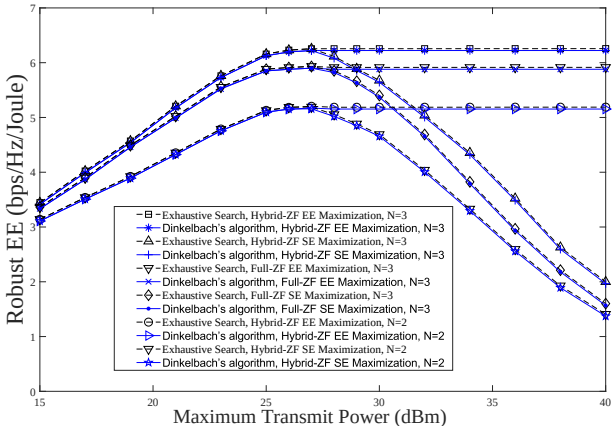


Fig. 3. Robust EE performance versus the maximum available power at the BS in hybrid-ZF and full-ZF schemes by using Dinkelbach's algorithm and exhaustive-search. System parameters are $K = 2$ clusters, $R^{\min} = 1$ and error bound $\varepsilon = 0.001$.

and (46a). These problems are in fact linear programming (LP) after applying the Dinkelbach's algorithm which turns the fractional program into a simple subtractive form. The complexity of solving an LP is $O\left(n_{LP}^2 m_{LP}\right)$, where m_{LP} is the number of linear constraints and n_{LP} is the dimension of optimization variables. For both problems in (27a) and (46a), we have $m_{LP} = 6K + 1$ and $n_{LP} = 7K$. Thus, the complexity of solving these problems is $O\left(49K^2(6K+1)\right)$. Furthermore, the complexity of alternating optimization-based solution is $O\left[L_I\left(L_D\left(49K^2(6K+1)\right)\right)\right]$, where L_D and L_I denote the numbers of iterations required for the Dinkelbach's algorithm in Algorithm 1 and alternating optimization iterations in Algorithm 2, respectively. The parameters L_D and L_I depend on the predefined tolerance set for the algorithms. L_I can be determined by a numerical analysis since no formula is available for the sequential method in Algorithm 2 to calculate the number of required iterations. From [55], the number of required iterations in the Dinkelbach's algorithm (i.e., L_D in Algorithm 1) to solve $\max \frac{f(x)}{g(x)}$ with tolerance ϵ can be expressed as $\log_2\left(\frac{U-L}{\epsilon}\right)$, where L and U are a lower-bound and an upper-bound for the objective function $\frac{f(x)}{g(x)}$, respectively.

VI. Simulation Results

We evaluate the performance of the proposed robust EE design for the MISO NOMA system by generating 1000 Monte-Carlo realizations of the flat fading channels. A downlink transmission is considered in a single cell with one BS equipped with N antennas and K clusters with two single-antenna users per cluster. The small-scale fading of the channels is assumed to be Rayleigh fading which represents an isotropic scattering environment. The large-scale fading effect is modelled by ${d_{lk}}^{-\beta}$ to incorporate the path-loss effects, where d_{lk} is the distance between $U_{l,k}$ and BS, measured in meters and β is the path-loss exponent . Hence, the channel coefficients between BS and user $U_{l,k}$ are generated using $\mathbf{h}_{l,k} = \chi\sqrt{{d_{lk}}^{-\beta}}$, where $\chi \sim \mathcal{CN}(0, \mathbf{I})$ and $\beta = 3.8$ [56]. Throughout the simulations, it is assumed that users are uniformly distributed within a circle with a radius of 50 meters around the BS, but no closer than 1 meter. In addition, we assume that the users' locations are fixed and the average is taken over the small-scale fading of the propagation channels. In addition, we assume that the noise power is $\sigma^2 = 0.01$ at each receiver, and the minimum QoS requirement for all users is the same. Herein, the term non-robust scheme refers to the scheme where the beamforming vectors are designed based on imperfect CSI without incorporating channel uncertainty information.

The achievable robust EE against maximum available transmit power at the BS is presented in Fig. 2 for both full-ZF and hybrid-ZF schemes and conventional OMA scheme. In this figure, the EE maximization represents the solution to the original optimization problems in (27a) and (46a), while SE maximization represents the EE obtained by maximizing the sum rate of the system. In other words, the sum rate maximization problem is solved and then the allocated power are used to calculate the EE of the defined SE problem. As shown in Fig. 2, the achievable EE reaches a maximum with a certain available power (referred to as green power in the literature) and then it remains constant for any available power which is more than the green power. Hence, one can conclude that just a portion of the power budget contributes achieving the maximum EE, and using more power will deteriorate the performance of the system in terms of EE, which is the case in the SE maximization-based design. In addition, it illustrates that NOMA outperforms the conventional OMA scheme in terms of EE by sharing resources in an efficient way.

For a given transmit power and with minimum required transmit antennas in each scheme (i.e. 2 antennas in hybrid-ZF scheme and 3 antennas in full-ZF scheme), the full-ZF can achieve more EE than that of the hybrid-ZF scheme. In fact, the full-ZF scheme can provide higher data rate by completely removing other clusters interference at the cost of more required transmit antennas at the BS.

In Fig. 3, we compare the performance of the Dinkelbach's algorithm with the exhaustive-search algorithm. As seen in this figure, the proposed algorithm can offer a similar performance

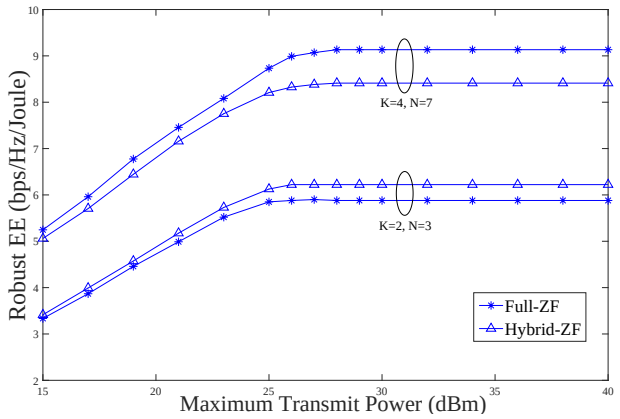


Fig. 4. Robust EE performance versus the maximum available power in hybrid-ZF and full-ZF schemes with equal number of transmit antennas at the BS. System parameters are $R^{\min} = 1$ and error bound $\varepsilon = 0.001$.

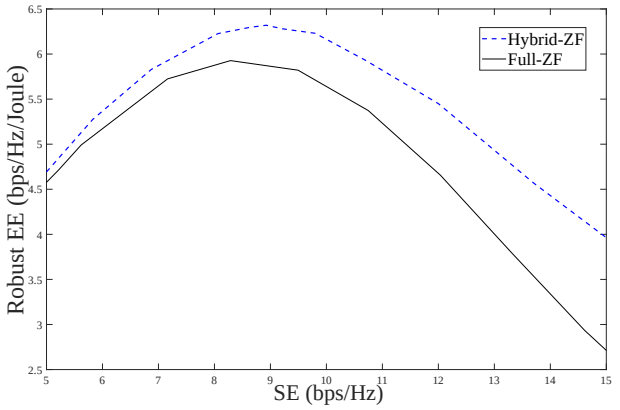


Fig. 5. The EE-SE trade-off for full-ZF and hybrid-ZF schemes. System parameters are $K = 2$ clusters, $N = 3$ antennas.

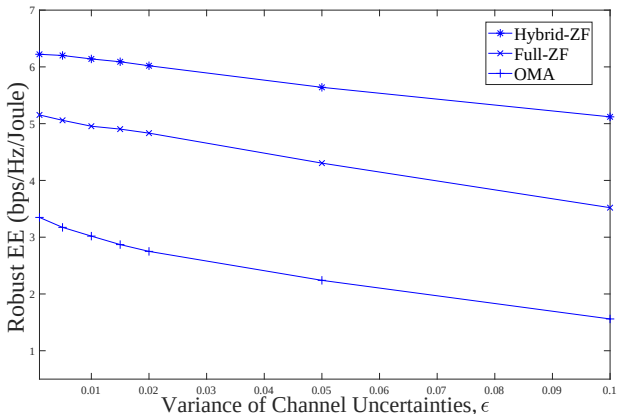


Fig. 6. Robust EE performance with different variance of channel uncertainty in full-ZF, hybrid-ZF and OMA schemes. System parameters are $K = 2$ clusters, $N = 3$ antennas and $R^{\min} = 1$.

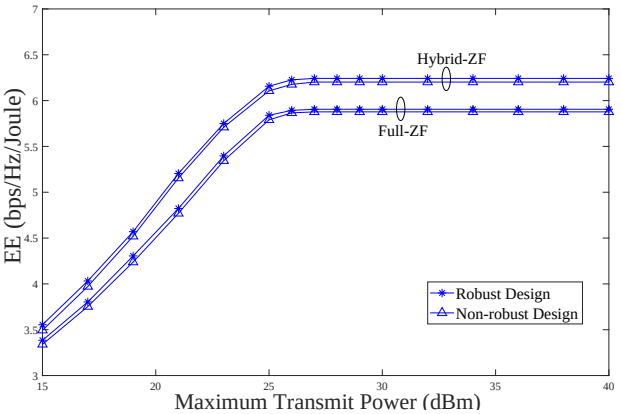


Fig. 7. EE performance versus maximum available power for the robust and the non-robust schemes with channel estimation error bound $\varepsilon = 0.001$ in hybrid-ZF and full-ZF schemes. System parameters are $K = 2$ clusters, $N = 3$ antennas and $R^{\min} = 1$.

to that of the exhaustive-search. Note that the complexity and computation time of exhaustive-search is significantly higher than that of the Dinkelbach's algorithm, particularly with a large number of variables.

To draw a fair comparison, it is assumed that an equal number of transmit antennas is employed for both hybrid-ZF and full-ZF schemes. As seen in Fig. 4, the hybrid-ZF scheme outperforms the full-ZF in terms of EE when there are a few clusters. This is due to the fact that the full-ZF requires more transmit power to completely remove the residual interference, while this type of interference has less impact in the systems with a few clusters. In other words, the rate improvement in full-ZF is not as much as the required power, which degrades the system performance in terms of EE. However, by increasing the number of clusters, the full-ZF scheme outperforms the hybrid-ZF scheme because the residual interference increases, which has a significant impact on the overall performance of the system.

Next we evaluate the trade-off between the SE and EE of the proposed schemes. Fig. 5 depicts the EE-SE trade-off of both full-ZF and hybrid-ZF schemes. As shown in Fig. 5, both SE and EE increase up to a maximum level which is known as the best trade-off point, and then EE decreases while SE increases. Beyond this best trade-off point, the EE should be sacrificed to achieve higher SE for which the BS requires more transmit power. On the other hand, the impact of different channel uncertainty on the achieved EE is represented in Fig. 6. It can be observed from Fig. 6 that the EE decreases for both schemes as the variance of the channel uncertainty in the CSI increases.

Next, we demonstrate the impact of the proposed robust design on the achievable EE and rate by comparing with the performance of the non-robust scheme. The achieved EE for robust and non-robust designs are depicted in Fig. 7 for different available transmit power at the BS. As shown, the results of the robust and non-robust schemes are almost identical for $\varepsilon = 0.001$. To have a fair comparison, we compare the performance of the robust and the non-robust schemes in term of rate-satisfaction ratio, which is defined as the ratio between the achieved rate and the target rate at each user. Hence, a rate-satisfaction ratio greater than 1 indicates that the rate requirement is satisfied at each user. Fig. 8 depicts the histogram of the rate-satisfaction ratio for the robust and non-robust schemes. The simulation result implies that the rate constraint in the robust design is satisfied all the time regardless of the channel uncertainties. However, the

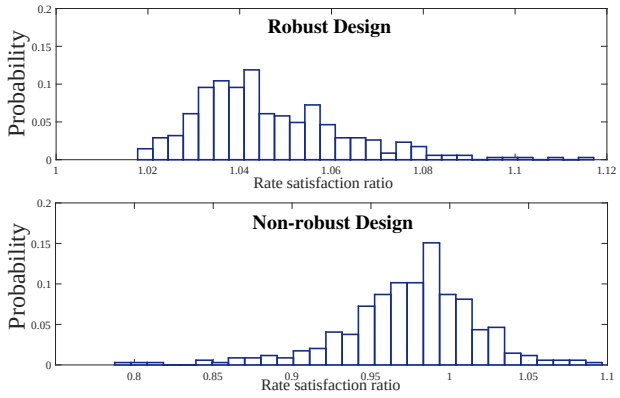


Fig. 8. Histogram for the rate-satisfaction ratio in the robust and non-robust NOMA scheme with channel estimation error bound $\varepsilon = 0.001$ and $R^{\min} = 1$.

non-robust design cannot satisfy the target rate requirement for many cases since it does not take channel uncertainties into account.

VII. Conclusions

In this paper, we have studied the robust EE maximization problem for a MISO NOMA systems with clustering, under total transmit power constraint and minimum rate requirement at each user. In these robust schemes, the inevitable channel uncertainties are taken into account to reduce their impact on the overall system performance. For beamforming design, the ZF approach is employed to mitigate the inter-cluster interference. In particular, we proposed two different ZF schemes, namely: I) hybrid-ZF and II) full-ZF. The objective function that defines the EE of the system is a non-convex and a non-linear function which formulates the original problem into a fractional programming. To deal with the non-convexity issues introduced by both objective function and constraints, an iterative algorithm which exploits the first order Taylor series approximations was applied to transform the original intractable problem into a more tractable and equivalent one. In each iteration, the Dinkelbach's algorithm was employed to convert the non-linear fractional programming problem into a simple subtractive form. Simulation results validated the performance of the proposed schemes in terms of the achieved EE and SE. Despite the fact that the full-ZF scheme can completely remove the interference between different clusters, it requires more transmit antennas than the hybrid-ZF scheme to serve the same number of users. However, by increasing the number of clusters, the inter-cluster interference increases, and consequently, the full-ZF approach shows a better performance in terms of EE. In addition, results confirmed that the proposed robust approach outperforms the non-robust scheme in terms of the rate-satisfaction ratio at each user.

Appendix A
Proof of Lemma 1

Let us assume that the numerator and denominator of $\text{SINR}_{i,k}^{j}$ are independent and derive their worst-case terms separately. Based on this assumption, we introduce a function $\varphi_{i,k}$ as a lower bound for $\inf_{\Delta\hat{\mathbf{h}}_{i,k}}(\text{SINR}_{i,k}^{j})$ in (A.1).

Invoking the triangle inequality followed by the Cauchy-Schwartz inequality, one can conclude that

$$\begin{aligned}\left|(\hat{\mathbf{h}}_{i,k}+\Delta\hat{\mathbf{h}}_{i,k})^H\mathbf{w}_k\right| &\geq \left|\hat{\mathbf{h}}_{i,k}^H\mathbf{w}_k\right| - \left|\Delta\hat{\mathbf{h}}_{i,k}^H\mathbf{w}_k\right| \\ &\geq \left|\hat{\mathbf{h}}_{i,k}^H\mathbf{w}_k\right| - \varepsilon\|\mathbf{w}_k\|, \qquad \text{(A.2)}\\ \left|(\hat{\mathbf{h}}_{i,k}+\Delta\hat{\mathbf{h}}_{i,k})^H\mathbf{w}_n\right| &\leq \left|\hat{\mathbf{h}}_{i,k}^H\mathbf{w}_n\right| + \left|\Delta\hat{\mathbf{h}}_{i,k}^H\mathbf{w}_n\right| \\ &\leq \left|\hat{\mathbf{h}}_{i,k}^H\mathbf{w}_n\right| + \varepsilon\|\mathbf{w}_n\|, \qquad \text{(A.3)}\end{aligned}$$

where it is assumed that the channel uncertainty is upper limited by $\|\Delta\hat{\mathbf{h}}_{i,k}\| \leq \varepsilon$. Then, after plugging (A.2) and (A.3) into the numerator and the denominator of (A.1), we obtain

$$\inf_{\|\Delta\hat{\mathbf{h}}_{i,k}\|\leq\varepsilon}\left(\left|(\hat{\mathbf{h}}_{i,k}+\Delta\hat{\mathbf{h}}_{i,k})^H\mathbf{w}_k\right|^2\right)=\left|\left(\left|\hat{\mathbf{h}}_{i,k}^H\mathbf{w}_k\right|-\varepsilon\|\mathbf{w}_k\|\right)^+\right|^2, \qquad \text{(A.4)}$$

$$\sup_{\|\Delta\hat{\mathbf{h}}_{i,k}\|\leq\varepsilon}\left(\left|(\hat{\mathbf{h}}_{i,k}+\Delta\hat{\mathbf{h}}_{i,k})^H\mathbf{w}_n\right|^2\right) = \left|\left|\hat{\mathbf{h}}_{i,k}^H\mathbf{w}_n\right|+\varepsilon\|\mathbf{w}_n\|\right|^2, \qquad \text{(A.5)}$$

$$\sup_{\|\Delta\hat{\mathbf{h}}_{i,k}\|\leq\varepsilon}\left(\left|\Delta\hat{\mathbf{h}}_{i,k}^H\mathbf{w}_m\right|^2\right) = \left(\varepsilon\|\mathbf{w}_n\|\right)^2, \qquad \text{(A.6)}$$

which completes the proof. ■

Appendix B
Dinkelbach's Algorithm

Dinkelbach's algorithm is a well-known technique to tackle the following concave-convex fractional problem (CCFP):

$$\begin{aligned}\max_{x} \quad & \frac{f(\mathbf{x})}{g(\mathbf{x})}, \qquad \text{(B.1)}\\ s.t. \quad & c_i(\mathbf{x}) \leq 0,\ \forall\ i = 1,\ldots,I, \qquad \text{(B.2)}\\ & h_j(\mathbf{x}) = 0,\ \forall\ j = 1,\ldots,J, \qquad \text{(B.3)}\end{aligned}$$

where $f(\mathbf{x})$ is a non-negative differentiable concave function, $g(\mathbf{x})$ is a positive differentiable convex function, c_i is convex for all $i = 1,\ldots,I$, and h_j is an affine function for all $j = 1,\ldots,J$.

Dinkelbach's algorithm has been originally introduced in [57], [58]. Furthermore, it belongs to the class of parametric algorithms. The fundamental concept of this algorithm is to obtain the solution of a CCFP by solving a sequence of simple subproblems which converge to the global optimal solution of the CCFP. The pseudo-code of Dinkelbach's algorithm is provided in Table I.

References

[1] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 36–43, May 2014.

[2] 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, January 2016.

$$\varphi_{i,k} = \frac{p_j \ \inf_{\Delta\hat{\mathbf{h}}_{i,k}} \left(|(\hat{\mathbf{h}}_{i,k} + \Delta\hat{\mathbf{h}}_{i,k})^H \mathbf{w}_k|^2 \right)}{\sum_n p_n \ \sup_{\Delta\hat{\mathbf{h}}_{i,k}} \left(|(\hat{\mathbf{h}}_{i,k} + \Delta\hat{\mathbf{h}}_{i,k})^H \mathbf{w}_n|^2 \right) + \sum_m p_m \ \sup_{\Delta\hat{\mathbf{h}}_{i,k}} \left(|\Delta\hat{\mathbf{h}}_{i,k}^H \mathbf{w}_m|^2 \right) + \sigma^2}. \tag{A.1}$$

[3] Y. Sun, D. W. K. Ng, Z. Ding, and R. Schober, "Optimal joint power and subcarrier allocation for full-duplex multicarrier non-orthogonal multiple access systems," *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1077–1091, March 2017.

[4] Z. Ding, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2181–2195, October 2017.

[5] F. Alavi, N. M. Yamchi, M. R. Javan, and K. Cumanan, "Limited feedback scheme for device-to-device communications in 5G cellular networks with reliability and cellular secrecy outage constraints," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 9, pp. 8072–8085, September 2017.

[6] M. Fozooni, H. Q. Ngo, M. Matthaiou, S. Jin, and G. C. Alexandropoulos, "Hybrid processing design for multipair massive MIMO relaying with channel spatial correlation," *IEEE Transactions on Communications*, vol. 67, no. 1, pp. 107–123, January 2019.

[7] A. Zappone and E. Jorswieck, "Energy efficiency in wireless networks via fractional programming theory," *Foundations and Trends in Communications and Information Theory*, vol. 11, no. 3, pp. 185–396, 2015.

[8] Y. Wu *et al.*, "Green transmission technologies for balancing the energy efficiency and spectrum efficiency trade-off," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 112–120, November 2014.

[9] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. U. Hassan, and C. Yuen, "Energy efficiency tradeoff mechanism towards wireless green communication: A survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 686–705, Firstquarter 2016.

[10] Q. Li, H. Niu, A. Papathanassiou, and G. Wu, "5G network capacity: key elements and technologies," *IEEE Vehicular Technology Magazine*, vol. 9, no. 1, pp. 71–78, March 2014.

[11] 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, December 2014.

[12] L. Dai, B. Wang, Y. Yuan, S. Han, C. L. I, and Z. Wang, "Non-orthogonal multiple access for 5G: solutions, challenges, opportunities, and future research trends," *IEEE Communications Magazine*, vol. 53, no. 9, pp. 74–81, September 2015.

[13] S. M. 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, October 2017.

[14] F. Alavi, K. Cumanan, Z. Ding, and A. G. Burr, "Beamforming techniques for non-orthogonal multiple access in 5G cellular networks," *IEEE Transactions on Vehicular Technology*, Early Access 2018.

[15] 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, October 2017.

[16] S. M. R. Islam, M. Zeng, O. A. Dobre, and K. Kwak, "Resource allocation for downlink NOMA systems: Key techniques and open issues," *IEEE Wireless Communications*, vol. 25, no. 2, pp. 40–47, April 2018.

[17] M. F. Hanif, Z. Ding, T. Ratnarajah, and G. K. Karagiannidis, "A minorization-maximization method for optimizing sum rate in the downlink of non-orthogonal multiple access systems," *IEEE Transactions on Signal Processing*, vol. 64, no. 1, pp. 76–88, January 2016.

[18] Y. Zhang, H. M. Wang, Q. Yang, and Z. Ding, "Secrecy sum rate maximization in non-orthogonal multiple access," *IEEE Communications Letters*, vol. 20, no. 5, pp. 930–933, May 2016.

[19] Z. Ding, Z. Zhao, M. Peng, and H. V. Poor, "On the spectral efficiency and security enhancements of NOMA assisted multicast-unicast streaming," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 3151–3163, July 2017.

[20] P. Xu and K. Cumanan, "Optimal power allocation scheme for non-orthogonal multiple access with α-fairness," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 10, pp. 2357–2369, October 2017.

[21] P. Xu, K. Cumanan, and Z. Yang, "Optimal power allocation scheme for NOMA with adaptive rates and alpha-fairness," in *Proc. IEEE Global Communications Conference*, December 2017.

[22] H. Al-Obiedollah, K. Cumanan, J. Thiyagalingam, A. G. Burr, Z. Ding, and O. A. Dobre, "Sum rate fairness trade-off-based resource allocation technique for MISO NOMA systems," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2019.

[23] Q. Sun, S. Han, C. L. I, and Z. Pan, "Energy efficiency optimization for fading MIMO non-orthogonal multiple access systems," in *Proc. IEEE International Conference on Communications (ICC)*, June 2015, pp. 2668–2673.

[24] 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.

[25] 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.

[26] ——, "Energy-efficient resource scheduling for NOMA systems with imperfect channel state information," in *Proc. IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–5.

[27] F. Fang, H. Zhang, J. Cheng, S. Roy, and V. C. M. Leung, "Joint user scheduling and power allocation optimization for energy-efficient NOMA systems with imperfect CSI," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 12, pp. 2874–2885, December 2017.

[28] J. Wang, H. Xu, L. Fan, B. Zhu, and A. Zhou, "Energy-efficient joint power and bandwidth allocation for NOMA systems," *IEEE Communications Letters*, vol. 22, no. 4, pp. 780–783, April 2018.

[29] P. Wu, J. Zeng, X. Su, H. Gao, and T. Lv, "On energy efficiency optimization in downlink MIMO-NOMA," in *IEEE International Conference on Communications Workshops (ICC Workshops)*, May 2017, pp. 399–404.

[30] 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.

[31] H. Al-Obiedollah, K. Cumanan, J. Thiyagalingam, A. G. Burr, Z. Ding, and O. A. Dobre, "Energy efficiency fairness beamforming design for MISO NOMA systems," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC)*, 2019.

[32] ——, "Energy efficient beamforming design for MISO non-orthogonal multiple access systems," *IEEE Transactions on Communications*, pp. 1–1, 2019.

[33] M. Zeng, W. Hao, O. A. Dobre, and V. Poor, "Energy-efficient power allocation in uplink mmwave massive MIMO with NOMA," *IEEE Transactions on Vehicular Technology*, pp. 1–1, 2019.

[34] M. Zeng, A. Yadav, O. Dobre, and H. V. Poor, "Energy-efficient joint user-RB association and power allocation for uplink hybrid NOMA-OMA," *IEEE Internet of Things Journal*, pp. 1–1, 02 2019.

[35] X. He and Y. C. Wu, "Tight probabilistic SINR constrained beamforming under channel uncertainties," *IEEE Transactions on Signal Processing*, vol. 63, no. 13, pp. 3490–3505, July 2015.

[36] K. Cumanan, R. Krishna, V. Sharma, and S. Lambotharan, "A robust beamforming based interference control technique and its performance for cognitive radios," in *Proc. International Symposium on Communications and Information Technologies*, October 2008, pp. 9–13.

[37] ——, "Robust interference control techniques for cognitive radios using worst-case performance optimization," in *Proc. International Symposium on Information Theory and Its Applications*, December 2008, pp. 1–5.

[38] K. Cumanan, Z. Ding, Y. Rahulamathavan, M. M. Molu, and H. Chen, "Robust MMSE beamforming for multiantenna relay networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 3900–3912, May 2017.

[39] C. Shen, T.-H. Chang, K.-Y. Wang, Z. Qiu, and C.-Y. Chi, "Distributed robust multicell coordinated beamforming with imperfect CSI: An ADMM approach," *IEEE Transactions on Signal Processing*, vol. 60, no. 6, pp. 2988–3003, June 2012.

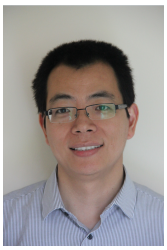
[40] S. K. Joshi, U. L. Wijewardhana, M. Codreanu, and M. Latva-aho, "Maximization of worst-case weighted sum-rate for MISO downlink

systems with imperfect channel knowledge," *IEEE Transactions on Communications*, vol. 63, no. 10, pp. 3671–3685, October 2015.
[41] S. Nasseri and M. R. Nakhai, "Robust interference management via outage-constrained downlink beamforming in multicell networks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, December 2013, pp. 3470–3475.
[42] C. Shen, T. H. Chang, K. Y. Wang, Z. Qiu, and C. Y. Chi, "Chance-constrained robust beamforming for multi-cell coordinated downlink," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, December 2012, pp. 4957–4962.
[43] 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. 10 231–10 236, December 2016.
[44] F. Alavi, K. Cumanan, Z. Ding, and A. G. Burr, "Robust beamforming techniques for non-orthogonal multiple access systems with bounded channel uncertainties," *IEEE Communications Letters*, vol. 21, no. 9, pp. 2033–2036, September 2017.
[45] ——, "Outage constraint based robust beamforming design for non-orthogonal multiple access in 5G cellular networks," in *Proc. IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, October 2017, pp. 1–5.
[46] M. Tian, Q. Zhang, S. Zhao, Q. Li, and J. Qin, "Robust beamforming in downlink MIMO NOMA networks using cutting-set method," *IEEE Communications Letters*, vol. 22, no. 3, pp. 574–577, March 2018.
[47] B. Kimy, S. Lim, H. Kim, S. Suh, J. Kwun, S. Choi, C. Lee, S. Lee, and D. Hong, "Non-orthogonal multiple access in a downlink multiuser beamforming system," in *Proc. IEEE Military Communications Conference MILCOM*, November 2013, pp. 1278–1283.
[48] Y. Li, M. Jiang, Q. Zhang, Q. Li, and J. Qin, "Secure beamforming in downlink MISO nonorthogonal multiple access systems," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7563–7567, August 2017.
[49] M. Fozooni, M. Matthaiou, E. Bjornson, and T. Q. Duong, "Performance limits of MIMO systems with nonlinear power amplifiers," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, December 2015, pp. 1–7.
[50] S. Shahbazpanahi, A. B. Gershman, Zhi-Quan Luo, and Kon Max Wong, "Robust adaptive beamforming using worst-case SINR optimization: a new diagonal loading-type solution for general-rank signal models," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, April 2003.
[51] A. Abdel-Samad, T. N. Davidson, and A. B. Gershman, "Robust transmit eigen beamforming based on imperfect channel state information," *IEEE Transactions on Signal Processing*, vol. 54, no. 5, pp. 1596–1609, May 2006.
[52] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt, "Zero-forcing methods for downlink spatial multiplexing in multiuser MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 461–471, February 2004.
[53] A. Ben-Tal and A. Nemirovski, "Lecture on modern convex optimization: analysis, algorithms, and engineering applications," *Philadelphia, PA, USA: MPS-SIAM*.
[54] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
[55] A. Zappone, P. Lin, and E. Jorswieck, "Energy efficiency of confidential multi-antenna systems with artificial noise and statistical CSI," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 8, pp. 1462–1477, December 2016.
[56] M. Fozooni, M. Matthaiou, S. Jin, and G. C. Alexandropoulos, "Massive MIMO relaying with hybrid processing," in *Proc. IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.
[57] R. Jagannathan, "On some properties of programming problems in parametric form pertaining to fractional programming," in *Management Science*, vol. 12, no. 7, 1966.
[58] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, March 1967.

Faezeh Alavi (S'12) received the B.Sc. degree from Shahid Beheshti University, Tehran, Iran, the M.Sc. degree from Tarbiat Modares University, Tehran, Iran, both in Electrical Engineering in 2011 and 2013, respectively, and the Ph.D. degree in Electrical Engineering from University of York, York, U.K in 2018. She is currently with the Algorithm Group at Combitech, Göteborg, Sweden. Her research interests include non-orthogonal multiple access (NOMA), MIMO, physical layer security, convex optimization techniques, and also applications of robust optimization theory on the resource allocation in wireless networks.

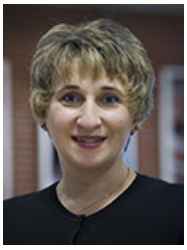
Kanapathippillai Cumanan (M'10) received the B.Sc. degree with first class honors in electrical and electronic engineering from the University of Peradeniya, Sri Lanka in 2006 and the PhD degree in signal processing for wireless communications from Loughborough University, Loughborough, UK, in 2009. He is currently a lecturer at the Department of Electronic Engineering, The University of York, UK. From March 2012 to November 2014, he was working as a research associate at School of Electrical and Electronic Engineering, Newcastle University, UK. Prior to this, he was with the School of Electronic, Electrical and System Engineering, Loughborough University, UK. In 2011, he was an academic visitor at Department of Electrical and Computer Engineering, National University of Singapore, Singapore. From January 2006 to August 2006, he was a teaching assistant with Department of Electrical and Electronic Engineering, University of Peradeniya, Sri Lanka. His research interests include non-orthogonal multiple access (NOMA), cell-free massive MIMO, physical layer security, cognitive radio networks, convex optimization techniques and resource allocation techniques. He has published more than 80 journal articles and conference papers which attracted more than 1200 Google scholar citations. He has been also recently appointed as an associate editor for IEEE Access journal. Dr. Cumanan was the recipient of an overseas research student award scheme (ORSAS) from Cardiff University, Wales, UK, where he was a research student between September 2006 and July 2007.

Milad Fozooni received the B.Sc. degree in electrical engineering from the K. N. Toosi University of Technology, Tehran, Iran, in 2008, the M.Sc. degree in electrical engineering from the University of Tehran, Tehran, in 2012, and the Ph.D. degree from Queen's University Belfast in 2017, with a focus on low-cost architectures for future multiple-input multiple-output (MIMO) systems. He is currently a Researcher with the Radio Group, Ericsson Research, Göteborg, Sweden. His main research interests include MIMO, massive MIMO systems, mm-wave communications, signal processing, and resource allocation techniques.



Zhiguo Ding (S'03-M'05) received his B.Eng in Electrical Engineering from the Beijing University of Posts and Telecommunications in 2000, and the Ph.D degree in Electrical Engineering from Imperial College London in 2005. From Jul. 2005 to Apr. 2018, he was working in Queen's University Belfast, Imperial College, Newcastle University and Lancaster University. Since Apr. 2018, he has been with the University of Manchester as a Professor in Communications. From Oct. 2012 to Sept. 2018, he has also been an academic visitor in Princeton University. Dr Ding' research interests are 5G networks, game theory, cooperative and energy harvesting networks and statistical signal processing. He is serving as an Editor for *IEEE Transactions on Communications*, *IEEE Transactions on Vehicular Technology*, and *Journal of Wireless Communications and Mobile Computing*, and was an Editor for *IEEE Wireless Communication Letters*, *IEEE Communication Letters* from 2013 to 2016. He received the best paper award in IET ICWMC-2009 and IEEE WCSP-2014, the EU Marie Curie Fellowship 2012-2014, the Top IEEE TVT Editor 2017, IEEE Heinrich Hertz Award 2018, IEEE Jack Neubauer Memorial Award 2018 and IEEE Best Signal Processing Letter Award 2018.

Sangarapillai Lambotharan is Professor of Digital Communications and the Head of Signal Processing and Networks Research Group in the Wolfson School Mechanical, Electrical and Manufacturing Engineering at Loughborough University, UK. He received the PhD degree in Signal Processing from Imperial College London, UK in 1997, where he remained until 1999 as a postdoctoral research associate. He was a visiting scientist at the Engineering and Theory Centre of Cornell University, USA in 1996. Between 1999 and 2002, he was with Motorola Applied Research Group, UK and investigated various projects including physical link layer modelling and performance characterization of GPRS, EGPRS and UTRAN. He was with King's College London and Cardiff University as a lecturer and senior lecturer respectively from 2002 to 2007. His current research interests include 5G networks, MIMO, blockchain, machine learning and network security. He has authored over 200 journal and conference articles in these areas.



Octavia A. Dobre received the Dipl. Ing. and Ph.D. degrees from Politehnica University of Bucharest (formerly Polytechnic Institute of Bucharest), Romania, in 1991 and 2000, respectively. Between 2002 and 2005, she was with New Jersey Institute of Technology, USA and Politehnica University of Bucharest. In 2005, she joined Memorial University, Canada, where she is currently Professor and Research Chair. She was a Visiting Professor with Massachusetts Institute of Technology, USA and Université de Bretagne Occidentale, France. Her research interests include enabling technologies for 5G and beyond, blind signal identification and parameter estimation techniques, as well as optical and underwater communications. She authored and co-authored over 250 refereed papers in these areas. Dr. Dobre serves as the Editor-in-Chief (EiC) of the IEEE Open Journal of the Communications Society, as well as an Editor of the IEEE Communications Surveys and Tutorials, IEEE Vehicular Communications Magazine, and IEEE Systems. She was the EiC of the Communications Letters, as well as Senior Editor, Editor, and Guest Editor for various prestigious journals and magazines. Dr. Dobre was the General Chair, Technical Program Co-Chair, Tutorial Co-Chair, and Technical Co-Chair of symposia at numerous conferences, including IEEE ICC and IEEE Globecom. Dr. Dobre was a Royal Society Scholar in 2000 and a Fulbright Scholar in 2001. She obtained Best Paper Awards as various conferences, including IEEE ICC and IEEE WCNC. Dr. Dobre is a Distinguished Lecturer of the IEEE Communications Society and a Fellow of the Engineering Institute of Canada.