Resource-Efficient NOMA Transmission via Joint Bandwidth and Rate Allocations

Yuan Wu, Xiaoding Wang, Li Ping Qian, and Kejie Ni

***Abstract*— In this letter, we investigate the joint channel bandwidth and rate allocations for downlink non-orthogonal multiple access with the objective of maximizing the resource utilization efficiency (RUE). In particular, we take into account the utility of base station (BS) in serving the mobile terminals' traffic and the cost in consuming channel bandwidth and measure the RUE as the ratio between the BS's utility and the total power consumption. Despite the non-convexity of the joint optimization problem, we propose a three-layered vertical decomposition and design an efficient algorithm to compute the optimal solution. Extensive numerical results are provided to validate the performance of our proposed algorithms.**

***Index Terms*— Non-orthogonal multiple access (NOMA), resource allocation, optimization.**

I. Introduction

NON-ORTHOGONAL multiple access (NOMA) has been considered as one of the enabling technologies for accommodating the explosive traffic growth in future cellular systems. By allowing a group of mobile users to simultaneously transmit over a same resource block and further using the successive interference cancellation (SIC) to mitigate the co-channel interference, NOMA is expected to significantly improve the spectrum efficiency and energy efficiency in radio access networks [1]–[5]. Zhang *et al.* [6] proposed a power allocation strategy for maximizing the energy efficiency of downlink multi-user NOMA. Liu *et al.* [7] considered a cooperative relaying system using NOMA and proposed a power allocation scheme from the perspective of global energy efficiency. Zeng *et al.* [8] proposed an energy-efficient power allocation for a multiple-input multiple-output non-orthogonal multiple access system. In [9], a joint power allocation and subchannel assignment for NOMA has been proposed for maximizing the energy efficiency. In [10], a joint power allocation and traffic scheduling scheme has been proposed for the NOMA-assisted multi-relay transmission.

Manuscript received September 3, 2018; revised October 5, 2018; accepted November 1, 2018. Date of publication November 9, 2018; date of current version February 11, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61572440, in part by the Zhejiang Provincial Natural Science Foundation of China under Grants LR16F010003 and LR17F010002, and in part by the open research fund of National Mobile Communications Research Laboratory, Southeast University under Grant 2019D11. The associate editor coordinating the review of this paper and approving it for publication was F. Wang. *(Corresponding author: Li Ping Qian.)*

Y. Wu is with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China, and also with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China.

X. Wang and K. Ni are with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China.

L. P. Qian is with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China, and also with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: lpqian@zjut.edu.cn).

Digital Object Identifier 10.1109/LCOMM.2018.2879945

In addition to the power allocation, properly sizing the channel bandwidth is important to NOMA transmission. In [11], a joint optimization of power and bandwidth allocations has been proposed for improving the energy-efficiency of NOMA (but without accounting for the bandwidth-usage cost in the objective). Different from [11], we focus on the single-channel (but with adjustable bandwidth), and analytically characterize the consequent structure property of the users' optimal rate allocations. Specifically, in this work, we propose a joint optimization of the channel bandwidth and rate allocation for downlink NOMA, with the objective of maximizing the resource-utilization efficiency (RUE). We model the utility of base station (BS) as the difference between the weighted total downlink throughput to the mobile terminals' (MTs') and the weighted cost for the bandwidth usage, and then measure the RUE as the ratio between the BS's utility and the total power consumption. To tackle with the non-convexity of the joint optimization problem, we exploit a three-layered vertical decomposition and identify the structural property of the MTs' optimal rate allocations. Exploiting the decomposition and the structural property, we propose an efficient algorithm to compute the optimal solution.

II. System Model and Problem Formulation

We consider the downlink of NOMA system in which one BS serves a group of MTs $\mathcal{I} = \{1, 2, \dots, I\}$. Since the NOMA requires the SIC to mitigate co-channel interference among the MTs, we assume that the MTs are ordered according to

$$g_{B1} > g_{B2} > \dots > g_{Bi} > g_{Bj} > \dots > g_{BI}, \quad (1)$$

where g_{Bi} denotes the channel power gain from the BS to MT i (we assume that the BS has the perfect channel state information). Based on (1), the downlink throughput from the BS to MT i is

$$R_i = W_B \log_2\left(1 + \frac{g_{Bi}p_{Bi}}{g_{Bi}\sum_{j=1}^{i-1} p_{Bj} + W_B n_0}\right), \quad \forall i \in \mathcal{I}. \quad (2)$$

where W_B denotes the BS's bandwidth allocation and p_{Bi} denotes the BS's transmit-power allocation for MT i. Exploiting (2), we have the following important result.

Proposition 1: Given the channel bandwidth W_B and the rate allocations $\{R_i\}_{i\in\mathcal{I}}$ for the MTs, the BS's minimum total power consumption can be compactly given by:

$$p_B^{\min}(W_B, \{R_i\}_{i\in\mathcal{I}}) = W_B \sum_{i=1}^{I} \left(\frac{n_0}{g_{Bi}} - \frac{n_0}{g_{Bi-1}}\right)\left(2^{\frac{1}{W_B}\sum_{m=i}^{I} R_m} - 1\right), \quad \forall i \in \mathcal{I}. \quad (3)$$

Proof: The proof is similar to that for [10, Proposition 1]. ∎

Based on Proposition 1, we formulate the following optimization problem

$$
\begin{aligned}
\text{(P1):}\quad & \max \frac{\alpha \sum_{i=1}^{I} R_i - \beta W_B}{p_B^{\min}(W_B, \{R_i\}_{i\in\mathcal{I}}) + p_0} \\
& \text{subject to: } p_B^{\min}(W_B, \{R_i\}_{i\in\mathcal{I}}) \leq P_B^{\text{tot}}, \qquad (4)\\
& \qquad R_i^{\text{req}} \leq R_i \leq R_i^{\max}, \quad \forall i \in \mathcal{I}, \qquad (5)\\
& \qquad 0 < W_B \leq W_B^{\text{tot}}, \qquad (6)\\
& \text{variables: } W_B, \quad \text{and } \{R_i\}_{i\in\mathcal{I}}.
\end{aligned}
$$

In the objective function, $\alpha \sum_{i=1}^{I} R_i - \beta W_B$ denotes the BS's utility in achieving the total throughput $\sum_{i=1}^{I} R_i$ when consuming channel-bandwidth W_B, where α denotes the weight for the BS's total throughput, and β denotes the weight for the bandwidth utilization cost. Parameter p_0 denotes the BS's fixed power consumption. The objective function represents the RUE measured by the ratio between the BS's utility and the total power consumption. Constraint (4) means that the BS's total power consumption cannot exceed its budget P_B^{tot}. Constraint (5) means that the rate allocation for MT i cannot be no smaller than its requirement R_i^{req} and no greater than its upper-bound $R_i^{\max}$. Constraint (6) means that the BS's bandwidth usage cannot exceed the bandwidth budget W_B^{tot}. Problem (P1) is difficult to solve due to its non-convexity. Thus, in the next section, we propose an efficient algorithm to solve it. We notice that in addition to the objective function used in Problem (P1), there exist other objective functions which can be used for improving the resource-utilization efficiency, e.g., $\alpha \sum_{i=1}^{I} R_i - \beta W_B - \gamma p_B^{\min}(W_B, \{R_i\}_{i\in\mathcal{I}})$ (with γ denoting the weight for the power consumption). With a slight modification, our proposed algorithm can also be used for solving the corresponding optimization problem.

III. Layered Structure of Problem (P1)

The key idea for solving Problem (P1) is to exploit its layered structure. Specifically, we introduce an auxiliary variable η which denotes the lower-bound of RUE as follows:

$$
\frac{\alpha \sum_{i=1}^{I} R_i - \beta W_B}{p_B^{\min}(W_B, \{R_i\}_{i\in\mathcal{I}}) + p_0} \geq \eta. \qquad (7)
$$

With η, we can equivalently transform Problem (P1) into:

$$
\begin{aligned}
\text{(P1-E):}\quad & \max \eta \\
& \text{subject to: } \alpha \sum_{i=1}^{I} R_i - \beta W_B \\
& \qquad - \eta\big(p_B^{\min}(W_B, \{R_i\}_{i\in\mathcal{I}}) + p_0\big) \geq 0, \qquad (8)\\
& \qquad \text{constraints } (4),\ (5),\ \text{and } (6),\\
& \text{variables: } W_B, \{R_i\}_{i\in\mathcal{I}}, \text{ and } \eta \geq 0.
\end{aligned}
$$

Problem (P1-E) aims at finding the maximum value of η which can ensure that the feasible region constructed by (4), (5), (6), and (8) is non-empty. The optimal solution of Problem (P1-E), denoted by η^*, is the maximum value of RUE for Problem (P1).

Given η, to determine whether the feasible region constructed by (4), (5), (6), and (8) is non-empty or not, we consider:

$$
\begin{aligned}
\text{(P1-E-Sub):}\ V_\eta = & \max \alpha \sum_{i=1}^{I} R_i - \beta W_B \\
& - \eta\big(p_B^{\min}(W_B, \{R_i\}_{i\in\mathcal{I}}) + p_0\big),\\
& \text{subject to: constraints } (4),\ (5),\ \text{and } (6),\\
& \text{variables: } W_B \text{ and } \{R_i\}_{i\in\mathcal{I}}.
\end{aligned}
$$

BiSec-Algorithm: to Solve Problem (P1-E-Top) and Determine η^*

1: **Initialization:** set the current upper-bound $\eta^{\text{upp}} = \eta^{\max}$, the current lower-bound $\eta^{\text{low}} = 0$, and the accuracy tol for the computational error.
2: **while** $|\eta^{\text{upp}} - \eta^{\text{low}}| \geq$ tol **do**
3: Set $\eta^{\text{temp}} = \frac{\eta^{\text{upp}} + \eta^{\text{low}}}{2}$.
4: Use Subroutine-forRUE to obtain V_η.
5: **if** $V_\eta < 0$ **then**
6: Set $\eta^{\text{upp}} = \eta^{\text{temp}}$.
7: **else**
8: Set $\eta^{\text{low}} = \eta^{\text{temp}}$.
9: **end if**
10: **end while**
11: **Output**: $\eta^* = \eta^{\text{temp}}$.

If $V_\eta \geq 0$, then it means that the feasible region constructed by (4), (5), (6), and (8) is non-empty under the given η. Otherwise (i.e., $V_\eta < 0$), the feasible region is empty.

Based on V_η output by Problem (P1-E-Sub), we can find η^* that solves Problem (P1-E) as follows:

$$
\text{(P1-E-Top):}\ \eta^* = \arg\max \left\{\eta \in [0, \eta^{\max}] | V_\eta \geq 0\right\},
$$

where the upper-bound $\eta^{\max} = \frac{\sum_{i\in\mathcal{I}} R_i^{\text{req}}}{p_0}$.

To find η^* efficiently, we identify the following result.

Proposition 2: For Problem (P1-E-Sub), the value of V_η decreases with respect to η.

Proof: In Problem (P1-E-Sub), all constraints (4), (5), and (6) do not depend on η, while the objective function decreases with respect to η. As a result, V_η decreases in η. ■

Based on Proposition 1, we propose BiSec-Algorithm, which is based on the bisection-search, to find η^* and solve Problem (P1-E-Top). Notice that in each round of iteration, given the temporary value of η^{temp}, we need to use a Subroutine-forRUE in Step 4 to solve Problem (P1-E-Sub) and determine the value of V_η. We will illustrate the details in Section IV.

IV. Proposed Algorithm to Solve Problem (P1-E-Sub)

We further propose Subroutine-forRUE to solve Problem (P1-E-Sub) and determine the value of V_η. However, Problem (P1-E-Sub) is still a non-convex optimization with respect to W_B and $\{R_i\}_{i\in\mathcal{I}}$. Our key is to adopt a further decomposition, namely, in addition to the given η, we further assume that the bandwidth allocation W_B is given, and thus optimize $\{R_i\}_{i\in\mathcal{I}}$ as:

$$
\begin{aligned}
\text{(RA-Sub):}\ Z_{\eta,W_B} = & \max \alpha \sum_{i\in\mathcal{I}} R_i - \beta W_B \\
& - \eta\big(p_B^{\min}(W_B, \{R_i\}_{i\in\mathcal{I}}) + p_0\big),\\
& \text{subject to: } p_B^{\min}(W_B, \{R_i\}_{i\in\mathcal{I}}) \leq P_B^{\text{tot}}, \qquad (9)\\
& \text{variables: } R_i^{\text{req}} \leq R_i \leq R_i^{\max}, \quad \forall i \in \mathcal{I}.
\end{aligned}
$$

In Problem (RA-Sub), both W_B and η are given in advance.

By treating Z_{η,W_B} (i.e., the output of Problem (RA-Sub)) as a function of W_B, we then continue to optimize W_B as

$$
\text{(BA-Top):}\ V_\eta = \max_{0 \leq W_B \leq W_B^{\text{tot}}} Z_{\eta,W_B}
$$

In particular, we identify the following result.

Proposition 3: Given η and W_B, Problem (RA-Sub) is a strictly convex optimization with respect to $\{R_i\}_{i\in\mathcal{I}}$.

Proof: The proof is based on the convex optimization theory. We skip the details due to the limited space here. ■

The convexity of Problem (RA-Sub) enables us to use the Karush-Kuhn-Tucker (KKT) conditions to determine the optimal solution. Let us use λ to denote the dual variable for (9). We thus can express the Lagrangian function as

$$\mathcal{L}(\{R_i\}_{i\in\mathcal{I}},\lambda) = \alpha\sum_{i=1}^{I} R_i - \beta W_B - (\eta+\lambda)p_B^{\min}(W_B,\{R_i\}_{i\in\mathcal{I}}) - \eta p_0 + \lambda P_B^{\text{tot}}.$$

Moreover, we can derive:

$$\mathcal{L}'_i(\{R_i\}_{i\in\mathcal{I}},\lambda) = \frac{\partial\mathcal{L}(\{R_i\}_{i\in\mathcal{I}},\lambda)}{\partial R_i} = \alpha - (\eta+\lambda)\ln 2\sum_{j=1}^{i}\Big(\frac{n_0}{g_{Bi}} - \frac{n_0}{g_{Bi-1}}\Big)2^{\frac{1}{W_B}\sum_{m=j}^{I}R_m}, \quad \forall i\in\mathcal{I}. \quad (10)$$

Eq. (10) shows $\mathcal{L}'_1(\{R_i\}_{i\in\mathcal{I}},\lambda) > \mathcal{L}'_2(\{R_i\}_{i\in\mathcal{I}},\lambda) > \ldots > \mathcal{L}'_I(\{R_i\}_{i\in\mathcal{I}},\lambda)$, which indicates that there exists at most one particular MT r with $\mathcal{L}'_r(\{R_i\}_{i\in\mathcal{I}},\lambda) = 0$, while all the other MTs $1\le i\le r-1$ with $\mathcal{L}'_i(\{R_i\}_{i\in\mathcal{I}},\lambda) > 0$ and all the other MTs $r+1\le i\le I$ with $\mathcal{L}'_i(\{R_i\}_{i\in\mathcal{I}},\lambda) < 0$.

Therefore, we identify the following results.

Proposition 4: The optimal solution of Problem (RA-Sub) occurs in one of the following possible cases.

- *(Case-I):* There exists a particular MT $r\in\mathcal{I}$, which has
$$\mathcal{L}'_i(\{R_i\}_{i\in\mathcal{I}},\lambda) = 0,$$
and for the other MTs in $\mathcal{I}$, we have
$$R_i^* = R_i^{\max}, \quad \forall i = 1,2,\ldots,r-1, \quad (11)$$
$$R_i^* = R_i^{\text{req}}, \quad \forall i = r+1,r+2,\ldots,I. \quad (12)$$
Meanwhile, for MT r, its R_r^* can be determined according to the following two possible subcases.
  - *(Subcase-I)* supposing $\lambda = 0$, R_r^* is determined by
$$\mathcal{L}'_r(\hat{\mathbf{R}},0) = 0, \quad (13)$$
where $\hat{\mathbf{R}} = \{\{R_i^{\max}\}_{1\le i\le r-1}, R_r, \{R_i^{\text{req}}\}_{r+1\le i\le I}\}$.
  - *(Subcase-II)* supposing $\lambda > 0$, R_r^* is determined by
$$p_B^{\min}(W_B,\hat{\mathbf{R}}) = P_B^{\text{tot}}. \quad (14)$$

  After knowing $\mathbf{R}^* = \{R_i^*\}_{i\in\mathcal{I}}$, the value of λ is determined according to $\mathcal{L}'_r(\mathbf{R}^*,\lambda) = 0$.
- *(Case-II):* the boundary-case that $R_i^* = R_i^{\text{req}}, \forall i\in\mathcal{I}$.
- *(Case-III):* the boundary-case that $R_i^* = R_i^{\max}, \forall i\in\mathcal{I}$.

Proof: We focus on proving Case I here. The key is based on the property $\mathcal{L}'_1(\{R_i\}_{i\in\mathcal{I}},\lambda) > \mathcal{L}'_2(\{R_i\}_{i\in\mathcal{I}},\lambda) > \ldots > \mathcal{L}'_I(\{R_i\}_{i\in\mathcal{I}},\lambda)$, which is implied by eq. (10).

- First, let us suppose that there exists a particular MT r with its $\mathcal{L}'_r(\{R_i\}_{i\in\mathcal{I}},\lambda) = 0$. Then, there exists $\mathcal{L}'_i(\{R_i\}_{i\in\mathcal{I}},\lambda) > 0$, for $i = 1,2,\ldots,r-1$, which leads to $R_i^* = R_i^{\max}, i = 1,2,\ldots,r-1$ in eq. (11). Meanwhile, there exists $\mathcal{L}'_i(\{R_i\}_{i\in\mathcal{I}},\lambda) < 0$, for $i = r+1,r+2,\ldots,I$, which leads to $R_i^* = R_i^{\text{req}}, i = r+1,r+2,\ldots,I$ in (12).
- Next, we determine the value of R_r^*. Suppose that $\lambda = 0$ (i.e., constraint (9) is slack). Then, we can exploit $\mathcal{L}'_r(\hat{\mathbf{R}},0) = 0$ to determine R_r^*, which leads to (13). Otherwise, supposing that $\lambda > 0$, we use $p_B^{\min}(W_B,\hat{\mathbf{R}}) = P_B^{\text{tot}}$ to determine R_r^*, which leads to (14).

We thus finish the proof. ■

Subroutine-forRA: to Compute Z_{η,W_B} and the Corresponding $\{R_i^*\}_{i\in\mathcal{I}}$

1: **Input** & **Initialization:** Input (η, W_B). Set CBV $= 0$, CBS $= \emptyset$, and $r = 1$.
2: **while** $r \le I$ **do**
3: Set $R_i^{\text{cur}} = R_i^{\max}, \forall i = 1,2,\ldots,r-1$ and set $R_i^{\text{cur}} = R_i^{\text{req}}, \forall i = r+1, r+2,\ldots,I$.
4: Set $\mathbf{R}^{\text{cur-low}} = \{\{R_i^{\text{cur}}\}_{i=1,2,\ldots,r-1}, R_r^{\text{req}}, \{R_i^{\text{cur}}\}_{i=r+1,r+2,\ldots,I}\}$.
5: Set $\mathbf{R}^{\text{cur-upp}} = \{\{R_i^{\text{cur}}\}_{i=1,2,\ldots,r-1}, R_r^{\max}, \{R_i^{\text{cur}}\}_{i=r,r+1,r+2,\ldots,I}\}$.
6: **if** $p_B^{\min}(W_B,\mathbf{R}^{\text{cur-low}}) > P_B^{\text{tot}}$ **then**
7: Turn to Step 37 directly.
8: **else**
9: **if** $\mathcal{L}'_r(\mathbf{R}^{\text{cur-low}},0) < 0$ **then**
10: Set $R_r^{\text{cur}} = R_r^{\text{req}}$.
11: **else**
12: **if** $\mathcal{L}'_r(\mathbf{R}^{\text{cur-upp}},0) > 0$ **then**
13: Set $R_r^{\text{cur}} = R_r^{\max}$.
14: **else**
15: Use the bisection search to find $R_r^{\text{cur}} \in [R_r^{\text{req}}, R_r^{\max}]$ such that $\mathcal{L}'_r(\{\{R_i^{\text{cur}}\}_{i=1,2,\ldots,r-1}, R_r^{\text{cur}}, \{R_i^{\text{cur}}\}_{i=r+1,r+2,\ldots,I}\}, 0) = 0$.
16: **end if**
17: **end if**
18: Set $\mathbf{R}^{\text{cur}} = \{\{R_i^{\text{cur}}\}_{i=1,2,\ldots,r-1}, R_r^{\text{cur}}, \{R_i^{\text{cur}}\}_{i=r+1,r+2,\ldots,I}\}$.
19: **if** $p_B^{\min}(W_B,\mathbf{R}^{\text{cur}}) < P_B^{\text{tot}}$ **then**
20: Set $Z^{\text{cur}} = \alpha\sum_{i\in\mathcal{I}}R_i^{\text{cur}} - \beta W_B - \eta(p_B^{\min}(W_B,\mathbf{R}^{\text{cur}}) + p_0)$.
21: **if** $Z^{\text{cur}} \ge$ CBV **then**
22: Set CBV $= Z^{\text{cur}}$ and CBS $= \mathbf{R}^{\text{cur}}$.
23: **end if**
24: **end if**
25: Set $R_r^{\text{cur}} = \emptyset$.
26: **if** $p_B^{\min}(W_B,\mathbf{R}^{\text{cur-upp}}) < P_B^{\text{tot}}$ **then**
27: Turn to Step 37 directly.
28: **else**
29: Use the bisection search to find $R_r^{\text{cur}} \in [R_i^{\text{req}}, R_r^{\max}]$ such that $p_B^{\min}(W_B, \{\{R_i^{\text{cur}}\}_{i=1,2,\ldots,r-1}, R_r^{\text{cur}}, \{R_i^{\text{cur}}\}_{i=r+1,r+2,\ldots,I}\}) = P_B^{\text{tot}}$.
30: Set $\mathbf{R}^{\text{cur}} = \{\{R_i^{\text{cur}}\}_{i=1,2,\ldots,r-1}, R_r^{\text{cur}}, \{R_i^{\text{cur}}\}_{i=r+1,r+2,\ldots,I}\}$.
31: Set $Z^{\text{cur}} = \alpha\sum_{i\in\mathcal{I}}R_i^{\text{cur}} - \beta W_B - \eta(p_B^{\min}(W_B,\mathbf{R}^{\text{cur}}) + p_0)$.
32: **if** $Z^{\text{cur}} \ge$ CBV **then**
33: Set CBV $= Z^{\text{cur}}$ and CBS $= \mathbf{R}^{\text{cur}}$.
34: **end if**
35: **end if**
36: **end if**
37: Update $r = r+1$.
38: **end while**
39: Set $\mathbf{R}^{\text{cur}} = \{R_i^{\text{req}}\}_{i\in\mathcal{I}}$.
40: Set $Z^{\text{cur}} = \alpha\sum_{i\in\mathcal{I}}R_i^{\text{cur}} - \beta W_B - \eta(p_B^{\min}(W_B,\mathbf{R}^{\text{cur}}) + p_0)$.
41: **if** $Z^{\text{cur}} \ge$ CBV and $p_B^{\min}(W_B,\mathbf{R}^{\text{cur}}) \le P_B^{\text{tot}}$ **then**
42: Set CBV $= Z^{\text{cur}}$ and CBS $= \mathbf{R}^{\text{cur}}$.
43: **end if**
44: Set $\mathbf{R}^{\text{cur}} = \{R_i^{\max}\}_{i\in\mathcal{I}}$.
45: Set $Z^{\text{cur}} = \alpha\sum_{i\in\mathcal{I}}R_i^{\text{cur}} - \beta W_B - \eta(p_B^{\min}(W_B,\mathbf{R}^{\text{cur}}) + p_0)$.
46: **if** $Z^{\text{cur}} \ge$ CBV and $p_B^{\min}(W_B,\mathbf{R}^{\text{cur}}) \le P_B^{\text{tot}}$ **then**
47: Set CBV $= Z^{\text{cur}}$ and set CBS $= \mathbf{R}^{\text{cur}}$.
48: **end if**
49: **Output**: $\{R_i^*\}_{i\in\mathcal{I}} =$ CBS, and $Z_{\eta,W_B} =$ CBV.

Based on Proposition 3, we propose Subroutine-for RA to compute Z_{η,W_B}. In Subroutine-forRA, the WHILE-LOOP from Steps 2 to 38 enumerates $r = 1,2,...,I$. Steps 9 to 24 exam Subcase-I in Proposition 3, and Steps 29 to 34 exam Subcase-II.

With Subroutine-forRA, we can find Z_{η,W_B} for each given (η, W_B). We thus continue to solve Problem (BA-Top) and present our Subroutine-forRUE (used in Step 4in BiSec-Algorithm). The difficulty in solving Problem (BA-Top) is that we cannot express Z_{η,W_B} analytically. Fortunately, Problem (BA-Top) is a single-variable optimization problem

Subroutine-forRUE: to Compute V_η and the Corresponding W_B^*

Input and Initialization: Input η. Set a small step-size Δ, and set $W_B^{\text{cur}} = \Delta$. Set CBV $= 0$ and CBS $= \emptyset$.
while $W_B^{\text{cur}} < W_B^{\text{tot}}$ **do**
  Given η and W_B^{cur}, use Subroutine-forRA to find $Z_{\eta, W_B^{\text{cur}}}$.
  if $Z_{\eta, W_B^{\text{cur}}} >$ CBV **then**
    Set CBV $= Z_{\eta, W_B}$ and CBS $= W_B^{\text{cur}}$.
  end if
  Update $W_B^{\text{cur}} = W_B^{\text{cur}} + \Delta$.
end while
Output: $W_B^* =$ CBS and $V_\eta =$ CBV.

which searches for the best W_B^* within a fixed interval $[0, W_B^{\text{tot}}]$. Thus, we adopt the linear-search method to find W_B^*. The details are shown in the following Subroutine-forRUE.

Until now we finish illustrating our Subroutine-forRUE to find V_η which is used in Step 4 in BiSec-Algorithm to solve Problem (P1-E). Thus, we finish solving the whole problem.

V. Numerical Results

We present the numerical results to validate the performance of our proposed algorithms. Specifically, we consider a scenario in which a targeted cluster of nearby MTs receive the downlink throughput from the BS. The cluster of nearby MTs are randomly distributed within a circle whose center is located at (100m,0m) and the radius is 20m. Meanwhile, the BS is located at the origin (0m,0m) and uses NOMA to transmit to the MTs simultaneously. We use the path-loss model to model the channel power gain from the BS to each MT, namely, the channel power gain from the BS to MT i is $g_{Bi} = \frac{\varrho_{Bi}}{l_{Bi}^{\kappa}}$, where l_{Bi} denotes the distance from the BS to MT i, and κ denotes the power-scaling factor for the path loss. Parameter ϱ_{Bi} follows an exponential distribution with unit mean. We set $P_B^{\text{tot}} = 3$W, $p_0 = 0.1$W. We set $R_i^{\text{req}} = 1$Mbps and $R_i^{\max} = 4$Mbps for each MT, and set $\alpha = 0.8$ and $\beta = 1$.

Figure 1 shows the advantage of our proposed joint channel bandwidth and rate allocations for improving the RUE. We compare our scheme with the fixed bandwidth allocation scheme but with the optimized $\{R_i\}_{i\in\mathcal{I}}$. For the fixed bandwidth scheme, we test three cases, namely, fixing $W_B = \frac{1}{4}W_B^{\text{tot}}, \frac{1}{2}W_B^{\text{tot}}$, and $\frac{3}{4}W_B^{\text{tot}}$. The left subplot shows the results under $W_B^{\text{tot}} = 12$MHz, and the right subplot shows the results under $W_B^{\text{tot}} = 16$MHz. All the results show that our proposed scheme can improve the REU compared with both the fixed bandwidth allocation scheme.

Figure 2 shows the advantage of our proposed joint optimization scheme for improving the RUE, in comparison with the fixed rate allocation scheme but with the optimized W_B. We test three cases, namely, fixing each MT i's rate $R_i = \frac{1}{4}(R_i^{\text{req}} + R_i^{\max}), \frac{1}{2}(R_i^{\text{req}} + R_i^{\max})$, and $\frac{3}{4}(R_i^{\text{req}} + R_i^{\max})$. The results again validate that our proposed joint optimization scheme can effectively improve the RUE, in comparison with the fixed rate allocation scheme and the FDMA scheme. This advantage essentially stems from our proposed joint optimization of the bandwidth and rate allocation for the NOMA transmission. In Figure 2, we also show the performance of the conventional frequency division multiple access (FDMA) scheme, in which the BS uses FDMA to send data to the MTs and we optimize the corresponding RUE. The results validate that our proposed NOMA scheme outperforms the conventional FDMA scheme. Notice that despite the advantage

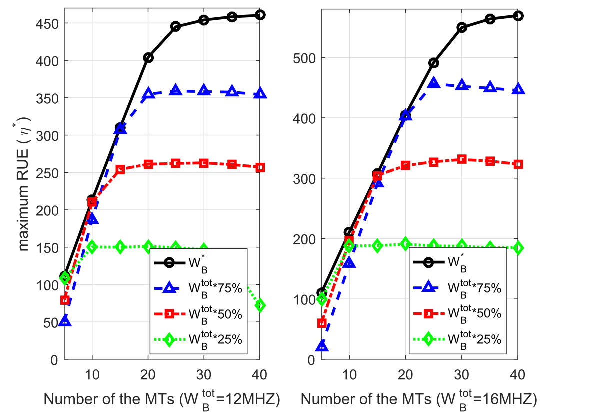

Fig. 1. Performance comparison with the fixed bandwidth scheme.

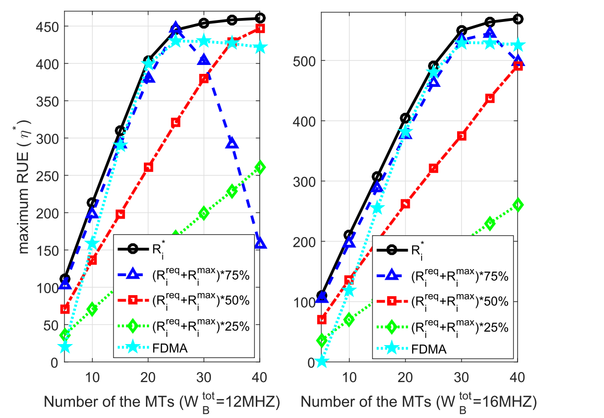

Fig. 2. Comparison with the fixed rate allocation scheme and the FDMA scheme.

of improving RUE, NOMA requires additional computational resources for the advanced signal processing (e.g., SIC) in comparison with FDMA. The computational resources, however, are not considered here.

References

[1] Y. Liu *et al.*, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[2] Y. Liu *et al.*, "Non-orthogonal multiple access in large-scale heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2667–2680, Dec. 2017.

[3] Z. Ding *et al.*, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[4] L. P. Qian *et al.*, "Joint uplink base station association and power control for small-cell networks with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5567–5582, Sep. 2017.

[5] L. Yang *et al.*, "On the impact of user scheduling on diversity and fairness in cooperative NOMA," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11296–11301, Nov. 2018.

[6] Y. Zhang *et al.*, "Energy-efficient transmission design in non-orthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2852–2857, Mar. 2017.

[7] Q. Liu *et al.*, "Energy-efficient transmission design in cooperative relaying systems using NOMA," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 594–597, Mar. 2018.

[8] M. Zeng *et al.*, "Energy-efficient power allocation for MIMO-NOMA with multiple users in a cluster," *IEEE Access*, vol. 6, pp. 5170–5181, 2018.

[9] F. Fang *et al.*, "Energy-efficient resource allocation for downlink non-orthogonal multiple access network," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3722–3732, Sep. 2016.

[10] Y. Wu *et al.*, "Optimal power allocation and scheduling for non-orthogonal multiple access relay-assisted networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 11, pp. 2591–2606, Nov. 2018.

[11] J. Wang *et al.*, "Energy-efficient joint power and bandwidth allocation for NOMA systems," *IEEE Commun. Lett.*, vol. 22, no. 4, pp. 780–783, Apr. 2018.