This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TCOMM.2022.3181673, IEEE Transactions on Communications

Computation Efficiency Optimization for RIS-Assisted Millimeter-Wave Mobile Edge Computing Systems

Xiangbin Yu, *Senior Member, IEEE*, Kai Yu, Xu Huang, Xiaoyu Dang, *Member, IEEE* Kezhi Wang, *Senior Member, IEEE*, Jiali Cai

***Abstract*—In this paper, we present the computation-efficient resource allocation (RA) schemes for millimeter-wave mobile edge computing (mmWave-MEC) system with the aid of reconfigurable intelligent surface (RIS), which is used to assist the uplink communication from the users to the base station (BS). By means of the theoretical analysis, the achievable rate and computation efficiency (CE) are derived. Then, the optimization problem for the CE maximization under the constraints of the minimum rate, maximum power consumption and local CPU frequency is formulated, where the joint design of the hybrid beamforming at the BS and the passive beamforming at the RIS as well as the local resource allocation of each user is carried out. An effective iterative algorithm based on the penalized inexact block coordinate descent (BCD) method is proposed to obtain the computation-efficient RA scheme. Next, a low-complexity suboptimal RA scheme based on the BCD method is proposed, and corresponding algorithm is presented. Simulation results show that the proposed schemes are effective, and high CE can be attained. Moreover, the second scheme can achieve the CE performance close to the first scheme but with lower complexity. Besides, it is effective to deploy the RIS scheme in mmWave-MEC system, which can strike a balance between the CE and energy consumption when compared to the conventional relay schemes.**

***Index Terms*—Millimeter-wave communication, mobile edge computing, reconfigurable intelligent surface, computation efficiency, hybrid beamforming.**

I. Introduction

MOBILE edge computing (MEC) is considered as a potential key technology in the future wireless communication, and it can improve the performance of edge network and meet the requirements of different computing services. Moreover, as one candidate of the next generation spectrum technology, millimeter-wave (mmWave) communication has the advantages of broadband and high speed, which can support various kinds of networks, e.g., Internet of Things [1]- [2]. The mmWave communication is also suitable for integrating into MEC system due to its higher capacity [3]-[4]. In [3] and [4], the delay optimization for multi-user and single-user mmWave-MEC systems were addressed, and the advantages of combining MEC and mmWave communication were presented. However, the mmWave channel is vulnerable to the blockage and subjected to high propagation loss. This issue may be addressed by deploying the reconfigurable intelligent surface (RIS) in the system. The RIS uses a number of low-cost passive reflectors to improve the system performance in wireless propagation environment, and has been regarded as a promising technology for the next generation wireless communication systems [5]–[7].

By studying the physical structure and electromagnetic properties of RIS, the authors in [8] established the free space path loss model of wireless communication assisted by RIS in different scenarios, and the experimental results verified the accuracy of the proposed model, which laid a foundation for theoretical research and practical application of RIS-assisted wireless communication. In [9], the optimization scheme for the energy efficiency of the multi-user downlink communication system based on RIS was studied, and the simulation results indicated that the energy efficiency of the proposed RIS scheme was much higher than that of the conventional multi-antenna relay scheme with amplify-and-forward protocol. By jointly optimizing the precoding matrix of the base station (BS) and the reflection coefficient of RIS, the optimization scheme of spectral efficiency in the multicast communication system was addressed in [10], and the simulation results showed the performance gain of RIS deployment in the system. In order to bypass the obstacles, the multiple RISs were introduced in [11] to improve the coverage of mmWave signals, and the received signal power can be maximized by jointly optimizing the transmit precoding of the BS and reflection coefficients of the RIS. The results verified the advantages of the proposed optimization scheme and showed that RIS can help mmWave communication improve the robustness to the blocking problems. In [12], the joint design of hybrid beamforming (BF) of BS and the reflection coefficients of RIS in multiuser mmWave communication system with RIS was addressed. The gradient projection method was used to minimize the mean square error between the received symbols and transmitted symbols. To increase the computational performance, the authors in [13] and [14] studied the application of RIS in MEC system. The impact of an RIS on the computational performance of MEC system was analyzed in [13]. By maximizing the sum computational

This work was supported in part by Natural Science Foundation of China (62031017, 61971220, 61971221). Open Research Fund of State Key Laboratory of MillimeterWaves of Southeast University under Grant K202215.

X.Yu, K.Yu, X. Huang, X. Dang and J.Cai are with College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China (e-mail:yxbxwy@gmail.com, yukai152163@163.com, 2958601767@qq.com, xydang@yahoo.com, 673696343@qq.com). K.Wang is with Department of Computer and Information Sciences, Northumbria University, United Kingdom (email: kezhi.wang@northumbria.ac.uk). X.Yu is also with the State Key Laboratory of Millimeter Waves, Southeast University, Nanjing, China.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TCOMM.2022.3181673, IEEE Transactions on Communications

bits, an iterative algorithm based on the Lagrange dual method and Karush-Kuhn-Tucker conditions and bisection method was proposed to enhance the system performance. In [14], by minimizing the computational latency of IRS-aided MEC systems, the iterative sophisticated algorithms were presented for optimizing both the computing and communications, and superior performance is attained. The results of both papers showed that the RIS can help achieve obvious performance improvement in MEC system.

Recently, the integration of MEC and other emerging mobile communication technologies has become one of research trends. However, there were few works to address the optimization design of RIS-assisted mmWave-MEC system due to the challenge in the optimization. To our best knowledge, the computation efficiency (CE) optimization for mmWave-MEC with RIS is not yet available in the literature. The CE, defined as the ratio of the computation bits (CBs) to the energy consumption [15], can be used to evaluate the efficiency of performing computation and communication per joule in MEC system. Motivated by the above reason, we study the CE optimization for RIS-assisted mmWave-MEC in this paper, and present two computationally efficient resource allocation schemes for the RIS-assisted mmWave-MEC to improve CE and ensure the user fairness. The main contributions of this paper are summarized as follows:

1) The RIS-assisted mmWave-MEC (RIS-mmWave-MEC) is presented and the RIS is used to assist the uplink communication from the users to the BS to improve the performance. With the theoretical analysis, the achievable rate and CE are firstly deduced for performance evaluation and optimization. Then, according to the max-min fairness criterion, an optimization problem for CE maximization is formulated subject to the constraints of the minimum computation-bit rate and maximum power consumption of the user and local CPU frequency. For this problem, the hybrid BF (HBF) of the BS and the passive BF (PBF) of RIS as well as the local resource allocation of each user are jointly optimized to increase the CE.

2) To tackle the optimization of CE maximization, the penalty function is firstly used to transform the original problem into the penalty form. Then, the inexact block coordinate descending (BCD) method is employed to decompose the penalty problem into several subproblems, which can be solved by means of the Riemannian manifold optimization (RMO), alternating optimization (AO), and successive convex approximation (SCA) methods. With these results, an effective iteration algorithm based on the penalty function and inexact BCD (IBCD) method is proposed to obtain the computation-efficient resource allocation scheme for the RIS-mmWave-MEC system.

3) Considering that the above scheme has high complexity, another low-complexity suboptimal scheme based on the BCD method is presented. By introducing HBF vectors, the original problem is transformed into four subproblems on power allocation (PA), PBF, CPU frequency and HBF. Firstly, given PA, PBF and CPU frequency, the closed-form HBF vectors are attained. Then given PBF, CPU frequency and HBF, closed-form PA is derived, and given CPU frequency, PBF and HBF, closed-form CPU frequency is deduced. After that, given PA, HBF and CPU frequency, using the penalty function as well as the BCD and AO methods, the suboptimal PBF is derived, and resultant closed-form PBF is obtained for each iteration. With the obtained HBF, the analog BF (ABF) and digital BF (DBF) are respectively derived by using the RMO method. Based on these results, a joint optimization algorithm with the BCD method is proposed to obtain the suboptimal PA, CPU frequency, PBF, ABF and DBF.

4) Simulation results present the effectiveness of the proposed CE optimization frameworks for the RIS-mmWave-MEC system. It is shown that the CE can be greatly increased by the proposed resource allocation schemes. Moreover, the RIS can improve the efficiency of offloading computing in mmWave communications. Furthermore, compared with the conventional relay schemes, the RIS scheme may obtain the tradeoff between the CE and energy consumption.

Notations: Vectors and matrices are respectively represented by boldface lower and upper case symbols, $\mathbf{I}_n$ is an identity matrix with n rows and n columns. $(\cdot)^*$, $(\cdot)^T$ and $(\cdot)^H$ stand for the complex conjugate, the transpose and conjugate transpose, respectively. $(\cdot)^\dagger$ denotes the pseudo inverse. $|\cdot|$ and $\|\cdot\|$ represent the absolute value and 2-norm, respectively. $\mathrm{Re}\{\cdot\}$ means taking the real part. $\odot$ denotes Hadamard product. $\mathcal{O}(\cdot)$ stands for the big-O notation. $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ denotes the complex Gaussian distribution with zero-mean and covariance matrix $\mathbf{R}$, and $\mathrm{U}[a, b]$ represents the uniform distribution in the interval $[a, b]$. $\mathrm{diag}\{\mathbf{a}\}$ is a diagonal matrix with elements of $\mathbf{a}$ on its main diagonal.

II. System model and problem formulation

A. System model

We consider a RIS-assisted mmWave-MEC system, as shown in Fig. 1, where there is K single-antenna users, a RIS including the reflecting element array and controller and an mmWave BS that connects to a high-performance MEC server. Due to low computation capacity, each user can offload the tasks to the MEC server for collaborative computation, and the RIS can be used for assisting the offloading from the users to the BS. The BS calculates the reflection coefficients of RIS, and then feed back to the RIS controller via the dedicated link. It is noted that the signals reflected by the RIS more than once are ignored. The BS adopts the fully-connected HBF architecture and is equipped with N antennas, N_{RF} radio frequency (RF) chains, N low-noise amplifiers (LNAs) and $N_{\mathrm{RF}}N$ phase shifters, where each antenna connects to all the RF chains through a low-noise amplifier (LNA) and N_{RF} phase shifters. The HBF architecture of the BS consists of the ABF matrix $\mathbf{A} \in \mathbb{C}^{N \times N_{\mathrm{RF}}}$ and the DBF matrix $\mathbf{D} \in \mathbb{C}^{N_{\mathrm{RF}} \times N_s}$, where N_s is the number of data streams. We define the sets $\mathcal{K} \triangleq \{1, .., K\}$, $\mathcal{N} \triangleq \{1, ..., N\}$ and $\mathcal{N}_{\mathrm{RF}} \triangleq \{1, .., N_{\mathrm{RF}}\}$. In particular, $\mathbf{A}$ needs to meet the constant modulus (CM) constraint, i.e., $|[\mathbf{A}]_{i,j}| = 1/\sqrt{N} (\forall i \in \mathcal{N}, j \in \mathcal{N}_{\mathrm{RF}})$. To ensure the spatial multiplexing gain of the users, it is assumed that $K = N_s \leq N_{\mathrm{RF}}$. As a result, $\mathbf{D}$ can be divided into $\mathbf{D} = [\mathbf{d}_1, .., \mathbf{d}_K]$, where $\mathbf{d}_k \in \mathbb{C}^{N_{\mathrm{RF}} \times 1} (\forall k \in \mathcal{K})$. Besides, RIS has M ideal passive reflection elements,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TCOMM.2022.3181673, IEEE Transactions on Communications

and the vector of reflection coefficients is denoted as $\boldsymbol{\theta} = [\exp(j\theta_1), ..., \exp(j\theta_M)]^T$, where the amplitude of reflection coefficients is set as one to maximize the signal reflection for simplicity [5], [11], [14], [16], [17]. Correspondingly, the PBF matrix is given by $\boldsymbol{\Theta} = \text{diag}\{\boldsymbol{\theta}\} \in \mathbb{C}^{M\times M}$, where $\theta_m \in [0, 2\pi](\forall m \in \mathcal{M})$ is the phase shift of reflection coefficients, and the set $\mathcal{M} \triangleq \{1, .., M\}$ is defined. The BS and RIS both employ a uniform linear array. It is supposed that the perfect channel state information can be available at the BS, which is typically assumed in [4], [11], [12], [16], [17].

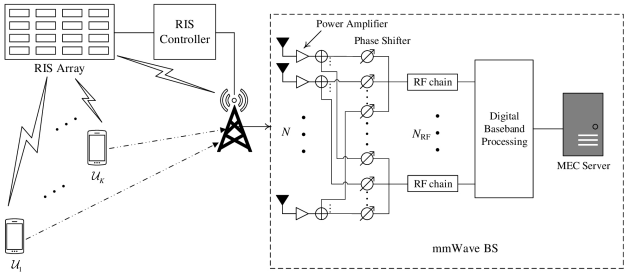

Fig. 1. RIS-assisted mmWave-MEC system.

Considering the uniform linear array, the mmWave channel $\mathbf{H} \in \mathbb{C}^{N\times M}$ between the BS and RIS can be modeled as

$$\mathbf{H} = \lambda_{\text{B}}\lambda_{\text{I}}\sqrt{\tfrac{MN}{L_B}} \sum_{l=0}^{L_B-1} \alpha_l \mathbf{a}_B(N, \psi_l^{(\text{AoA})})\mathbf{a}_I^H(M, \phi_l^{(\text{AoD})}), \tag{1}$$

where λ_{B} and λ_{I} are the receive antenna gain of the BS and the reflection gain of the RIS elements, respectively, L_B is the number of mmWave channel paths between the RIS and the BS, α_l, $\psi_l^{(\text{AoA})}$ and $\phi_l^{(\text{AoD})}$ represent the complex gain, the angle of arrival at the BS and the angle of department at the RIS of the l-th path, respectively, and $\psi_l^{(\text{AoA})} \sim \text{U}[0, 2\pi]$, $\phi_l^{(\text{AoD})} \sim \text{U}[0, 2\pi]$. $\mathbf{a}_B(N, \psi_l^{(\text{AoA})})$ and $\mathbf{a}_I(M, \phi_l^{(\text{AoD})})$ denote the array steering vector of the BS and RIS, respectively, in which $\mathbf{a}(N, \vartheta) = \frac{1}{\sqrt{N}}[1, \exp(j\pi\sin\vartheta), ..., \exp(j\pi(N-1)\sin\vartheta)]^T$, and $\vartheta \sim \text{U}[0, 2\pi]$.

The mmWave channel $\mathbf{h}_k \in \mathbb{C}^{M\times 1}$ between the k-th user $\mathcal{U}_k$ and the RIS can be modeled as

$$\mathbf{h}_k = \lambda_{\text{I}}\lambda_{\text{U}}\sqrt{\tfrac{M}{L_{k,I}}} \sum_{l=0}^{L_{k,I}-1} \beta_{k,l}\mathbf{a}_I(M, \phi_{k,l}^{(\text{AoA})}), \tag{2}$$

where λ_{U} is the transmit antenna gain. $L_{k,I}$ denotes the number of mmWave channel paths between $\mathcal{U}_k$ and the RIS, $\beta_{k,l}$ and $\phi_{k,l}^{(\text{AoA})} \sim \text{U}[0, 2\pi]$ represent the complex gain coefficient, the angle of arrival at the RIS of the l-th path, respectively. Similarly, the mmWave channel $\mathbf{h}_{d,k} \in \mathbb{C}^{N\times 1}$ between the k-th user $\mathcal{U}_k$ and the BS can be modeled as

$$\mathbf{h}_{d,k} = \lambda_{\text{B}}\lambda_{\text{U}}\sqrt{\tfrac{N}{L_{k,B}}} \sum_{l=1}^{L_{k,B}} \Delta_{k,l}\mathbf{a}_B(N, \varphi_{k,l}^{(\text{AoA})}), \tag{3}$$

where $L_{k,B}$ denotes the number of mmWave channel paths between $\mathcal{U}_k$ and the BS, $\Delta_{k,l}$ and $\varphi_{k,l}^{(\text{AoA})} \sim \text{U}[0, 2\pi]$ represent the complex gain coefficient, the angle of arrival at the BS of the l-th path, respectively. Specifically, in (1), (2) and (3), $l = 0$ denotes the light-of-sight (LOS) path, while $l > 0$ denotes the non-light-of-sight (NLOS) path.

Taking the partial offloading mode into account, the length of each frame in the system is set as T [18]. For local computing of $\mathcal{U}_k$, the number of computation bits L_k^{loc} and the energy consumption E_k^{loc} are respectively given by

$$L_k^{\text{loc}} = Tf_k/C_k,\ E_k^{\text{loc}} = T\xi_k f_k^3 + TP_{k,c}, \tag{4}$$

where C_k, f_k, ξ_k and $P_{k,c}$ denote the number of CPU cycles per bit, CPU frequency, CPU chip coefficient and the fixed circuit power consumption, respectively.

B. Problem formulation

In this subsection, we give the derivation of the achievable rate and present the CE optimization scheme. Firstly, the achievable rate for the computation offloading of $\mathcal{U}_k$ is derived. Then, the CE optimization problem is formulated. At the BS, the detected signal of $\mathcal{U}_k$ after the processing of HBF can be expressed as

$$\begin{aligned} y_k &= (\mathbf{A}\mathbf{d}_k)^H \textstyle\sum_{i=1}^K (\mathbf{H}\boldsymbol{\Theta}\mathbf{h}_i + \mathbf{h}_{d,i})\sqrt{p_i}x_i + (\mathbf{A}\mathbf{d}_k)^H\mathbf{n} \\ &= \mathbf{d}_k^H\mathbf{A}^H \textstyle\sum_{i=1}^K (\mathbf{G}_i\boldsymbol{\theta} + \mathbf{h}_{d,i})\sqrt{p_i}x_i + \mathbf{d}_k^H\mathbf{A}^H\mathbf{n}, \end{aligned} \tag{5}$$

where $x_i \sim \mathcal{CN}(0, 1)$ is the transmit signal of $\mathcal{U}_i(\forall i \in \mathcal{K})$, p_i is the transmit power, $\mathbf{G}_i = \mathbf{H}\text{diag}\{\mathbf{h}_i\} \in \mathbb{C}^{N\times M}$ is the cascaded mmWave channel, $\mathbf{n}_k \in \mathbb{C}^{N\times 1} \sim \mathcal{CN}(0, \sigma^2\mathbf{I}_N)$ is the Gaussian white noise and $\sigma^2 = n_0 B$, in which n_0 is the single-side power spectral density and B is the system bandwidth. With (5), the signal-to-interference-plus-noise-ratio (SINR) of $\mathcal{U}_k$ is derived as

$$\text{SINR}_k = \frac{p_k\left|\mathbf{d}_k^H\mathbf{A}^H(\mathbf{G}_k\boldsymbol{\theta}+\mathbf{h}_{d,k})\right|^2}{\sum_{j=1,\neq k}^K p_j\left|\mathbf{d}_k^H\mathbf{A}^H(\mathbf{G}_j\boldsymbol{\theta}+\mathbf{h}_{d,j})\right|^2+\|\mathbf{A}\mathbf{d}_k\|^2\sigma^2}. \tag{6}$$

Accordingly, the achievable rate can be written as

$$\bar{R}_k = B\log_2(1 + \text{SINR}_k). \tag{7}$$

Similar to [18] and [19], the computation time and the time of results feedback are neglected since the MEC server has a strong computation capability compared with users and the number of the bits related to the computation results is relatively smaller. Hence, for the computation offloading of $\mathcal{U}_k$, the number of computation bits L_k^{off} and the energy consumption E_k^{off} can be respectively given by

$$L_k^{\text{off}} = T\bar{R}_k, \text{and} \quad E_k^{\text{off}} = T\zeta_k p_k, \tag{8}$$

where ζ_k is the LNA coefficient.

Hence, with (4) and (8), the CE of $\mathcal{U}_k$ can be expressed as

$$\eta_k = \frac{L_k^{\text{loc}}+L_k^{\text{off}}}{E_k^{\text{loc}}+E_k^{\text{off}}} = \frac{\bar{R}_k+f_k/C_k}{\zeta_k p_k+\xi_k f_k^3+P_{k,c}}. \tag{9}$$

Substituting (6) and (7) into (9) yields

$$\eta_k = \frac{B\log_2(1+\text{SINR}_k)+f_k/C_k}{\zeta_k p_k+\xi_k f_k^3+P_{k,c}}. \tag{10}$$

In order to enhance the CE of the system and ensure the fairness of all users, the CE optimization problem based on the

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TCOMM.2022.3181673, IEEE Transactions on Communications

max-min fairness criterion for the RIS-mmWave-MEC system can be formulated as

$$\begin{aligned}(\mathrm{P0}):\ &\max_{\{\mathbf{D},\mathbf{A},\boldsymbol{\theta},p_k,f_k\}}\ \min_{k\in\mathcal{K}}\{\eta_k\}\\ \text{s.t.}\ \ &C_1:|[\mathbf{A}_{i,j}]|=1/\sqrt{N},\forall i\in\mathcal{N},j\in\mathcal{N}_{RF},\\ &C_2:|[\boldsymbol{\theta}]_m|=1,\forall m\in\mathcal{M},\\ &C_3:\bar{R}_k+f_k/C_k\geq R_k^{\min},\forall k\in\mathcal{K},\\ &C_4:\zeta_k p_k+\xi_k f_k^3+P_{k,c}\leq P_k^{\max},\forall k\in\mathcal{K},\\ &C_5:0\leq f_k\leq f_k^{\max},\forall k\in\mathcal{K},\\ &C_6:P_k^{\min}\leq p_k\leq P_k^{\max},\forall k\in\mathcal{K},\end{aligned}\tag{11}$$

where C_1 and C_2 denote the CM constraints of $\mathbf{A}$ and $\boldsymbol{\theta}$, respectively, the constraint C_3 denotes the minimum rate constraint of each user, $R_k^{\min}$ is the the minimum rate of $\mathcal{U}_k$, the constraint C_4 denotes the power consumption constraint of each user, $P_k^{\max}$ is the maximum power consumption of $\mathcal{U}_k$, the constraint C_5 denotes the local CPU frequency constraint of each user, $f_k^{\max}$ is the maximum local CPU frequency of $\mathcal{U}_k$, and the constraint C_6 denotes the transmit power constraint of each user, and $P_k^{\min}$ is the minimum power consumption of $\mathcal{U}_k$ that can take the value close to zero.

III. Resource allocation scheme for CE optimization

In this section, we develop a resource allocation scheme to improve the CE by solving the optimization problem (P0). Since the problem (P0) is non-convex and difficult to solve directly, we present an equivalent form of problem (P0) in the following **Theorem** 1.

***Theorem** 1: By introducing auxiliary variables* $\left\{\mathbf{w}_k,u_{k,\tilde{k}},q_k,\gamma_k,z_k,R_k,P_k,\eta\right\}(\forall k,\tilde{k}\in\mathcal{K})$*, the problem* (*P0*) *can be transformed equivalently as*

$$\begin{aligned}(\tilde{\mathrm{P}}0):\ &\max_{\{\mathbf{D},\mathbf{A},\boldsymbol{\theta},\mathcal{V},\eta\}}\ \eta\\ \text{s.t.}\ \ &C_1,C_2,C_5,\\ &\tilde{C}_{3,a}:q_k^{-1}|u_{k,k}|^2\geq\gamma_k^{-1},\forall k\in\mathcal{K},\\ &\tilde{C}_{3,b}:\textstyle\sum_{j\neq k}^{K}q_j^{-1}|u_{k,j}|^2+\|\mathbf{w}_k\|^2\sigma^2\leq z_k,\forall k\in\mathcal{K},\\ &\bar{C}_{3,c}:\log_2(1+\gamma_k^{-1}z_k^{-1})+\tfrac{f_k}{BC_k}\geq\tfrac{R_k}{B},\forall k\in\mathcal{K},\\ &\tilde{C}_{4,a}:\zeta_k q_k^{-1}+\xi_k f_k^3+P_{k,c}\leq P_k,\forall k\in\mathcal{K},\\ &\tilde{C}_{4,b}:P_k\leq P_k^{\max},\forall k\in\mathcal{K},\\ &\tilde{C}_6:1/P_k^{\max}\leq q_k\leq 1/P_k^{\min},\forall k\in\mathcal{K},\\ &C_7:\mathbf{w}_k=\mathbf{A}\mathbf{d}_k,\forall k\in\mathcal{K},\\ &C_8:u_{k,\tilde{k}}=\mathbf{w}_k^H(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta}+\mathbf{h}_{d,\tilde{k}}),\forall k,\tilde{k}\in\mathcal{K},\\ &C_9:R_k\geq R_k^{\min},\forall k\in\mathcal{K},\\ &C_{10}:R_k\geq\eta P_k,\forall k\in\mathcal{K},\end{aligned}\tag{12}$$

where $\mathcal{V}\triangleq\left\{\mathbf{w}_k,u_{k,\tilde{k}},q_k,f_k,\gamma_k,z_k,R_k,P_k\right\}(\forall k,\tilde{k}\in\mathcal{K})$.

Proof: Please see Appendix I.

Therefore, problem (P0) is equivalent to problem ($\tilde{\mathrm{P}}0$). To tackle the equality constraints $\{C_7,C_8\}$, we use the penalty function method to reformulate the problem ($\tilde{\mathrm{P}}0$) as (13), which is shown at the top of next page. where $\varrho^{(l-1)}>0$ is the penalty parameter at the $(l-1)$-th iteration, which is updated by $\varrho^{(l)}=\varpi\varrho^{(l-1)}(\varpi>1)$ and ϖ is the penalty increasing factor. Considering that the optimized variables in problem ($\hat{\mathrm{P}}0$) have block structure, we can decompose the problem ($\hat{\mathrm{P}}0$) into four subproblems {P1, P2, P3, P4} by using the IBCD method.

By setting the value of $\{\mathbf{D},\mathbf{A},\boldsymbol{\theta},\mathcal{V},\eta\}$ at the $(r-1)$-th iteration of IBCD method as $\{\mathbf{D}^{(r-1)},\mathbf{A}^{(r-1)},\boldsymbol{\theta}^{(r-1)},\mathcal{V}^{(r-1)},\eta^{(r-1)}\}$, where $\mathcal{V}^{(r-1)}=\left\{\mathbf{w}_k^{(r-1)},u_{k,\tilde{k}}^{(r-1)},q_k^{(r-1)},f_k^{(r-1)},\gamma_k^{(r-1)},z_k^{(r-1)},R_k^{(r-1)},P_k^{(r-1)}\right\}(\forall k,\tilde{k}\in\mathcal{K})$, we need to perform the following steps at the r-th iteration of IBCD method:

1) Solve the subproblem (P1) w.r.t $\mathbf{D}$ for the fixed $\{\mathbf{A},\boldsymbol{\theta},\mathcal{V},\eta\}$.

The subproblem w.r.t $\mathbf{D}$ is

$$(\mathrm{P1}):\min_{\mathbf{D}}\ \ \textstyle\sum_{k=1}^{K}\|\mathbf{w}_k-\mathbf{A}\mathbf{d}_k\|^2.\tag{14}$$

This problem can be divided into K independent subproblems, and the k-th ($k\in\mathcal{K}$) subproblem can be given by

$$(\mathrm{P1.1}):\min_{\mathbf{d}_k}\ \ \|\mathbf{w}_k-\mathbf{A}\mathbf{d}_k\|^2.\tag{15}$$

For the problem (P1.1), the optimal solution is $\mathbf{d}_k^{\text{opt}}=\left(\mathbf{A}^H\mathbf{A}\right)^{\dagger}\mathbf{A}^H\mathbf{w}_k$. Thus, $\mathbf{D}^{(r)}$ is updated by

$$\mathbf{D}^{(r)}=\left[\mathbf{d}_1^{\text{opt}},...,\mathbf{d}_K^{\text{opt}}\right].\tag{16}$$

2) Solve the subproblem (P2) w.r.t $\mathbf{A}$ for the fixed $\{\mathbf{D},\boldsymbol{\theta},\mathcal{V},\eta\}$.

The subproblem w.r.t $\mathbf{A}$ is

$$\begin{aligned}(\mathrm{P2}):\min_{\mathbf{A}}\ \ &\textstyle\sum_{k=1}^{K}\|\mathbf{w}_k-\mathbf{A}\mathbf{d}_k\|^2\\ \text{s.t.}\ \ &|\,[\mathbf{A}]_{i,j}\,|=\tfrac{1}{\sqrt{N}},\forall i\in\mathcal{N},j\in\mathcal{N}_{RF}.\end{aligned}\tag{17}$$

We can employ the RMO method [20] [21] to solve the problem (P2). Let $\tilde{\mathbf{A}}=\sqrt{N}\mathbf{A}$, then the problem (P2) can be changed to

$$\begin{aligned}(\mathrm{P2.1}):\min_{\tilde{\mathbf{A}}}\ \ &f(\tilde{\mathbf{A}})=\textstyle\sum_{k=1}^{K}||\mathbf{w}_k-\tilde{\mathbf{A}}\mathbf{d}_k/\sqrt{N}||^2\\ \text{s.t.}\ \ &\tilde{\mathbf{A}}\in\mathcal{R},\end{aligned}\tag{18}$$

where $\mathcal{R}\triangleq\{\tilde{\mathbf{A}}\in\mathbb{C}^{N\times N_{\mathrm{RF}}}\big||[\tilde{\mathbf{A}}]_{i,j}|=1,\forall i\in\mathcal{N},j\in\mathcal{N}_{RF}\}$.

With (P2.1), the Euclidean gradient of $f(\tilde{\mathbf{A}})$ is given by

$$\nabla_{\mathcal{E}}f(\tilde{\mathbf{A}})=\textstyle\sum_{k=1}^{K}(-2\mathbf{w}_k\mathbf{d}_k^H/\sqrt{N}+2\tilde{\mathbf{A}}\mathbf{d}_k\mathbf{d}_k^H/N).\tag{19}$$

Based on (19), according to [20], the Riemannian gradient of $f(\tilde{\mathbf{A}})$ is written as

$$\nabla_{\mathcal{R}}f(\tilde{\mathbf{A}})=\nabla_{\mathcal{E}}f(\tilde{\mathbf{A}})-\mathrm{Re}\{\nabla_{\mathcal{E}}f(\tilde{\mathbf{A}})\odot\tilde{\mathbf{A}}\}\odot\tilde{\mathbf{A}}.\tag{20}$$

At the v-th iteration of RMO algorithm, $\tilde{\mathbf{A}}$ is updated by

$$\tilde{\mathbf{A}}^{(v)}=\mathrm{Ret}(\phi^{(v-1)};\nabla_{\mathcal{R}}f(\tilde{\mathbf{A}}^{(v-1)})),\tag{21}$$

where $\tilde{\mathbf{A}}^{(v-1)}$ is the value of $\tilde{\mathbf{A}}$ at the $(v-1)$-th iteration, $\phi^{(v-1)}$ is the Armijo step size at the v-th iteration, and the function $\mathrm{Ret}(\phi;\nabla_{\mathcal{R}}f(\tilde{\mathbf{A}}))$ w.r.t $\{\phi,\tilde{\mathbf{A}}\}$ is given by

$$\mathrm{Ret}(\phi;\nabla_{\mathcal{R}}f(\tilde{\mathbf{A}}))\triangleq(\tilde{\mathbf{A}}-\phi\nabla_{\mathcal{R}}f(\tilde{\mathbf{A}}))\odot|\tilde{\mathbf{A}}-\phi\nabla_{\mathcal{R}}f(\tilde{\mathbf{A}})|^{-1}.\tag{22}$$

Therefore, the RMO algorithm for solving the subproblem (P2) is summarized as the following Algorithm 1.

3) Solve the subproblem (P3) w.r.t $\boldsymbol{\theta}$ for the fixed $\{\mathbf{A},\mathbf{D},\mathcal{V},\eta\}$.

$$(\hat{\text{P0}}): \max_{\{\mathbf{D},\mathbf{A},\boldsymbol{\theta},\mathcal{V},\eta\}} \quad \eta - \tfrac{1}{2}\varrho^{(l-1)}\left\{\sum_{k=1}^{K}\|\mathbf{w}_k - \mathbf{A}\mathbf{d}_k\|^2 + \sum_{k=1}^{K}\sum_{\tilde{k}=1}^{K}|u_{k,\tilde{k}} - \mathbf{w}_k^H(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta} + \mathbf{h}_{d,\tilde{k}})|^2\right\} \quad \text{s.t.} \quad C_1, C_2, \tilde{C}_{3,a}, \tilde{C}_{3,b}, \bar{C}_{3,c}, \tilde{C}_{4,a}, \tilde{C}_{4,b}, C_5, \tilde{C}_6, C_9, C_{10}, \tag{13}$$

Algorithm 1 RMO Algorithm for solving subproblem (P2)

1: **Initialize**: tolerance $\varepsilon > 0$, the iteration number $v = 0$, maximum iteration $v_{\max}$ and initial point $\tilde{\mathbf{A}}^{(v)} = \sqrt{N}\mathbf{A}^{(r-1)}$, $\bar{\phi} > 0$, $0 < \mu < 1$, $0 < \vartheta < 1$;
2: **repeat**
3: $v = v + 1$;
4: Compute the Armijo step size $\phi^{(v-1)} = \bar{\phi}\mu^Z$, where Z is the smallest integer satisfying $f(\tilde{\mathbf{A}}^{(v-1)}) - f\left(\text{Ret}\left(\bar{\phi}\mu^Z; \nabla_{\mathcal{R}} f(\tilde{\mathbf{A}}^{(v-1)})\right)\right) \geq -\vartheta\bar{\phi}\mu^Z \text{Re}\left\{\text{tr}\left(\nabla_{\mathcal{E}}^H f(\tilde{\mathbf{A}}^{(v-1)})\nabla_{\mathcal{R}} f(\tilde{\mathbf{A}}^{(v-1)})\right)\right\}$;
5: $\tilde{\mathbf{A}}^{(v)} = \text{Ret}\left(\phi^{(v-1)}; \nabla_{\mathcal{R}} f(\tilde{\mathbf{A}}^{(v-1)})\right)$;
6: **until** $|f(\tilde{\mathbf{A}}^{(v)}) - f(\tilde{\mathbf{A}}^{(v-1)})| < \varepsilon$ or $v > v_{\max}$;
7: **Output**: $\mathbf{A}^{(r)} = \tilde{\mathbf{A}}^{(v)}/\sqrt{N}$.

The subproblem w.r.t $\boldsymbol{\theta}$ is

$$(\text{P3}): \min_{\boldsymbol{\theta}} \quad \sum_{k=1}^{K}\sum_{\tilde{k}=1}^{K}|u_{k,\tilde{k}} - \mathbf{w}_k^H\left(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta} + \mathbf{h}_{d,\tilde{k}}\right)|^2 \quad \text{s.t.} \quad |[\boldsymbol{\theta}]_m| = 1, \forall m \in \mathcal{M}. \tag{23}$$

Let $\tilde{u}_{k,\tilde{k}} = u_{k,\tilde{k}} - \mathbf{w}_k^H\mathbf{h}_{d,\tilde{k}}$, then the subproblem can be transformed into

$$(\text{P3.1}): \min_{\boldsymbol{\theta}} \quad \sum_{k=1}^{K}\sum_{\tilde{k}=1}^{K}|\tilde{u}_{k,\tilde{k}} - \mathbf{w}_k^H\mathbf{G}_{\tilde{k}}\boldsymbol{\theta}|^2 \quad \text{s.t.} \quad |[\boldsymbol{\theta}]_m| = 1, \forall m \in \mathcal{M}. \tag{24}$$

Since $|\tilde{u}_{k,\tilde{k}} - \mathbf{w}_k^H\mathbf{G}_{\tilde{k}}\boldsymbol{\theta}|^2 = \boldsymbol{\theta}^H(\mathbf{G}_{\tilde{k}}^H\mathbf{w}_k\mathbf{w}_k^H\mathbf{G}_{\tilde{k}})\boldsymbol{\theta} - 2\text{Re}\{\boldsymbol{\theta}^H(\mathbf{G}_{\tilde{k}}^H\mathbf{w}_k\tilde{u}_{k,\tilde{k}})\} + |\tilde{u}_{k,\tilde{k}}|^2$, the problem (P3.1) can be rewritten as

$$(\text{P3.2}): \min_{\boldsymbol{\theta}} \quad \boldsymbol{\theta}^H\mathbf{V}\boldsymbol{\theta} - 2\text{Re}\left\{\boldsymbol{\theta}^H\tilde{\mathbf{v}}\right\} + \bar{v} \quad \text{s.t.} \quad |[\boldsymbol{\theta}]_m| = 1, \forall m \in \mathcal{M}, \tag{25}$$

where $\mathbf{V} = \sum_{k=1}^{K}\sum_{\tilde{k}=1}^{K}\mathbf{G}_{\tilde{k}}^H\mathbf{w}_k\mathbf{w}_k^H\mathbf{G}_{\tilde{k}}$, $\tilde{\mathbf{v}} = \sum_{k=1}^{K}\sum_{\tilde{k}=1}^{K}\mathbf{G}_{\tilde{k}}^H\mathbf{w}_k\tilde{u}_{k,\tilde{k}}$, $\bar{v} = \sum_{k=1}^{K}\sum_{\tilde{k}=1}^{K}|\tilde{u}_{k,\tilde{k}}|^2$. Considering $\mathbf{V} = \mathbf{V}^H$, we have:

$$\begin{aligned}\boldsymbol{\theta}^H\mathbf{V}\boldsymbol{\theta} &= [\mathbf{V}]_{m,m}|[\boldsymbol{\theta}]_m|^2 + \sum_{i=1,\neq m}^{M}\sum_{j=1,\neq m}^{M}[\boldsymbol{\theta}]_i^*[\mathbf{V}]_{i,j}[\boldsymbol{\theta}]_j \\ &\quad + 2\text{Re}\{[\boldsymbol{\theta}]_m^*\sum_{j=1,\neq m}^{M}[\mathbf{V}]_{m,j}[\boldsymbol{\theta}]_j\}, \\ \boldsymbol{\theta}^H\tilde{\mathbf{v}} &= [\boldsymbol{\theta}]_m^*[\tilde{\mathbf{v}}]_m + \sum_{i=1,\neq m}^{M}[\boldsymbol{\theta}]_i^*[\tilde{\mathbf{v}}]_i.\end{aligned} \tag{26}$$

Substituting (26) into (25), the problem (P3.2) can be reformulated as the problem (P3.3), which is shown at the top of next page.

With the aid of AO algorithm in [22], the problem (P3.3) can be divided into M subproblems to be solved iteratively, namely, we can solve one of the elements of $\boldsymbol{\theta}$ given that others are fixed. Based on this, the subproblems for solving $[\boldsymbol{\theta}]_m$ are shown as

$$(\text{P3.4}): \max_{[\boldsymbol{\theta}]_m} \quad \text{Re}\left\{[\boldsymbol{\theta}]_m^*([\tilde{\mathbf{v}}]_m - \sum_{j=1,\neq m}^{M}[\mathbf{V}]_{m,j}[\boldsymbol{\theta}]_j)\right\} \quad \text{s.t.} \quad |[\boldsymbol{\theta}]_m| = 1, \forall m \in \mathcal{M}. \tag{28}$$

The problem (P3.4) can be rewritten as

$$(\text{P3.5}): \max_{\theta_m} \quad \cos(\kappa_m - \theta_m) \quad \text{s.t.} \quad 0 \leq \theta_m \leq 2\pi, \tag{29}$$

where κ_m and θ_m are the phases of $[\tilde{\mathbf{v}}]_m - \sum_{j\neq m}^{M}[\mathbf{V}]_{m,j}[\boldsymbol{\theta}]_j$ and $[\boldsymbol{\theta}]_m$, respectively. Obviously, the optimal solution of (P3.5) is

$$([\boldsymbol{\theta}]_m)^{\text{opt}} = \exp(j\kappa_m). \tag{30}$$

Therefore, the AO algorithm for solving subproblem (P3) is summarized as Algorithm 2.

Algorithm 2 AO Algorithm for solving subproblem (P3)

1: **Initialize**: tolerance $\varepsilon_1 > 0$, iteration number t=0, maximum iteration $t_{\max}$ and initial point $\tilde{\boldsymbol{\theta}}^{(t)} = \tilde{\boldsymbol{\theta}}^{(r-1)}$;
2: **repeat**
3: $t = t + 1$.
4: **for** $m = 1 : M$ **do**
5: Update $[\tilde{\boldsymbol{\theta}}^{(t)}]_m$ according to (30);
6: **end for**
7: **until** $\left\|\tilde{\boldsymbol{\theta}}^{(t)} - \tilde{\boldsymbol{\theta}}^{(t-1)}\right\| < \varepsilon_1$ or $t > t_{\max}$;
8: **Output**: $\tilde{\boldsymbol{\theta}}^{(r)} = \tilde{\boldsymbol{\theta}}^{(t)}$.

4) Solve the subproblem (P4) w.r.t $\{\mathcal{V}, \eta\}$ for the fixed $\{\mathbf{A}, \mathbf{D}, \boldsymbol{\theta}\}$.

The subproblem w.r.t $\{\mathcal{V}, \eta\}$ is given by (P4), which is shown at the top of next page.

We next use the SCA method to solve the problem above. Firstly, the following two lemmas are introduced to provide suitable convex set to approximate the non-convex constraints $\tilde{C}_{3,a}, \tilde{C}_{3,c}, C_{10}$. i.e.,

Lemma 1: For a convex function $f(\mathbf{x})$ whose domain of definition is $\mathbf{x} \in \Omega$, under given $\mathbf{x_0} \in \Omega$, we have [23]:

$$f(\mathbf{x}) \geq f(\mathbf{x}_0) + \nabla f(\mathbf{x}_0)^T(\mathbf{x} - \mathbf{x}_0), \forall \mathbf{x} \in \Omega. \tag{32}$$

Lemma 2: For given $\tilde{x}, \tilde{y} > 0$, we can obtain [24]:

$$xy \leq \tfrac{\tilde{y}}{2\tilde{x}}x^2 + \tfrac{\tilde{x}}{2\tilde{y}}y^2, \forall x, y > 0. \tag{33}$$

According to **Lemma** 1, for the non-convex constraint $\tilde{C}_{3,a}$, $q_k^{-1}|u_{k,k}|^2$ is a convex function w.r.t $\{q_k, u_{k,k}\}$, thus we have:

$$\begin{aligned}q_k^{-1}|u_{k,k}|^2 \geq\ & 2\text{Re}\{(u_{k,k}^{(r-1)})^* u_{k,k}\}/q_k^{(r-1)} \\ & -|u_{k,k}^{(r-1)}|^2 q_k/(q_k^{(r-1)})^2.\end{aligned} \tag{34}$$

Thus, $\tilde{C}_{3,a}$ can be approximated as

$$\begin{aligned}\bar{C}_{3,a}: & \ 2\text{Re}\{(u_{k,k}^{(r-1)})^* u_{k,k}\}/q_k^{(r-1)} \\ & -|u_{k,k}^{(r-1)}|^2 q_k/(q_k^{(r-1)})^2 \geq \gamma_k^{-1}.\end{aligned} \tag{35}$$

$$\begin{aligned}(\text{P3.3}): \min_{\boldsymbol{\theta}} \quad & [\mathbf{V}]_{m,m} + \sum_{i\neq m}^{M}\sum_{j\neq m}^{M} [\boldsymbol{\theta}]_i^* [\mathbf{V}]_{i,j} [\boldsymbol{\theta}]_j + 2\mathrm{Re}\{[\boldsymbol{\theta}]_m^* (\sum_{j\neq m}^{M} [\mathbf{V}]_{m,j} [\boldsymbol{\theta}]_j - [\tilde{\mathbf{v}}]_m) - \sum_{i\neq m}^{M} [\boldsymbol{\theta}]_i^* [\tilde{\mathbf{v}}]_i\} \\ \text{s.t.} \quad & |[\boldsymbol{\theta}]_m| = 1, \forall m \in \mathcal{M}.\end{aligned} \tag{27}$$

$$\begin{aligned}(\text{P4}): \max_{\{\mathcal{V},\eta\}} \quad & \eta - \tfrac{1}{2}\varrho^{(l-1)} \left\{ \sum_{k=1}^{K} \|\mathbf{w}_k - \mathbf{A}\mathbf{d}_k\|^2 + \sum_{k=1}^{K}\sum_{\tilde{k}=1}^{K} \left| u_{k,\tilde{k}} - \mathbf{w}_k^H \left(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta} + \mathbf{h}_{d,\tilde{k}}\right)\right|^2 \right\} \\ \text{s.t.} \quad & \tilde{C}_{3,a}, \tilde{C}_{3,b}, \bar{C}_{3,c}, \tilde{C}_{4,a}, \tilde{C}_{4,b}, C_5, \tilde{C}_6, C_9, C_{10}.\end{aligned} \tag{31}$$

Similarly, the non-convex constraint $\bar{C}_{3,c}$ can be approximated as

$$\begin{aligned}\hat{C}_{3,c}: \log_2(1+(\gamma_k^{(r-1)} z_k^{(r-1)})^{-1}) - \frac{\gamma_k - \gamma_k^{(r-1)}}{\gamma_k^{(r-1)}\left(\gamma_k^{(r-1)} z_k^{(r-1)}+1\right)\ln 2} \\ - \frac{z_k - z_k^{(r-1)}}{z_k^{(r-1)}\left(\gamma_k^{(r-1)} z_k^{(r-1)}+1\right)\ln 2} + \frac{f_k}{BC_k} \geq \frac{R_k}{B}.\end{aligned} \tag{36}$$

Then, by applying **Lemma** 2, the non-convex constraint C_{10} can be approximated as

$$\tilde{C}_{10}: R_k \geq \frac{\eta^{(r-1)}}{2P_k^{(r-1)}} P_k^2 + \frac{P_k^{(r-1)}}{2\eta^{(r-1)}}\eta^2. \tag{37}$$

Based on the above results, the problem P4 can be transformed approximately into the convex optimization problem (P4.1), which is shown at the top of next page.

For the above convex optimization problem, we can use the standard convex optimization tools to obtain the optimal solution. Let $\{\mathcal{V}^{\mathrm{opt}}, \eta^{\mathrm{opt}}\}$ be the optimal solution of problem P4.1, then $\left\{\mathcal{V}^{(r)}, \eta^{(r)}\right\}$ can be updated by

$$\mathcal{V}^{(r)} = \mathcal{V}^{\mathrm{opt}}, \eta^{(r)} = \eta^{\mathrm{opt}}. \tag{39}$$

According to the analysis above, an effective iterative algorithm for CE optimization based on the penalty function and the IBCD method is proposed to solve the original problem $(\tilde{\text{P}}0)$, i.e, the penalized IBCD (PIBCD) algorithm. Correspondingly, the algorithm realization is summarized as Algorithm 3.

For comparison, "MaxMinCE", "MaxMinCB" and "MaxSumCE" are used to refer to maximizing the minimum CE among all users, the minimum computation bits among all users, and the sum of CE of all users, respectively. Although Algorithm 3 is designed for the MaxMinCE scheme, it can also be applied to the MaxMinCB and MaxSumCE schemes. For the MaxMinCB scheme, the optimization objective is $\max\min\{R_k\}$, where we can introduce the auxiliary variable $\tilde{R}$ (such that $\min\{R_k\} \geq \tilde{R}$) so as to apply the framework of Algorithm 3. While for the MaxSumCE scheme, the optimization objective is $\max \sum_{k=1}^{K} R_k/P_k$, where we can introduce the auxiliary variable $\tilde{\eta}$ (such that $R_k \geq \tilde{\eta} P_k$) so as to apply the framework of Algorithm 3.

Algorithm 3 PIBCD-based algorithm for solving problem $(\tilde{\text{P}}0)$

Initialize: outer tolerance $\epsilon_1 > 0$, number of outer iterations l=0, maximum outer iteration $l_{\max}$, outer convergence flag **flag** =0, inter tolerance $\epsilon_2 > 0$, number of inter iterations r, maximum inter iteration $r_{\max}$, penalty parameters $\{\varrho^{(0)} > 0, \varpi > 1\}$, constraint violation $\Xi^{(0)} = +\infty$;
repeat
  $l = l + 1$;
  initial point $\{\mathbf{D}^{(0)}, \mathbf{A}^{(0)}, \boldsymbol{\theta}^{(0)}, \mathcal{V}^{(0)}, \eta^{(0)}\}$, the value of objective function $\Gamma^{(0)} = 0$, and $r = 0$;
  repeat
    $r = r + 1$;
    Update $\mathbf{D}^{(r)}$ for fixed $\{\mathbf{A}^{(r-1)}, \boldsymbol{\theta}^{(r-1)}, \mathcal{V}^{(r-1)}, \eta^{(r-1)}\}$ according to (16);
    Update $\mathbf{A}^{(r)}$ for fixed $\{\mathbf{D}^{(r)}, \boldsymbol{\theta}^{(r-1)}, \mathcal{V}^{(r-1)}, \eta^{(r-1)}\}$ according to Algorithm 1;
    Update $\boldsymbol{\theta}^{(r)}$ for fixed $\{\mathbf{D}^{(r)}, \mathbf{A}^{(r)}, \mathcal{V}^{(r-1)}, \eta^{(r-1)}\}$ according to Algorithm 2;
    Update $\{\mathcal{V}^{(r)}, \eta^{(r)}\}$ for fixed $\{\mathbf{D}^{(r)}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r)}\}$ according to (39);
    Calculate $\Gamma^{(r)}$ according to (40);
  until $\left|\Gamma^{(r)} - \Gamma^{(r-1)}\right| < \epsilon_2$ or $r > r_{\max}$
  Calculate $\Xi^{(r)}$ according to (40);
  if $\Xi^{(r)} \leq \epsilon_1$ **then**
    flag $= 1$;
  else
    $\varrho^{(l)} = \varpi \varrho^{(l-1)}$;
  end if
until flag $= 1$ or $l > l_{\max}$
Output: A suboptimal solution $\{\mathbf{D}^{(r)}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r)}, \mathcal{V}^{(r)}, \eta^{(r)}\}$ of problem $(\tilde{\text{P}}0)$.

IV. Low-complexity suboptimal resource allocation scheme

In this section, we present the low-complexity suboptimal scheme for solving the problem (P0) based on the BCD method.

By introducing the HBF vectors $\mathbf{w}_k = \mathbf{A}\mathbf{d}_k$, the SINR in (6) can be rewritten as

$$\widetilde{\mathrm{SINR}}_k = \frac{p_k \left|\mathbf{w}_k^H(\mathbf{G}_k\boldsymbol{\theta} + \mathbf{h}_{d,k})\right|^2}{\sum_{j=1,\neq k}^{K} p_j \left|\mathbf{w}_k^H(\mathbf{G}_j\boldsymbol{\theta} + \mathbf{h}_{d,j})\right|^2 + \|\mathbf{w}_k\|^2\sigma^2}. \tag{41}$$

Based on this, the problem $(\text{P}0)$ can be transformed into

$$\begin{aligned}(\mathcal{P}0): \quad & \max_{\{\boldsymbol{\theta},\mathbf{w}_k,p_k,f_k\}} \min_{k\in\mathcal{K}} \left\{ \frac{B\log_2(1+\widetilde{\mathrm{SINR}}_k) + f_k/C_k}{\zeta_k p_k + \xi_k f_k^3 + P_{k,c}} \right\} \\ \text{s.t.} \quad & C_1, C_2, C_4, C_5, C_6 \\ & \hat{C}_3: \log_2(1+\widetilde{\mathrm{SINR}}_k) + \frac{f_k}{BC_k} \geq \frac{R_k^{\min}}{B}, \forall k \in \mathcal{K}.\end{aligned} \tag{42}$$

Thus, the original problem is divided into the problem $\mathcal{P}0$, and the subproblem w.r.t $\mathbf{A}$ and $\mathbf{d}_k$ based on the obtained $\mathbf{w}_k$ (which can be solved by Algorithm 1 and (16)). In the following, we firstly solve the problem $\mathcal{P}0$. Considering that the optimized variables $\{\mathbf{w}_k, p_k, f_k, \boldsymbol{\theta}\}$ in $\mathcal{P}0$ have block structure, we can use the BCD method to decompose the $\mathcal{P}0$

$$\begin{aligned}(\text{P4.1}): \max_{\{\mathcal{V},\eta\}} \quad & \eta - \tfrac{1}{2}\varrho^{(l-1)} \left\{ \sum_{k=1}^{K} \|\mathbf{w}_k - \mathbf{A}\mathbf{d}_k\|^2 + \sum_{k=1}^{K}\sum_{\tilde{k}=1}^{K} \left| u_{k,\tilde{k}} - \mathbf{w}_k^H \left(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta} + \mathbf{h}_{d,\tilde{k}} \right) \right|^2 \right\} \\ \text{s.t.} \quad & C_2, \bar{C}_{3,a}, \tilde{C}_{3,b}, \hat{C}_{3,c}, \tilde{C}_{4,a}, \tilde{C}_{4,b}, C_5, \tilde{C}_6, C_9, \tilde{C}_{10}. \end{aligned} \tag{38}$$

$$\begin{aligned} \Gamma^{(r)} &= \eta^{(r)} - \tfrac{1}{2}\varrho^{(l-1)} \left(\sum_{k=1}^{K} \|\mathbf{w}_k^{(r)} - \mathbf{A}^{(r)}\mathbf{d}_k^{(r)}\|^2 + \sum_{k=1}^{K}\sum_{\tilde{k}=1}^{K} |u_{k,\tilde{k}}^{(r)} - (\mathbf{w}_k^{(r)})^H (\mathbf{G}_{\tilde{k}}\boldsymbol{\theta}^{(r)} + \mathbf{h}_{d,\tilde{k}})|^2 \right), \\ \Xi^{(r)} &= \max \left\{ \max_{k\in\mathcal{K}} \left\{ \|\mathbf{w}_k^{(r)} - \mathbf{A}^{(r)}\mathbf{d}_k^{(r)}\|_\infty \right\}, \max_{k,\tilde{k}\in\mathcal{K}} \left\{ |u_{k,\tilde{k}}^{(r)} - (\mathbf{w}_k^{(r)})^H (\mathbf{G}_{\tilde{k}}\boldsymbol{\theta}^{(r)} + \mathbf{h}_{d,\tilde{k}})| \right\} \right\}. \end{aligned} \tag{40}$$

into four subproblems $\{\mathcal{P}1, \mathcal{P}2, \mathcal{P}3, \mathcal{P}4\}$ as follows, and the following steps are performed for solving the subproblems at the t-th iteration of BCD method:

1) Solve the subproblem $(\mathcal{P}1)$ w.r.t $\mathbf{w}_k$ for the fixed $\{p_k, f_k, \boldsymbol{\theta}\}$

Given $\{p_k, f_k, \boldsymbol{\theta}\}$, the problem $(\mathcal{P}1)$ w.r.t $\mathbf{w}_k$ is equivalent to

$$(\mathcal{P}1): \max_{\mathbf{w}_k} \quad J_1 = \widetilde{\text{SINR}}_k = \frac{\mathbf{w}_k^H \mathbf{R}_k \mathbf{w}_k}{\mathbf{w}_k^H \mathbf{R}_s \mathbf{w}_k} \tag{43}$$

where $\mathbf{R}_k = p_k \left(\mathbf{G}_k\boldsymbol{\theta} + \mathbf{h}_{d,k}\right)\left(\mathbf{G}_k\boldsymbol{\theta} + \mathbf{h}_{d,k}\right)^H$ and $\mathbf{R}_s = \sum_{j\neq k}^{K} p_j \left(\mathbf{G}_j\boldsymbol{\theta} + \mathbf{h}_{d,j}\right)\left(\mathbf{G}_j\boldsymbol{\theta} + \mathbf{h}_{d,j}\right)^H + \mathbf{I}_N\sigma^2$. J_1 is a generalized Rayleigh quotient, so the optimal solution of the problem $(\mathcal{P}1)$ can be given by

$$\mathbf{w}_k^{\text{o}} = \mathbf{R}_s^{-1}\sqrt{p_k}\left(\mathbf{G}_k\boldsymbol{\theta} + \mathbf{h}_{d,k}\right). \tag{44}$$

2) Solve the subproblem $(\mathcal{P}2)$ w.r.t p_k for the fixed $\{\mathbf{w}_k, f_k, \boldsymbol{\theta}\}$

Given $\{\mathbf{w}_k, f_k, \boldsymbol{\theta}\}$, the problem $(\mathcal{P}2)$ w.r.t p_k is

$$\begin{aligned}(\mathcal{P}2): \quad & \max_{\{p_k\}} \quad \min_{k\in\mathcal{K}} \left\{ \eta_k = \tfrac{B\log_2(1+p_kT_k)+f_k/C_k}{\zeta_k p_k + \xi_k f_k^3 + P_{k,c}} \right\} \\ & \text{s.t.} \;\; \hat{C}_3, C_4, C_6, \end{aligned} \tag{45}$$

where $T_k = \frac{\left|\mathbf{w}_k^H(\mathbf{G}_k\boldsymbol{\theta}+\mathbf{h}_{d,k})\right|^2}{\sum_{j\neq k}^{K} p_j \left|\mathbf{w}_k^H(\mathbf{G}_j\boldsymbol{\theta}+\mathbf{h}_{d,j})\right|^2 + \|\mathbf{w}_k\|^2\sigma^2}$. Based on the BCD method, the problem $(\mathcal{P}2)$ can be divided into K subproblems, and we can alternately solve p_k for given $p_j, \forall j \in \mathcal{K}, j \neq k$. Thus, the $(\mathcal{P}2)$ is changed to

$$\begin{aligned}(\mathcal{P}2.1): \max_{p_k} \quad & J_2 = \tfrac{B\log_2(1+p_kT_k)+f_k/C_k}{\zeta_k p_k + \xi_k f_k^3 + P_{k,c}} \\ \text{s.t.} \quad & C_4,\; C_6,\; \check{C}_3 : p_k \geq \hat{P}_k^{\min}, \end{aligned} \tag{46}$$

where $\hat{P}_k^{\min} = \frac{1}{T_k}(2^{(R_k^{\min} - f_k C_k^{-1})/B} - 1)$. Since the numerator of J_2 is a concave function on p_k, and the denominator of J_2 is an affine function, the objective J_2 is strictly pseudo-concave. Based on this, considering that the constraints are linear, the optimization problem $\mathcal{P}2.1$ is strictly pseudo-convex, and it has a global solution. By evaluating $\partial J_2/\partial p_k = 0$, we can obtain the optimal closed-form solution of p_k as

$$p_k^{\text{o}} = \tfrac{\exp\{W(be^a) - a\} - 1}{T_k}, \tag{47}$$

where $a = f_k C_k^{-1} B^{-1} \ln 2 - 1, b = T_k \zeta_k^{-1}\left(\xi_k f_k^3 + P_{k,c}\right) - 1$, $W(\cdot)$ is the Lambert W function [25]. Considering the constraints C_4 and C_6 , we have:

$$p_k^{\text{o}} = \min\{\max\{p_k^{\text{o}}, \hat{P}_k^{\min}, P_k^{\min}\}, \hat{P}_k^{\max}, P_k^{\max}\}. \tag{48}$$

where $\hat{P}_k^{\max} = \left(P_k^{\max} - P_{k,c} - \xi_k f_k^3\right)\zeta_k^{-1}$. Based on the obtained p_k, following the above calculation method, we update the PA calculations of other users $\{p_j\}$ until the PA coefficients of all users are computed.

3) Solve the subproblem $(\mathcal{P}3)$ w.r.t f_k for the fixed $\{\boldsymbol{\theta}, p_k, \mathbf{w}_k\}$

Given $\{\boldsymbol{\theta}, p_k, \mathbf{w}_k\}$, the problem $(\mathcal{P}3)$ w.r.t f_k is formulated as

$$\begin{aligned}(\mathcal{P}3): \quad & \max_{\{f_k\}} \quad \min_{k\in\mathcal{K}} \left\{ \eta_k = \tfrac{B\log_2(1+\widetilde{\text{SINR}}_k)+f_k/C_k}{\zeta_k p_k + \xi_k f_k^3 + P_{k,c}} \right\} \\ & \text{s.t.} \;\; \hat{C}_3, C_4, C_5. \end{aligned} \tag{49}$$

The above problem is equivalent to

$$\begin{aligned}(\tilde{\mathcal{P}}3): \max_{f_k} \quad & J_3 = \tfrac{B\log_2(1+\widetilde{\text{SINR}}_k)+f_k/C_k}{\zeta_k p_k + \xi_k f_k^3 + P_{k,c}} \\ \text{s.t.} \quad & \hat{C}_3,\; C_5,\; \hat{C}_4 : f_k \leq \hat{f}_k^{\max}, \end{aligned} \tag{50}$$

where $\hat{f}_k^{\max} = \sqrt[3]{\left(P_k^{\max} - P_{c,k} - \zeta_k p_k\right)/\xi_k}$ is from the constraint C_4. Since the numerator of J_3 is a linear function on f_k, and the denominator of J_2 is a convex function, the objective J_3 is strictly pseudo-concave. Based on this, considering that the constraints are linear, the optimization problem $\tilde{\mathcal{P}}3$ is strictly pseudo-convex, and it has a global solution. By calculating $\partial J_3/\partial p_k = 0$, we can obtain:

$$2\xi_k f_k^3 + 3\xi_k C_k (B\log_2(1 + \widetilde{\text{SINR}}_k)) f_k^2 - (\zeta_k p_k + P_{k,c}) = 0. \tag{51}$$

Equation (51) is a cubic equation w.r.t f_k, and it is shown that this equation has only positive real-valued solution. Hence, we can employ the Cardan formula to obtain the optimal solution of f_k, i.e., f_k^{o}. Considering the constraints $\hat{C}_3, C_5$, we have:

$$f_k^{\text{o}} = \min\{f_k^{\max}, \hat{f}_k^{\max}, \max\{f_k^{\text{o}}, f_k^{\min}, 0\}\}, \tag{52}$$

where $f_k^{\min} = C_k[R_k^{\min} - B\log_2(1 + \widetilde{\text{SINR}}_k)]$.

4) Solve the subproblem $(\mathcal{P}4)$ w.r.t $\boldsymbol{\theta}$ for the fixed $\{\mathbf{w}_k, p_k, f_k\}$

Given $\{\mathbf{w}_k, p_k, f_k\}$, the problem $(\mathcal{P}4)$ w.r.t $\boldsymbol{\theta}$ is formulated as

$$\begin{aligned}(\mathcal{P}4): \quad & \max_{\boldsymbol{\theta}} \quad \min_{k\in\mathcal{K}} \left\{ \eta_k = \tfrac{B\log_2(1+\widetilde{\text{SINR}}_k)+f_k/C_k}{P_k} \right\} \\ & \text{s.t.} \;\; C_2, \hat{C}_3, \end{aligned} \tag{53}$$

where $P_k = \zeta_k p_k + \xi_k f_k^3 + P_{k,c}$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TCOMM.2022.3181673, IEEE Transactions on Communications

By introducing the variable $\eta = \min\{\eta_k\}$ and $u_{k,\tilde{k}} = \mathbf{w}_k^H \left(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta} + \mathbf{h}_{d,\tilde{k}}\right), \forall k, \tilde{k} \in \mathcal{K}$, the problem $(\mathcal{P}4)$ is equivalently transformed into

$$\begin{aligned}(\tilde{\mathcal{P}}4): \max_{\{\boldsymbol{\theta},\eta,\{u_{k,\tilde{k}}\}\}} \quad & \eta \\ \text{s.t.} \quad & C_2, \\ & \bar{C}_3: \eta \geq \max\left\{R_k^{\min}/P_k\right\}, \forall k \in \mathcal{K}, \\ & \bar{C}_4: \eta_k = \frac{B\log_2(1+\frac{p_k\left|u_{k,k}\right|^2}{\sum_{j=1,\neq k}^K p_j\left|u_{k,j}\right|^2+\|\mathbf{w}_k\|^2\sigma^2})+f_k C_k^{-1}}{P_k} \geq \eta, \\ & C_5: u_{k,\tilde{k}} = \mathbf{w}_k^H \left(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta} + \mathbf{h}_{d,\tilde{k}}\right), \forall k, \tilde{k} \in \mathcal{K}.\end{aligned} \tag{54}$$

Then we use the penalty function method to deal with the constraint C_5, and correspondingly, the problem $(\tilde{\mathcal{P}}4)$ is decomposed into the subproblem $\hat{\mathcal{P}}4$, which is shown at the top of next page, and the subproblem w.r.t $\boldsymbol{\theta}$ based on obtained $\{u_{k,\tilde{k}}\}$, where $\psi_k(\eta) = 2^{(P_k\eta - f_k C_k^{-1})/B} - 1$, $\sigma_k^2 = \|\mathbf{w}_k\|^2\sigma^2$, and $[\mathbf{U}]_{k,\tilde{k}} = u_{k,\tilde{k}}$. μ_k is the Lagrange multiplier, the value of $\boldsymbol{\theta}$ is from the previous iteration, and $\rho^{(l-1)} > 0$ is the penalty parameter at the $(l-1)$-th iteration.

Considering that the optimized variables in $(\hat{\mathcal{P}}4)$ have block structure, we employ the BCD method to solve the problem $(\hat{\mathcal{P}}4)$, that is, optimize $\mathbf{U}$ for given η and optimize η for given $\mathbf{U}$.

Firstly, given η, we evaluate $\mathbf{U}$. With (55), we can calculate $\partial J_4/\partial u_{k,k}^*$ and $\partial J_4/\partial u_{k,j}^*$ as

$$\begin{aligned}\frac{\partial J}{\partial u_{k,k}^*} &= -\tfrac{1}{2}\rho^{(l-1)}u_{k,k} + \tfrac{1}{2}\rho^{(l-1)}\mathbf{w}_k^H\left(\mathbf{G}_k\boldsymbol{\theta} + \mathbf{h}_{d,k}\right) \\ &\quad + \mu_k p_k u_{k,k}, \\ \frac{\partial J}{\partial u_{k,j}^*} &= -\tfrac{1}{2}\rho^{(l-1)}u_{k,j} + \tfrac{1}{2}\rho^{(l-1)}\mathbf{w}_k^H\left(\mathbf{G}_j\boldsymbol{\theta} + \mathbf{h}_{d,j}\right) \\ &\quad - \psi_k(\eta)\mu_k p_j u_{k,j}.\end{aligned} \tag{56}$$

With (56), we can obtain that $\frac{\partial^2 J_4}{\partial u_{k,k}^* \partial u_{k,k}}$ and $\frac{\partial^2 J_4}{\partial u_{k,j}^* \partial u_{k,j}}$ are negative, and $\frac{\partial^2 J_4}{\partial u_{k,k}^* \partial u_{k,j}}$ and $\frac{\partial^2 J_4}{\partial u_{k,j}^* \partial u_{k,k}}$ $(k \neq j)$ are zero. Hence, given η, the optimization problem w.r.t $\{u_{k,\tilde{k}}\}$ is convex, and thus we can obtain the optimal solution of $u_{k,k}$ and $u_{k,j}$ by setting $\partial J_4/\partial u_{k,k}^*$=0 and $\partial J_4/\partial u_{k,j}^*$=0, i.e.,

$$u_{k,k}^{\mathrm{o}} = \frac{\frac{1}{2}\rho^{(l-1)}\mathbf{w}_k^H(\mathbf{G}_k\boldsymbol{\theta}+\mathbf{h}_{d,k})}{\frac{1}{2}\rho^{(l-1)}-\mu_k p_k}, \quad u_{k,j}^{\mathrm{o}} = \frac{\frac{1}{2}\rho^{(l-1)}\mathbf{w}_k^H(\mathbf{G}_j\boldsymbol{\theta}+\mathbf{h}_{d,j})}{\frac{1}{2}\rho^{(l-1)}+\psi_k(\eta)\mu_k p_j}. \tag{57}$$

Next, we solve η given $\mathbf{U}$. Under this case, using (55), we can calculate $\partial J_4/\partial\eta$ as

$$\frac{\partial J_4}{\partial \eta} = 1 - \ln 2 \sum_{k=1}^K \mu_k 2^{\frac{P_k\eta - f_k C_k^{-1}}{B}} \frac{P_k}{B}\left(\sum_{j\neq k}^K p_j\left|u_{k,j}\right|^2 + \sigma_k^2\right). \tag{58}$$

With (58), it is shown that $\partial^2 J_4/\partial\eta^2$ is negative. Thus, we can obtain that the optimal solution η^o by solving $\partial J_4/\partial\eta = 0$, which can be realized by the bisection method. Considering the constraints $\bar{C}_3$ and $\bar{C}_4$, η^{o} is updated as

$$\eta^{\mathrm{o}} = \min\left\{\max\left\{\eta^o, \eta^{(lo)}\right\}, \eta^{(up)}\right\}. \tag{59}$$

where $\eta^{(lo)}$=$\max\{R_{\min,k}/P_k\}$ is from $\bar{C}_3$, and $\eta^{(up)}$=$\min\left\{\frac{B\log_2(1+\widetilde{\mathrm{SINR}}_k)+f_k C_k^{-1}}{P_k}\right\}$ is from $\bar{C}_4$.

With the obtained η and $\mathbf{U}$, the multiplier μ_k is updated by the sub-gradient method [23] as

$$\mu_k^{(q+1)} = \left[\mu_k^{(q)} - v^{(q)}(p_k\left|u_{k,k}\right|^2 - \psi(\eta)(\sum_{j\neq k}^K p_j|u_{k,j}|^2 + \sigma_k^2))\right]^+, \tag{60}$$

where $\mu_k^{(q)}$ is q-th iteration of μ_k, and $v^{(q)}$ is small positive step size for the q-th iteration.

Based on the obtained $\mathbf{U}$, we can solve the $\boldsymbol{\theta}$, and corresponding optimization problem is

$$\begin{aligned}(\mathcal{P}4.2): \min_{\boldsymbol{\theta}} \quad & \textstyle\sum_{k=1}^K\sum_{\tilde{k}=1}^K |u_{k,\tilde{k}} - \mathbf{w}_k^H\left(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta} + \mathbf{h}_{d,\tilde{k}}\right)|^2 \\ \text{s.t.} \quad & C_2.\end{aligned} \tag{61}$$

It is the same as problem (P3), so that problem $(\mathcal{P}4.2)$ can be solved by Algorithm 2, and resultant optimized $\boldsymbol{\theta}$ is attained. Based on the analysis above, the algorithm for solving the problem $(\mathcal{P}4)$ can be summarized as Algorithm 4.

Algorithm 4 Penalty BCD-based algorithm for solving problem $(\mathcal{P}4)$

1: **Initialize**: tolerance ϵ_3, ϵ_4, the iteration number l, q=0, the maximum iteration $l_{\max}, q_{\max}$, the convergence flag **flag** =0 and the penalty parameters $\{\varrho^{(0)} > 0, \varpi > 1\}$;
2: **repeat**
3: $l = l + 1$;
4: initial point $\{\mathbf{U}^{(0)}, \eta^{(0)}\}$ and $q = 0$;
5: **repeat**
6: $q = q + 1$;
7: Update $\mathbf{U}^{(q)}$ for fixed $\eta^{(q-1)}$ according to (57);
8: Update $\eta^{(q)}$ for fixed $\mathbf{U}^{(q)}$ according to (58) and (59);
9: Update $\mu_k{}^{(q)}$ according to (60);
10: **until** $|\mu_k^{(q)} - \mu_k^{(q-1)}| < \epsilon_4$ or $q > q_{\max}$
11: Calculate $\Xi_1^{(q)} = \max_{k,\tilde{k}\in\mathcal{K}}\{|u_{k,\tilde{k}}^{(q)} - (\mathbf{w}_k^{(t)})^H(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta} + \mathbf{h}_{d,\tilde{k}})|\}$;
12: **if** $\Xi_1^{(q)} \leq \epsilon_3$ **then**
13: **flag** $= 1$;
14: **else**
15: $\varrho^{(l)} = \varpi\varrho^{(l-1)}$;
16: **end if**
17: **until flag** $= 1$ or $l > l_{\max}$
18: Update $\boldsymbol{\theta}$ for the obtained $\{\mathbf{U}^{(q)}\}$ according to Algorithm 2;
19: **Output**: A suboptimal solution $\boldsymbol{\theta}$ of problem $(\mathcal{P}4)$.

In conclusion, the above procedure for obtaining the joint resource allocations is summarized in Algorithm 5. The Algorithm 5 can achieve the CE close to that of the Algorithm 3, but it has lower complexity than the latter since the standard optimization tool is not used.

V. Convergence and Complexity Analysis

In this section, we firstly analyze the convergence behaviour of Algorithm 3 and Algorithm 5, and then give the complexity analysis of these two algorithms.

For Algorithm 3, considering $\eta \leq \min_{k\in\mathcal{K}}\{R_k/P_k\}$ and limited power, the objective function of the problem $(\tilde{\mathrm{P}}0)$ is upper-

$$
\begin{aligned}
(\hat{\mathcal{P}}4): \max_{\{\mathbf{U},\eta\}} \quad & J_4 = \eta - \tfrac{1}{2}\rho^{(l-1)} \left(\sum_{k=1}^{K} \sum_{j=1}^{K} \left| u_{k,j} - \mathbf{w}_k^H \left(\mathbf{G}_j \boldsymbol{\theta} + \mathbf{h}_{d,j} \right) \right|^2 \right) \\
& + \sum_{k=1}^{K} \mu_k \left\{ p_k \left| u_{k,k} \right|^2 - \psi_k(\eta) \left(\sum_{j \neq k}^{K} p_j \left| u_{k,j} \right|^2 + \sigma_k^2 \right) \right\} \\
\text{s.t.} \quad & \bar{C}_3, \bar{C}_4
\end{aligned} \tag{55}
$$

Algorithm 5 BCD-based algorithm for solving problem $(\mathcal{P}0)$

1: **Initialize**: tolerance ϵ_5, ϵ_6, initial point $\{f_k^{(0)}, p_k^{(0)}, \mathbf{A}^{(0)}, \mathbf{D}^{(0)}, \boldsymbol{\theta}^{(0)}\}$, the value of objective function $\Gamma_2^{(0)} = 0$, and the iteration number $t, t^{'}$=0, the maximum iteration $t_{\max}, t_{\max}^{'}$.
2: **repeat**
3: $t = t + 1$;
4: Update $\{\mathbf{w}_k^{(t)}\}$ for fixed $\{f_k^{(t-1)}, p_k^{(t-1)}, \boldsymbol{\theta}^{(t-1)}, k \in \mathcal{K}\}$ according to (44);
5: Update $\{p_k^{(t)}\}$ for fixed $\{\mathbf{w}_k^{(t)}, f_k^{(t)}, \boldsymbol{\theta}^{(t-1)}, k \in \mathcal{K}\}$ according to (48);
6: Update $\{f_k^{(t)}\}$ for fixed $\{\mathbf{w}_k^{(t)}, p_k^{(t-1)}, \boldsymbol{\theta}^{(t-1)}, k \in \mathcal{K}\}$ according to (52);
7: Update $\boldsymbol{\theta}^{(t)}$ for fixed $\{\mathbf{w}_k^{(t)}, f_k^{(t)}, p_k^{(t)}\}$ according to Algorithm 4;
8: Calculate $\Gamma_2^{(t)} = \frac{B \log_2(1+\widetilde{\mathrm{SINR}}_k^{(t)}) + f_k^{(t)}/C_k}{\zeta_k p_k^{(t)} + \xi_k \left(f_k^{(t)}\right)^3 + P_{k,c}}$;
9: **until** $|\Gamma_2^{(t)} - \Gamma_2^{(t-1)}| < \epsilon_5$ or $t > t_{\max}$
10: **repeat**
11: $t^{'} = t^{'} + 1$;
12: Update $\mathbf{A}^{(t^{'})}$ for fixed $\mathbf{D}^{(t^{'}-1)}$ according to Algorithm 1;
13: Update $\mathbf{D}^{(t^{'})}$ for fixed $\mathbf{A}^{(t^{'})}$ according to (16);
14: **until** $\max_{k \in \mathcal{K}} \{\|\mathbf{w}_k^{(t)} - \mathbf{A}^{(t^{'})} \mathbf{d}_k^{(t^{'})}\|_\infty\} < \epsilon_6$ or $t^{'} > t_{\max}^{'}$
15: **Output**: A suboptimal solution $\{\boldsymbol{\theta}^{(t)}, f_k^{(t)}, p_k^{(t)}, \mathbf{A}^{(t^{'})}, \mathbf{D}^{(t^{'})}, \mathbf{w}_k^{(t)}\}$ of problem $(\mathcal{P}0)$.

bounded. Moreover, for given $\varrho^{(l-1)}$, it can be concluded that

$$
\begin{aligned}
\Upsilon_1: \quad & f\left(\mathbf{D}^{(r)}, \mathbf{A}^{(r-1)}, \boldsymbol{\theta}^{(r-1)}, \mathcal{V}^{(r-1)}, \eta^{(r-1)}\right) \\
& \geq f\left(\mathbf{D}^{(r-1)}, \mathbf{A}^{(r-1)}, \boldsymbol{\theta}^{(r-1)}, \mathcal{V}^{(r-1)}, \eta^{(r-1)}\right), \\
\Upsilon_2: \quad & f\left(\mathbf{D}^{(r)}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r-1)}, \mathcal{V}^{(r-1)}, \eta^{(r-1)}\right) \\
& \geq f\left(\mathbf{D}^{(r)}, \mathbf{A}^{(r-1)}, \boldsymbol{\theta}^{(r-1)}, \mathcal{V}^{(r-1)}, \eta^{(r-1)}\right), \\
\Upsilon_3: \quad & f\left(\mathbf{D}^{(r)}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r)}, \mathcal{V}^{(r-1)}, \eta^{(r-1)}\right) \\
& \geq f\left(\mathbf{D}^{(r)}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r-1)}, \mathcal{V}^{(r-1)}, \eta^{(r-1)}\right), \\
\Upsilon_4: \quad & f\left(\mathbf{D}^{(r)}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r)}, \mathcal{V}^{(r)}, \eta^{(r)}\right) \\
& \geq f\left(\mathbf{D}^{(r)}, \mathbf{A}^{(r)}, \boldsymbol{\theta}^{(r)}, \mathcal{V}^{(r-1)}, \eta^{(r-1)}\right),
\end{aligned} \tag{62}
$$

where Υ_1 holds because (16) is the optimal solution of the subproblem (P1), Υ_2 holds for the convergence of the RMO algorithm [21], Υ_3 holds for the convergence of the AO algorithm [22], and Υ_4 holds for the convergence of the SCA algorithm. If the feasible region of the problem $(\tilde{\mathrm{P}}0)$ is not empty, Algorithm 3 can converge to a local optimal solution of the problem $(\tilde{\mathrm{P}}0)$.

In what follows, the complexity of Algorithm 3 is analyzed. For Algorithm 3, its complexity mainly comes from Steps 7 to 10. In Step 7, the complexity of the inverse calculation of $\mathbf{A}^H \mathbf{A}$ is $\mathcal{O}\left(N_{\mathrm{RF}}^3\right)$. In Step 8, the complexity of Algorithm 1 is $\mathcal{O}\left(I_1 K N N_{\mathrm{RF}}\right)$, where I_1 is the iterative number of Algorithm 1. In Step 9, the complexity of Algorithm 2 is $\mathcal{O}\left(I_2 M\right)$, where I_2 is the iterative number of Algorithm 2. In Step 10, the standard convex optimization tools based on the interior point method (e.g. CVX in [26]) is employed to solve the problem (P4.1). Because the number of real optimization variables in (P4.1) is $2NK + 2K^2 + 6K + 1$, the complexity of the interior point method can be approximated as $\mathcal{O}((2NK + 2K^2 + 6K + 1)^{3.5} \ln(1/\delta))$, where δ is the solution accuracy. Therefore, the complexity of Algorithm 3 is approximated as $\mathcal{O}(I_3 I_4 (K N_{\mathrm{RF}}^3 + I_1 K N N_{\mathrm{RF}} + I_2 M + (2NK + 2K^2 + 6K + 1)^{3.5} \ln(1/\delta)))$, where I_3 and I_4 are the numbers of outer and inter iterations of Algorithm 3, respectively.

For Algorithm 5, it can be concluded in the t-th iteration that

$$
\begin{aligned}
\tilde{\Upsilon}_1: \quad & f\left(\mathbf{W}^{(t)}, \mathbf{p}^{(t-1)}, \mathbf{f}^{(t-1)}, \boldsymbol{\theta}^{(t-1)}\right) \\
& \geq f\left(\mathbf{W}^{(t-1)}, \mathbf{p}^{(t-1)}, \mathbf{f}^{(t-1)}, \boldsymbol{\theta}^{(t-1)}\right), \\
\tilde{\Upsilon}_2: \quad & f\left(\mathbf{W}^{(t)}, \mathbf{p}^{(t)}, \mathbf{f}^{(t-1)}, \boldsymbol{\theta}^{(t-1)}\right) \\
& \geq f\left(\mathbf{W}^{(t)}, \mathbf{p}^{(t-1)}, \mathbf{f}^{(t-1)}, \boldsymbol{\theta}^{(t-1)}\right), \\
\tilde{\Upsilon}_3: \quad & f\left(\mathbf{W}^{(t)}, \mathbf{p}^{(t)}, \mathbf{f}^{(t)}, \boldsymbol{\theta}^{(t-1)}\right) \\
& \geq f\left(\mathbf{W}^{(t)}, \mathbf{p}^{(t)}, \mathbf{f}^{(t-1)}, \boldsymbol{\theta}^{(t-1)}\right), \\
\tilde{\Upsilon}_4: \quad & f\left(\mathbf{W}^{(t)}, \mathbf{p}^{(t)}, \mathbf{f}^{(t)}, \boldsymbol{\theta}^{(t)}\right) \\
& \geq f\left(\mathbf{W}^{(t)}, \mathbf{p}^{(t)}, \mathbf{f}^{(t)}, \boldsymbol{\theta}^{(t-1)}\right),
\end{aligned} \tag{63}
$$

where $\mathbf{p} = [p_1, ..., p_K]^T$, $\mathbf{f} = [f_1, ..., f_K]^T$, and $\mathbf{W} = [\mathbf{w}_1, ..., \mathbf{w}_K]$. $\tilde{\Upsilon}_1, \tilde{\Upsilon}_2$ and $\tilde{\Upsilon}_3$ hold because (44), (48) and (52) are the optimal solutions of the subproblems. $\tilde{\Upsilon}_4$ holds because the convergence of Algorithm 4 and Algorithm 2, whose convergence analysis is shown as follows. Given $\varrho^{(l-1)}$, it can be concluded that

$$
\begin{aligned}
\bar{\Upsilon}_1: \quad & f\left(\mathbf{U}^{(q)}, \eta^{(q-1)}\right) \geq f\left(\mathbf{U}^{(q-1)}, \eta^{(q-1)}\right), \\
\bar{\Upsilon}_2: \quad & f\left(\mathbf{U}^{(q)}, \eta^{(q)}\right) \geq f\left(\mathbf{U}^{(q)}, \eta^{(q-1)}\right),
\end{aligned} \tag{64}
$$

where $\bar{\Upsilon}_1$ and $\bar{\Upsilon}_2$ hold because (57) and (59) are the optimal solutions.

The above results show that the objective value is increasing after each iteration. Moreover, because of the limited power supply, the rate is also limited. Thus, the objective value is upper-bounded. Hence, the Algorithm 5 can be guaranteed to converge.

Regarding the Algorithm 5, the calculation burden mainly comes from Steps 4-7 and 12-13. In Step 4 of Algorithm 5, the complexity of the inverse calculation of $\mathbf{R}_s$ is $\mathcal{O}\left(N^3\right)$. In Step 5, the complexity is $\mathcal{O}\left(K\right)$. In Step 6, the complexity is $\mathcal{O}\left(K\right)$. In Step 7, Algorithm 4 is used. For Algorithm 4, its complexity involves the penalty BCD and Algorithm 2. The complexity of Algorithm 2 is $\mathcal{O}(M \tilde{I}_1)$, where $\tilde{I}_1$ is the iterative number of Algorithm 2. The penalty BCD has two loops, including the inner BCD iteration $\tilde{I}_2$ and outer iteration of the penalty function method $\tilde{I}_3$. Correspondingly, its complexity is $\mathcal{O}\left(M \tilde{I}_1 + (K^2 + K) \tilde{I}_2 \tilde{I}_3\right)$. Besides, the complexity for solving $\{\mathbf{A}, \mathbf{D}\}$ is $\mathcal{O}\left(\tilde{I}_4 K N N_{\mathrm{RF}} + K N_{\mathrm{RF}}^3\right)$,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TCOMM.2022.3181673, IEEE Transactions on Communications

where $\tilde{I}_4$ is the iterative number of Algorithm 1. Based on the above analysis, the complexity of Algorithm 5 is approximated as $\mathcal{O}((KN^3+2K+M\tilde{I}_1+(K^2+K)\tilde{I}_2\tilde{I}_3)\tilde{I}_5+(\tilde{I}_4KNN_{\mathrm{RF}}+KN_{\mathrm{RF}}^3)\tilde{I}_6)$, where $\tilde{I}_5$ and $\tilde{I}_6$ are the outer BCD iteration of Steps 4-7 and BCD iteration of Steps 12-13, respectively.

VI. Simulation Results

In this section, we evaluate the CE performance of the proposed resource allocation schemes for RIS-mmWave-MEC system by simulations. For the simulation setup, it is assumed that the BS and RIS are located at (0m, 0m) and (80m, 5m), respectively, and all users are uniformly distributed in the circular area with (150m, 0m) as the center and the radius of 5m. Correspondingly, the central distance of users to the BS d_c=150m. Besides, $\lambda_{\mathrm{U}} = 0$ dBi and $\lambda_{\mathrm{B}} = 9.82$ dBi, and the relative gain of the RIS is $\nu = \frac{\lambda_{\mathrm{I}}}{\sqrt{\lambda_{\mathrm{B}}\lambda_{\mathrm{U}}}} = 10$ dB [27] [17]. The carrier frequency is 28 GHz, $L_B = L_{k,I} = L_{k,B} = 4(\forall k \in \mathcal{K})$, and $n_0 = -174$ dBm. Based on the [28] and [11], the complex gains α_l follows a complex Gaussian distribution, i.e., $\alpha_l \sim \mathcal{CN}(0, 10^{-0.1\mu_l(d)})$ and so are $\beta_{k,l}$ and $\Delta_{k,l}$, where $\mu_l(d) = a+10b\log_{10}(d)+\xi$ dB, d is the distance in meters between the transmitter and receiver and defined as $\sqrt{(x_t-x_r)^2+(y_t-y_r)^2}$, where (x_t, y_t) and (x_r, y_r) are the locations of transmitter and receiver, respectively. The values of a, b and ξ are set to as $a = 61.4, b = 2, \xi = 5.8$ dB for LOS path and $a = 72, b = 2.92, \xi = 8.7$ dB for NLOS path [28]. Unless otherwise specified, other main parameters are listed in Table I, which are based on [15], [18], [29]. Simulation results are shown in Figs. 2-11, where 'MaxMinCE', 'MaxMinCB', and 'MaxSumCE' are defined in Section III, and the proposed two suboptimal schemes are referred as 'sub-scheme1' and 'sub-scheme2', respectively.

TABLE I
Simulation parameters for RIS-assisted mmWave-MEC

Parameter	Value	Parameter	Value
K	3	M	100
N	16	ζ_k	1/0.38
N_{RF}	3	ξ_k	10^{-28}
B	2 MHz	C_k	10^3 cycles/bit
$P_k^{\max}$	$P_{\max}$	$R_k^{\min}$	10^4 bits/s
$P_{k,c}$	50 mW	$f_k^{\max}$	1 GHz

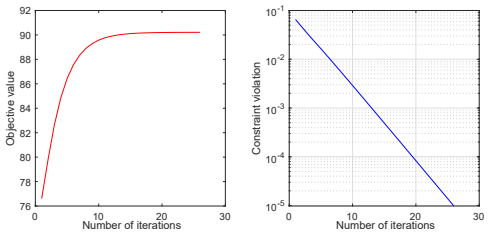

Fig. 2. The convergence of Algorithm 3.

Fig. 2 shows the convergence behaviour of Algorithm 3, where Fig. 2 (a) and (b) respectively illustrate the iterative process of the objective function value and constraint violation $\Xi^{(l)}$ of problem $(\tilde{\mathrm{P}}0)$ in Algorithm 3 under the partial offloading mode, where $P_{\max} = 0.052$W. As it can be seen from Fig. 2, on the one hand, the objective function of the problem $(\tilde{\mathrm{P}}0)$ increases gradually to a stable value with the increase of the number of iterations. On the other hand, after some iterations, the constraint violation $\Xi^{(l)}$ tends to be smaller and finally decreases to a predefined acceptable level, 10^{-5}, which means that the solution obtained by this algorithm is a feasible solution of the problem $(\tilde{\mathrm{P}}0)$. The above results verify the convergence of Algorithm 3.

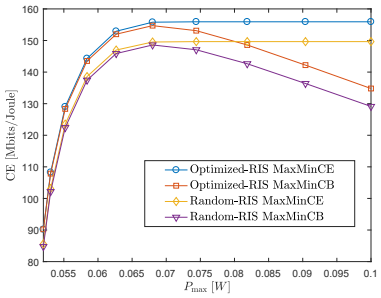

Fig. 3. CE of the system with different RIS configurations under the partial offloading mode.

Fig. 3 gives the CE comparison of the system with the sub-scheme1 for two different RIS configurations (i.e., Optimized-RIS and Random-RIS) under the partial offloading mode, where MaxMinCE and MaxMinCB schemes are considered, and Optimized-RIS and Random-RIS mean that RIS employs the reflection phase vector obtained by Algorithm 3 and RIS uses the random reflection phase vector, respectively. From Fig. 3, it is observed that the MaxMinCE and MaxMinCB schemes have almost the same CE when $P_{\max}$ is small, and increase as $P_{\max}$ increases. However, when $P_{\max}$ becomes large, the CE of the MaxMinCB scheme start to decrease. This is because in the MaxMinCB scheme, the number of computation bits increases as $P_{\max}$ increases, but the power consumption increases even more. As a result, the CE keeps decreasing. In contrast, the CE of the MaxMinCE scheme tends to be stable when $P_{\max}$ is large. Besides, it is obvious that the system with Optimized-RIS can achieve higher CE than that with Random-RIS since the former employs the optimized reflection phase vector, which is attained by optimizing the CE, while the reflection phase vector in the latter is randomly produced.

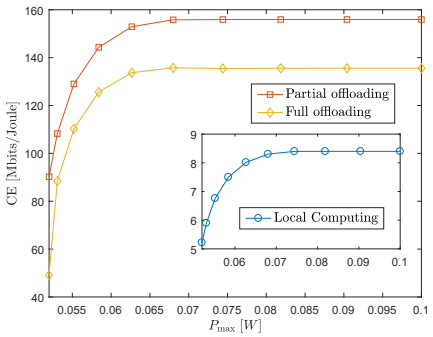

Fig. 4. CE of the system with different computing modes.

Fig. 4 illustrates the CE of the system with the sub-scheme1

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TCOMM.2022.3181673, IEEE Transactions on Communications

for different computing modes, where the local computing, the partial offloading and the full offloading modes are compared. The optimal resource allocation scheme for the local computing mode can be found from [30], while the resource allocation schemes for the partial offloading and the full offloading modes are attained by using the framework of Algorithm 3. As shown in Fig. 4, the system with partial offloading mode has the highest CE, and the CE of the system with full offloading mode is obviously higher than that with local computing mode. These results above indicate that RIS can be well applied to the mmWave-MEC system to improve the efficiency of offloading computing, and help to solve the propagation loss problem in mmWave communications.

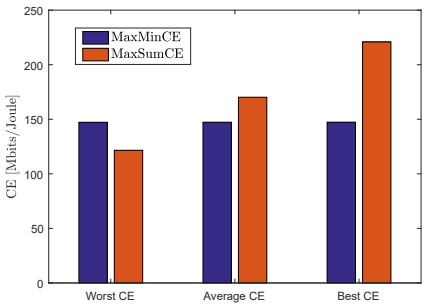

Fig. 5. User fairness comparison of MaxMinCE scheme and MaxSumCE scheme under partial offloading mode.

Fig. 5 displays the user fairness of MaxMinCE and MaxSumCE schemes under partial offloading mode, where "Worst CE", "Average CE" and "Best CE" refer to the worst CE, average CE and best CE of all users respectively, and $P_{\max} = 0.06$W. It is found that the advantage of MaxMinCE scheme is that it can guarantee the CE of the worst user, but it may sacrifice the CE of other users. In contrast, MaxSumCE scheme can improve the overall CE of all users, especially the Best CE, but it can not guarantee the CE of the worst user. The above results reflect the advantage and disadvantage of MaxMinCE scheme and MaxSumCE scheme in the aspect of user fairness. Hence, how to choose these two schemes depend on the actual situation and user fairness requirements.

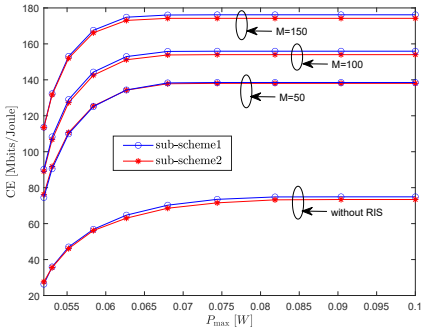

Fig. 6. CE of the system with different numbers of RIS reflection elements under partial offloading mode.

Fig. 6 gives the CE of the system with different numbers of RIS reflection elements under partial offloading mode, where the sub-scheme1 and sub-scheme2 are compared, and $M \in \{0, 50, 100, 150\}$. From Fig. 6, it is found that the sub-scheme2 can achieve the CE performance close to that of the sub-scheme1, but the former has lower complexity than the latter because the latter needs to use the complicated CVX, which can also be seen from the complexity analysis in section V and iteration comparison in the following figures, that is, the former only needs less iteration. Besides, the CE is increased as the number of reflection elements increases. Specifically, the system with M=150 has higher CE than that with M=100, and the system with M=100 has higher CE than that with M=50. This is because the design of RIS reflection phase provides the PBF gain of RIS, and increasing the number of RIS reflection elements can bring higher PBF gain. Furthermore, the system without RIS has the worst performance, and its CE is obviously lower than that with RIS, which shows that RIS can effectively improve the CE.

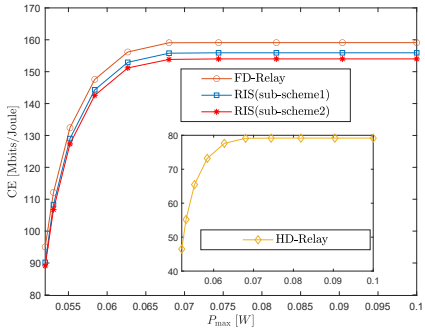

Fig. 7. CE comparison of mmWave-MEC systems with RIS and conventional relay schemes.

In Fig. 7, we compare the CE of mmWave-MEC systems with RIS scheme and and conventional relay schemes (i.e., FD-Relay and HD-Relay), where FD-Relay and HD-Relay refer to the ideal full-duplex amplify-and-forward (AF) relay and half-duplex AF relay, respectively. For RIS scheme, sub-scheme1 and sub-scheme2 are considered. Besides, for FD-Relay and HD-Relay, the number of antennas is the same as the number of reflection elements in RIS, and the location is the same as RIS [9]. As shown in Fig. 7, the sub-scheme2 still obtains slightly lower CE than the sub-scheme1 due to better approximation. It can be observed that the CE of RIS is obviously higher than that of HD-Relay since RIS works on the full-duplex mode. Furthermore, the CE of RIS is slightly lower than that of FD-relay. This is mainly because RIS is passive and lacks the RF chain compared to FD-relay. In general, however, the energy consumption of RIS is low and the number of reflection elements is more than the antenna number of the BS, while the energy consumption of the conventional relay schemes is high and the antenna number is less than that of the BS [31]. Therefore, compared with the convectional relay schemes, the RIS scheme may obtain the effective tradeoff between the CE and energy consumption.

Fig. 8 shows the CE of mmWave-MEC systems with sub-scheme1 and sub-scheme2 under different numbers of antenna and RFC, where N=16, 32, and N_{RF}=3, 4. With the increase

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI10.1109/TCOMM.2022.3181673, IEEE Transactions on Communications

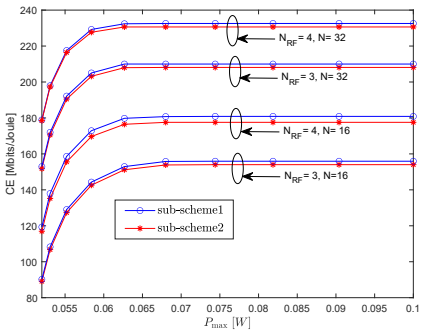

Fig. 8. CE of mmWave-MEC systems with different antenna numbers and RFC numbers.

of N and/or N_{RF}, the system CE can be effectively improved. Namely, the system with N=32 has higher CE than that with N=16 for the same N_{RF}, and the system with N_{RF}=4 has higher CE than that with N_{RF}=3 for the same N. The reason is that with the increase of the number of antennas, mmWave-MEC system can obtain higher spatial diversity gain. Moreover, more array gain can be exploited by increasing the number of RFCs. As a result, the CE performance of the system is increased, as expected. Besides, the proposed two schemes are still valid and have near performance.

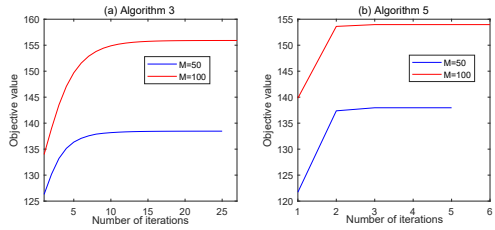

Fig. 9. Convergence behavior of Algorithms 3 and 5 for different M.

Fig. 9 illustrates the convergence performance of Algorithm 3 and Algorithm 5 with different number of RIS reflection elements, where M=50, 100, and $P_{\max}$ = 0.1W. From Fig. 9, it is found that the CE is gradually increasing and finally saturated as the iteration increases. Namely, these two algorithms can converge to their respective stable points after some iterations. Thus, the convergence of Algorithm 1 and Algorithm 3 are guaranteed for different M. However, their required iterative numbers are different, the Algorithm 3 needs about 25 iterations to converge, while the Algorithm 5 only needs about 5 iterations. Thus, the latter has lower complexity. Moreover, with M increasing, the CE is obviously increased. Besides, in Fig. 10, we give the convergence behaviors of Algorithm 3 and Algorithm 5 with different numbers of antennas, where $N = 16, 32$, $P_{\max}$ = 0.1W. From Fig. 10, we can observe the results similar to those in Fig. 9. Namely, for different numbers of antennas, Algorithms 3 and 5 can still converge to their respective optimized values after some numbers of iteration, and the Algorithm 5 needs less iteration than the Algorithm 3. Moreover, the CE with $N = 32$ is higher that with $N = 16$ after convergence, as expected. The results above further confirm that the proposed two algorithms can converge well under different system parameters.

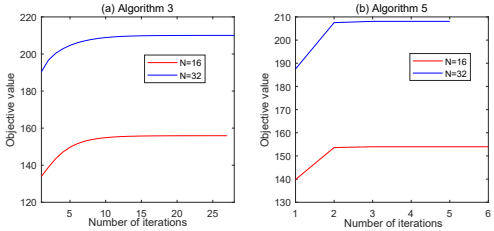

Fig. 10. Convergence behavior of Algorithms 3 and 5 for different N.

In Fig. 11, we evaluate the impact of the distance of the users to BS on the CE performance of the system with and without RIS, where the CE versus the d_c are provided, and $P_{\max}$ = 0.1W. As shown in Fig. 11, with the increase of the distance d_c, the CE performance of the system without RIS rapidly decreases because of the increasing large path loss. While for the system with RIS, its CE decreases slowly due to the assistance of RIS, and it can obtain much higher CE than the system without RIS. Especially as d_c grows big, their CE gap becomes large. Moreover, when the d_c is close to the 80m (i.e., in the proximity of the RIS), the signals can experience less path loss and the CE performance can be improved. However, when the distance d_c is increased to about 160m, the performance gain brought by RIS is no longer increased significantly and starts to become stable. This is because the users are far away from the RIS and the path loss becomes severe, and the corresponding advantage brought by RIS is not very significant. Based on the above analysis, the application of RIS can improve the performance greatly by providing the PBF gain. Furthermore, the RIS can assist the users in offloading the tasks to the MEC server to increase the CE when direct link is weak or not available due to the obstacles. Besides, the RIS may be deployed near the users to improve the CE performance.

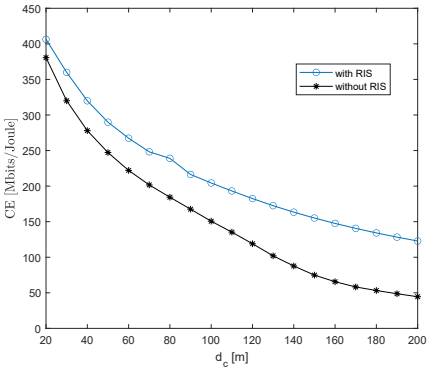

Fig. 11. Performance of CE versus the distance d_c.

VII. Conclusions

We have addressed the computation-efficient resource allocation scheme design for RIS-aided mmWave-MEC systems. Considering the HBF, the achievable rate and CE of the system are derived. In terms of the max-min fairness criterion, and

under the constraints of the minimum rate and maximum power consumption of the users, the optimization problem of jointly optimizing the HBF of BS and the passive BF of RIS as well as the transmit power and local CPU frequency of each user is formulated to maximize the CE. By means of theoretical analysis and mathematical transformation, an effective iterative algorithm based on the PIBCD method has been proposed to solve the optimization problem, and obtain the suboptimal resource allocation scheme. Then, the suboptimal scheme with low complexity is also presented based on the BCD method, and it has the CE performance near to that of the first suboptimal scheme. The convergence and complexity of two schemes are also analyzed. Simulation results verify the effectiveness of the proposed two resource allocation schemes, and show that RIS is feasible and valid when it is used in mmWave-MEC systems.

APPENDIX I
PROOF OF THEOREM 1

In this appendix, we give the proof of Theorem 1. Let $\mathcal{X}^{\text{opt}}$ denote the optimal value of variable $\mathcal{X}$, where $\mathcal{X}$ represents the arbitrary optimization variable appearing in the proof of **Theorem** 1.

Firstly, by introducing $\mathbf{w}_k = \mathbf{A}\mathbf{d}_k$, $u_{k,\tilde{k}} = \mathbf{w}_k^H\left(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta}+\mathbf{h}_{d,\tilde{k}}\right)$, and $q_k = p_k^{-1}(\forall k,\tilde{k}\in\mathcal{K})$, (6) can be rewritten as

$$\overline{\text{SINR}}_k = \frac{q_k^{-1}|u_{k,k}|^2}{\sum_{j\neq k}^{K} q_j^{-1}|u_{k,j}|^2+\|\mathbf{w}_k\|^2\sigma^2}. \tag{65}$$

Thus, the problem (P0) can be equivalently transformed into

$$\begin{aligned}\max_{\{\mathbf{D},\mathbf{A},\boldsymbol{\theta},\mathbf{w}_k,u_{k,\tilde{k}},q_k,f_k\}} \quad & \min_{k\in\mathcal{K}}\left\{\frac{B\log_2(1+\overline{\text{SINR}}_k)+f_k/C_k}{\zeta_k q_k^{-1}+\xi_k f_k^3+P_{k,c}}\right\}\\ \text{s.t.}\quad & C_1, C_2, C_5,\\ & \tilde{C}_3: \log_2\left(1+\overline{\text{SINR}}_k\right)+\tfrac{f_k}{BC_k}\geq\tfrac{R_k^{\min}}{B},\forall k\in\mathcal{K},\\ & \tilde{C}_4: \zeta_k q_k^{-1}+\xi_k f_k^3+P_{k,c}\leq P_k^{\max},\forall k\in\mathcal{K},\\ & \tilde{C}_6: 1/P_k^{\max}\leq q_k\leq 1/P_k^{\min},\forall k\in\mathcal{K},\\ & C_7: \mathbf{w}_k=\mathbf{A}\mathbf{d}_k,\forall k\in\mathcal{K},\\ & C_8: u_{k,\tilde{k}}=\mathbf{w}_k^H(\mathbf{G}_{\tilde{k}}\boldsymbol{\theta}+\mathbf{h}_{d,\tilde{k}}),\forall k,\tilde{k}\in\mathcal{K}.\end{aligned} \tag{66}$$

Secondly, by introducing the auxiliary variables $\{\gamma_k, z_k\}(\forall k\in\mathcal{K})$, we have the following constraints:

$$\begin{aligned}&\tilde{C}_{3,a}: q_k^{-1}|u_{k,k}|^2\geq\gamma_k^{-1},\forall k\in\mathcal{K},\\ &\tilde{C}_{3,b}: \textstyle\sum_{j=1,\neq k}^{K} q_j^{-1}|u_{k,j}|^2+\|\mathbf{w}_k\|^2\sigma^2\leq z_k,\forall k\in\mathcal{K}.\end{aligned} \tag{67}$$

Correspondingly, the problem (66) is equivalently transformed into

$$\begin{aligned}\max_{\{\mathbf{D},\mathbf{A},\boldsymbol{\theta},\mathbf{w}_k,u_{k,\tilde{k}},q_k,f_k,\gamma_k,z_k\}} \quad & \min_{k\in\mathcal{K}}\left\{\frac{B\log_2(1+\gamma_k^{-1}z_k^{-1})+f_k/C_k}{\zeta_k q_k^{-1}+\xi_k f_k^3+P_{k,c}}\right\}\\ \text{s.t.}\quad & C_1, C_2, \tilde{C}_{3,a}, \tilde{C}_{3,b}, \tilde{C}_4, C_5, \tilde{C}_6, C_7, C_8,\\ & \tilde{C}_{3,c}: \log_2\left(1+\gamma_k^{-1}z_k^{-1}\right)+\tfrac{f_k}{BC_k}\geq\tfrac{R_k^{\min}}{B},\forall k\in\mathcal{K}.\end{aligned} \tag{68}$$

Since the objective function of problem (68) is monotonically increasing w.r.t γ_k^{-1} and monotonically decreasing w.r.t z_k, it is concluded from the constraints $\tilde{C}_{3,a}$ and $\tilde{C}_{3,b}$ in (67) that

$$\begin{aligned}&\left(\gamma_k^{\text{opt}}\right)^{-1}=\left(q_k^{\text{opt}}\right)^{-1}\left|u_{k,k}^{\text{opt}}\right|^2,\forall k\in\mathcal{K},\\ &z_k^{\text{opt}}=\sum_{j=1,\neq k}^{K} q_j^{-1}\left|u_{k,j}^{\text{opt}}\right|^2+\left\|\mathbf{w}_k^{\text{opt}}\right\|^2\sigma^2,\forall k\in\mathcal{K}.\end{aligned} \tag{69}$$

Substituting (69) into (68), it is easily obtained that the problem (68) is equivalent to (66).

Thirdly, by introducing the auxiliary variables R_k and $P_k(\forall k\in\mathcal{K})$, we can obtain the transformed constraints as

$$\begin{aligned}&\bar{C}_{3,c}: \log_2\left(1+\gamma_k^{-1}z_k^{-1}\right)+\tfrac{f_k}{BC_k}\geq\tfrac{R_k}{B},\forall k\in\mathcal{K},\\ &\tilde{C}_{4,a}: \zeta_k q_k^{-1}+\xi_k f_k^3+P_{k,c}\leq P_k,\forall k\in\mathcal{K},\\ &\tilde{C}_{4,b}: P_k\leq P_k^{\max},\forall k\in\mathcal{K},\\ &C_9: R_k\geq R_k^{\min},\forall k\in\mathcal{K}.\end{aligned} \tag{70}$$

Hence, the problem (68) can be further transformed into

$$\begin{aligned}\max_{\{\mathbf{D},\mathbf{A},\boldsymbol{\theta},\mathbf{w}_k,u_{k,\tilde{k}},q_k,f_k,\gamma_k,z_k,R_k,P_k\}} \quad & \min_{k\in\mathcal{K}}\{R_k/P_k\}\\ \text{s.t.}\quad & C_1, C_2, \tilde{C}_{3,a}, \tilde{C}_{3,b}, \bar{C}_{3,c}, \tilde{C}_{4,a}, \tilde{C}_{4,b},\\ & C_5, \tilde{C}_6, C_7, C_8, C_9.\end{aligned} \tag{71}$$

Since the objective function of problem (71) is monotonically increasing w.r.t R_k and monotonically decreasing w.r.t P_k, it can be concluded from the constraints $\bar{C}_{3,c}$ and $\tilde{C}_{4,a}$ that

$$\begin{aligned}&R_k^{\text{opt}}=B\log_2\left(1+\left(\gamma_k^{\text{opt}}z_k^{\text{opt}}\right)^{-1}\right)+\tfrac{f_k^{\text{opt}}}{C_k},\forall k\in\mathcal{K},\\ &P_k^{\text{opt}}=\zeta_k\left(q_k^{\text{opt}}\right)^{-1}+\xi_k\left(f_k^{\text{opt}}\right)^3+P_{k,c},\forall k\in\mathcal{K}.\end{aligned} \tag{72}$$

With (71) and (72), the problem (71) will be equivalent to the problem (68).

Finally, by introducing the auxiliary variable η, the problem (71) will be transformed equivalently into

$$\begin{aligned}\max_{\{\mathbf{D},\mathbf{A},\boldsymbol{\theta},\mathbf{w}_k,u_{k,\tilde{k}},q_k,f_k,\gamma_k,z_k,R_k,P_k,\eta\}} \quad & \eta\\ \text{s.t.}\quad & C_1, C_2, \tilde{C}_{3,a}, \tilde{C}_{3,b}, \bar{C}_{3,c}, \tilde{C}_{4,a}, \tilde{C}_{4,b}, C_5, \tilde{C}_6, C_7, C_8, C_9, C_{10}.\end{aligned} \tag{73}$$

According to the analysis above, the problem (73) is equivalent to the original problem (11). Namely, the problem (12) is equivalent to the problem (11).

REFERENCES

[1] X. Wang, L. kong, F. Kong, F. Qiu, M. Xia, S. Arnon, G. Chen, "Millimeter wave communication: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 20, no. 3, pp. 1616-1653, 2018.

[2] X. Yu, F. Xu, K. Yu and N. Li, "Joint energy-efficient power allocation and beamforming for uplink mmWave-NOMA system, *IEEE Trans. Veh. Technol.*, vol. 69, no. 10, pp. 12291-12295, Oct. 2020.

[3] C. Zhao, Y. Cai, A. Liu, et al., "Mobile edge computing meets mmWave communications: joint beamforming and resource allocation for system delay minimization," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2382-2396, Apr. 2020.

[4] C. Zhao, Y. Cai, M. Zhao and Q. Shi, "Joint hybrid beamforming and offloading for mmWave mobile edge computing systems," *2019 IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, Morocco, pp. 1-6, 2019.

[5] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment:Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106-112, Jan. 2020.

[6] Q. Tao, J. Wang and C. Zhong, "Performance analysis of intelligent reflecting surface aided communication systems," *IEEE Commun. Lett.*, vol. 24, no. 11, pp.2464-2468, Nov. 2020.

[7] Z.Peng, Z. Zhang, C. Pan, et al., "Multiuser full-duplex two-way communications via intelligent reflecting surface," *IEEE Trans. Signal Process.*, vol.69, pp. 837-851, Jan. 2021.

[8] W. Tang, M. Z. Chen, X. Chen, et al., "Wireless communications with reconfigurable intelligent surface: path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421-439, Jan. 2021.

[9] C. Huang, A. Zappone, G. C. Alexandropoulos, et al., "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157-4170, Aug. 2019.

[10] G. Zhou, C. Pan, H. Ren, et al., "Intelligent reflecting surface aided multigroup multicast MISO communication systems," *IEEE Trans. Signal Process.*, vol. 68, pp. 3236-3251, Apr. 2020.

[11] P. Wang, J. Fang, X. Yuan, Z. Chen and H. Li, "Intelligent reflecting surface-assisted millimeter wave communications: joint active and passive precoding design," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 14960-14973, Dec. 2020.

[12] C. Pradhan, A. Li, L. Song, B. Vucetic and Y. Li, "Hybrid precoding design for reconfigurable intelligent surface aided mmWave communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 7, pp. 1041-1045, Jul. 2020.

[13] Z. Chu, P. Xiao, M. Shojafar, et al., "Intelligent reflecting surface assisted mobile edge computing for internet of things," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 619-623, Mar. 2021.

[14] T. Bai, C. Pan, Y. Deng, et al., "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2666-2682, Nov. 2020.

[15] H. Sun, F. Zhou and R. Q. Hu, "Joint offloading and computation energy efficiency maximization in a mobile edge computing system," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3052-3056, Mar. 2019.

[16] Q. Wu and R. Zhang, "Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp.1735-1748, Aug. 2020.

[17] Y. Cao, T. Lv, Z. Lin, et al., "Delay-constrained joint power control, user detection and passive beamforming in intelligent reflecting surface-assisted uplink mmWave System," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 482-495, 2021.

[18] F. Zhou and R. Q. Hu, "Computation efficiency maximization in wireless-powered mobile edge computing networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3170-3184, May. 2020.

[19] Z. Li, M. Chen, Z.Yang, et al., "Energy efficient reconfigurable intelligent surface enabled mobile edge computing networks with NOMA," *IEEE Trans. Cognitive Commun and Network*. Vol.7, no.2, pp. 427-440. June 2021.

[20] J. Chen, "Manifold optimization approach for data detection in massive multiuser MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3652-3657, Apr. 2018.

[21] P. A. Absil, R. Mahony, R. Sepulchre, "Optimization algorithms on matrix manifolds," *Princeton University Press*, 2009.

[22] J. C. Bezdek and R. J. Hathaway, "Convergence of alternating optimization," *Neural, Parallel, Sci. Comput.*, vol. 11, no. 4, pp. 351-368, Dec. 2003.

[23] S. Boyd, L. Vandenberghe, "Convex optimization," *Cambridge University Press*, 2004.

[24] W. Hao, M. Zeng, G. Sun, O. MutaH. Gacanin, "Codebook-cased max-min energy-efficient resource allocation for uplink mmWave MIMO-NOMA system," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8303-8314, Dec. 2019.

[25] R. M. Corless, G. H. Gonnet, D. E. G. Hare, et. al., "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, no. 1, pp. 329-359, 1996.

[26] M. Grant and S. Boyd, "CVX: MATLAB software for disciplined convex programming," Version 2.1. [Online]. Available: http://cvxr.com/cvx, 2017.

[27] H. Guo, Y.-C. Liang, J. Chen, and E. G. Larsson, "Weighted sumrate optimization for intelligent reflecting surface enhanced wireless networks," 2019. [Online]. Available: arXiv:1905.07920.

[28] M. R. Akdeniz, Y. Liu, M. K. Samimi, et al., "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164-1179, Jun. 2014.

[29] Y. Wu, Y. Wang, F. Zhou and R. Q. Hu, "Computation efficiency maximization in OFDMA-based mobile edge computing networks," *IEEE Commun. Lett.*, vol. 24, no. 1, pp. 159-163, Jan. 2020.

[30] F. Xu, X. Yu, J. Cai, et al., "Computation efficiency optimization in UAV-enabled mobile edge computing system with multi-carrier non-orthogonal multiple access," *EURASIP J. Wireless Commun. Netw.*, pp. 1-22, 2020.

[31] M. R. Zamani, M. Eslami, M. Khorramizadeh and Z. Ding, "Energy-efficient power allocation for NOMA with imperfect CSI," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 1009-1013, Jan. 2019.