IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON ADVANCES IN SIGNAL PROCESSING FOR NON-ORTHOGONAL MULTIPLE ACCESS

Received January 7, 2019, accepted January 20, 2019, date of publication February 6, 2019, date of current version February 22, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2896252

Energy Efficiency Maximization Design for Full-Duplex Cooperative NOMA Systems With SWIPT

HE HUANG[1] AND MIN ZHU[2]

[1]Department of Information Technology, Wenzhou Vocational and Technical College, Wenzhou 325000, China
[2]College of Information Science and Technology, Zhejiang Shuren University, Hangzhou 310009, China

Corresponding author: Min Zhu (zhumin@zjsru.edu.cn)

This work was supported in part by the Wenzhou Science and Technology Bureau Program under Grant G20140010, in part by the State Administration of Work Safety's Key Technology Projects under Grant zhejiang-0005-2015AQ, and in part by the National Science Foundation of China under Grant 61571241.

ABSTRACT This paper investigates non-orthogonal multiple access (NOMA) system, where the near NOMA user is able to communicate with a base station (BS) directly while the far NOMA user needs to resort to a full-duplex (FD) relay. Wireless power transfer is applied to our system where the FD relay integrated with power splitting architecture can be powered wirelessly by the ambient radio signals simultaneously. Moreover, we assume that self-interference cancellation at the relay and inter-user interference cancellation at the near NOMA user are imperfect. For the purpose of maximizing the energy efficiency of the NOMA system, the power splitting ratio and transmit beamforming vectors are researched for the BS and relay with the guarantee of QoS requirements for two users. The formulated problem appears to be nonlinear fractional programming, which is highly nonconvex and the global optimality is hard to obtain. By introducing Charnes–Cooper's transformation and inner approximation method, we develop an iterative algorithm to gain the near optimal solution. The simulation results show that the proposed joint optimizing algorithm outperforms the existing scheme in terms of energy efficiency as well as the effect of SI cancellation on performance gain.

INDEX TERMS Non-orthogonal multiple access (NOMA), full-duplex relay, power splitting, fractional programming, iterative algorithm.

I. INTRODUCTION

State of the art, the application of non-orthogonal multiple access (NOMA) in fifth generation (5G) networks has drawn wide attention due to its potential to improve spectral efficiency [1]–[3] and energy efficiency [4], [5]. Different from conventional orthogonal multiple access (OMA), NOMA enables multiple users to be served at the same time and frequency by superimposing multiple users in the power domain at the transmitter and using successive interference cancellation (SIC) at the receiver [6], [7]. In NOMA system, the users are divided into different kinds of categories according to their channel conditions, i.e., the near user and the far user. However, the existence of the near user indeed influences the recept reliability of the far user [8]. To ensure user fairness, the far user with poor channel condition is allocated more power than the near user with good channel condition. However, to guarantee the performance of the far user, most of the power is allocated to the far user blindly. This will not only destroy quality-of-service (QoS) requirement of the near user, but also lower the reception reliability of the far user [8]. Recently, cooperative techniques have been widely applied to NOMA system to achieve a balance between the performance of two users. In cooperative NOMA system, two types of relay are employed, i.e., in-band relay [9], [10] and out-band relay [11]–[13]. The former is to regard the near user as a relay to help the far user, and the latter is a dedicated relay utilized to assist the transmission between the BS and the far user. Particularly, [10] combined wireless power transfer with cooperative NOMA networks. In the proposed system, the near user harvested energy and used it for relay. With the help of a dedicated relay, [11] proved

The associate editor coordinating the review of this manuscript and approving it for publication was Miaowen Wen.

the superiority of NOMA in terms of outage performance. Meanwhile, [12] considered the application of the relay to the cooperative NOMA system and showed the remarkable outage performance gains of the proposed scheme compared to the existing OMA schemes. In addition, [13] proposed two optimal relay selection schemes for downlink cooperative NOMA system with multiple out-band relays.

To our knowledge, the above works [11]–[13] are based on half-duplex (HD) relays. Without doubt, HD mode leads to the reduction of spectral efficiency due to the requirement of additional time resources. Hence, full-duplex (FD) relay, which allows the relay to simultaneously transmit and receive signals on the same frequency channel, can be used to make up for this loss [14], [15]. However, there exist some challenges caused by FD mode in out-band relay systems, i.e., inter-user interference and self-interference (SI) at the FD relay, which would significantly influence the system performance. To take the advantage of FD operation, many effective methods for SI cancellation have been proposed in [16] and [17], such as antenna isolation and analog SI cancellation. Motivated by this, some researchers began to come up with the combination of NOMA and FD relay [18], [19]. Reference [18] proposed a FD device-to-device aided cooperative NOMA scheme where the near user is required to be the FD relay for assisting the far user. Reference [19] adopted the FD relay to help the far user in cooperative NOMA and proved the superiority of the FD NOMA scheme by analyzing the outage probability. Moreover, numerical results also show that FD operation can bring more performance gain compared to the HD one. To compare FD and HD cooperative NOMA schemes with or without direct link between the BS and the far user, [20] discussed outage probability, ergodic rate and energy efficiency respectively, and verified that FD NOMA is superior to HD NOMA towards outage probability and ergodic rate in low signal-to-noise radio region.

To cope with the challenge that the energy-constrained relay will be threatened by power scarcity, simultaneous wireless information and power transfer (SWIPT) is proposed as a promising approach to prolong the lifetime of energy-restricted relays [10], [21], [22], [24]. In general, two practical receiver architectures, i.e., power splitting (PS) and time switching (TS), are used to perform SWIPT. The former allows the relay to split the received RF signal into two streams for energy harvesting and information transfer, respectively. As for the latter one, the relay is allowed to harvest energy and transmit information at different time slots. In [10] and [24], the near user in the proposed NOMA system can relay the far users' information messages with the harvested energy, such that the batteries of the relay would not be exhausted by collaboration behavior. Considering the fact that the near user sacrifices its data rate to help the far user, [25] maximized the data rate of the near user while satisfying the QoS requirement of the far user in HD cooperative SWIPT NOMA system. In addition, some efforts about SWIPT-assisted relay systems for amplify-and-forward (AF) and decode-and-forward (DF) schemes have been studied in terms of outage probability and ergodic capacity, respectively [26], [27].

Although the aforementioned works have provided the exhaustive analysis of NOMA, SWIPT and FD relay respectively, the related works about comprehensive combination of NOMA, SWIPT and FD relay are still few. Motivated by the above discussion, [28] studied FD cooperative NOMA with SWIPT where the near user is regarded as FD relay. Alasaba *et al.* [28] showed that the proposed scheme achieves higher sum rate than both of the conventional OMA scheme and non-cooperative NOMA scheme.

A. CONTRIBUTION

In this paper, we consider the energy efficiency maximization in downlink SWIPT-assisted cooperative NOMA system. The contributions are summarized in the following.

1) we adopt the dedicated relay to help transmission between the BS and the far user instead of selecting the near user as the relay under the sacrificing of data rate. To improve recept reliability, the FD relay harvests energy from received signal and decodes information simultaneously by means of power splitting protocol. The aim is to maximize system energy efficiency (EE) while satisfying the QoS constraints of two NOMA users, which is still an open issue on the proposed system. Accordingly, we formulate the joint optimization of beamforming vectors and power splitting ratio.
2) The formulated problem appears to be a highly non-convex fractional programming, which is hard to achieve global optimal solution. To tackle it, we first convert the problem into convex form by utilizing Charnes-Cooper's transformation and inner approximation method. Then, we propose an iterative algorithm with fast convergence performance to reach at a Karush-Kuhn-Tucker (KKT) point.
3) Numerical results demonstrate that the proposed scheme outperforms other schemes in terms of energy efficiency. In addition, we also verify that the impact of SI cancellation level on the performance gain by simulation and provide a detailed analysis about the simulation result.

B. ORGANIZATION & NOTATIONS

The remainder of this paper is organized as follows. In Section II, we introduce the system model of cooperative NOMA system with full-duplex relay, and then formulate the energy-efficient optimization problem. Section III analyses the structure of the formulated problem and proposes the iterative algorithm for solving it. Simulation results and conclusions are presented in Section IV and Section V, respectively.

Notations: bold upper case letters and bold lower case letters are used to denote matrices and vectors, respectively; $\|\cdot\|$ and $\mathbb{E}|\cdot|$ represent the Euclidean norm of the vector and the expectation, respectively; the conjugate transpose

of a matrix or a vector is denoted by $(\cdot)^H$; $\mathbb{C}^{m\times n}$ stands for the space of complex matrices of dimensions given in superscript; $\mathcal{CN}\left(0,\sigma^2\right)$ denotes a complex Gaussian random variable with zero mean and variance σ^2; $Re\{\cdot\}$ represents the real part of complex number.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider a two-user NOMA system, where U1 directly communicates with the BS, while U2 receive signals from the BS with the assistance of a dedicated relay, as depicted in Fig. 1. Note that the transmitter with a narrow beam pattern can efficiently focus its signal/energy on a target receive antenna [29], [30]. Hence, we assume that U1 and U2 are both single-antenna devices, and the BS is equipped with $M_b > 1$ transmit antennas. The relay works in full-duplex mode by a set of a single receive antenna and $N_t > 1$ transmit antennas. Assume that loop interference brought by FD mode still remains in the relay due to imperfect SI cancellation. To prolong its battery lifetime, the relay adopts power splitting architecture, such that it can harvest energy from a part of the received signal and forward the remaining information using DF protocol with the harvested energy. We Assume that direct link from the BS and U2 is so negligible that can be ignored [19], [31]. Herein, we assume that the channel gain of U1 is better than that of U2.

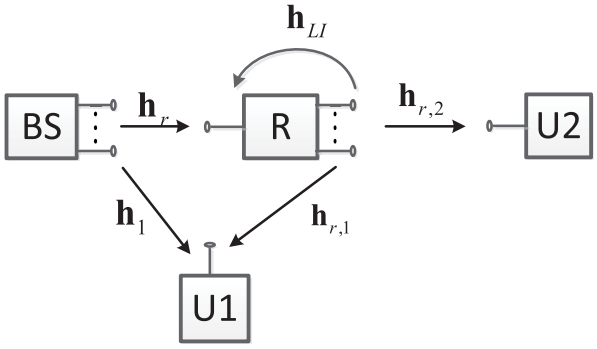

FIGURE 1. A downlink SWIPT-based NOMA with full-duplex relay.

A. CHANNEL MODEL

Let h_r, h_1, h_{r1} and $\mathbf{h}_{r2}$ represent the channel vectors from the BS to R, from the BS to U1, from R to U1 and from R to U2, repectively, and be modeled as $\mathbf{h}_i = g_i d_i^{-\frac{\upsilon}{2}}$, where $i \in \{r, 1, r1, r2\}$, $g_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$, is the normalized small-scale fading coefficient, υ is the path loss exponent, and d_i is the distance between the two nodes, i.e., d_r is the distance between the BS and the relay, d_1 is the distance between BS and U1, d_{r1} is the distance between the relay and U1, and d_{r2} is the distance between the relay and U2. $h_{LI} \sim \mathcal{CN}(0, \theta_{LI} I)$ is the loop-channel between transmit and receive antennas at the relay. Note that direct link between the BS and the far user is supposed to be neglected due to the shielding effect or long distance. Moreover, the perfect channel state information (CSI) is assumed in this paper [32], [33].

B. SIGNAL MODEL

The BS transmits the superimposed signals intended for U1 and U2 as

$$\mathbf{x}(t) = \mathbf{w}_1 x_1(t) + \mathbf{w}_2 x_2(t), \tag{1}$$

where $x_1(t)$ and $x_2(t)$ are assumed to be the normalized information message to U1 and U2 respectively, i.e., $\mathbb{E}|x_1(t)|^2 = \mathbb{E}|x_2(t)|^2 = 1$, $\mathbf{w}_1 \in \mathbb{C}^{M_b}$ and $\mathbf{w}_2 \in \mathbb{C}^{M_b}$ are the corresponding transmit beamforming, and satisfy the power constraint, i.e., $\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 \le P_s$, P_s is the power budget at the BS. The relay operates in FD mode, and thus can simultaneously receive $x(t)$ and forward $s(t)$ to the U2. Hence, the received signal at relay is

$$y_r(t) = \mathbf{h}_r^H \mathbf{x}(t) + \mathbf{h}_{LI}^H \mathbf{w}_r s(t) + n_r(t), \tag{2}$$

where $\hat{x}_2(t-\tau)$ is decoded version of $x_2(t)$ at the relay, τ denotes the processing delay at the relay caused by FD mode, $\mathbf{w}_r \in \mathbb{C}^{N_t\times 1}$ represents the transmit beamforming at the relay, $n_r \sim \mathcal{CN}\left(0,\sigma_r^2\right)$ is the additive white Gaussian noise (AWGN) at the relay. To perform SWIPT, the power splitting architecture is employed at the relay. The received power signal at the relay is split for energy harvesting (EH) and information decoding (ID). With power splitting ratio $\alpha \in (0,1)$, the signal for ID is $y_r^{\mathrm{ID}}(t) = \sqrt{\alpha} y_r(t)$. After self-interference cancellation, the based signal of the relay can be given by

$$\hat{y}_r(t) = \sqrt{\alpha}\left(\mathbf{h}_r^H \mathbf{x}(t) + \hat{\mathbf{h}}_{LI}^H \mathbf{w}_r s(t) + n_r(t)\right) + z_1(t), \tag{3}$$

where $z_1 \sim \mathcal{CN}\left(0,\delta_1^2\right)$ is the noise caused by ID. $\hat{\mathbf{h}}_{LI} = \sqrt{\rho}\mathbf{h}_{LI}$, $\rho \in [0,1]$ denotes the level of residual self-interference. This parameter models the effect of passive LI suppression such as antenna isolation. Particularly, $\rho = 0$ refers to perfect interference cancellation. The relay tries to decode the information intended for U2 while treating the signal $x_1(t)$ of U1 as interference. Thus, the received signal to interference and noise ratio (SINR) at relay can be expressed as

$$\begin{aligned}\gamma_{r,2} &= \frac{\alpha\left|\mathbf{h}_r^H \mathbf{w}_2\right|^2}{\alpha\left(\left|\mathbf{h}_r^H \mathbf{w}_1\right|^2 + \rho\left|\mathbf{h}_{LI}^H \mathbf{w}_r\right|^2 + \sigma_r^2\right) + \delta_r^2} \\ &= \frac{\mathbf{w}_2^H \mathbf{H}_r \mathbf{w}_2}{\mathbf{w}_1^H \mathbf{H}_r \mathbf{w}_1 + \rho \mathbf{w}_r^H \mathbf{H}_{LI} \mathbf{w}_r + \sigma_r^2 + \frac{\delta_r^2}{\alpha}}.\end{aligned} \tag{4}$$

Suppose that noise power is negligible compared with the power of the received signals [36], the harvested power at the relay can be expressed as

$$p_r = \frac{\zeta(1-\alpha)\left(\sum_{i=1}^{2} \mathbf{w}_i^H \mathbf{H}_r \mathbf{w}_i + \mathbf{w}_r^H \mathbf{H}_{LI} \mathbf{w}_r\right) T}{T-\tau}, \tag{5}$$

where $\zeta \in (0,1]$ is the energy conversion efficiency. For briefness, T is set by $T = 1$.

According to the concept of NOMA, U1 first decodes the message of U2 and then removes this part by applying

the successive interference cancellation. After that, U1 can decode its own message. In practice, U1 cannot perfectly remove the message $\hat{x}_2(t-\tau)$ from the relay and by invoking the parameter μ the received signal of U1 is given by

$$y_1(t) = \mathbf{h}_1^H \mathbf{x}(t) + \sqrt{\mu}\mathbf{h}_{r1}^H \mathbf{w}_r \hat{x}_2(t-\tau) + n_1(t), \quad (6)$$

where $\mu \in [0, 1]$ represents the level of interference cancellation.

Thus, the received SINR of U2 observed at U1 can be written as

$$\gamma_{1,2} = \frac{\left|\mathbf{h}_1^H \mathbf{w}_2\right|^2}{\left|\mathbf{h}_1^H \mathbf{w}_1\right|^2 + \mu\left|\mathbf{h}_{r1}^H \mathbf{w}_r\right|^2 + \sigma_1^2} = \frac{\mathbf{w}_2^H \mathbf{H}_1 \mathbf{w}_2}{\mathbf{w}_1^H \mathbf{H}_1 \mathbf{w}_1 + \mu \mathbf{w}_r^H \mathbf{H}_{r1} \mathbf{w}_r + \sigma_1^2}. \quad (7)$$

And the received SINR to detect its own message at U1 can be represented as

$$\gamma_{1,1} = \frac{\left|\mathbf{h}_1^H \mathbf{w}_1\right|^2}{\mu\left|\mathbf{h}_{r1}^H \mathbf{w}_r\right|^2 + \sigma_1^2} = \frac{\mathbf{w}_1^H \mathbf{H}_1 \mathbf{w}_1}{\mu \mathbf{w}_r^H \mathbf{H}_{r1} \mathbf{w}_r + \sigma_1^2}. \quad (8)$$

The received signal of U2 can be given by

$$y_2(t) = \mathbf{h}_{r2}^H \mathbf{w}_r \hat{x}_2(t-\tau) + n_2(t), \quad (9)$$

where $n_2(t) \sim \mathcal{CN}\left(0, \sigma_1^2\right)$ is the AWGN at the U2. Hence, the received SINR to detect signal $x_2(t)$ of U2 is given by

$$\gamma_{2,2} = \frac{\left|\mathbf{h}_{r2}^H \mathbf{w}_r\right|^2}{\sigma_2^2} = \frac{\mathbf{w}_r^H \mathbf{H}_{r2} \mathbf{w}_r}{\sigma_2^2}. \quad (10)$$

According to [34] and [35], the achievable rates for U1 and U2 can be respectively computed by

$$R_1 = \log\left(1 + \gamma_{1,1}\right) \quad (11)$$

and

$$R_2 = \min\left\{\log\left(1+\gamma_{r,2}\right), \log\left(1+\gamma_{1,2}\right), \log\left(1+\gamma_{2,2}\right)\right\}. \quad (12)$$

C. PROBLEM FORMULATION

We consider the energy efficiency maximization (EEmax) of the proposed system. Energy efficiency is equivalently defined as the ratio between the system sum rate and the system power consumption [36], [37], i.e.,

$$\eta_{EE} = \frac{R_{sum}}{P_{sum}}. \quad (13)$$

In the following, we first describe the two relevant terms, i.e., sum rate and power consumption. After that, the problem formulation is presented.

D. SUM RATE

The sum data rate of the system can be written as

$$R_{sum} = R_1 + R_2. \quad (14)$$

E. POWER CONSUMPTION

We consider the system power consumption. In fact, the system power consumption is counteracted by the harvested energy. The energy efficiency of a wireless communication system can be indeed improved by energy harvesting. As in [36], we take the harvested energy into consideration in the formulation of the system power consumption. Thus, the energy consumption of the whole system can be denoted by

$$p_{sum} = \|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 + \|\mathbf{w}_r\|^2 + p_c - p_r. \quad (15)$$

where $\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2$ represents the transmit power at the BS and $\|\mathbf{w}_r\|^2$ stands for transmit power at the relay, p_c is a constant power consumption of circuits and p_r is the harvested power at the relay. Note that the circuit power at the relay is assumed to be independent from data rate and supplied by a battery, which can be charge by renewable resources like wind or solar [22], [23].

F. ENERGY EFFICIENCY MAXIMIZATION

The goal is to jointly optimize the PS ratio and beamforming vectors such that energy efficiency maximization is achieved subject to QoS requirement of the far user and power constraints at the relay and BS. Based on the above definition and analysis, the energy-efficient power allocation problem is then formulated as problem P1:

$$\max_{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_r, \alpha} \ \eta_{EE}, \quad (16a)$$

$$s.t.\ R_1 \geq \bar{R}_1, \quad R_2 \geq \bar{R}_2, \quad (16b)$$

$$\|\mathbf{w}_r\|^2 \leq p_r, \quad (16c)$$

$$\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 \leq P_s, \quad (16d)$$

$$0 \leq \alpha \leq 1, \quad (16e)$$

where $\eta_{EE} = \min\left\{\frac{\log(1+\gamma_{r,2})}{p_{sum}}, \frac{\log(1+\gamma_{1,2})}{p_{sum}}, \frac{\log(1+\gamma_{2,2})}{p_{sum}}\right\} + \frac{\log(1+\gamma_{1,1})}{p_{sum}}$, $\bar{R}_1$ and $\bar{R}_2$ in (16b) are the minimum data rate requirements of the near NOMA user and the far NOMA user respectively, p_r and P_s are the maximum allowed transmit power at the relay and the BS, respectively, which are characterized by constraints (16c) and (16d). In addition, assume that the initial energy stored at the battery of relay is large enough to support the initial operation of the FD U1 [38], (16e) is the inherent constraint for PS ratio.

The formulated problem appears as a nonlinear fractional programming, which is highly non-convex due to that the nonconvex constraints (16b) and objective function (16a). For this challenging EEmax problem, most works would solve it based on Dinkelbach's approach [39], of which the computational complexity would be high. In the following, we propose an efficient beamforming design based on inner approximation method in [40] and Charnes-Cooper's transformation.

III. PROPOSED JOINT OPTIMIZATION FOR ENERGY EFFICIENCY MAXIMIZATION

In this section, we develop an iterative algorithm to solve problem (16). The core idea is to transform (16) into an equivalent formulation in which its convexity is more exposed, and then apply convex approximation to the nonconvex parts. For concise presentation, we set $\mathbf{w} = \{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_r\}$. Then, by introducing the parameters $\mathbf{v} = \left\{v_{1,1}, v_{r,2}, v_{1,2}, v_{2,2}\right\}$ and β, the problem (16) can be recast as

$$\max_{\mathbf{w},\mathbf{s},\alpha,\beta} \frac{s_1+s_2}{\hat{p}_{sum}} \tag{17a}$$

$$s.t.\ \ \gamma_{1,1} \geq v_{1,1}, \tag{17b}$$

$$\gamma_{r,2} \geq v_{r,2}, \quad \gamma_{1,2} \geq v_{1,2},\ \gamma_{2,2} \geq v_{2,2}, \tag{17c}$$

$$\log\left(1+v_{1,1}\right) \geq s_1 \tag{17d}$$

$$\min\left\{\log\left(1+v_{r,2}\right), \log\left(1+v_{1,2}\right), \log\left(1+v_{2,2}\right)\right\} \geq s_2, \tag{17e}$$

$$s_1 \geq \bar{R}_1, s_2 \geq \bar{R}_2, (1-\alpha)\,\beta \geq 1, \tag{17f}$$

$$\|\mathbf{w}_r\|^2 \leq \hat{p}_r, \tag{17g}$$

$$(16d), (16e), \tag{17h}$$

where $\hat{p}_r = \frac{\zeta\left(\mathbf{w}_1^H\mathbf{H}_r\mathbf{w}_1+\mathbf{w}_2^H\mathbf{H}_r\mathbf{w}_2+\mathbf{w}_r^H\mathbf{H}_{LI}\mathbf{w}_r\right)}{(1-\tau)\beta}$, $\hat{p}_{sum} = \|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 + \|\mathbf{w}_r\|^2 + p_c - \hat{p}_r$. The equivalence between (16) and (17) is due to the fact that all the constraints from (17b) to (17f) hold with quality at optimality, which can be proved by contradiction. Take (17b) as example, suppose that the optimal $v_{1,1}^*$, s_1^* satisfy $\gamma_{1,1} > v_{1,1}^*$ and $\log(1+v_{1,1}^*) = s_1^*$. Then, there exists larger $v_{1,1}$ such that $\gamma_{1,1} = v_{1,1}$ and $\log(1+v_{1,1}) = s_1$, resulting in larger objective. The larger objective s_1 is also feasible to (17), which contradicts the optimal assumption.

Using (4), (7), (8) and (10), the problem (17) can be further rewritten explicitly as

$$\max_{\mathbf{w},\mathbf{v},\mathbf{s},\alpha,\,\beta} \frac{s_1+s_2}{\hat{p}_{sum}\left(\mathbf{w},\beta\right)}, \tag{18a}$$

$$s.t.\ \frac{\mathbf{w}_1^H\mathbf{H}_1\mathbf{w}_1}{v_{1,1}} \geq \mu\mathbf{w}_r^H\mathbf{H}_{r,1}\mathbf{w}_r + \sigma_1^2, \tag{18b}$$

$$\frac{\mathbf{w}_2^H\mathbf{H}_r\mathbf{w}_2}{v_{r,2}} \geq \varphi_{r,2}\left(\mathbf{w}_1, \mathbf{w}_r, \alpha\right),$$

$$\frac{\mathbf{w}_2^H\mathbf{H}_1\mathbf{w}_2}{v_{1,2}} \geq \varphi_{1,2}\left(\mathbf{w}_1, \mathbf{w}_r\right), \ \frac{\mathbf{w}_r^H\mathbf{H}_{r,2}\mathbf{w}_r}{v_{2,2}} \geq \sigma_2^2, \tag{18c}$$

$$\|\mathbf{w}_r\|^2 \leq \frac{\zeta\left(\mathbf{w}_1^H\mathbf{H}_r\mathbf{w}_1+\mathbf{w}_2^H\mathbf{H}_r\mathbf{w}_2+\mathbf{w}_r^H\mathbf{H}_{LI}\mathbf{w}_r\right)}{(1-\tau)\,\beta}, \tag{18d}$$

$$(17d)-(17f), (16d), (16e), \tag{18e}$$

where $\varphi_{r,2}\left(\mathbf{w}_1, \mathbf{w}_r, \alpha\right) = \mathbf{w}_1^H\mathbf{H}_r\mathbf{w}_1 + \rho\mathbf{w}_r^H\mathbf{H}_{LI}\mathbf{w}_r + \sigma_r^2 + \frac{\delta_r^2}{\alpha}$, $\varphi_{1,2}\left(\mathbf{w}_1, \mathbf{w}_r\right) = \mathbf{w}_1^H\mathbf{H}_1\mathbf{w}_1 + \mu\mathbf{w}_r^H\mathbf{H}_{r,1}\mathbf{w}_r + \sigma_1^2$. Nevertheless, this optimization problem is still intractable. Firstly, the objective in (18a) is a nonlinear fractional function. Secondly, except for the convex constraints in (18e), constraints (18b) -(18d) are still nonconvex. This is because the functions in both sides of the constraints in (18b)-(18d) are convex, which motivates us to solve (18) by using inner approximate method. Thus, the fact that quadratic-over-linear function, e.g. $\frac{\mathbf{w}_i^H\mathbf{H}_i\mathbf{w}_i}{v_i}$, is convex [40], lays solid foundation for the following mathematical transformation.

By applying Charnes-Cooper's transformation [41], the objective can be simplified into linear function. However, the non-convexity of constraints in problem (18) are still unchanged. Before we handle these nonconvex constraints, we first present the following theorem, which guarantee the equivalence between (18) and (19). For concise presentation, we denote $s = \{s_1, s_2\}$.

Theorem 1: The optimal solution of problem (18) is denoted by $\left(\mathbf{w}^*, \mathbf{v}^*, \mathbf{s}^*, \alpha^*, \beta^*\right)$, then the optimal solution of the following problem

$$\max_{\tilde{\mathbf{w}},\tilde{\mathbf{v}},\tilde{\mathbf{s}},\tilde{\alpha},\tilde{\beta}} \tilde{s}_1+\tilde{s}_2 \tag{19a}$$

$$s.t.\ \ \frac{\tilde{\mathbf{w}}_1^H\mathbf{H}_1\tilde{\mathbf{w}}_1}{\tilde{v}_{1,1}} \geq \mu\frac{\tilde{\mathbf{w}}_r^H\mathbf{H}_{r1}\tilde{\mathbf{w}}_r}{t} + t^2\sigma_1^2, \tag{19b}$$

$$\frac{\tilde{\mathbf{w}}_2^H\mathbf{H}_r\tilde{\mathbf{w}}_2}{\tilde{v}_{r,2}} \geq \frac{\tilde{\mathbf{w}}_1^H\mathbf{H}_r\tilde{\mathbf{w}}_1}{t} + \frac{\rho\tilde{\mathbf{w}}_r^H\mathbf{H}_{LI}\tilde{\mathbf{w}}_r}{t} + t\sigma_r^2 + \frac{\delta_r^2t^2}{\tilde{\alpha}}, \tag{19c}$$

$$\frac{\tilde{\mathbf{w}}_2^H\mathbf{H}_1\tilde{\mathbf{w}}_2}{\tilde{v}_{1,2}} \geq \frac{\tilde{\mathbf{w}}_1^H\mathbf{H}_1\tilde{\mathbf{w}}_1}{t} + \mu\frac{\tilde{\mathbf{w}}_r^H\mathbf{H}_{r1}\tilde{\mathbf{w}}_r}{t} + t\sigma_1^2,$$

$$\frac{\tilde{\mathbf{w}}_r^H\mathbf{H}_{r2}\tilde{\mathbf{w}}_r}{\tilde{v}_{2,2}} \geq t\sigma_2^2, \tag{19d}$$

$$\frac{\left\|\tilde{\mathbf{w}}_r\right\|^2}{t} \leq \frac{\zeta\left(\tilde{\mathbf{w}}_1^H\mathbf{H}_r\tilde{\mathbf{w}}_1+\tilde{\mathbf{w}}_2^H\mathbf{H}_r\tilde{\mathbf{w}}_2+\tilde{\mathbf{w}}_r^H\mathbf{H}_{LI}\tilde{\mathbf{w}}_r\right)}{(1-\tau)\,\tilde{\beta}}, \tag{19e}$$

$$\frac{\left\|\tilde{\mathbf{w}}_1\right\|^2+\left\|\tilde{\mathbf{w}}_2\right\|^2+\left\|\tilde{\mathbf{w}}_r\right\|^2}{t} + tp_c - \frac{\zeta\left(\tilde{\mathbf{w}}_1^H\mathbf{H}_r\tilde{\mathbf{w}}_1+\tilde{\mathbf{w}}_2^H H_r\tilde{\mathbf{w}}_2+\tilde{\mathbf{w}}_r^H\mathbf{H}_{LI}\tilde{\mathbf{w}}_r\right)}{(1-\tau)\,\tilde{\beta}} \leq 1, \tag{19f}$$

$$\left\|\tilde{\mathbf{w}}_1\right\|^2+\left\|\tilde{\mathbf{w}}_2\right\|^2 \leq t^2p_s, \tag{19g}$$

$$t\log\left(1+\frac{\tilde{v}_{1.1}}{t}\right) \geq \tilde{s}_1,\ 0 \leq \tilde{\alpha} \leq t,\ \tilde{s}_1 \geq t\bar{R}_1,$$

$$\tilde{s}_2 \geq t\bar{R}_2,\ t-\tilde{\alpha} \geq \frac{t^2}{\tilde{\beta}}, \tag{19h}$$

is written as $\left(\tilde{\mathbf{w}}=\mathbf{w}^*/\hat{p}_{sum}\left(\mathbf{w}^*,\beta^*\right), \tilde{\mathbf{v}} = \mathbf{v}^*/\hat{p}_{sum}\left(\mathbf{w}^*,\beta^*\right), \tilde{\mathbf{s}} = \mathbf{s}^*/\hat{p}_{sum}\left(\mathbf{w}^*,\beta^*\right), \tilde{\alpha} = \alpha^*/\hat{p}_{sum}\left(\mathbf{w}^*,\beta^*\right), \tilde{\beta} = \beta^*/\hat{p}_{sum}\left(\mathbf{w}^*,\beta^*\right), t = 1/\hat{p}_{sum}\left(\mathbf{w}^*,\beta^*\right)\right)$.

In other words, let $(\mathbf{w}^*, \mathbf{v}^*, \mathbf{s}^*, \alpha^*, \beta^*)$ be optimal of (19), then $\left(\mathbf{w}=\tilde{\mathbf{w}}^*/t^*,\ \mathbf{s} = \tilde{\mathbf{s}}^*/t^*,\ \mathbf{v}=\tilde{\mathbf{v}}^*/t^*,\ \mathbf{s}=\tilde{\mathbf{s}}^*/t^*,\ \alpha = \tilde{\alpha}^*/t^*,\ \beta = \tilde{\beta}^*/t^*\right)$ is also an optimal solution of (18). The detailed proof is presented as follows.

Proof: Assume that the feasible sets of (18) and (19) are Θ and $\tilde{\Theta}$ respectively. For any given $\{\mathbf{w}, \mathbf{v}, \mathbf{s}, \alpha, \beta\} \in \Theta$, it is easily found that $\hat{p}_{sum}\left(\mathbf{w}, \beta\right) > 0$. Map from Θ to $\tilde{\Theta}$, we can define $\tilde{\mathbf{w}} = \mathbf{w}/\hat{p}_{sum}\left(\mathbf{w}, \beta\right)$, $\tilde{\mathbf{v}} = \mathbf{v}/\hat{p}_{sum}\left(\mathbf{w}, \beta\right)$,

$\tilde{\mathbf{s}} = \mathbf{s}/\hat{p}_{sum}(\mathbf{w}, \beta)$, $\tilde{\alpha} = \alpha/\hat{p}_{sum}(\mathbf{w}, \beta)$, $\tilde{\beta} = \beta/\hat{p}_{sum}(\mathbf{w}, \beta)$, $t = 1/\hat{p}_{sum}(\mathbf{w}, \beta)$. By sample computing, it is not difficult to conclude that the optimization problem (19) for $\left\{\tilde{\mathbf{w}}, \tilde{\mathbf{v}}, \tilde{\mathbf{s}}, \tilde{\alpha}, \tilde{\beta}\right\} \in \tilde{\Theta}$ obtains the same objective $\tilde{s}_1 + \tilde{s}_2 = \frac{s_1+s_2}{\hat{p}_{sum}(\mathbf{w},\beta)}$. On the other hand, if map from $\left\{\tilde{\mathbf{w}}, \tilde{\mathbf{v}}, \tilde{\mathbf{s}}, \tilde{\alpha}, \tilde{\beta}\right\} \in \tilde{\Theta}$ to $\left\{\mathbf{w} = \tilde{\mathbf{w}}/t, \mathbf{v} = \tilde{\mathbf{v}}/t, \mathbf{s} = \tilde{\mathbf{s}}/t, \alpha = \tilde{\alpha}/t, \beta = \tilde{\beta}/t\right\} \in \Theta$, the problems (18) and (19) yield the same objective. To sum up, the Charnes-Cooper's transformation is a one-to-one mapping between Θ and $\tilde{\Theta}$. The proof is completed. ■

According to the mentioned equivalence relationship between problem (18) and (19), we can solve (19) to obtain optimal $\left(\mathbf{w}^*, \mathbf{v}^*, \mathbf{s}^*, \alpha^*, \beta^*\right)$. Next, let us focus on the non-convex constraints (19b)-(19e). By employing the inner approximation method, we obtain local optimality by the iterative procedure. To be specific, at the k-th iteration, we can get the next feasible point for (19) by solving the following problem

$$\max_{\tilde{\mathbf{w}},\tilde{\mathbf{v}},\tilde{\mathbf{s}},\tilde{\alpha},\tilde{\beta},t} \quad \tilde{s}_1 + \tilde{s}_2 \tag{20a}$$

$$s.t. f_{1,1}^{(k)}\left(\tilde{\mathbf{w}}_1, \tilde{v}_{1.1}\right) \geq \mu \frac{\tilde{\mathbf{w}}_r^H \mathbf{H}_{r,1} \tilde{\mathbf{w}}_r}{t} + t^2\sigma_1^2, \tag{20b}$$

$$f_{r,2}^{(k)}\left(\tilde{\mathbf{w}}_2, \tilde{v}_{r.2}\right) \geq \frac{\tilde{\mathbf{w}}_1^H \mathbf{H}_r \tilde{\mathbf{w}}_1}{t} + \frac{\rho \tilde{\mathbf{w}}_r^H \mathbf{H}_{LI} \tilde{\mathbf{w}}_r}{t} + t\sigma_r^2 + \frac{\delta_r^2 t^2}{\tilde{\alpha}},$$

$$f_{1,2}^{(k)}\left(\tilde{\mathbf{w}}_2, \tilde{v}_{1.2}\right) \geq \frac{\tilde{\mathbf{w}}_1^H \mathbf{H}_1 \tilde{\mathbf{w}}_1}{t} + \mu \frac{\tilde{\mathbf{w}}_r^H \mathbf{H}_{r,1} \tilde{\mathbf{w}}_r}{t} + t\sigma_1^2,$$

$$f_{2,2}^{(k)}\left(\tilde{\mathbf{w}}_1, \tilde{v}_{r.2}\right) \geq t\sigma_2^2, \frac{\left\|\tilde{\mathbf{w}}_r\right\|^2}{t} \leq g^{(k)}\left(\tilde{\mathbf{w}}, \tilde{\beta}\right), \tag{20c}$$

$$\frac{\left\|\tilde{\mathbf{w}}_1\right\|^2 + \left\|\tilde{\mathbf{w}}_2\right\|^2 + \left\|\tilde{\mathbf{w}}_r\right\|^2}{t} + tp_c - g^{(k)}\left(\tilde{\mathbf{w}}, \tilde{\beta}\right) \leq 1, \tag{20d}$$

$$\left\|\tilde{\mathbf{w}}_1\right\|^2 + \left\|\tilde{\mathbf{w}}_2\right\|^2 \leq h^{(k)}(t) p_s, (19g), \tag{20e}$$

where

$$f_{1,1}^{(k)}\left(\tilde{\mathbf{w}}_1, \tilde{v}_{1.1}\right) = \frac{2R\left\{\left(\tilde{\mathbf{w}}_1^{(k)}\right)^H \mathbf{H}_1 \tilde{\mathbf{w}}_1\right\}}{\tilde{v}_{1,1}^{(k)}} - \frac{\left(\tilde{\mathbf{w}}_1^{(k)}\right)^H H_1 \tilde{\mathbf{w}}_1^{(k)}}{\left(\tilde{v}_{1,1}^{(k)}\right)^2} \tilde{v}_{1.1},$$

$$f_{r,2}^{(k)}\left(\tilde{\mathbf{w}}_2, \tilde{v}_{r.2}\right) = \frac{2R\left\{\left(\tilde{\mathbf{w}}_2^{(k)}\right)^H \mathbf{H}_r \tilde{\mathbf{w}}_2\right\}}{\tilde{v}_{r,2}^{(k)}} - \frac{\left(\tilde{\mathbf{w}}_2^{(k)}\right)^H \mathbf{H}_r \tilde{\mathbf{w}}_2^{(k)}}{\left(\tilde{v}_{r,2}^{(k)}\right)^2} \tilde{v}_{r.2},$$

$$f_{2,2}^{(k)}\left(\tilde{\mathbf{w}}_1, \tilde{v}_{r.2}\right) = \frac{2R\left\{\left(\tilde{\mathbf{w}}_2^{(k)}\right)^H H_r \tilde{\mathbf{w}}_2\right\}}{\tilde{v}_{r,2}^{(k)}} - \frac{\left(\tilde{\mathbf{w}}_2^{(k)}\right)^H \mathbf{H}_r \tilde{\mathbf{w}}_2^{(k)}}{\left(\tilde{v}_{r,2}^{(k)}\right)^2} \tilde{v}_{r.2},$$

$$f_{1,2}^{(k)}\left(\tilde{\mathbf{w}}_2, \tilde{v}_{1.2}\right) = \frac{2R\left\{\left(\tilde{\mathbf{w}}_2^{(k)}\right)^H \mathbf{H}_1 \tilde{\mathbf{w}}_2\right\}}{\tilde{v}_{1,2}^{(k)}} - \frac{\left(\tilde{\mathbf{w}}_2^{(k)}\right)^H \mathbf{H}_1 \tilde{\mathbf{w}}_2^{(k)}}{\left(\tilde{v}_{1,2}^{(k)}\right)^2} \tilde{v}_{1.2},$$

and

$$g^{(k)}(\tilde{\mathbf{w}}, \tilde{\beta}) = \frac{2\zeta\left(\sum_{i=1}^{2} Re\left\{(\tilde{\mathbf{w}}_i^{(k)})^H \mathbf{H}_r \tilde{\mathbf{w}}_i\right\} + Re\left\{(\tilde{\mathbf{w}}_r^{(k)})^H \mathbf{H}_{LI} \tilde{\mathbf{w}}_r\right\}\right)}{(1-\tau)\tilde{\beta}^{(k)}} - \frac{\zeta\left(\sum_{i=1}^{2} (\tilde{\mathbf{w}}_i^{(k)})^H \mathbf{H}_r \tilde{\mathbf{w}}_i^{(k)} + (\tilde{\mathbf{w}}_r^{(k)})^H \mathbf{H}_{LI} \tilde{\mathbf{w}}_r^{(k)}\right)}{1 - \tau\left(\tilde{\beta}^{(k)}\right)^2} \tilde{\beta}$$

are first-order approximation of $\frac{\mathbf{w}_1^H \mathbf{H}_1 \mathbf{w}_1}{v_{1,1}}$, $\frac{\mathbf{w}_2^H \mathbf{H}_r \mathbf{w}_2}{v_{r,2}}$, $\frac{\mathbf{w}_r^H \mathbf{H}_{r2} \mathbf{w}_r}{v_{2,2}}$, $\frac{\mathbf{w}_2^H \mathbf{H}_1 \mathbf{w}_2}{v_{1,2}}$ at feasible point $(\tilde{\mathbf{w}}, \tilde{\mathbf{v}}, \tilde{\mathbf{s}}, \tilde{\beta})$ respectively. $g^{(k)}(\mathbf{w}, \beta)$ and $h^{(k)}(t) = 2t^{(k)}t - \left(t^{(k)}\right)^2$ are the first-order approximation of $\hat{p}_r$ and t^2 around t, respectively.

The proposed method is outlined in Algorithm 1. Similar to the argument in [42], it is proved that the iterative algorithm converges to a KKT point of (16). For brevity, the detail is omitted in this paper.

Algorithm 1 The proposed algorithm for EE maximization

1: **Initilization**: Set $n = 0$ and select an initial feasible solution $(\tilde{\mathbf{w}}^{(k)}, \tilde{\mathbf{v}}^{(k)}, \tilde{\mathbf{s}}^{(k)}, \tilde{\alpha}^{(k)}, \tilde{\beta}^{(k)}, \tilde{t}^{(k)})$ by solving (20)
2: **While**
3: Solve (15) and obtain the optimal solution $(\tilde{\mathbf{w}}^*, \tilde{\mathbf{v}}^*, \tilde{\mathbf{s}}^*, \tilde{\alpha}^*, \tilde{\beta}^*, \tilde{t}^*)$
4: Set $n = n + 1$
5: Update $(\tilde{\mathbf{w}}^{(k+1)}, \tilde{\mathbf{v}}^{(k+1)}, \tilde{\mathbf{s}}^{(k+1)}, \tilde{\alpha}^{(k+1)}, \tilde{\beta}^{(k+1)}, \tilde{t}^{(k+1)}) = (\tilde{\mathbf{w}}^*, \tilde{\mathbf{v}}^*, \tilde{\mathbf{s}}^*, \tilde{\alpha}^*, \tilde{\beta}^*, \tilde{t}^*)$
6: **Until** Convergence
7: End while
8: **Output:** $\mathbf{w} = \tilde{\mathbf{w}}^{(k)}/t^{(k)}$ **and** $\alpha = \tilde{\alpha}^{(k)}/t^{(k)}$

Generation of feasible point: The feasible point of (16) can be obtained easily. Firstly, we set α among the internal (0, 1) and generate $\mathbf{w}$ by solving (17). Then, $(\mathbf{v}, \mathbf{s}, \beta)$ are computed when the constraints (16b)-(16f) hold with equality. Utilizing the transformation in problem (19), we can further obtain $\left(\tilde{\mathbf{w}}, \tilde{\mathbf{v}}, \tilde{\mathbf{s}}, \tilde{\alpha}, \tilde{\beta}, \tilde{t}\right)$. To obtain the initial feasible point, we set $\alpha = 0.4$. For the given α, the constraints (16b) and (16c) are still non-convex. To tackle it, let $y_1(\mathbf{w}_r) = \mu \mathbf{w}_r^H \mathbf{H}_{r,1} \mathbf{w}_r + \sigma_1^2$, $y_2(\mathbf{w}_1, \mathbf{w}_r) = \mathbf{w}_1^H \mathbf{H}_r \mathbf{w}_1 + \rho \mathbf{w}_r^H \mathbf{H}_{LI} \mathbf{w}_r + \sigma_r^2 + \frac{\delta_r^2}{\alpha}$ and $y_3(\mathbf{w}_1, \mathbf{w}_r) = \mathbf{w}_1^H \mathbf{H}_1 \mathbf{w}_1 + \mu \mathbf{w}_r^H \mathbf{H}_{r,1} \mathbf{w}_r + \sigma_1^2$, which are the denominators of (8), (4) and (7), respectively. Then, we can solve the convex problem to obtain the feasible points of the

original problem (16).

$$\max_{\mathbf{w}} \frac{h_1^{(k)}(\mathbf{w}_1, \mathbf{w}_r)}{\gamma_1} \tag{21a}$$

$$s.t.\ \min\left\{h_2^{(k)}(\mathbf{w}), h_3^{(k)}(\mathbf{w}), h_4^{(k)}(\mathbf{w}_r)\right\} \geq \gamma_2, \tag{21b}$$

$$\|\mathbf{w}_r\|^2 \leq \frac{\zeta(1-\alpha)\left(\sum_{i=1}^{2} h_5^{(k)}(\mathbf{w}_i) + h_6^{(k)}(\mathbf{w}_r)\right)}{1-\tau}, \tag{21c}$$

$$\|\mathbf{w}_r\|^2 \leq \frac{\zeta(1-\alpha)\left(\sum_{i=1}^{2} h_5^{(k)}(\mathbf{w}_i) + h_6^{(k)}(\mathbf{w}_r)\right)}{1-\tau}, \tag{21d}$$

$$\|\mathbf{w}_1\|^2 + \|\mathbf{w}_2\|^2 \leq P_s, \tag{21e}$$

where

$$\begin{aligned}
&\gamma_1 = 2^{\bar{R}_1} - 1, \gamma_2 = 2^{\bar{R}_2} - 1, h_1^{(k)}(\mathbf{w}_1, \mathbf{w}_r) \\
&= \frac{2Re\left\{\left(\mathbf{w}_1^{(k)}\right)^H \mathbf{H}_1 \mathbf{w}_1\right\}}{y_1^{(k)}} - \frac{\left(\mathbf{w}_1^{(k)}\right)^H \mathbf{H}_1 \mathbf{w}_1^{(k)}}{\left(y_1^{(k)}\right)^2} y_1, h_2^{(k)}(\mathbf{w}) \\
&= \frac{2Re\left\{(\mathbf{w}_2^{(k)})^H \mathbf{H}_r \mathbf{w}_2\right\}}{y_2^{(k)}} - \frac{(\mathbf{w}_2^{(k)})\mathbf{H}_r \mathbf{w}_2^{(k)}}{\left(y_2^{(k)}\right)^2} y_2, h_3^{(k)}(\mathbf{w}) \\
&= \frac{2Re\left\{\left(\mathbf{w}_2^{(k)}\right)^H \mathbf{H}_r \mathbf{w}_2\right\}}{y_2^{(k)}} - \frac{\left(\mathbf{w}_2^{(k)}\right)^H \mathbf{H}_r \mathbf{w}_2^{(k)}}{\left(y_2^{(k)}\right)^2} y_2, h_4^{(k)}(\mathbf{w}_r) \\
&= \frac{2\mathrm{Re}\left\{\left(\mathbf{w}_r^{(k)}\right)^H \mathbf{H}_{r,2} \mathbf{w}_r\right\}}{\sigma_2^2} - \frac{\left(\mathbf{w}_r^{(k)}\right)^H \mathbf{H}_{r,2} \mathbf{w}_r^{(k)}}{\sigma_2^2}, h_5^{(k)}(\mathbf{w}_i) \\
&= 2Re\left\{\left(\mathbf{w}_i^{(k)}\right)^H \mathbf{H}_r \mathbf{w}_i\right\} - \left(\mathbf{w}_i^{(k)}\right)^H \mathbf{H}_r \mathbf{w}_i^{(k)}
\end{aligned}$$

and

$$h_6^{(k)}(\mathbf{w}_r) = 2Re\left\{\left(\mathbf{w}_r^{(k)}\right)^H \mathbf{H}_{LI} \mathbf{w}_r\right\} - \left(\mathbf{w}_r^{(k)}\right)^H \mathbf{H}_{LI} \mathbf{w}_r^{(k)} \tag{22}$$

are the first-order approximation of $\frac{\mathbf{w}_1^H \mathbf{H}_1 \mathbf{w}_1}{y_1(\mathbf{w}_r)}$, $\frac{\mathbf{w}_2^H \mathbf{H}_r \mathbf{w}_2}{y_2(\mathbf{w}_1, \mathbf{w}_r)}$, $\frac{\mathbf{w}_2^H \mathbf{H}_1 \mathbf{w}_2}{y_3(\mathbf{w}_1, \mathbf{w}_r)}$, $\frac{\mathbf{w}_r^H \mathbf{H}_{r,2} \mathbf{w}_r}{\sigma_2^2}$, $(\mathbf{w}_i)^H \mathbf{H}_r \mathbf{w}_i$ and $(\mathbf{w}_r)^H \mathbf{H}_{LI} \mathbf{w}_r$ around feasible point $(\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_r)$, respectively.

IV. NUMERICAL RESULTS

This section numerically evaluates the EE performance of the proposed cooperative NOMA scheme, which is labeled as "NOMA+FD relay". For comparison, we present two other schemes, i.e., conventional OMA scheme and HD cooperative NOMA scheme, which are labeled as "OMA+FD relay" and "NOMA+HD relay", respectively. In conventional OMA scheme, we adopt a time-division multiple access (TDMA) system as a benchmark, where same time slots are allocated to users respectively. In HD cooperative NOMA scheme, the relay works in HD mode, i.e., receives signal and harvests energy in the first phase and forwards information using the harvested energy, which is different from FD relay in the proposed scheme.

A. PARAMETER SETTING

In the simulations, we solve the convex problem by utilizing the state-of-the-start SOCP solver MOSEK [43]. The average energy efficiency is obtained over 1000 random channel realizations. Without loss of generality, the pass loss of loop interference channel $\mathbf{h}_{LI}$ is set by $\theta_{LI} = -10$ dB and $\mathbf{h}_i = g_i d_i^{-\frac{\alpha}{2}}$ is generated with $g_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$ and fixed d_i. The constant circuit power consumption is assumed to be $p_c = 45$ dBm. Unless otherwise noted, the default parameter setting can be seen in Table 1.

TABLE 1. Parameter setting.

Parameters	Value
System bandwidth	1 [MHz]
Noise variances (σ_1^2, σ_r^2, δ_1^2)	-70 [dBm]
The error tolerance (ε)	10^{-3}
Distances d_{b1},d_{br},d_{r1},d_{r2}	8[m], 8[m], 6[m], 15[m]
The requested minimum rate of U1 and U2 ($\bar{R}_1$, $\bar{R}_2$)	$R_1 = R_2 = 1$ [Mbits/J]
The number of transmit antennas at BS and U1 (M_b, N_t)	$M_b = N_t = 2$
The available power at the BS (P_s)	45 [dBm]
SI cancellation level (ρ)	-30 [dB]
Inter-user Interference cancellation level at U1(μ)	-50 [dB]
Energy conversion efficiency (η)	0.8
Processing delay at relay (τ)	0.5

B. CONVERGENCE RESULT

The convergence behavior of Algorithm 1 is illustrated in Fig. 2. As shown in the figure, the proposed algorithm converges very fast, which stops iterating within fifteen steps. Moreover, we also study the effect of the number of transmit antenna at relay N_t on convergence performance. Obviously, The convergence rate is not sensitive to the problem complexity which is proportional to N_t. As expected, the objectives are improved when N_t is increased.

C. ENERGY EFFICIENCY MAXIMIZATION

Fig. 3 plots the average energy efficiency as a function of the transmit power at the BS, P_s under different transmission strategies. We can see that the EE performance of all schemes increases with P_s and the proposed scheme yields the best EE performance among three schemes. Moreover, the EE performance of all schemes converges at $P_s = 35$ dBm, indicating that full power transmission is not necessarily optimal in terms of energy efficiency. That is, there exists a tradeoff between EE performance and transmit power at the BS when $P_s = 35$ dBm. From the perspective of EE,

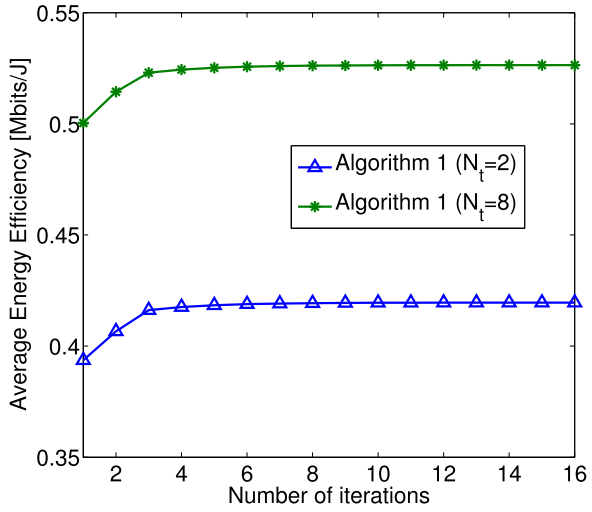

FIGURE 2. Convergence of Algorithm 1 for different numbers of antennas at the relay with $P_S = 35$ dBm.

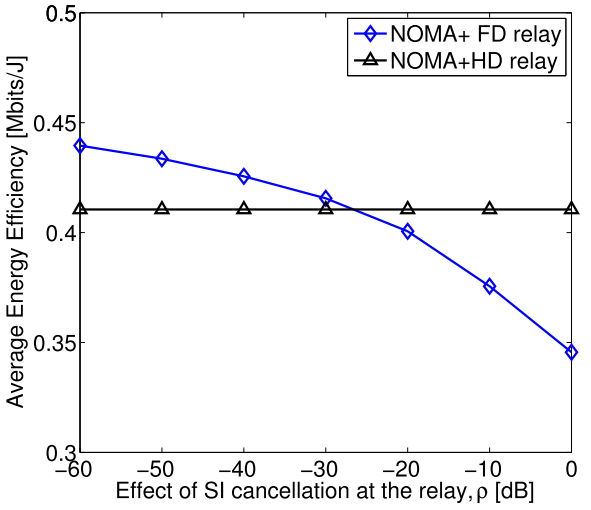

FIGURE 4. Average energy efficiency versus effect of SI cancellation at the relay with $P_S = 35$ dBm.

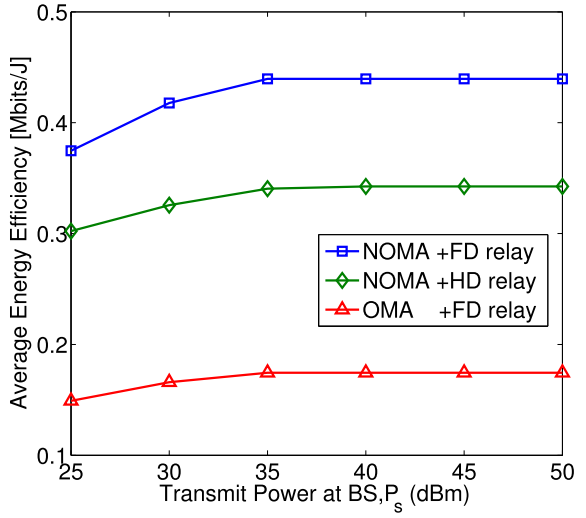

FIGURE 3. Average energy efficiency versus transmit power at the BS.

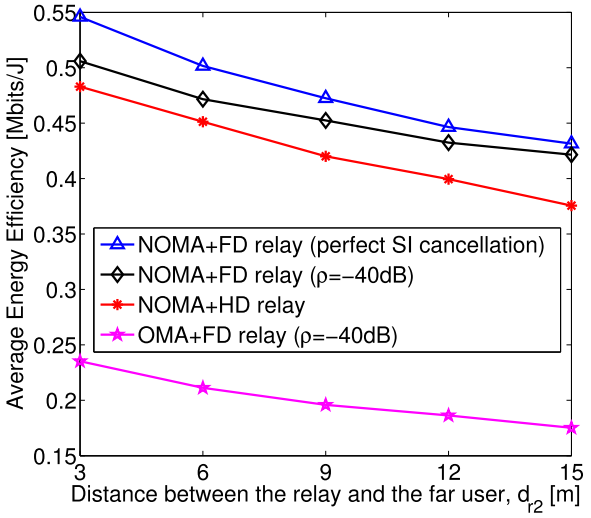

FIGURE 5. Average energy efficiency versus distance between the relay and the far user with $P_S = 35$ dBm.

NOMA can achieve better performance gain than conditional OMA. The reason is that NOMA can enable two users to be served simultaneously, which can improve diversity gain and spectrum utilization. In addition, the FD scheme can offer more performance gain than the HD scheme.

Fig. 4 illustrates the impact of SI cancellation level on the energy efficiency of system at the relay. The energy efficiency of the proposed scheme decreases with SI cancellation level, i.e., less amount of SI cancellation at the relay leads to the worse energy efficiency. Moreover, as shown in this figure, FD scheme is not always superior to the HD scheme towards energy efficiency in the case of imperfect SI cancellation. There exists a critical point, i.e., EE performance of the proposed FD scheme is better than the one of the HD scheme for $\rho \leq -28$ dB, while is inferior to one of the HD scheme for $\rho \geq -28$ dB.

Fig. 5 represents the average energy efficiency versus distance between the relay and the far user. Herein, we also consider perfect SI cancellation case. Obviously, the EE performance of all schemes decrease with the increase of path-loss effect between the relay and the far user. The main reason of the phenomenon is that the data rate at the far user U2 will decrease with the increase of the path-loss effect, and the BS will allocate more power to U2 such that its QoS requirement can be guaranteed. A more important observation is that perfect SI cancellation ($\rho = 0$) can bring the better EE performance than which imperfect perfect SI cancellation does.

V. CONCLUSION

In this paper, a cooperative NOMA system has been studied, which consists of two users and a dedicated relay for assisting transmission from the BS to the far NOMA user. By introducing Charnes-Cooper's transformation and inner approximation methods, the near optimal power splitting ratio and transmit beamforming are obtained. As a result, the energy efficiency of the proposed cooperative NOMA

scheme is maximized subject to the QoS requirements of two users by the joint design of three beamforming vectors and power splitting ratio. The simulation results proved that the proposed scheme can significantly enhance the energy efficiency performance compared to the two benchmark schemes as well as the effect of SI cancellation on performance gain.

REFERENCES

[1] S. M. R. Islam, N. Avazov, O. A. Dobre, and K.-S. Kwak, "Power-domain non-orthogonal multiple access (NOMA) in 5G systems: Potentials and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 721–742, 2nd Quart., 2017.

[2] F. Zhou, Y. Wu, Y.-C. Liang, Z. Li, Y. Wang, and K.-K. Wong, "State of the art, taxonomy, and open issues on cognitive radio networks with NOMA," *IEEE Wireless Commun.*, vol. 25, no. 2, pp. 100–108, Apr. 2017.

[3] S. M. R. Islam, M. Zeng, and O. A. Dobre, "NOMA in 5G systems: Exciting possibilities for enhancing spectral efficiency," *IEEE 5G Tech Focus*, vol. 1, no. 2, pp. 1–6, Jun. 2017. [Online]. Available: http://5g.ieee.org/tech-focus

[4] W. Hao, M. Zeng, Z. Chu, and S. Yang, "Energy-efficient power allocation in millimeter wave massive MIMO with non-orthogonal multiple access," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 782–785, Dec. 2017.

[5] F. Zhou, Y. Wu, R. Q. Hu, Y. Wang, and K.-K. Wong, "Energy-efficient NOMA enabled heterogeneous cloud radio access networks," *IEEE Netw.*, vol. 32, no. 2, pp. 152–160, Mar./Apr. 2018.

[6] S. Chen, B. Ren, Q. Gao, S. Kang, S. Sun, and K. Niu, "Pattern division multiple access—A novel nonorthogonal multiple access for fifth-generation radio networks," *IEEE Trans. Veh. Technol*, vol. 66, no. 4, pp. 3185–3196, Apr. 2017.

[7] B. Zheng, M. Wen, F. Chen, J. Tang, and F. Ji, "Secure NOMA based full-duplex two-way relay networks with artificial noise against eavesdropping," in *Proc. IEEE Int. Conf. Commun.(ICC)*, Kansas City, MO, USA, May 2018, pp. 1–6.

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

[9] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[10] Y. Liu, Z. Ding, M. Elkashlan, and H. V. Poor, "Cooperative non-orthogonal multiple access with simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 938–953, Apr. 2016.

[11] J. Men and J. Ge, "Non-orthogonal multiple access for multiple-antenna relaying networks," *IEEE Commun. Lett.*, vol. 19, no. 10, pp. 1686–1689, Oct. 2015.

[12] X. Liang, Y. Wu, D. W. K. Ng, Y. Zuo, S. Jin, and H. Zhu, "Outage performance for cooperative NOMA transmission with an AF relay," *IEEE Commun. Lett.*, vol. 21, no. 11, pp. 2428–2431, Nov. 2017.

[13] P. Xu, Z. Yang, Z. Ding, and Z. Zhang, "Optimal relay selection schemes for cooperative NOMA," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7851–7855, Aug. 2018.

[14] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.

[15] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-driven characterization of full-duplex wireless systems," *IEEE Trans. Wireless Commun.*, vol. 11, no. 12, pp. 4296–4307, Dec. 2012.

[16] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983–5993, Dec. 2011.

[17] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sep. 2014.

[18] Z. Zhang, Z. Ma, M. Xiao, Z. Ding, and P. Fan, "Full-duplex device-to-device-aided cooperative nonorthogonal multiple access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4467–4471, May 2017.

[19] C. Zhong and Z. Zhang, "Non-orthogonal multiple access with cooperative full-duplex relaying," *IEEE Commun. Lett.*, vol. 20, no. 12, pp. 2478–2481, Dec. 2016.

[20] X. Yue, Y. Liu, S. Kang, A. Nallanathan, and Z. Ding, "Exploiting full/half-duplex user relaying in NOMA systems," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 560–575, Feb. 2018.

[21] H. Liu, K. J. Kim, K. S. Kwak, and H. V. Poor, "Power splitting-based SWIPT with decode-and-forward full-duplex relaying," *IEEE Trans. Wireless Commun.*, vol. 15, no. 11, pp. 7561–7577, Nov. 2016.

[22] L. Zhao, X. Wang, and T. Riihonen, "Transmission rate optimization of full-duplex relay systems powered by wireless energy transfer," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6438–6450, Oct. 2017.

[23] W. Tan, G. Xu, E. De Carvalho, D. Xie, L. Fan, and C. Li, "Low cost and high effciency hybrid architecture massive MIMO systems based on DFT processing," *Wireless Commun. Mobile Comput.*, vol. 10, pp. 1–11, May 2018.

[24] B. Zheng, X. Wang, M. Wen, and F. Chen, "NOMA-based multi-pair two-way relay networks with rate splitting and group decoding," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2328–2341, Oct. 2017.

[25] Y. Xu *et al.*, "Joint beamforming and power-splitting control in downlink cooperative SWIPT NOMA systems," *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4874–4886, Sep. 2017.

[26] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Relaying protocols for wireless energy harvesting and information processing," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3622–3636, Jul. 2013.

[27] Y. Ye, Y. Li, D. Wang, F. Zhou, R. Q. Hu, and H. Zhang, "Optimal transmission schemes for DF relaying networks using SWIPT," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7062–7072, Aug. 2018.

[28] Y. Alasaba, C. Y. Leow, and S. K. A. Rahim, "Full-duplex cooperative non-orthogonal multiple access with beamforming and energy harvesting," *IEEE Access*, vol. 6, pp. 19726–19738, 2018.

[29] W. Wu, S. Wu, and B. Wang, "Robust multi-objective beamforming design for power efficient and secure communication in MU-MISO networks," *IEEE Access*, vol. 5, pp. 13277–13285, 2017.

[30] F. Zhou, Z. Chu, H. Sun, R. Q. Hu, and L. Hanzo, "Artificial noise aided secure cognitive beamforming for cooperative MISO-NOMA using SWIPT," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 4, pp. 918–931, Apr. 2018.

[31] Y. Zeng and R. Zhang, "Full-duplex wireless-powered relay with self-energy recycling," *IEEE Wireless Commun. Lett.*, vol. 4, no. 2, pp. 201–204, Apr. 2015.

[32] W. Duan, M. Wen, Z. Xiong, and M. H. Lee, "Two-stage power allocation for dual-hop relaying systems with non-orthogonal multiple access," *IEEE Access*, vol. 5, pp. 2254–2261, 2017.

[33] D. Wan, M. Wen, F. Ji, Y. Liu, and Y. Huang, "Cooperative NOMA systems with partial channel state information over Nakagami-*m* fading channels," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 947–958, Mar. 2018.

[34] H. Xie, F. Gao, S. Jin, J. Fang, and Y.-C. Liang, "Channel estimation for TDD/FDD massive MIMO systems with channel covariance computing," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 4206–4218, Jun. 2018.

[35] B. Wang, F. Gao, S. Jin, H. Lin, and G. Y. Li, "Spatial- and frequency-wideband effects in millimeter-wave massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 66, no. 13, pp. 3393–3406, Jul. 2018.

[36] Q. Shi, C. Peng, W. Xu, M. Hong, and Y. Cai, "Energy efficiency optimization for MISO SWIPT systems with zero-forcing beamforming," *IEEE Trans. Signal Process.*, vol. 64, no. 4, pp. 842–854, Feb. 2016.

[37] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3170–3184, Apr. 2017.

[38] W. Wu, B. Wang, Y. Zeng, H. Zhang, Z. Yang, and Z. Deng, "Robust secure beamforming for wireless powered full-duplex systems with self-energy recycling," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 10055–10069, Nov. 2017.

[39] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, pp. 492–498, Mar. 2017.

[40] S. Boyd and L. Vandenberghe, *Convex Optimization*, S. Boyd, Ed., 1st ed. Cambridge, U.K.: Cambridge Univ. Press., 2014.

[41] A. Charnes and W. W. Cooper, "Programming with linear fractional functionals," *Naval Res. Logistics Quart.*, vol. 9, nos. 3–4, pp. 181–186, 1962.

[42] B. R. Marks and G. P. Wright, "Technical note—A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, 1978.

[43] (2014). *MOSEK ApS*. [Online]. Available: http://www.mosek.com

HE HUANG received the B.S. and M.S. degrees from Hangzhou Dianzi University, China, in 2001 and 2006, respectively. He is currently with the Department of Information Technology, Wenzhou Vocational and Technical College, China. His research interests include optimization algorithm design and artificial intelligence.

MIN ZHU received the B.S. degree from the Nanjing University of Posts and Telecommunications, China, in 2002, the M.S. degree from the Beijing University of Posts and Telecommunications, China, in 2005, and the Ph.D. degree from the Nanjing University of Posts and Telecommunications, in 2018. She is currently with the College of Information Science and Technology, Zhejiang Shuren University, China. Her research interests include routing protocol, optimization algorithm design, and key technologies in 5G.