Robust Beamforming Design in a NOMA Cognitive Radio Network Relying on SWIPT

Haijian Sun, *Student Member, IEEE*, Fuhui Zhou, *Member, IEEE*, Rose Qingyang Hu, *Senior Member, IEEE*, and Lajos Hanzo, *Fellow, IEEE*

Abstract

This paper studies a multiple-input single-output non-orthogonal multiple access cognitive radio network relying on simultaneous wireless information and power transfer. A realistic non-linear energy harvesting model is applied and a power splitting architecture is adopted at each secondary user (SU). Since it is difficult to obtain perfect channel state information (CSI) in practice, instead either a bounded or gaussian CSI error model is considered. Our robust beamforming and power splitting ratio are jointly designed for two problems with different objectives, namely minimizing the transmission power of the cognitive base station and maximizing the total harvested energy of the SUs. The optimization problems are difficult to solve mainly because of the non-linear structure of the energy harvesting and CSI errors models. We converted them into convex forms by using semi-definite relaxation. For the minimum transmission power problem, we obtain the rank-2 solution under the bounded CSI error model, while for the maximum energy harvesting problem, a two-loop procedure using a one-dimensional search is proposed. Our simulation results show that the proposed scheme significantly outperforms its traditional orthogonal multiple access counterpart. Furthermore, the performance using the gaussian CSI error model is generally better than that using the bounded CSI error model.

Index Terms

Robust beamforming, non-orthogonal multiple access, non-linear energy harvesting model, cognitive radio, imperfect channel state information.

Haijian Sun, Fuhui Zhou and Rose Qingyang Hu are with the Department of Electrical and Computer Engineering, Utah State University, USA (e-mail: rosehu@ieee.org). Lajos Hanzo is with the University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: lh@ecs.soton.ac.uk).

I. INTRODUCTION

Non-orthogonal multiple access (NOMA) has been recognized as one of the most promising techniques for next-generation wireless communication systems due to its capability of supporting a high spectral efficiency (SE) and massive connectivity [1]. Since its design philosophy may be combined with diverse transceivers, it has drawn tremendous attention in multiple-antenna systems [2]-[4], in cooperative networks [5], [6], in device-to-device (D2D) networks [7], as well as in downlink and uplink multi-cell networks [8]. In contrast to classic orthogonal multiple access (OMA), NOMA provides simultaneous access to multiple users at the same time and on the same frequency band, for example by using power-domain multiplexing. In order to decrease the mutual interference among different users of power-domain NOMA, successive interference cancellation (SIC) may be applied by the receivers [1]. It has been shown that NOMA is capable of achieving a higher SE and energy efficiency (EE) than OMA [2]-[8].

As another promising technique of improving the SE, cognitive radio (CR) techniques have also been investigated for decades, where the secondary users (SUs) may access the spectrum bands of the primary users (PUs), as long as the interference caused by SUs is tolerable [9]. According to [10], in order to implement CR in practice, three operational models have been proposed, namely, opportunistic spectrum access, spectrum sharing, and sensing-based enhanced spectrum sharing. It is envisioned that the combination of NOMA with CR is capable of further improving the SE. As a benefit of its low implementational complexity, spectrum sharing has been widely applied. In [11]-[13], the authors analyzed the performance of a spectrum sharing CR combined with NOMA. It was shown that the SE can be significantly improved by using NOMA in CR compared to that achieved by using OMA in CR.

On the other hand, the increasing greenhouse gas emissions have become a major concern also in the design of wireless communication networks. According to [14], cellular networks world-wide consume approximately 60 billion kWh energy per year. Moreover, this energy consumption is explosively increasing due to the unprecedented expansion of wireless networks to support ubiquitous coverage and connectivity. Furthermore, because of the rapid proliferation of Internet of Things (IoT) applications, most battery driven power limited IoT devices become useless if their battery power is depleted. Thus it is critical to use energy in an efficient way or to harness renewable energy sources. As remedy, energy harvesting (EH) exploits the pervasive frequency

radio signals for replenishing the batteries [15]. There have been two research thrusts on EH using RF technology. One focuses on wirelessly powered networks, where a so-called harvest-then-transmit protocol is applied [16]. The other one uses simultaneous wireless information and power transfer (SWIPT) [18]-[20], which is the focus of this paper. In literature, SWIPT in CR has been extensively studied. Specifically, the work in [21] considered the optimal beamforming design in a MISO CR downlink network where a similar power splitting structure is applied in the user receiver. Hu *et. al* [22], on the other hand, investigated the objective of EH energy maximization, and the resource allocation problem is formulated to address that goal. Additionally, [23] considered the underlay scheme in CR network and proposed the optimal beamforming design. To address both the SE and EE, a multiple-input single-output (MISO) NOMA CR using SWIPT is considered based on a practical non-linear EH model. Robust beamforming design problems are studied under a pair of channel state information (CSI) error models. The related contributions and the motivation of our work are summarized as follows.

A. Related Work and Motivation

The prior contributions related to this paper can be divided into two categories based on the EH model adopted, i.e. the linear [20]-[36] and the non-linear EH model [16], [37]-[41]. In the linear EH model, the power harvested increases linearly with the input power, while the EH under the non-linear model exhibits more realistic non-linear characteristics especially at the power-tail.

Linear EH model: In [24], Liu *et al*. analyzed the performance of a cooperative NOMA system relying on SWIPT, which outperformed OMA. Do *et al*. [25] extended [24] and studied the beneficial effect of the user selection scheme on the performance of a cooperative NOMA system using SWIPT. In [26], Yang *et al*. presented a theoretical analysis of two power allocation schemes conceived for a cooperative NOMA system with SWIPT. It was shown that the outage probability achieved under NOMA is lower than that obtained under OMA. Diamantoulakis *et al*. [27] studied the optimal resource allocation design of wireless-powered NOMA systems. The optimal power and time allocation were designed for maximizing the max-min fairness among users. In their following work [28], a joint downlink and uplink scheme was considered in the wirelessly powered network. Comparisons are performed between NOMA and TDMA. The results show that NOMA is more energy-efficient in the downlink SWIPT network. In order to

improve the EE, multiple antennas were applied in a NOMA system associated with SWIPT, and the transmit beamforming and the power splitting factor were jointly optimized for maximizing the transmit rate of users [29].

The contributions in [24]-[29] investigated conventional wireless NOMA systems, which did not consider the interference between the secondary network and the primary network. Recently, authors of [20], [30]-[34] studied optimal resource allocation problems in CR associated with SWIPT. In [20], an optimal transmit beamforming scheme was proposed in a multi-objective optimization framework. It was shown that there are several tradeoffs in CR-aided SWIPT. Based on the work in [20], the authors proposed a jointly optimal beamforming and power splitting scheme to minimize the transmit power of the base station in multiple-user CR-aided SWIPT [30]. Considering the practical imperfect CSI, Zhou *et al*. [31] studied robust beamforming design problems in MISO CR-aided SWIPT, where the bounded and the gaussian CSI error models were applied. It was shown that the performance achieved under the gaussian CSI error model is better than that obtained under the bounded CSI error model. The work in [31] was then extended to multiple-input multiple-output (MIMO) CR-aided SWIPT in [32] and [33], where the bounded CSI error model was applied in [32] and the gaussian CSI error model was used in [33] and [35]. In contrast to [20], [30]-[33], Zhou *et al*. [34] studied robust resource allocation problems in CR-aided SWIPT under opportunistic spectrum access.

Non-linear EH model: In [16], robust resource allocation schemes were proposed for maximizing the sum transmission rate or the max-min transmission rate of MIMO-assisted wireless powered communication networks, where a practical non-linear EH model is considered. It was shown that a performance gain can be obtained under a practical non-linear EH model over that attained under the linear EH model. In order to maximize the power-efficient and sum-energy harvested by SWIPT systems, Boshkovska *et al*. designed optimal beamforming schemes in [37] and [38]. Recently, under the idealized perfect CSI assumption, the rate-energy region was quantified in MIMO systems relying on SWIPT and the practical non-linear EH model in [39]. In order to improve the security of a SWIPT system, a robust beamforming design problem was studied under a bounded CSI error model in [40]. The investigations in [16], [37]-[40] were performed in the context of conventional SWIPT systems. Recently, Wang *et al*. [41] extended a range of classic resource allocation problems into a wireless powered CR counterpart. The optimal channel and power allocation scheme were proposed for maximizing

the sum transmission rate.

The resource allocation schemes proposed in [24]-[29] investigated a conventional NOMA system with SWIPT. The mutual interference should be considered and the quality of service (QoS) of the PUs should be protected in NOMA CR. Moreover, the resource allocation schemes proposed in [20], [30]-[34] are based on the classic OMA scheme. Thus, these schemes are not applicable to NOMA CR with SWIPT due to the difference between OMA and NOMA. Furthermore, an idealized linear EH model was applied in [20]-[34], which is impractical since the practical power conversion circuit results in a non-linear end-to-end wireless power transfer. Therefore, it is of great importance to design optimal resource allocation schemes for NOMA CR-aided SWIPT based on the practical non-linear EH model.

Although the practical non-linear EH model was applied in [16], [37]-[41], the authors of [16], [37]-[40] considered conventional OMA systems using SWIPT. Moreover, the resource allocation scheme proposed in [34] is based on OMA and cannot be directly introduced in NOMA CR-aided SWIPT. However, at the time of writing, there is a scarcity of investigations on robust resource allocation design for NOMA CR-aided SWIPT under the practical non-linear EH model. Several challenges have to be addressed to design robust resource allocation schemes for NOMA CR-aided SWIPT. For example, the impact of the CSI error and of the residual interference due to the imperfect SIC should be considered, which makes the robust resource allocation problem quite challenging. Thus, we study robust resource allocation problems in NOMA CR-aided SWIPT.

B. Contributions of the Paper

Our contribution expands [16] in three major contexts. Firstly, in this paper, a NOMA MISO CR-aided SWIPT is considered, while a OMA MIMO wireless powered network was used in [16]. Secondly, the work in [16] relies on the bounded CSI error model, while both the bounded and the gaussian CSI error model are applied in our work. Thirdly, part of our work considers the minimum transmit power as the optimization objective, which is not considered in [16]. Notice that this paper is also an extension from our conference one [17] which only considered minimizing transmission power under bounded imperfect CSI model. The contributions of our work are hence summarized as follows.

    1)   A minimum transmission power problem is formulated under both the bounded and the

gaussian CSI error models in a NOMA MISO CR network. The robust beamforming weights and the power splitting ratio are jointly designed. The original problem is difficult to solve owing to its non-convex nature from the non-linear EH model as well as to the imperfect CSI. Hence we transform this problem to a convex one. Finally, we prove that the robust beamforming weights can be found and the rank is lower than two under the bounded CSI error model.

2) We also consider another optimization problem, where the objective function is based on maximizing the harvested energy. Similarly, this problem is formulated under the above pair of imperfect CSI error models. The non-linear EH model makes the original problem even harder to solve. Nevertheless, we managed to transform it to an equivalent form and applied a two-loop procedure for solving it. The inner loop solves a convex problem, while the outer loop iteratively adjusts the parameters. Furthermore, to decouple the coupled variables, a one-dimensional search algorithm is proposed as well.

3) Simulation results show the superiority of the proposed scheme over the traditional OMA scheme; the performance gain of NOMA becomes higher when the required data rate at each SU is higher. Moreover, the results also demonstrate that under gaussian CSI error model, the performance is generally better than that under the bounded CSI error model.

The remainder of this paper is organized as follows. The system model is presented in Section II. Section III details our robust beamforming design in the context of our power minimization problems under a pair of imperfect CSI error models. Robust beamforming design in EH maximization problems under both imperfect CSI error models are presented in Section IV. Our simulation results are discussed in Section V. Finally, Section VI concludes the paper.

Notations: Boldface capital letters and boldface lower case letters denote matrices and vectors, respectively. The identity matrix is denoted by $\mathbf{I}$; vec($\mathbf{A}$) represents the vectorization of matrix $\mathbf{A}$ and it is attained by stacking its column vectors. The Hermitian (conjugate) transpose, trace, and rank of a matrix $\mathbf{A}$ are represented respectively by $\mathbf{A}^{\mathbf{H}}$, Tr($\mathbf{A}$) and Rank($\mathbf{A}$). $\mathbf{x}^{\dagger}$ denotes the conjugate transpose of a vector $\mathbf{x}$. $\mathbb{C}^{M \times N}$ denotes a M-by-N dimensional complex matrix set. $\mathbf{A} \succeq \mathbf{0} \, (\mathbf{A} \succ \mathbf{0})$ represents that $\mathbf{A}$ is a Hermitian positive semi-definite (definite) matrix. $\|\cdot\|$ represents the Euclidean norm of a vector. $|\cdot|$ denotes the absolute value of a complex scalar. $\mathbf{x} \sim \mathcal{CN}(\mathbf{u}, \boldsymbol{\Sigma})$ represents that $\mathbf{x}$ is a random vector, which follows a complex Gaussian distribution with mean $\mathbf{u}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbb{E}[\cdot]$ represents the expectation operator.

Re $(\mathbf{a})$ extracts the real part of vector $\mathbf{a}$. $\mathbb{R}_+$ represents the set of all non-negative real numbers.

II. System and Energy Harvesting Models

A. System Model

We consider a downlink CR system with one cognitive base station (CBS), one primary base station (PBS), N PUs and K SUs. The CBS is equipped with M antennas, while each user and PBS have a single antenna. It is assumed that the SUs are energy-constrained and energy harvest circuits are used. Specifically, the receiver architecture relies on a power splitting design. Once the signal is detected by the receiver, it will be divided into two parts. One part is used for information detection, while the other part for energy harvesting. Similar structures can be found in [24], [29]. To better utilize the radio resources, all UEs are allowed to access the same resource simultaneously. To be specific, the PBS sends messages to all PUs, while the CBS communicates with all SUs simultaneously by applying NOMA principles by controlling the interference from the CBS to PUs below a certain level [11]. Let us denote the set of SUs and
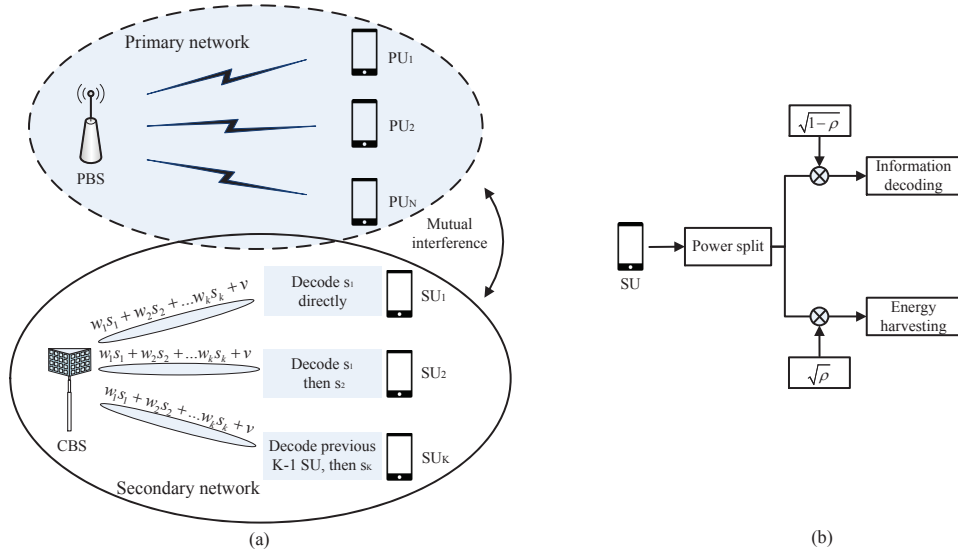

Fig. 1: (a) an illustration of the system model. (b) the power splitting architecture of SUs.

PUs as $\mathcal{K} = \{1, 2, \ldots, K\}$ and $\mathcal{N} = \{1, 2, \ldots, N\}$, respectively. The signal received by the kth SU can be expressed as

$$y_k^S = \mathbf{h}_k^\dagger \mathbf{x} + n_k^S, \ \ k \in \mathcal{K}, \tag{1}$$

where $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$ is the channel gain between the CBS and the kth SU, while n_k^S is the joint effect of additive white Gaussian noise (AWGN) and interference from the PBS. $n_k^S \sim \mathcal{CN}(0, \sigma_{k,S}^2)$, where $\sigma_{k,S}^2$ is the power. This interference model represents a worst-case scenario [20]. Furthermore, $\mathbf{x}$ is the message transmitted to SUs after precoding. According to the NOMA principle, we have:

$$\mathbf{x} = \sum_{k=1}^{K} \mathbf{w}_k s_k + \mathbf{v}, \tag{2}$$

where $\mathbf{w}_k \in \mathbb{C}^{M \times 1}$ is the precoding vector for the k-th UE and s_k is the corresponding intended message. Furthermore, $\mathbf{v} \in \mathbb{C}^{M \times 1}$ is the energy vector allowing us to improve the energy harvesting efficiency at the SUs. We assume that s_k is unitary, i.e. $\mathbb{E}[|s_k|^2] = 1$, and $\mathbf{v}$ obeys the complex Gaussian distribution, i.e. $\mathbf{v} \sim \mathcal{CN}(0, \mathbf{V})$, where $\mathbf{V}$ is the covariance matrix of $\mathbf{v}$.

Likewise, the extra interference arriving from the CBS to the n-th PU is

$$y_n^P = \mathbf{g}_n^\dagger \mathbf{x}, \ n \in \mathcal{N}, \tag{3}$$

where $\mathbf{g}_n^\dagger \in \mathbb{C}^{M \times 1}$ is the channel gain between the CBS and the n-th PU [31].

B. Non-linear EH Model

Most of the existing literature considered an idealized linear energy harvesting model, where the energy collected by the k-th SU is expressed as $E_k^{\text{Linear}} = \eta \, E_k^{\text{In}}$, $E_k^{\text{In}} = \rho \left(\mathbf{h}_k^\dagger (\sum_{j=1}^K \mathbf{w}_j \mathbf{w}_j^\dagger + \mathbf{V}) \mathbf{h}_k + \sigma_{k,S}^2 \right)$ is the input power, where ρ is the power splitting factor that controls the amount of received energy allocated to energy harvesting, $0 < \rho < 1$, while η is the energy conversion efficiency factor, $0 < \eta \leq 1$. However, measurements relying on real-world testbeds show that a typical energy harvesting model exhibits a non-linear end-to-end characteristic. To be specific, the harvested energy first grows almost linearly with the increase of the input power, and then saturates when the input power reaches a certain level. Several models have been proposed in the literature and one of the most popular ones is [16], which is formulated as follows:

$$E_k^{\text{Practical}} = \frac{\Psi_k^{\text{Practical}} - M_k \Omega_k}{1 - \Omega_k}, \quad \Omega_k = \frac{1}{1 + \exp(a_k b_k)}, \tag{4a}$$

$$\Psi_k^{\text{Practical}} = \frac{M_k}{1 + \exp\left(-a_k (E_k^{\text{In}} - b_k) \right)}, \tag{4b}$$

where $E_k^{\text{Practical}}$ is the actual energy harvested from the circuit. Furthermore, $\Psi_k^{\text{Practical}}$ represents a function of the input power E_k^{In}. Additionally, M_k is the maximum power that a receiver can

harvest, while a_k together with b_k characterizes the physical hardware in terms of its circuit sensitivity, limitations, and leakage currents [16].

On the other hand, the signal received in the k-th SU information decoding circuit is

$$y_k^D = \sqrt{1-\rho}(\mathbf{h}_k^\dagger \mathbf{x} + n_k^S) + n_k^D, \tag{5}$$

where n_k^D is the AWGN imposed by the information decoding receiver.

III. POWER MINIMIZATION BASED PROBLEM FORMULATION

Since $\mathbf{x}$ is a composite signal consisting of all SUs' messages, SIC is applied at the receiver side to detect the received signal. The detection is carried out in the same order of the channel gains, i.e. the SUs with lower channel gain will be decoded first. A pair of imperfect CSI error models are considered, namely a bounded and a gaussian model. We adopt both of these in this paper and assume that all SUs have a perfect knowledge of their own CSI.

A. Bounded CSI Error Model

In this model, we consider a bounded error imposed on the estimated CSI, which can be treated as the worst-case scenario. Specifically, the channels can be modeled as follows.

$$\mathbf{h}_k = \widehat{\mathbf{h}}_k + \Delta\mathbf{h}_k, \ \mathbf{\Gamma}_k \triangleq \left\{ \Delta\mathbf{h}_k \in \mathbb{C}^{M \times 1} : ||\Delta\mathbf{h}_k||^2 \leq \varphi_k^2 \right\}, \ \forall k \in \mathcal{K}, \tag{6a}$$

$$\mathbf{g}_n = \widehat{\mathbf{g}}_n + \Delta\mathbf{g}_n, \ \mathbf{\Theta}_n \triangleq \left\{ \Delta\mathbf{g}_n \in \mathbb{C}^{M \times 1} : ||\Delta\mathbf{g}_n||^2 \leq \psi_n^2 \right\}, \ \forall n \in \mathcal{N}, \tag{6b}$$

where $\widehat{\mathbf{h}}_k$ and $\widehat{\mathbf{g}}_n$ are the estimated channel vectors for $\mathbf{h}_k$ and $\mathbf{g}_n$, respectively, while $\mathbf{\Gamma}_k$ and $\mathbf{\Theta}_n$ define the set of channel variations due to estimation errors. The model defines all the uncertainty regions that are confined by power constraints. Furthermore, we use block Rayleigh fading channels, which remain constant within each block, but change from block to block independently.

1) NOMA Transmission: Without loss of generality, we sort the estimated channels of SUs in the ascending order, i.e., $||\hat{\mathbf{h}}_1||^2 \leq ||\hat{\mathbf{h}}_2||^2 \leq \ldots ||\hat{\mathbf{h}}_K||^2$. According to the SIC principle, SU i can detect and remove SU k's signal, for $1 \leq k < i \leq K$. Thus, when SU i decodes signal s_k, the signals for the previously decoded $(k-1)$ SUs have been subtracted from the received composite signal. Due to channel estimation errors, however, these $(k-1)$ signals may not be

completely removed, leaving some residual signals as interference. Therefore, the signal at UE i when decoding s_k becomes

$$y_{i,k}^S = \sqrt{1-\rho}\Big(\mathbf{h}_i^\dagger \mathbf{w}_k s_k + \sum_{j=1}^{k-1} \Delta\mathbf{h}_i^\dagger \mathbf{w}_j s_j + \sum_{j=k+1}^{K} \mathbf{h}_i^\dagger \mathbf{w}_j s_j + \mathbf{h}_i^\dagger \mathbf{v} + n_k^S\Big) + n_k^D. \tag{7}$$

Here, the first term is the desired received signal, the second term is the interference due to imperfect channel estimation, and the third term represents the NOMA interference. The corresponding signal-to-interference-plus-noise ratio (SINR) for the i-th SU after SIC is applied at the receiver is given as

$$\text{SINR}_i^k = \frac{\mathbf{h}_i^\dagger \mathbf{w}_k \mathbf{w}_k^\dagger \mathbf{h}_i}{\sum_{j=1}^{k-1} \Delta\mathbf{h}_i^\dagger \mathbf{w}_j \mathbf{w}_j^\dagger \Delta\mathbf{h}_i + \sum_{j=k+1}^{K} \mathbf{h}_i^\dagger \mathbf{w}_j \mathbf{w}_j^\dagger \mathbf{h}_i + \mathbf{h}_i^\dagger \mathbf{V}\mathbf{h}_i + \sigma_{k,S}^2 + \frac{\sigma_D^2}{(1-\rho)}}. \tag{8}$$

Since signal s_k can be detected at every SU i, as long as $k < i$, there are a set of SINRs for signal s_k. For CBS, the maximum data rate for SU k should be $R_k = \log_2(1 + \min_{k \leq i \leq K} \text{SINR}_i^k)$. Furthermore, the channel estimation error should be considered. The worst-case rate of SU k is

$$R_k = \log_2\Big(1 + \min_{\Delta\mathbf{h}_i \in \mathbf{\Gamma}_i} \{\min_{k \leq i \leq K} \text{SINR}_i^k\}\Big). \tag{9}$$

2) Problem Formulation: In this sub-section, we seek to find the precoding vectors $\mathbf{w}_k$, $k \in \mathcal{K}$, the energy vector $\mathbf{v}$, and the power split ratio ρ, which altogether achieve a satisfactory quality of service (QoS) for all users, and at the same time, they can harvest part of the energy for their future usage. Thus, the problem can be formulated as follows:

$$\mathbf{P}_1: \quad \min_{\mathbf{w}_k \in \mathbb{C}^{M \times 1}, \mathbf{V} \in \mathbb{C}^{M \times M}, \rho} \text{Tr}(\sum_{k=1}^{K} \mathbf{w}_k^\dagger \mathbf{w}_k) + \text{Tr}(\mathbf{V}) \tag{10a}$$

$$\text{s.t. } C1 \ : \ R_k \geq R_{k,\min}, \ \forall k \in \mathcal{K}, \tag{10b}$$

$$C2 \ : \ E_k^{\text{Practical}} \geq P_{k,s}, \ \forall \Delta\mathbf{h}_k \in \mathbf{\Gamma}_k, \ \forall k \in \mathcal{K}, \tag{10c}$$

$$C3 \ : \ \mathbf{g}_n^\dagger\Big(\sum_{j=1}^{K} \mathbf{w}_j \mathbf{w}_j^\dagger + \mathbf{V}\Big)\mathbf{g}_n \leq P_{n,p}, \ \forall \Delta\mathbf{g}_n \in \mathbf{\Theta}_n, \ \forall n \in \mathcal{N}, \tag{10d}$$

$$C4 \ : \ \sum_{j=1}^{K} \mathbf{w}_j^\dagger \mathbf{w}_j + \text{Tr}(\mathbf{V}) \leq P_B, \tag{10e}$$

$$C5 \ : \ 0 < \rho < 1, \tag{10f}$$

$$C6 \ : \ \mathbf{V} \succeq \mathbf{0}. \tag{10g}$$

Our goal is to minimize the total transmitted power. The constraint $C1$ ensures that SU k does attain the predefined minimum data rate; $C2$ allows each SU to harvest the amount of energy that at least compensates the static power dissipation $P_{k,s}$; $C3$ is the interference limit for the n-th PU; $C4$ represents the maximum transmit power constraint of the BS; in $C5$, the power split factor should be in the range of $(0, 1)$. The optimization problem $\mathbf{P}_1$ is hard to solve due to its non-convexity constraints $C1$ and $C2$. Moreover, the realistic imperfect CSI imposes another challenge on the original problem. In the following, we transform the variables.

Let us introduce $\mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^\dagger$ and $\gamma_{k,\min} \triangleq (2^{R_{k,\min}} - 1)$. Then $C1$ in (10b) becomes

$$\min_{\Delta \mathbf{h}_i \in \mathbf{\Gamma}_i} \frac{\mathbf{h}_i^\dagger \mathbf{W}_k \mathbf{h}_i}{\sum_{j=1}^{k-1} \Delta \mathbf{h}_i^\dagger \mathbf{W}_j \Delta \mathbf{h}_i + \sum_{j=k+1}^{K} \mathbf{h}_i^\dagger \mathbf{W}_j \mathbf{h}_i + \mathbf{h}_i^\dagger \mathbf{V} \mathbf{h}_i + \sigma_{k,S}^2 + \frac{\sigma_D^2}{(1-\rho)}} \geq \gamma_{k,\min}, \quad (11)$$
$$i = \{k, k+1, \ldots, K\}, \forall k \in \mathcal{K}.$$

For the notational simplicity, we denote the above constraint as $\Xi_{i,k}$. Thus, $\mathbf{P}_1$ becomes

$$\mathbf{P}_2 : \min_{\mathbf{W}_k \in \mathbb{C}^{M \times M}, \mathbf{V} \in \mathbb{C}^{M \times M}, \rho} \mathrm{Tr}\left(\sum_{k=1}^{K} \mathbf{W}_k + \mathbf{V}\right) \quad (12a)$$

$$\text{s.t. } C1 \ : \ \Xi_{i,k} \quad (12b)$$

$$C2 \ : \ E_k^{\mathrm{Practical}} \geq P_{k,s}, \ \forall \Delta \mathbf{h}_k \in \mathbf{\Gamma}_k, \ \forall k \in \mathcal{K}, \quad (12c)$$

$$C3 \ : \ \mathbf{g}_n^\dagger \left(\sum_{j=1}^{K} \mathbf{W}_j + \mathbf{V}\right) \mathbf{g}_n \leq P_{n,p}, \ \forall \Delta \mathbf{g}_n \in \mathbf{\Theta}_n, \ \forall n \in \mathcal{N}, \quad (12d)$$

$$C4 \ : \ \mathrm{Tr}\left(\sum_{k=1}^{K} \mathbf{W}_k + \mathbf{V}\right) \leq P_B, \quad (12e)$$

$$C5 \ : \ 0 < \rho < 1, \quad (12f)$$

$$C6 \ : \ \mathbf{V} \succeq \mathbf{0}, \mathbf{W}_k \succeq \mathbf{0}, \quad (12g)$$

$$C7 \ : \ \mathrm{Rank}(\mathbf{W}_k) = 1, \ \forall k \in \mathcal{K}. \quad (12h)$$

Here, $C6$ comes from the fact that both $\mathbf{V}$ and $\mathbf{W}_k$ are positive semi-definite matrices. The extra constraint that the rank of $\mathbf{W}_k$ should be 1 is also non-convex. In what follows, we first reformulate $C1$ in (12b) according to the $\mathcal{S}$-Procedure of [42].

Lemma 3.1: $C1$ in (12b) can be reformulated as

$$\begin{bmatrix} \alpha_{i,k} \mathbf{I} + \mathbf{C}_k - \gamma_{k,\min} \sum_{j=1}^{k-1} \mathbf{W}_j & \mathbf{C}_k \widehat{\mathbf{h}}_i \\ \widehat{\mathbf{h}}_i^\dagger \mathbf{C}_k & -\alpha_{i,k} \varphi_k^2 + \mathbf{\Phi}_k \end{bmatrix} \succeq \mathbf{0}, \ \forall k \in \mathcal{K}, \ i = \{k, k+1, \ldots, K\}, \quad (13)$$

where $\mathbf{C}_k = \mathbf{W}_k - \gamma_{k,\min}(\sum_{j=k+1}^{K} \mathbf{W}_j + \mathbf{V})$ and $\boldsymbol{\Phi}_k = \widehat{\mathbf{h}}_i^\dagger \mathbf{C}_k \widehat{\mathbf{h}}_i - \gamma_{k,\min}\left(\sigma_{k,S}^2 + \frac{\sigma_D^2}{(1-\rho)}\right)$, and $\alpha_{i,k}$ is a slack variable conditioned on $\alpha_{i,k} \geq 0$.

Proof: Given $\mathbf{h}_i = \widehat{\mathbf{h}}_i + \Delta\mathbf{h}_i$ and (11), we have

$$\Delta\mathbf{h}_i^\dagger \left(\gamma_{k,\min}(\sum_{j \neq k} \mathbf{W}_j + \mathbf{V}) - \mathbf{W}_k\right)\Delta\mathbf{h}_i + 2\,\mathrm{Re}\left\{\widehat{\mathbf{h}}_i^\dagger \left(\gamma_{k,\min}(\sum_{j=k+1}^{K} \mathbf{W}_j + \mathbf{V}) - \mathbf{W}_k\right)\Delta\mathbf{h}_i\right\} \quad (14)$$

$$+\widehat{\mathbf{h}}_i^\dagger \left(\gamma_{k,\min}(\sum_{j=k+1}^{K} \mathbf{W}_j + \mathbf{V}) - \mathbf{W}_k\right)\widehat{\mathbf{h}}_i + \gamma_{k,\min}\left(\sigma_{k,S}^2 + \frac{\sigma_D^2}{(1-\rho)}\right) \leq 0.$$

From the fact that $\Delta\mathbf{h}_i^\dagger \Delta\mathbf{h}_i - \varphi_k^2 \leq 0$ and according to the $\mathcal{S}$-Procedure, the lemma is proved. ∎

Similarly, $C3$ in (12d) can be transformed into

$$\begin{bmatrix} \beta_n\mathbf{I} - \boldsymbol{\Sigma} & -\boldsymbol{\Sigma}\widehat{\mathbf{g}}_n \\ -\widehat{\mathbf{g}}_n^\dagger\boldsymbol{\Sigma} & -\beta_n\psi_n^2 - \widehat{\mathbf{g}}_n^\dagger\boldsymbol{\Sigma}\widehat{\mathbf{g}}_n + P_{n,p} \end{bmatrix} \succeq \mathbf{0}, \ \forall n \in \mathcal{N}, \quad (15)$$

where $\boldsymbol{\Sigma} = \sum_{j=1}^{K} \mathbf{W}_j + \mathbf{V}$, and $\beta_n \geq 0$ is also a slack variable.

Next, we apply similar manipulations to (12c), which becomes

$$\min_{\Delta\mathbf{h}_k \in \boldsymbol{\Gamma}_k} \rho\left(\mathbf{h}_k^\dagger\boldsymbol{\Sigma}\mathbf{h}_k + \sigma_{k,S}^2\right) \geq D_k, \quad (16)$$

where $D_k = -\ln\left(\frac{1}{P_{k,s}(1-\Omega_k)/M_k+\Omega_k} - 1\right)/a_k + b_k$ is a constant. This condition holds, provided that $a_k > 0$, which is always true in real systems.

Then, applying the $\mathcal{S}$-Procedure to (16), we have the following

$$\begin{bmatrix} \theta_k\mathbf{I} + \boldsymbol{\Sigma} & \boldsymbol{\Sigma}\widehat{\mathbf{h}}_k \\ \widehat{\mathbf{h}}_k^\dagger\boldsymbol{\Sigma} & -\theta_k\varphi_k^2 + \widehat{\mathbf{h}}_k^\dagger\boldsymbol{\Sigma}\widehat{\mathbf{h}}_k + \sigma_{k,S}^2 - \frac{D_k}{\rho} \end{bmatrix} \succeq \mathbf{0}, \ \forall k \in \mathcal{K}, \quad (17)$$

where $\theta_k \geq 0$.

Therefore, $\mathbf{P}_2$ becomes

$$\mathbf{P}_3: \quad \min_{\mathbf{W}_k, \mathbf{V}, \rho, \{\alpha_{i,k}\}, \{\beta_n\}, \{\theta_k\}} \mathrm{Tr}(\sum_{k=1}^{K} \mathbf{W}_k + \mathbf{V}) \quad (18a)$$

$$\text{s.t.} \quad (13), (15), (17), (12e), (12f), (12g), \quad (18b)$$

$$\alpha_{i,k}, \beta_n, \theta_k \geq 0, \ \forall k \in \mathcal{K}, \ i = \{k, k+1, \ldots, K\}, \ \forall n \in \mathcal{K}. \quad (18c)$$

Observe that we drop (12h), since it is not a convex term. This relaxation is commonly referred to as the semi-definite relaxation (SDR) technique. For the specific problem in $\mathbf{P}_2$, the following theorem proves that the optimal $\mathbf{W}_k$ has a limited rank.

Theorem 3.2: If $\mathbf{P}_2$ is feasible, the rank of $\mathbf{W}_k, k \in \mathcal{K}$ is always less than or equal to 2.

Proof: See Appendix. ∎

The transformed problem $\mathbf{P}_3$ is not convex because of the coupling variables ρ in (17) and $(1 - \rho)$ in the denominator of (13). To be able to take advantage of the *CVX* software package, we introduce a pair of auxiliary variables. Specifically, let $p = \frac{1}{1-\rho}$ and $q = \frac{1}{\rho}$. In this way, (13), (15), and (17) become convex terms. Then, we have additional constraints for p and q:

$$p \geq \frac{1}{1 - \rho} \quad \text{and} \quad q \geq \frac{1}{\rho}. \tag{19}$$

It may be readily verified that this transformation does not change the optimal solution of $\mathbf{P}_3$.

B. Complexity Analysis

$\mathbf{P}_3$ has $\frac{K(K+1)}{2}$ linear matrix inequality (LMI) constraints of size $(M + 1)$ in (13), due to the higher decoding complexity. There are N LMI of size $(M+1)$ in (15), K LMI of size $(M+1)$ in (17). Besides, in (12g), there are $(K + 1)$ LMI with size M, and a total of $\frac{K(K+1)}{2} + 2N + K + 2$ linear constraints. Thus, according to [31] and reference therein, the total complexity is

$$\ln(\tau^{-1})n\sqrt{\Psi_{\text{complexity}}}\left((\frac{K(K + 1)}{2} + N + 2K + 1)[(M + 1)^3 + n(M + 1)^2] \right. \tag{20}$$

$$\left. + (K + 1)(M^3 + nM^2) + \frac{K(K + 1)}{2} + 2N + K + 2 + n^2 \right),$$

where $n = (K + 1)M^2 + N + K + \frac{K(K+1)}{2}$ is the number of of decision variables. $\Psi_{\text{complexity}} = \frac{K(K+1)}{2} + N + 2K + 1)M + K^2 + 4N + 3K + 4$, and τ is the accuracy of iteration.

C. Matrix Decomposition

Now we proceed to find the solution of the problem $\mathbf{P}_2$, after which there is one more step to get the original solution for $\mathbf{w}_k$. If $\mathbf{W}_k$ yields rank 1, we can simply write $\mathbf{W}_k^\star = \mathbf{w}_k^\star \mathbf{w}_k^{\star\dagger}$. Otherwise, if $\text{Rank}(\mathbf{W}_k^\star) = 2$, we have several optional approaches to extract $\mathbf{w}_k^\star$. To name a few, we list two methodologies here.

1) *Eigen-decomposition.* Let us denote two eigenvalues of $\mathbf{W}_k^\star$ by λ_1 and λ_2, where $\lambda_1 > \lambda_2 \geq 0$. Clearly, $\mathbf{W}_k^\star = \lambda_1 \mathbf{w}_{1k} \mathbf{w}_{1k}^\dagger + \lambda_2 \mathbf{w}_{2k} \mathbf{w}_{2k}^\dagger$, $\mathbf{w}_{ik}, i = \{1, 2\}$ are the corresponding eigenvectors. To get the rank 1 approximation from a rank 2 matrix, we can let the solution of the original problem be $\widehat{\mathbf{w}}_k = \sqrt{\lambda_1} \mathbf{w}_{1k} \mathbf{w}_{1k}^\dagger$, provided it is feasible.

2) *Randomization technique.* Similar to eigen-decomposition, we first decompose $\mathbf{W}_k^\star$ according to $\mathbf{W}_k^\star = \mathbf{U}_k \mathbf{T}_k \mathbf{U}_k^\dagger$. Then, we let $\widehat{\mathbf{w}}_k = \mathbf{U}_k \mathbf{T}_k^{1/2} \mathbf{e}_k$, where the *m*-th element of $\mathbf{e}_k$ is $[\mathbf{e}_k]_m = e^{j\theta_{k,m}}$ and $\theta_{k,m}$ obeys an independent and uniform distribution within $[0, 2\pi)$.

The above two methods are essentially the same. If we want to get a more precise result, another scaling factor can be added. Specifically, let us define c_k as the scaling factor yet to be determined. Certainly, the problem can be transformed in terms of $\mathbf{W}_k$ and c_k, once we get the optimal value, we can apply either one of the above methods to get a better result. Another point worth noting here is that when the rank of $\mathbf{W}_k$ is 2, there only exists the approximation result of $\mathbf{w}_k^\star$, and this approximation always provides an upper bound.

D. Gaussian CSI Error Model

In Section III-*A*, we introduced a bounded channel model, which defines a confined region for the channel variations, which provides a worst-case estimation. Another commonly used more realistic estimation model assumes that the channel estimation error obeys the Gaussian distribution [31][36][41], which is formulated as follows:

$$\mathbf{h}_k = \widehat{\mathbf{h}}_k + \Delta\mathbf{h}_k, \ \Delta\mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{H}_k), \ \forall k \in \mathcal{K}, \tag{21a}$$

$$\mathbf{g}_n = \widehat{\mathbf{g}}_n + \Delta\mathbf{g}_n, \ \Delta\mathbf{g}_n \sim \mathcal{CN}(0, \mathbf{G}_n), \ \forall n \in \mathcal{N}, \tag{21b}$$

where $\Delta\mathbf{h}_k$ and $\Delta\mathbf{g}_n$ are the channel estimation error vectors, while $\widehat{\mathbf{h}}_k$ and $\widehat{\mathbf{g}}_n$ are the channel vectors estimated at the BS side. Furthermore, $\mathbf{H}_k$ and $\mathbf{G}_n$ are the covariance matrices of the estimation error vectors.

Even though we apply different channel models, the residual interference due to imperfect CSI estimation affects the message detection similarly to the bounded error model. Thus the achievable data rate expression of SU k remains the same except that $\Delta\mathbf{h}_k$ is in a new set. In contrast to the existing NOMA contributions on imperfect CSI [4], in this paper we use the

above-mentioned gaussian estimation error model to form an optimization problem as follows:

$$\mathbf{P}_4: \quad \min_{\mathbf{W}_k \in \mathbb{C}^{M \times M}, \mathbf{V} \in \mathbb{C}^{M \times M}, \rho} \quad \text{Tr}(\sum_{k=1}^{K} \mathbf{W}_k + \mathbf{V}) \tag{22a}$$

$$\text{s.t. } C1 \quad : \quad \Pr\{R_k \geq R_{k,\min}\} \geq 1 - \xi_k, \ \forall k \in \mathcal{K}, \tag{22b}$$

$$C2 \quad : \quad \Pr\{E_k^{\text{Practical}} \geq P_{k,s}\} \geq 1 - \xi_{k,s}, \ \forall \Delta \mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{H}_k), \ \forall k \in \mathcal{K}, \tag{22c}$$

$$C3 \quad : \quad \Pr\left\{\mathbf{g}_n^\dagger \mathbf{\Sigma} \mathbf{g}_n \leq P_{n,p}\right\} \geq 1 - \xi_{n,p}, \forall \Delta \mathbf{g}_n \sim \mathcal{CN}(0, \mathbf{G}_n), \forall n \in \mathcal{N}, \tag{22d}$$

$$C4 \quad : \quad (12e) - (12h). \tag{22e}$$

Here, we assume that the probability of having a rate of R_k is higher than $R_{k,\min}$, which is a predefined value, and we use the threshold ξ_k to control the probability. Likewise, $\xi_{k,s}$ and $\xi_{n,p}$, where $k \in \mathcal{K}$ and $n \in \mathcal{N}$, are used for controlling the outage probability of harvested energy of the kth SU and the interference experienced by the n-th PU, respectively. $\mathbf{P}_4$ is hard to solve owing to its non-convexity, together with constraints $C1 - C3$, which involve probability and uncertainty. Inspired by [31], we solve the resulted optimization problem with the aid of approximations by applying Bernstein-type inequalities [43].

1) Bernstein-type Inequality I [43]: Let $f(\mathbf{z}) = \mathbf{z}^\dagger \mathbf{A} \mathbf{z} + 2\text{Re}\{\mathbf{z}^\dagger \mathbf{b}\} + c$, where $\mathbf{A} \in \mathbb{H}^N$, $\mathbf{b} \in \mathbb{C}^{N \times 1}$, $c \in \mathbb{R}$, and $\mathbf{z} \sim \mathcal{CN}(0, \mathbf{I})$. For any $\xi \in (0, 1]$, an approximate and convex form of

$$\Pr\{f(\mathbf{z}) \geq 0\} \geq 1 - \xi \tag{23}$$

can be written as

$$\text{Tr}(\mathbf{A}) - \sqrt{-2\ln(\xi)}\upsilon_1 + \ln(\xi)\upsilon_2 + c \geq 0, \tag{24a}$$

$$\left\| \begin{bmatrix} \text{vec}(\mathbf{A}) \\ \sqrt{2}\mathbf{b} \end{bmatrix} \right\| \leq \upsilon_1, \tag{24b}$$

$$\upsilon_2 \mathbf{I} + \mathbf{A} \succeq \mathbf{0}, \upsilon_2 \geq 0. \tag{24c}$$

Here, υ_1 and υ_2 are slack variables.

In order to use the above Lemma, we have to transform $\Delta \mathbf{h}_i$ to a standard complex Gaussian vector. Let $\Delta \mathbf{h}_i = \mathbf{H}_i^{1/2} \tilde{\mathbf{h}}_i$, where $\tilde{\mathbf{h}}_i \sim \mathcal{CN}(0, \mathbf{I})$. Substituting it into (11), the convex

approximation becomes

$$\text{Tr}\Big(\mathbf{H}_i^{1/2}(\mathbf{C}_k - \gamma_{k,\min}\sum_{j=1}^{k-1}\mathbf{W}_j)\mathbf{H}_i^{1/2}\Big) - \sqrt{-2\ln(\xi_k)}\upsilon_{1i,k} + \ln(\xi_k)\upsilon_{2i,k} + c_{i,k} \geq 0, \quad (25a)$$

$$c_{i,k} = \widehat{\mathbf{h}}_i^\dagger \mathbf{C}_k \widehat{\mathbf{h}}_i - r_{k,\min}\Big(\sigma_{k,S}^2 + \frac{\sigma_D^2}{1-\rho}\Big), \quad (25b)$$

$$\left\|\left[\begin{array}{c} \text{vec}\Big(\mathbf{H}_i^{1/2}(\mathbf{C}_k - \gamma_{k,\min}\sum_{j=1}^{k-1}\mathbf{W}_j)\mathbf{H}_i^{1/2}\Big) \\ \sqrt{2}\mathbf{H}_i^{1/2}\mathbf{C}_k\widehat{\mathbf{h}}_i \end{array}\right]\right\| \leq \upsilon_{1i,k}, \quad (25c)$$

$$\upsilon_{2i,k}\mathbf{I} + \Big(\mathbf{H}_i^{1/2}(\mathbf{C}_k - \gamma_{k,\min}\sum_{j=1}^{k-1}\mathbf{W}_j)\mathbf{H}_i^{1/2}\Big) \succeq \mathbf{0}, \upsilon_{2i,k} \geq 0, \ \forall k \in \mathcal{K}, i = \{k,\ldots,K\}, \quad (25d)$$

where $\upsilon_{1i,k}$ and $\upsilon_{2i,k}$ are slack variables.

For (22c), we use a simple transformation similar as that in (16), which leads to:

$$\Pr\Big\{\rho(\mathbf{h}_k^\dagger \mathbf{\Sigma}\mathbf{h}_k + \sigma_{k,S}^2) \geq D_k\Big\} \geq 1 - \xi_{k,s}. \quad (26)$$

Furthermore, by applying the inequalities in (24), (26) can be expressed as

$$\text{Tr}\Big(\mathbf{H}_k^{1/2}\mathbf{\Sigma}\mathbf{H}_k^{1/2}\Big) - \sqrt{-2\ln(\xi_{k,s})}\upsilon_{1k,s} + \ln(\xi_{k,s})\upsilon_{2k,s} + c_{k,s} \geq 0, \quad (27a)$$

$$c_{k,s} = \widehat{\mathbf{h}}_k^\dagger \mathbf{\Sigma}\widehat{\mathbf{h}}_k + \sigma_{k,S}^2 - \frac{D_k}{\rho}, \quad (27b)$$

$$\left\|\left[\begin{array}{c} \text{vec}\Big(\mathbf{H}_k^{1/2}\mathbf{\Sigma}\mathbf{H}_k^{1/2}\Big) \\ \sqrt{2}\mathbf{H}_k^{1/2}\mathbf{\Sigma}\widehat{\mathbf{h}}_k \end{array}\right]\right\| \leq \upsilon_{1k,s}, \quad (27c)$$

$$\upsilon_{2k,s}\mathbf{I} + \Big(\mathbf{H}_k^{1/2}\mathbf{\Sigma}\mathbf{H}_k^{1/2}\Big) \succeq \mathbf{0}, \upsilon_{2k,s} \geq 0, \ \forall k \in \mathcal{K}, \quad (27d)$$

where $\upsilon_{1k,s}$ and $\upsilon_{2k,s}$, $k \in \mathcal{K}$, are slack variables.

2) Bernstein-type Inequality II [44]: Let $f(\mathbf{z}) = \mathbf{z}^\dagger \mathbf{A}\mathbf{z} + 2\text{Re}\{\mathbf{z}^\dagger \mathbf{b}\} + c$, where $\mathbf{A} \in \mathbb{H}^N$, $\mathbf{b} \in \mathbb{C}^{N\times 1}$, $c \in \mathbb{R}$, and $\mathbf{z} \sim \mathcal{CN}(0, \mathbf{I})$. For any $\xi \in (0, 1]$, an approximate and convex form for

$$\Pr\{f(\mathbf{z}) \leq 0\} \geq 1 - \xi \quad (28)$$

can be written as

$$\text{Tr}(\mathbf{A}) + \sqrt{-2\ln(\xi)}\upsilon_1 - \ln(\xi)\upsilon_2 + c \geq 0, \quad (29a)$$

$$\left\|\left[\begin{array}{c} \text{vec}(\mathbf{A}) \\ \sqrt{2}\mathbf{b} \end{array}\right]\right\| \leq \upsilon_1, \quad (29b)$$

$$\upsilon_2\mathbf{I} - \mathbf{A} \succeq \mathbf{0}, \upsilon_2 \geq 0, \quad (29c)$$

where υ_1 and υ_2 are slack variables.

We apply Bernstein-type Inequality II to (22d), and let $\Delta \mathbf{g}_n = \mathbf{G}_n^{1/2} \tilde{\mathbf{g}}_n$, where $\tilde{\mathbf{g}}_n \sim \mathcal{CN}(0, \mathbf{I})$ is a standard Gaussian vector. We can have the following convex-form approximation.

$$\text{Tr}(\mathbf{G}_n^{1/2} \boldsymbol{\Sigma} \mathbf{G}_n^{1/2}) + \sqrt{-2\ln(\xi_{n,p})}\upsilon_{1,n} - \ln(\xi_{n,p})\upsilon_{2,n} + c_n \geq 0, \tag{30a}$$

$$c_n = \widehat{\mathbf{g}}_n^\dagger \boldsymbol{\Sigma} \widehat{\mathbf{g}}_n - P_{n,p}, \tag{30b}$$

$$\left\| \begin{bmatrix} \text{vec}(\mathbf{G}_n^{1/2} \boldsymbol{\Sigma} \mathbf{G}_n^{1/2}) \\ \sqrt{2}\mathbf{G}_n^{1/2} \boldsymbol{\Sigma} \widehat{\mathbf{g}}_n \end{bmatrix} \right\| \leq \upsilon_{1,n}, \tag{30c}$$

$$\upsilon_{2,n}\mathbf{I} - \mathbf{G}_n^{1/2} \boldsymbol{\Sigma} \mathbf{G}_n^{1/2} \succeq \mathbf{0}, \upsilon_{2,n} \geq 0, \forall n \in \mathcal{N}, \tag{30d}$$

where $\upsilon_{1,n}$ and $\upsilon_{2,n}$ are slack variables.

Lastly, we relax $\mathbf{P}_4$ by dropping the constraint that $\mathbf{W}_k$ should have rank 1 for now, since it is not a convex one. The relaxed version of the problem is

$$\mathbf{P}_5 : \min_{\mathbf{W}_k, \mathbf{V}, \rho, \{\upsilon_{1i,k}\}, \{\upsilon_{2i,k}\}, \{\upsilon_{1k,s}\}, \{\upsilon_{2k,s}\}, \{\upsilon_{1,n}\}, \{\upsilon_{2,n}\}} \text{Tr}(\sum_{k=1}^K \mathbf{W}_k + \mathbf{V}) \tag{31a}$$

$$\text{s.t.} \quad (25), (27), (30), (12e), (12f), (12g). \tag{31b}$$

Likewise, the coupling variables in (25b) and (27b) make $\mathbf{P}_5$ a non-convex problem. Thus we can still use the transformation in (19), which converts $\mathbf{P}_5$ into an equivalent optimization problem that can be efficiently solved by *CVX*.

IV. MAXIMUM HARVESTED ENERGY PROBLEM FORMULATION

In contrast to Sections III, where the minimum transmission power problem is considered, in the following we consider the optimization problem of maximizing the total harvested energy. This problem has important real-world applications, since most of the consumer electronics products are battery-driven and thus their energy efficiency is critical. In this section, we first formulate the problem, then we transform it in a convex way so that an existing software package can solve it efficiently. A one-dimensional search algorithm will be used. Furthermore, we also consider our previous pair of channel models.

A. Bounded CSI Error Model

Upon considering the imperfect CSI model used in (6), the maximum total harvested energy of all SUs can be formulated as follows:

$$\mathbf{P}_6: \max_{\mathbf{W}_k \in \mathbb{C}^{M \times M}, \mathbf{V} \in \mathbb{C}^{M \times M}, \rho, \{\alpha_{i,k}\}, \{\beta_n\}, \{\theta_k\}} \sum_{k=1}^{K} E_k^{\text{Practical}} \tag{32a}$$

$$\text{s.t.} \quad (13), (15), (12e), (12f), (12g), (12h), \tag{32b}$$

$$\alpha_{i,k}, \beta_n, \theta_k \geq 0, \ \forall k \in \mathcal{K}, \ i = \{k, k+1, \ldots, K\}, \ \forall n \in \mathcal{K}. \tag{32c}$$

The rank operation is not convex, thus we drop the constraint (12h) first, as previously in $\mathbf{P}_3$. Additionally, the objective function relies on a realistic non-linear energy harvesting model, and it is not convex either. Essentially, it is a sum-of-ratio problem, and its global optimization is possible by applying the following transformations:

$$\max_{\mathbf{W}_k \in \mathbb{C}^{M \times M}, \mathbf{V} \in \mathbb{C}^{M \times M}, \rho, \{\alpha_{i,k}\}, \{\beta_n\}, \{\theta_k\}, \{\tau_k\}} \sum_{k=1}^{K} \frac{M_k}{1 + \exp\left(-a_k(\tau_k - b_k)\right)}, \tag{33a}$$

$$E_k^{\text{In}} \geq \tau_k, \ \forall \Delta\mathbf{h}_k. \ \forall k \in \mathcal{K}. \tag{33b}$$

After applying the $\mathcal{S}$-Procedure of [42] to (33b), it becomes

$$\begin{bmatrix} \theta_k \mathbf{I} + \mathbf{\Sigma} & \mathbf{\Sigma}\widehat{\mathbf{h}}_k \\ \widehat{\mathbf{h}}_k^\dagger \mathbf{\Sigma} & -\theta_k \varphi_k^2 + \widehat{\mathbf{h}}_k^\dagger \mathbf{\Sigma}\widehat{\mathbf{h}}_k + \sigma_{k,S}^2 - \frac{\tau_k}{\rho} \end{bmatrix} \succeq \mathbf{0}, \ \forall k \in \mathcal{K}. \tag{34}$$

Furthermore, according to [37], [45], if $\mathbf{P}_6$ has the optimal solutions $\mathbf{W}_k^\star$ and $\mathbf{V}^\star$, there exist two sets of vectors $\boldsymbol{\mu} = \{\mu_1, \mu_2, \ldots, \mu_K\}$ and $\boldsymbol{\epsilon} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_K\}$ such that the solutions are also optimal for the following equivalent parametric optimization problem:

$$\mathbf{P}_7: \max_{\mathbf{W}_k \in \mathbb{C}^{M \times M}, \mathbf{V} \in \mathbb{C}^{M \times M}, \rho, \{\alpha_{i,k}\}, \{\beta_n\}, \{\theta_k\}, \{\tau_k\}} \sum_{k=1}^{K} \mu_k \Big\{ M_k - \epsilon_k \Big(1 + \exp(-a_k(\tau_k - b_k))\Big) \Big\}. \tag{35}$$

The optimal solutions and the vectors should satisfy

$$\epsilon_k \Big(1 + \exp(-a_k(\tau_k^\star - b_k))\Big) - M_k = 0, \tag{36a}$$

$$\mu_k \Big(1 + \exp(-a_k(\tau_k^\star - b_k))\Big) - 1 = 0, \forall k \in \mathcal{K}, \tag{36b}$$

where $E_k^{\text{In},\star} = \rho^\star \left(\mathbf{h}_k^\dagger (\sum_{j=1}^{K} \mathbf{W}_j^\star + \mathbf{V}^\star)\mathbf{h}_k + \sigma_{k,S}^2 \right) \geq \tau_k^\star$.

Now, the objective function has the *log-concave* form and it can be solved given the sets $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$. The iterative update of the vector sets can be carried out in the following way. Let us define

the function $\mathcal{F}(\boldsymbol{\mu}, \boldsymbol{\epsilon}) = \left[\epsilon_k \left(1 + \exp(-a_k(\tau_k^\star - b_k)) \right) - M_k, \ldots, \mu_k \left(1 + \exp(-a_k(\tau_k^\star - b_k)) \right) - 1 \right]$, $\forall k \in \mathcal{K}$. The next set of values of $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$ can be updated by solving $\mathcal{F}(\boldsymbol{\mu}, \boldsymbol{\epsilon}) = \mathbf{0}$. Specifically, in the q-th iteration, we update them as:

$$\boldsymbol{\mu}^{q+1} = \boldsymbol{\mu}^q + \varpi^q \mathbf{p}^q, \ \boldsymbol{\epsilon}^{q+1} = \boldsymbol{\epsilon}^q + \varpi^q \mathbf{p}^q, \tag{37}$$

where $\mathbf{p}^q = [\mathcal{F}'(\boldsymbol{\mu}, \boldsymbol{\epsilon})]\mathcal{F}(\boldsymbol{\mu}, \boldsymbol{\epsilon})$, $\mathcal{F}'(\boldsymbol{\mu}, \boldsymbol{\epsilon})$ is the Jacobian matrix of $\mathcal{F}(\boldsymbol{\mu}, \boldsymbol{\epsilon})$, ϖ^q is the largest ϖ^l that satisfies $||\mathcal{F}(\boldsymbol{\mu}^q + \varpi^l \mathbf{p}^q, \boldsymbol{\epsilon}^q + \varpi^l \mathbf{p}^q)|| \leq (1 - t\varpi^l)||\mathcal{F}(\boldsymbol{\mu}, \boldsymbol{\epsilon})||$, $l = 1, 2, \ldots$, $0 < \varpi^l < 1$, and $0 < t < 1$ [37] [45].

A two-loop algorithm is proposed for solving the problem. The outer loop gives $\boldsymbol{\mu}$ and $\boldsymbol{\epsilon}$ as the inputs of the inner loop, while the inner loop finds $\mathbf{W}_k^\star$ and $\mathbf{V}^\star$. Observe that in (34), there is a coupling variable $\frac{\tau_k}{\rho}$, which is convex with a given ρ. Therefore, in the inner loop, we have to perform a one-dimensional search for ρ as well. The detailed algorithm is formulated in Algorithm 1.

B. Gaussian CSI Error Model

In this section, we formulate the maximum harvested energy under the gaussian CSI error model formulated is as follows:

$$\mathbf{P}_8: \quad \max_{\mathbf{W}_k \in \mathbb{C}^{M \times M}, \mathbf{V} \in \mathbb{C}^{M \times M}, \rho} \sum_{k=1}^{K} E_k^{\text{Practical}} \tag{38a}$$

$$\text{s.t.} \quad (22b), (22d), (12e), (12f), (12g), (12h). \tag{38b}$$

We first simplify the objective function and then a new approximation will be formulated based on the *Bernstein-type Inequality* [43][44]. By involving a simple transformation, we arrive at:

$$\mathbf{P}_9: \quad \max_{\mathbf{W}_k, \mathbf{V}, \rho} \sum_{k=1}^{K} \mu_k \left\{ M_k - \epsilon_k \left(1 + \exp(-a_k(\tau_k - b_k)) \right) \right\}, \tag{39a}$$

$$\text{s.t.} \quad \Pr(E_k^{\text{In}} \geq \tau_k) \geq 1 - \varpi, \ \ \forall \Delta \mathbf{h}_k \sim \mathcal{CN}(0, \mathbf{H}_k), \ \forall k \in \mathcal{K}, \tag{39b}$$

$$(22b), (22d), (12e), (12f), (12g), (12h). \tag{39c}$$

Observe however that the transformation from (38a) to (39a) and (39b) is not exactly equivalent. The equivalent form should let $E_k^{\text{In}} \geq \tau_k$ in (39b). However, by setting ϖ to be a very small value, the transformation can be valid and it is also consistent with our gaussian CSI error model.

Algorithm 1 Robust Precoding Design for EH Maximization Problem

---

1: **Input**: Minimum required data rate R_k of SU k, noise power $\sigma_{k,S}^2$ and σ_D^2, channel uncertainty φ_k^2 and ψ_n^2, maximum allowed interference power $P_{n,p}$ for PU n, maximum BS transmitted power P_B, and randomly generated estimated channel $\widehat{\mathbf{h}}_k$ and $\widehat{\mathbf{g}}_n$.

2: **Initialisation**: Iteration number $q = 0, p = 1$, initial value of ρ as ρ_{start}, step s, end value ρ_{end}, $\boldsymbol{\mu}^0$, and $\boldsymbol{\epsilon}^0$, loop stop criteria m_{th}.

3: **One-dimensional Search**:

4: **for** $\rho = \rho_{\text{start}} : s : \rho_{\text{end}}$ **do**

5:    **repeat**: {Outer Loop}

6:       Solve for the optimization problem $\mathbf{P}_7$: {Inner Loop}

7:       **if** ($\mathbf{P}_7$ is feasible) **then**

8:          Obtain $\mathbf{W}_k^q$ and $\mathbf{V}^q$.

9:       **else**

10:         Break from the outer loop.

11:       **end if**

12:       Update $\boldsymbol{\mu}^{q+1}$ and $\boldsymbol{\epsilon}^{q+1}$ according to (37), then let $q = q + 1$.

13:    **until** $\left| \mu_k^{q+1} \left\{ M_k - \epsilon_k^{q+1} \left(1 + \exp(-a_k(\tau_k - b_k)) \right) \right\} \right| < m_{th}$

14:    Calculate $E_{\text{sum}}^i = \sum_k E_k^{\text{Practical}}$, then let $i = i + 1, q = 0$.

15: **end for**

16: Find the maximum value among all E_{sum}^i, and the precoding and energy matrix.

17: **Output**: Use either of the methods to get the precoding vector $\mathbf{w}_k^{\text{opt}}$ and $\mathbf{V}^{\text{opt}}$.

By applying the *Bernstein-type Inequality I* [43], (39b) becomes,

$$\text{Tr}\left(\mathbf{H}_k^{1/2} \boldsymbol{\Sigma} \mathbf{H}_k^{1/2} \right) - \sqrt{-2\ln(\varpi)}\upsilon_{1k,s} + \ln(\varpi)\upsilon_{2k,s} + c_{k,s} \geq 0, \tag{40a}$$

$$c_{k,s} = \widehat{\mathbf{h}}_k^\dagger \boldsymbol{\Sigma} \widehat{\mathbf{h}}_k + \sigma_{k,S}^2 - \frac{\tau_k}{\rho}, \tag{40b}$$

$$\left\| \begin{bmatrix} \text{vec}\left(\mathbf{H}_k^{1/2} \boldsymbol{\Sigma} \mathbf{H}_k^{1/2} \right) \\ \sqrt{2}\mathbf{H}_k^{1/2} \boldsymbol{\Sigma} \widehat{\mathbf{h}}_k \end{bmatrix} \right\| \leq \upsilon_{1k,s}, \tag{40c}$$

$$\upsilon_{2k,s}\mathbf{I} + \left(\mathbf{H}_k^{1/2} \boldsymbol{\Sigma} \mathbf{H}_k^{1/2} \right) \succeq \mathbf{0}, \upsilon_{2k,s} \geq 0, \ \forall k \in \mathcal{K}, \tag{40d}$$

where $\upsilon_{1k,s}$ and $\upsilon_{2k,s}$, $k \in \mathcal{K}$ are slack variables.

We also relax the problem by dropping the constraint that the rank of $\mathbf{W}_k$ must be 1, and the optimization problem becomes

$$\mathbf{P}_{10} : \max_{\substack{\mathbf{W}_k,\mathbf{V},\rho,\{v_{1i,k}\}, \\ \{v_{2i,k}\},\{v_{1k,s}\},\{v_{2k,s}\},\{v_{1,n}\},\{v_{2,n}\}}} \sum_{k=1}^{K} \mu_k \Big\{ M_k - \epsilon_k \Big(1 + \exp(-a_k(\tau_k - b_k)) \Big) \Big\}, \quad (41a)$$

$$\text{s.t.} \quad (40), (25), (30), \quad (41b)$$

$$(12e), (12f), (12g). \quad (41c)$$

Still, the coupling variable in (40) can be tackled by fixing ρ. A similar one-dimensional search for ρ, together with a two-loop algorithm can solve $\mathbf{P}_{10}$, the detailed step will be omitted here for space considerations.

V. SIMULATION RESULTS

In this section, we present our simulation results for characterizing the performance of the proposed robust beamforming conceived with NOMA under both the bounded and the gaussian CSI estimation error models. Unless otherwise stated, the parameters are chosen as in Table.I.

Parameters	Values
Number of SUs and PUs	$K = 3$, $N = 2$
Noise powers	$\sigma_{k,S}^2 = 0.1$, $\sigma_D^2 = 0.01$
Minimum required EH power	$P_{k,s} = 0.01$ Watt
Maximum tolerable interference of PUs	$P_{n,p} = -18$ dBm
Estimated channel gains	$\hat{\mathbf{h}}_k \sim \mathcal{CN}(0, 0.8\mathbf{I})$, $\hat{\mathbf{g}}_n \sim \mathcal{CN}(0, 0.1\mathbf{I})$
Outage probability threshold	$\xi_k = \xi_{k,s} = \xi_{n,p} = 0.05$
Gaussian CSI estimation	$\varpi_k^2 = 0.001$, $\varpi_n^2 = 0.0001$. [31]
Non-linear EH model	$M_k = 24$ mW, $a_k = 150$, and $b_k = 0.014$ [46]

TABLE I: Simulation Parameters

To achieve a fair comparison between the two channel estimation error models. If the covariance matrices of the channel estimation error vector $\Delta\mathbf{h}_k$ and $\Delta\mathbf{g}_n$ under the gaussian model are $\varpi_k^2\mathbf{I}$ and $\varpi_n^2\mathbf{I}$, respectively, then the bounded CSI radius under the worst-case scenario of φ_k and ψ_n should be [41]

$$\varphi_k = \sqrt{\frac{\varpi_k^2 F_{2M}^{-1}(1 - \xi_k)}{2}}, \ \psi_n = \sqrt{\frac{\varpi_n^2 F_{2M}^{-1}(1 - \xi_{n,p})}{2}}, \quad (42a)$$

where $F_{2M}^{-1}(\cdot)$ represents the complimentary cumulative distribution function (CCDF) of the Chi-square distribution with $2M$ degrees of freedom.

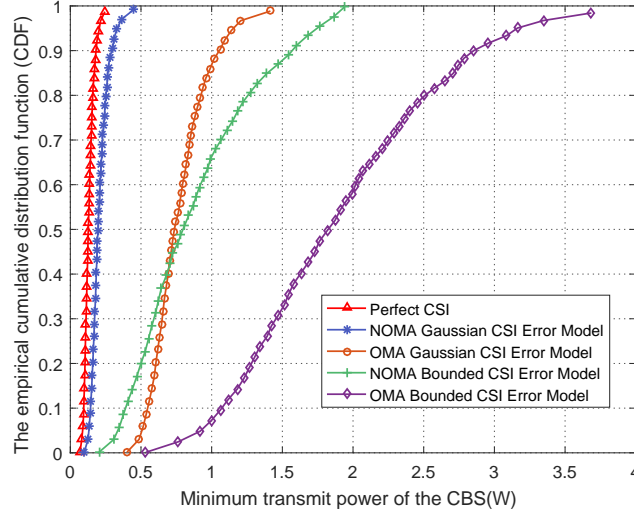## A. Power Minimization Problem

Fig. 2: The empirical CDF of the minimum transmit power of the CBS under different channel conditions. CBS antenna number $M = 10$, $P_B = 2$ Watts, $R_{\min} = 1$ bit/s/Hz.

Fig. 2 shows the empirical CDFs of the minimum transmit power of the CBS for both the imperfect CSI estimation error models. The maximum power P_B is set to 2 Watts. For comparison purposes, we also include the case of OMA, as it represents the traditional access technology. Notice that in order to reduce the inter-user interference, each OMA user exclusively uses one time slot. Thus, a total of K time slots are required instead of just one in our scheme. To obtain a fair comparison, each SU's achievable data rate should be averaged over all K time slots, which becomes $R_k^{\text{OMA}} = \frac{1}{K} \log_2(1 + \text{SINR}_k^{\text{OMA}})$. The reduced interference can be achieved at the cost of lower spectral and energy efficiency. We also observe that under both channel error models, the performance of NOMA is better than that with OMA. The reason is that, with OMA, the lower spectral efficiency makes the constraint of SU data rate harder to be satisfied, and hence the CBS has to apply a higher transmission power to compensate for that, which leads to a much higher energy consumption. Fig. 2 is generated from 1,000 independent realizations of
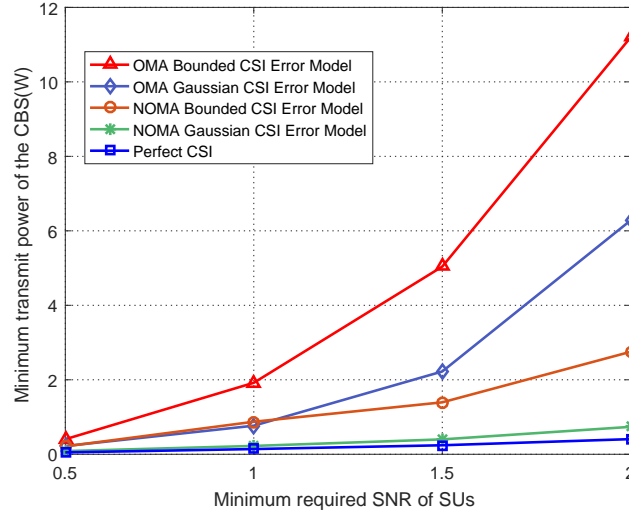
Fig. 3: The minimum transmit power of the CBS vs. the required SNR of SUs for $M = 10$, $P_B = 8$ Watts.

different channel conditions. As expected, the performance under perfect CSI is the best, since no additional power is used to compensate for the channel uncertainties. Furthermore, in both the OMA and NOMA schemes, the performance under the gaussian CSI channel estimation is better than that under the bounded CSI channel estimations, as bounded CSI represents the worst-case scenario. Observe that the minimum power in the OMA bounded CSI is over 2 Watts since we only limit the power of each time slot to 2 Watts and it is very likely that the total power over K slots will beyond that limit.

Fig. 3 shows the minimum transmit power of the CBS as a function of the minimum required SNR of SUs, $\gamma_{k,\min}$. As the SNR increases, the power increases under all CSI cases. Also, perfect CSI requires the least power, followed by NOMA relying on the gaussian CSI error model, NOMA in the bounded CSI model, OMA gaussian CSI model, and OMA bounded CSI model. Besides, compared to OMA, the CBS power in NOMA grows more slowly. In the parameter setting, $\gamma_{k,\min}$ plays a more important role in the constraints. For $\gamma_{k,\min} = 2$ in the NOMA case, the equivalent SNR for OMA will be 26. Thus, the gap between OMA and NOMA further increases with the required SNR.

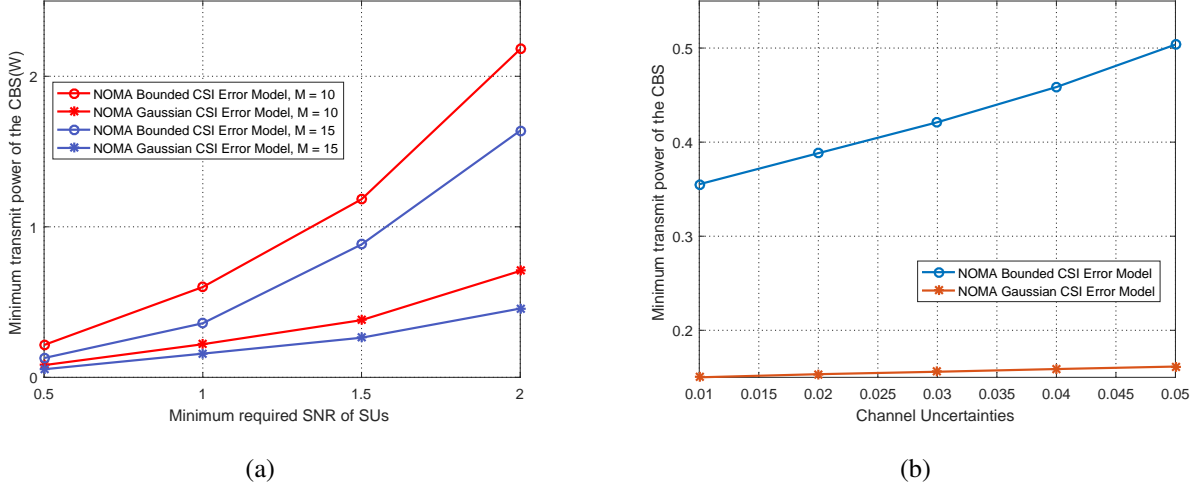The impact of the CBS antenna number is illustrated in Fig. 4(a), where the performance

Fig. 4: (a) Impact of the number of CBS antennas on the minimum transmitted power required in two imperfect CSI scenarios. (b) Impact of channel uncertainties ψ_n and φ_k on the overall minimum transmit power of the CBS, $M = 15$, $R_{\min} = 1$ bit/s/Hz, $P_B = 8$ Watts.

with different CBS antenna numbers and channel uncertainties are plotted. Specifically, Fig. 4(a) illustrates how the number of antennas affects the overall performance. The power required increases, when the SNR of SUs grows, regardless of how many antennas are mounted at the CBS. It is also observed that the minimum power required decreases when the number of antennas increases, since a larger number of antennas results in a higher degree of freedom (DoF). Besides, we also notice that the performance under the gaussian error model is better than that under the bounded channel error case. In Fig. 4(b), the impact of channel uncertainties is illustrated. We set $\psi_n^2 = \varphi_k^2 = [0.01 : 0.05]$, the corresponding covariance matrices in gaussian CSI estimation error scenario also change according to (42). Clearly, channel estimation error affects the bounded CSI scenario the most, since under worst-case CSI, the channel estimation error channel becomes worse, thus it needs more power to meet the data rate constraints. Nevertheless, the channel estimation error does not have much impact on the gaussian channel estimation error scenario.
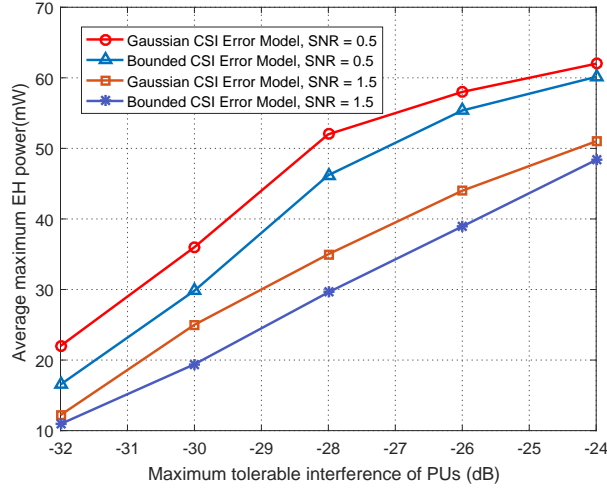
Fig. 5: Average maximum EH power under different interferences tolerated by the PUs, $M = 10$.

B. Energy Harvesting Maximization Problem

In this subsection, we present results for the maximum EH as our objective function. The CBS power is $P_B = 2$ Watts. Fig. 5 characterizes the average maximum EH power vs. the interference tolerated by the to PUs. One can observe that the energy harvested monotonically increases, when the maximum interference tolerated by the PUs grows, where a higher $P_{n,p}$ allows for a larger transmission power, leading to the increase of the harvested energy. Additionally, we can see that under the gaussian channel estimation error, the performance is better than that under the bounded channel estimation error case. When the channel conditions are better, less power is required for satisfying the data rate requirements. Hence more power can be reserved for EH. This also explains that when the required SNR is low, a high EH power can be achieved.

The impact of minimum SNRs required by the SUs is illustrated in Fig. 6. The number of CBS antennas is $M = 10$ and the interference threshold $P_{n,p}$ is set to -24 dBm. We also list the results for the OMA cases. As expected, the average maximum EH power decreases, when the required SNR increases. Similar observations show that under perfect CSI, the performance is the best, while the OMA bounded CSI estimation scenario is the worst. Moreover, we can see that the maximum EH power decreases significantly when the SNR grows. This is because more power has to be used for information detection, which leaves less power for energy harvesting.
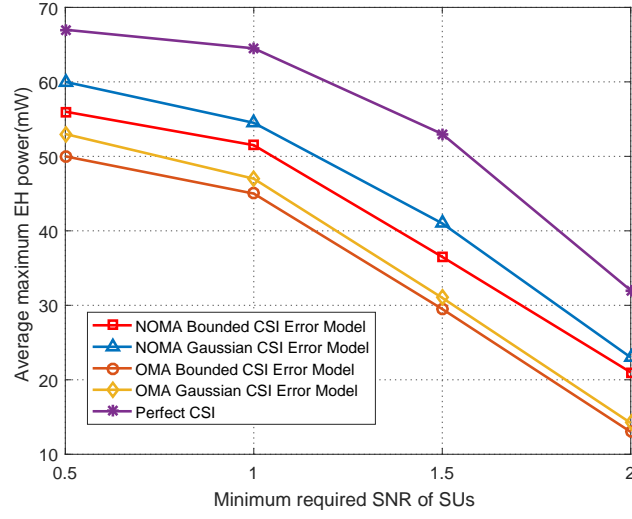
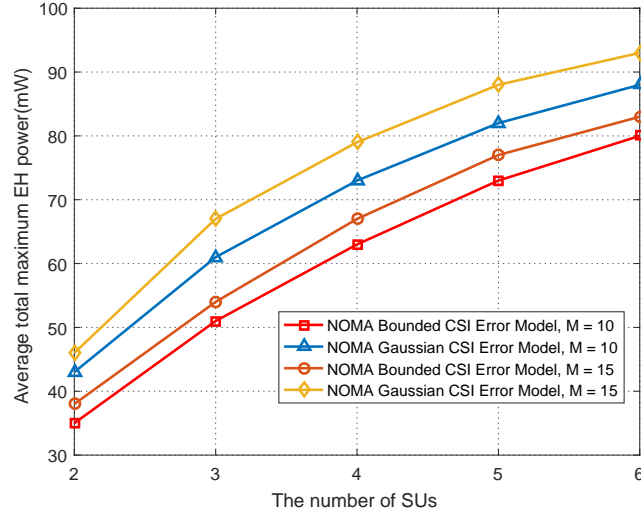Fig. 6: Average maximum EH power vs. the minimum SNR required by the SUs, $M = 10$.

Fig. 7: Average total EH power vs. the number of SUs for $P_{n,p} = -24$ dBm, $r_{\min} = 1$ bit/s/Hz.

Fig. 7 shows the average total EH power vs. the number of SUs. It can be observed that the total EH power grows, when the number of SUs increases, since more nodes participate in the harvesting process. Additionally, we can see that when the number of antennas is higher, more

EH power can be achieved. This is because more antennas give a higher system DoF, therefore less power is sufficient for information detection.

VI. CONCLUSIONS

In this paper, we considered MISO-NOMA CR-aided SWIPT under both the bounded and the Gaussian CSI estimation error model. To make the energy harvesting investigations more realistic, a non-linear EH model was applied. Robust beamforming and power splitting control were jointly designed for achieving the minimum transmission power and maximum EH. We transformed the non-convex minimum transmission power optimization problems into a convex form while applying a one-dimensional search algorithm to solve the maximum EH problem. Our simulation results showed that the performance achieved by using NOMA is better than that obtained by using the traditional OMA. Furthermore, a performance gain can be obtained under the gaussian CSI estimation error model over the bounded CSI error model. For the future research direction, the system model can be generalized to account for more use cases, for example, the physical layer security and the interference from multi-cell. Additionally, for the Gaussian CSI error model, the rank of the solution is not fully characterized in this work yet, which leaves room for improvement in the future.

APPENDIX A

PROOF OF THEOREM 3.2

Proof: To prove the Lemma, we first consider the Karush-Kuhn-Tucker (KKT) conditions of $\mathbf{P}_3$. Specifically, after some simple algebraic manipulations, (13) can be rewritten as

$$\begin{bmatrix} \alpha_{i,k}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & t_{i,k} \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ \widehat{\mathbf{h}}_i^\dagger \end{bmatrix} \mathbf{C}_k \begin{bmatrix} \mathbf{I} & \widehat{\mathbf{h}}_i \end{bmatrix} + \begin{bmatrix} -\gamma_{k,\min}\sum_{j=1}^{k-1}\mathbf{W}_j & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix} \succeq \mathbf{0}, \qquad (43)$$
$$\forall k \in \mathcal{K}, \ i = \{k, k+1, \ldots, K\},$$

where we have $t_{i,k} = -\alpha_{i,k}\varphi_k^2 - \gamma_{k,\min}\left(\sigma_{k,S}^2 + \frac{\sigma_D^2}{(1-\rho)}\right)$.

Similarly, (15) and 17 can be rewritten as

$$\begin{bmatrix} \beta_n\mathbf{I} & \mathbf{0} \\ \mathbf{0} & -\beta_n\psi_n^2 + P_{n,p} \end{bmatrix} - \begin{bmatrix} \mathbf{I} \\ \widehat{\mathbf{g}}_n^\dagger \end{bmatrix} \mathbf{\Sigma} \begin{bmatrix} \mathbf{I} & \widehat{\mathbf{g}}_n \end{bmatrix} \succeq \mathbf{0}, \ \forall n \in \mathcal{N}, \qquad (44)$$

and

$$\begin{bmatrix} \theta_k \mathbf{I} & \mathbf{0} \\ \mathbf{0} & m_k \end{bmatrix} + \begin{bmatrix} \mathbf{I} \\ \widehat{\mathbf{h}}_k^\dagger \end{bmatrix} \boldsymbol{\Sigma} \begin{bmatrix} \mathbf{I} & \widehat{\mathbf{h}}_k \end{bmatrix} \succeq \mathbf{0}, \ \forall k \in \mathcal{K}, \tag{45}$$

respectively, where $m_k = -\theta_k \varphi_k^2 + \sigma_{k,S}^2 - \frac{\tau_k}{\rho}$.

For notational simplicity, we let $\mathbf{X}_i = \begin{bmatrix} \mathbf{I} & \widehat{\mathbf{h}}_i \end{bmatrix}$ and $\mathbf{Y}_n = \begin{bmatrix} \mathbf{I} & \widehat{\mathbf{g}}_n \end{bmatrix}$. Furthermore, we denote $\mathbf{A}_{i,k} \in \mathbb{C}_+^{(M+1) \times (M+1)}$, $\mathbf{B}_k \in \mathbb{C}_+^{(M+1) \times (M+1)}$, $\mathbf{D}_n \in \mathbb{C}_+^{(M+1) \times (M+1)}$, $z \in \mathbb{R}_+$, and $\mathbf{E}_k \in \mathbb{C}_+^{(M) \times (M)}$ as the KKT multiplier. Then the Lagrange dual function $\mathcal{L}$ can be expressed as

$$\mathcal{L}(\mathbf{W}_k, \mathbf{V}, \mathbf{A}_{i,k}, \mathbf{B}_k, \mathbf{D}_n, z, \kappa) = \text{Tr}(\boldsymbol{\Sigma}) - \sum_{i,k} \text{Tr}(\mathbf{A}_{i,k} \mathbf{X}_i^\dagger \mathbf{C}_k \mathbf{X}_i) - \sum_{i,k} \text{Tr}(\mathbf{A}_{i,k} \mathbf{M}_k) \tag{46}$$

$$+ \sum_n \text{Tr}(\mathbf{D}_n \mathbf{Y}_n^\dagger \boldsymbol{\Sigma} \mathbf{Y}_n) - \sum_k \text{Tr}\Big(\mathbf{B}_k \mathbf{X}_k^\dagger \boldsymbol{\Sigma} \mathbf{X}_k\Big) + z\Big(\text{Tr}(\boldsymbol{\Sigma}) - P_B\Big) - \sum_k \text{Tr}(\mathbf{E}_k \mathbf{W}_k) + \kappa,$$

where $\mathbf{M}_k = \begin{bmatrix} -\gamma_{k,\min} \sum_{j=1}^{k-1} \mathbf{W}_j & \mathbf{0} \\ \mathbf{0} & 0 \end{bmatrix}$ and κ are the terms independent of $\mathbf{W}_k$. Taking the partial derivative of the dual function with respect to $\mathbf{W}_k$, we have

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}_k} = \mathbf{I} - \sum_i \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger + \gamma_{k,\min} \sum_i \sum_{j=1}^{k-1} \mathbf{X}_i \mathbf{A}_{i,j} \mathbf{X}_i^\dagger + \sum_i \gamma_{k,\min} \sum_{j=k+1}^{K} \mathbf{A}_{i,j} \tag{47}$$

$$+ \sum_n \mathbf{Y}_n \mathbf{D}_n \mathbf{Y}_n^\dagger - \sum_k \mathbf{X}_k \mathbf{B}_k \mathbf{X}_k^\dagger + z\mathbf{I} - \mathbf{E}_k = \mathbf{0}.$$

Additionally, the dual problem has to satisfy the completeness slackness

$$\Big(\begin{bmatrix} \alpha_{i,k} \mathbf{I} & \mathbf{0} \\ \mathbf{0} & t_{i,k} \end{bmatrix} + \mathbf{X}_i^\dagger \mathbf{C}_k \mathbf{X}_i + \mathbf{M}_k \Big) \mathbf{A}_{i,k} = \mathbf{0}, \tag{48a}$$

$$\mathbf{E}_k \mathbf{W}_k = \mathbf{0}, \forall k \in \mathcal{K}, i = \{k+1, \dots, K\}, \forall n \in \mathcal{N}. \tag{48b}$$

Right-multiplying $\mathbf{W}_k$ with (47), and substituting (48b), we arrive at:

$$\Big(\sum_i \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger + \sum_k \mathbf{X}_k \mathbf{B}_k \mathbf{X}_k^\dagger \Big) \mathbf{W}_k = \Big[(1+z)\mathbf{I} + \gamma_{k,\min} \sum_i \sum_{j=1}^{k-1} \mathbf{X}_i \mathbf{A}_{i,j} \mathbf{X}_i^\dagger \tag{49}$$

$$+ \gamma_{k,\min} \sum_i \sum_{j=k+1}^{K} \mathbf{A}_{i,j} + \sum_n \mathbf{Y}_n \mathbf{D}_n \mathbf{Y}_n^\dagger \Big] \mathbf{W}_k.$$

Since all the KKT multipliers are positive numbers or positive semi-definite matrices, we can readily verify $\left\{ (1+z)\mathbf{I} + \gamma_{k,\min} \sum_i \sum_{j=1}^{k-1} \mathbf{X}_i \mathbf{A}_{i,j} \mathbf{X}_i^\dagger + \gamma_{k,\min} \sum_i \sum_{j=k+1}^{K} \mathbf{A}_{i,j} + \sum_n \mathbf{Y}_n \mathbf{D}_n \mathbf{Y}_n^\dagger \right\} \succeq \mathbf{0}$. Thus it is non-singular. Left-multiplying a non-singular matrix with $\mathbf{W}_k$ does not change the rank of $\mathbf{W}_k$. Therefore, we have

$$\begin{aligned}
\text{Rank}(\mathbf{W}_k) &= \text{Rank}\Big((\sum_i \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger + \sum_k \mathbf{X}_k \mathbf{B}_k \mathbf{X}_k^\dagger) \mathbf{W}_k \Big) \quad (50) \\
&\leq \min\Big\{ \text{Rank}\Big(\sum_i \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger + \sum_k \mathbf{X}_k \mathbf{B}_k \mathbf{X}_k^\dagger \Big), \text{Rank}(\mathbf{W}_k) \Big\}.
\end{aligned}$$

Next, we show that the rank of $(\sum_i \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger)$ is 1. By summing (48a) in terms of the index i, then left-multiplying $\begin{bmatrix} \mathbf{I}_M & \mathbf{0} \end{bmatrix}$ and right-multiplying $\mathbf{X}_i^\dagger$, we have

$$\sum_i \alpha_{i,k} \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger - \sum_i \alpha_{i,k} \begin{bmatrix} \mathbf{0}_M & \mathbf{h}_i \end{bmatrix} \mathbf{A}_{i,k} \mathbf{X}_i^\dagger + \sum_i \mathbf{C}_k \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger \quad (51)$$

$$+ \sum_i (-\gamma_{k,\min} \sum_{j=1}^{k-1} \mathbf{W}_j) \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger - \sum_i (-\gamma_{k,\min} \sum_{j=1}^{k-1} \mathbf{W}_j) \begin{bmatrix} \mathbf{0}_M & \mathbf{h}_i \end{bmatrix} \mathbf{A}_{i,k} \mathbf{X}_i^\dagger = \mathbf{0}.$$

After a simple transformation, we have

$$\sum_i \Big(\alpha_{i,k} \mathbf{I} + \mathbf{C}_k - \gamma_{k,\min} \sum_{j=1}^{k-1} \mathbf{W}_j \Big) \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger = \sum_i (\alpha_{i,k} \mathbf{I} - \gamma_{k,\min} \sum_{j=1}^{k-1} \mathbf{W}_j) \begin{bmatrix} \mathbf{0}_M & \mathbf{h}_i \end{bmatrix} \mathbf{A}_{i,k} \mathbf{X}_i^\dagger. (52)$$

From the fact that (13) is a positive semidefinite matrix, $\Big(\alpha_{i,k} \mathbf{I} + \mathbf{C}_k - \gamma_{k,\min} \sum_{j=1}^{k-1} \mathbf{W}_j \Big)$ would be a non-singular matrix, thus the rank of the left term of the above equation is the same as $\sum_i \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger$. Also, it is easy to verify that the right term has a rank 1.

Similarly, we can prove that $\text{Rank}(\sum_n \mathbf{Y}_n \mathbf{D}_n \mathbf{Y}_n^\dagger) = 1$. Then, the following equation holds:

$$\begin{aligned}
\text{Rank}\Big(\sum_i \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger + \sum_k \mathbf{X}_k \mathbf{B}_k \mathbf{X}_k^\dagger \Big) &\leq \text{Rank}(\sum_i \mathbf{X}_i \mathbf{A}_{i,k} \mathbf{X}_i^\dagger) + \text{Rank}(\sum_n \mathbf{Y}_n \mathbf{D}_n \mathbf{Y}_n^\dagger) \\
&\leq 2, \quad (53)
\end{aligned}$$

which proves the theorem. ∎

REFERENCES

[1] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan and L. Hanzo, "Nonorthogonal Multiple Access for 5G and Beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347-2381, Dec. 2017.

[2] D. Zhang, Y. Liu, Z. Ding, Z. Zhou, A. Nallanathan and T. Sato, "Performance Analysis of Non-Regenerative Massive-MIMO-NOMA Relay Systems for 5G," *IEEE Trans. Commun.*, vol. 65, no. 11, pp. 4777-4790, Nov. 2017.

[3] X. Chen, Z. Zhang, C. Zhong, and D.W. K. Ng, "Exploitation multiple-antennas techniques for non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2207-2220, Oct. 2017.

[4] H. Sun, B. Xie, R. Q. Hu, and G. Wu, "Non-orthogonal multiple access with SIC error propagation in downlink wireless MIMO networks," in *Proc. IEEE VTC Fall 2016, invited paper*. Sept. 2016.

[5] Z. Ding, Z. Zhao, M. Peng and H. V. Poor, "On the Spectral Efficiency and Security Enhancements of NOMA Assisted Multicast-Unicast Streaming," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 3151-3163, July 2017.

[6] Y. Xu, H. Sun, R. Q. Hu, and Y. Qian, "Cooperative nonorthogonal multiple access in heterogeneous networks," in *Proc. IEEE GlOBECOMM 2015*, San Diego, Dec. 2015.

[7] H. Sun, Y. Xu, and R. Q. Hu, "A NOMA and MU-MIMO supported cellular network with underlaid D2D communications", in *Proc. IEEE VTC spring 2016*, May. 2016.

[8] Z. Zhang, H. Sun, and R. Q. Hu, "Downlink and uplink non-orthogonal multiple access in a dense wireless network", *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2771-2784, Dec. 2017.

[9] F. Zhou, N. C. Beaulieu, Z. Li, J. Si, and P. Qi, "Energy-efficient optimal power allocation for fading cognitive radio channels: Ergodic capacity, outage capacity and minimum-rate capacity," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2741-2755, Apr. 2016.

[10] A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proc. IEEE*, vol. 97, no. 5, pp. 894-914, May 2009.

[11] Y. Liu, Z. Ding, M. Elkashlan, and J. Yuan, "Nonorthogonal multiple access in large-scale underlay cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 10152-10157, Dec. 2016.

[12] L. Lv, J. Chen, Q. Li, and Z. Ding, "Design of cooperative non-orthogonal multicast cogntive radio multiple access for 5G systems: User scheduling and performance anaysis," *IEEE Trans. Commun.*, vol. 65, no. 6, pp. 2641-2656, June 2017.

[13] L. Lv, Q. Ni, Z. Ding and J. Chen, "Application of non-orthogonal multiple access in cooperative spectrum-sharing networks over Nakagami-m fading channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5506-5511, June 2017.

[14] X. Huang, T. Han, and N. Ansari, "On green energy powered cognitive radio networks," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 827-842, Second Quarter, 2015.

[15] X. Lu, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Wireless networks with RF energy harvesting: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 17, pp. 757-789, Second Quarter, 2015.

[16] E. Boshkovska, D. W. K. Ng, N. Zlatanov, A. Koelpin, and R. Schober, "Robust resource allocation for MIMO wireless powered communication networks based on a non-linear EH model," *IEEE Trans. Commun.*, vol. 65, no. 5, pp. 1984-1999, May 2017.

[17] H. Sun, F. Zhou, and Z. Zhang, "Robust Beamforming Design in a NOMA Cognitive Radio Network Relying on SWIPT", in *Proc. IEEE ICC*, Kansas City, MO, USA, 2018.

[18] G. Pan et al., "On Secrecy Performance of MISO SWIPT Systems With TAS and Imperfect CSI," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3831-3843, Sept. 2016.

[19] Z. Chu, H. Xing, M. Johnston, and S. Le Goff, "Secrecy rate optimizations for a MISO secrecy channel with multiple multi-antenna eavesdroppers," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 283-297, Jan. 2016.

[20] D. W. K. Ng, E. S. Lo, and R. Schober, "Multi-objective resource allocation for secure communication in cognitive radio

networks with wireless information and power transfer," *IEEE Trans. Veh. Technol.*, vol. 20, no. 2, pp. 328-331, Feb. 2016.

[21] P. V. Tuan and I. Koo, "Optimal multiuser MISO beamforming for power-splitting SWIPT cognitive radio networks," *IEEE Access*, vol. 5, pp. 14141-14153, Jul. 2017.

[22] Z. Hu, N. Wei, and Z. Zhang, "Optimal resource allocation for harvested energy maximization in wideband cognitive radio network with SWIPT. " *IEEE Access* vol. 5, 23383-23394, Aug. 2017.

[23] L. Mohjazi, I. Ahmed, S. Muhaidat, M. Dianati and M. Al-Qutayri, "Downlink beamforming for SWIPT multi-user MISO underlay cognitive radio networks," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 434-437, Feb. 2017.

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

[25] N. T. Do, D. B. da Costa, T. Q. Duong, and B. An, "A BNBF user selection scheme for NOMA-based cooperative relaying systems with SWIPT," *IEEE Commun. Lett.*, vol. 21, no. 3, pp. 664-667, Mar. 2017.

[26] Z. Yang, Z. Ding, P. Zhi, and N. A. Dhahir, "The impact of power allocation on cooperative non-orthogonal multiple access networks with SWIPT," *IEEE Trans. Wireless Commun.*, to be published, 2017.

[27] P. D. Diamantoulakis, K. N. Pappi, Z. Ding, and G. K. Karagiannidis, "Wireless-powered communications with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8422-8436, Oct. 2016.

[28] P. D. Diamantoulakis, K. N. Pappi, G. K. Karagiannidis, H. Xing, and A. Nallanathan, "Joint downlink/uplink design for wireless powered networks with interference, "*IEEE Access*, vol. 5, pp. 1534-1547, 2017.

[29] Y. Xu, C. Shen, Z. Ding, X. Sun, S. Yan, G. Zhu, and Z. Zhong, "Joint beamforming and power-splitting control in downlink cooperative SWIPT NOMA systems", *IEEE Trans. Signal Process.*,vol 65, no. 18, pp. 4874-4886. Sept. 2017.

[30] L. Mohjazi, I. Ahmed, S. Muhaidat, M. Dianati, and M. Al-Qutayri, "Downlink beamforming for SWIPT multi-user MISO underlay cogntive radio networks," *IEEE Commun. Lett.*, vol. 21, no.2, pp. 434-437, Feb. 2017.

[31] F. Zhou, Z. Li, J. Cheng, Q. Li, and J. Si, "Robust AN-aided beamforming and power splitting design for secure MISO cognitive radio with SWIPT," *IEEE Trans. Wireless Commun.*, vol. 16, no. 4, pp. 2450-2464, April 2017.

[32] B. Fang, Z. Qian, W. Zhang, and W. Shao, "AN-aided secrecy precoding for SWIPT in cognitive MIMO broadcast channels," *IEEE Commun. Lett.*, vol. 19, no. 9, pp. 1632-1635, Sept. 2015.

[33] C. Xu, Q. Zhang, Q. Li, Y. Tan, and J. Qin, "Robust transceiver design for wireless information and power transmission in underlay MIMO cognitive radio networks," *IEEE Commun. Lett.*, vol. 18, no. 9, pp. 1665-1668, Sept. 2014.

[34] F. Zhou, N. C. Beaulieu, J. Cheng, Z. Chu, and Y. Wang, "Robust max-min fairness resource allocation in sensing based wideband cognitive radio with SWIPT: Imperfect channel sensing," *IEEE Syst. Journal*, vol. PP, no. 99, pp. 1-12, Jun. 2017.

[35] T. A. Le Q.-T. Vien H. X. Nguyen D. W. K. Ng R. Schober "Robust chance-constrained optimization for power-efficient and secure SWIPT systems" *IEEE Trans. Green Commun. Netw.* vol. 1 no. 3 pp. 333-346 Sep. 2017.

[36] Y. Yuan and Z. Ding, "Outage constrained secrecy rate maximization design with SWIPT in MIMO-CR systems," *IEEE Trans. Veh. Technol.*, to be published, 2017.

[37] E. Boshkovska, D. W. K. Ng, L. Dai and R. Schober, "Power-Efficient and Secure WPCNs with Hardware Impairments and Non-Linear EH Circuit," *IEEE Trans. Commun.*, to be published.

[38] E. Boshkovska, A. Keolpin, D. W. K. Ng, N. Zlatanov, and R. Schober, "Robust beamforming for SWIPT systems with non-linear energy harvesting model," in *Proc. IEEE Inter. Signal Processing Adances in Wireless Communications*, Edinburgh, UK, 2016.

[39] K. Xiong, B. Wang, and K. J. Ray Liu, "Rate-energy region of SWIPT for MIMO broadcasting under nonlinear energy harvesting model," *IEEE Trans. Wireless Commun.*, to be published, 2017.

[40] E. Boshkovska, N. Zlatanov, L. Dai, D. W. K. Ng, and R. Schober, "Secure SWIPT networks based on a non-linear energy harvesting model," in *Proc. IEEE WCNC 2017*, San Francisco, CA, USA, 2017.

[41] Y. Wang, Y. Wang, F. Zhou, Y. Wu, and H. Zhou, "Resource allocation in wireless powered cognitive radio networks based on a practical non-linear energy harvesting model," *IEEE Access*, to be published, 2017.

[42] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[43] K. Y. Wang, A. M. C. So, T. H. Chang, W. K. Ma, and C. Y. Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: tractable approximations by conic optimization," *IEEE Trans. Signal Process.*, vol. 62, no. 21, pp. 5690-5705, Nov. 2014.

[44] I. Bechar, "A Bernstein-type inequality for stochastic processes of quadratic forms of Gaussian variables." [Online]. Avaliable: http://arxiv.org/abs/0909.3595.

[45] Y. Jong, "An efficient global optimization algorithm for nonlinear sum-of-ratios problem," May 2012. [Online]. Available: http://www.optimization-online.org/DB_FILE/2012/08/3586.pdf

[46] E. Boshkovska, X. Chen, L. Dai, D. W. K. Ng, and R. Schober, "Max-min fair beamforming for SWIPT systems with non-linear EH model", [Online]. Available: https://arxiv.org/pdf/1705.05029.pdf