Cognitive Non-orthogonal Multiple Access With Energy Harvesting: An Optimal Resource Allocation Approach

Fudong Li, Hai Jiang, *Senior Member, IEEE,* Rongfei Fan, *Member, IEEE*, and Peng Tan, *Member, IEEE*

***Abstract*—We consider a cognitive radio system, in which a secondary transmitter harvests energy from a primary transmitter's RF signal. The secondary transmitter, which provides decode-and-forward relaying service for the primary system, transmits its own data by using downlink non-orthogonal multiple access (NOMA). A time-switching protocol is used by the secondary transmitter to harvest energy and decode the primary transmitter's information. Our objective is to achieve maximal secondary throughput, by optimally selecting the time portion used for energy harvesting and the secondary transmitter's power allocation in NOMA transmission. Two optimization problems are formulated, in which the secondary receiver performs or does not perform successive interference cancellation (SIC), respectively. Although the two problems are nonconvex, we devise a method to transform the problems into equivalent problems under difference cases. Then we theoretically prove that the objective functions of the equivalent problems are quasiconcave, based on which we develop two-level bisection search algorithms to solve the equivalent problems. Interestingly, we show that performing SIC at the secondary receiver does not always guarantee a higher secondary throughput than the case without performing SIC. Computer simulation demonstrates that our algorithm performs better than an equal power allocation algorithm and an orthogonal multiple access algorithm.**

***Index Terms*—Cognitive radio, cooperation, relay, decode-and-forward, non-orthogonal multiple access, energy harvesting.**

I. Introduction

Communications over wireless channels are technically restricted by the bottleneck of spectrum efficiency, on which research efforts have been made for decades, and many state-of-the-art techniques have been devised. Among the techniques, cognitive radio (CR) and non-orthogonal multiple access (NOMA) are quite promising. CR alleviates the conflict between explosively increased mobile applications and scarce frequency resources, and significantly improves the spectrum efficiency mainly in three ways of spectrum access: interweave, underlay, and overlay [2]–[6]. As a novel paradigm of transmission, NOMA is recognized to be a key enabling technology to further improve spectrum efficiency in 5G and beyond [7]–[12]. NOMA does not have the limitation of transmitting only one user's signal in one resource block over traditional orthogonal multiple access (OMA) systems. The power domain, treated as a new degree of freedom of resource management, is inherently exploited in NOMA systems. Specifically, at the transmitter side, different users' signals are assigned with distinctive power levels to enable concurrent transmissions. At the receiver side, successive interference cancellation (SIC) is adopted to perform user information recovery.

Naturally, the interaction of NOMA with CR, termed *cognitive NOMA*, yields even more spectrum-efficient communications [13]–[20]. The CR inspired NOMA is conceptually introduced in [13], in which two users with different priorities are paired on one resource block. The weak user who undergoes poor channel gain from the transmitter has a higher priority, and thus, is deemed as the primary user (PU) with a target reception quality. The strong user who experiences good channel gain from the transmitter has a lower priority, and thus, is deemed as the secondary user (SU). Under the condition that the PU's target reception quality is guaranteed, the transmitter uses NOMA to send simultaneously to the PU and SU. The work in [14] extends the CR inspired NOMA concept to the scenario with multiple PUs and one single SU, in which the SU, having better channel gain than those of the PUs, is served together with all the PUs simultaneously by using NOMA. The transmission strategy of the multiple-antenna transmitter is designed such that the energy efficiency is maximized. By further extending CR inspired NOMA with multiple PUs and multiple SUs, the work in [15] proposes a distributed matching method, which pairs an SU with a PU and assigns transmit power levels for them, targeting system throughput maximization. In [16], underlaid with the primary network, the secondary network sets up an interference guard zone to limit the primary interference, and transmits information from one secondary transmitter (ST) to multiple secondary receivers (SRs) by employing NOMA. For secondary transmissions, the outage performance and diversity order are analyzed. A two-user underlay cognitive NOMA system is investigated in [17], in which a dedicated full-duplex relay is resorted to help a far user to forward information from the base station (BS). The power allocation, beamformer design and the outage performance are investigated. The work in [18] considers an overlay cognitive NOMA system, in which

Manuscript received September 27, 2018; revised February 20, 2019 and May 1, 2019; accepted May 5, 2019. This work was supported by the Natural Sciences and Engineering Research Council of Canada under Grant RGPIN-2018-06307. This paper was presented in part at the IEEE 86th Vehicular Technology Conference, Toronto, Canada, September 24–27, 2017 [1]. The review of this paper was coordinated by Dr. Z. Ding. *(Corresponding authors: Fudong Li and Hai Jiang.)*

F. Li and H. Jiang are with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6G 1H9, Canada (e-mail: fudong1@ualberta.ca; hai1@ualberta.ca).

R. Fan is with the School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China (e-mail: fanrongfei@bit.edu.cn).

P. Tan is with the TELUS Communications Inc., Scarborough, ON M1H 3J3, Canada (e-mail: peng.tan@telus.com).

the primary network is assisted by the secondary network that applies NOMA principe in its transmissions. In the first phase, a primary transmitter (PT) transmits signals to its primary receiver (PR) and an ST, while in the second phase, the ST cooperatively transmits the PR's and the SRs' signals utilizing NOMA. The outage performance of the primary and secondary systems are characterized. The work in [19] considers an overlay cognitive NOMA system with a number of STs, in which one ST is scheduled to use NOMA to forward primary signals and send its own signals to its receivers. Two scheduling schemes, targeting reliability and fairness, respectively, are proposed. A survey of cognitive NOMA techniques and future research trends is given in [20].

Energy efficiency is another major concern in wireless communications. Conventional wireless devices, powered by batteries, may experience energy depletion and inconvenience of battery replacement. Energy harvesting provides a promising and viable solution [21], [22], in which energy from various sources is harvested and used for transmissions. Apart from natural energy sources (such as wind and solar), wireless RF signal is a widely studied energy source since it can be obtained almost everywhere and at any time. Moreover, the simultaneous wireless information and power transfer (SWIPT) technique has the advantage of transferring energy and delivering information simultaneously [22]–[24]. To achieve SWIPT, two typical implementation protocols, time-switching (TS) and power-splitting (PS), are designed [24].

By integrating SWIPT into cognitive NOMA framework, a greener and more sustainable communication is expected. In the literature, some research efforts on SWIPT-based cognitive NOMA are conducted, such as [25], [26]. The work in [25] studies an underlay cognitive NOMA system, in which multiple STs harvest energy from one energy transmitter and transmit to one common BS by using NOMA. The secure energy efficiency maximization problem for secondary system is investigated. The work in [26] can be viewed as a CR inspired NOMA system enhanced with SWIPT. To serve a weak user, a strong user uses PS protocol to harvest energy from the transmitter's signals and decode its own and the weak user's information during the first phase. Afterwards, during the second phase, the strong user forwards the weak user's information. With a condition that the reception quality of the weak user is satisfied, the strong user's throughput is maximized, by optimally designing the beamformers of the multiple-antenna transmitter and the PS ratio at the strong user.

In this paper, we investigate an overlay cognitive NOMA scenario equipped with SWIPT, in which an ST uses TS protocol for energy harvesting and information decoding, and uses NOMA to help a PT and deliver its own information. The optimal solution, deriving the time ratio for energy harvesting as well as the NOMA power allocation of the ST, is provided in our paper. Our main contributions are summarized as follows. 1) A new SWIPT-enhanced cognitive NOMA framework: In the literature, SWIPT is integrated with underlay cognitive NOMA in [25], and with CR-inspired NOMA in [26]. However, SWIPT-enhanced overlay cognitive NOMA is not investigated in the literature, and its performance is unclear. To address this research gap, we investigate a SWIPT-enhanced overlay cognitive NOMA framework. 2) Optimal solution: The formulated problems are nonconvex, and are generally hard to solve. We devise a method to transform the formulated problems to equivalent problems under different cases. We theoretically prove that the objective functions of the equivalent problems are quasiconcave. We then develop an effective algorithm, by using a two-level bisection search, to find the optimal solution of each equivalent problem. We also develop a method that could reduce the number of iterations in the inner bisection search. 3) Interesting insights: Interestingly, different from existing NOMA works in which SIC is always applied, performing SIC in our work does not guarantee a better performance than the scenario without performing SIC. The insight behind this observation is also discussed.

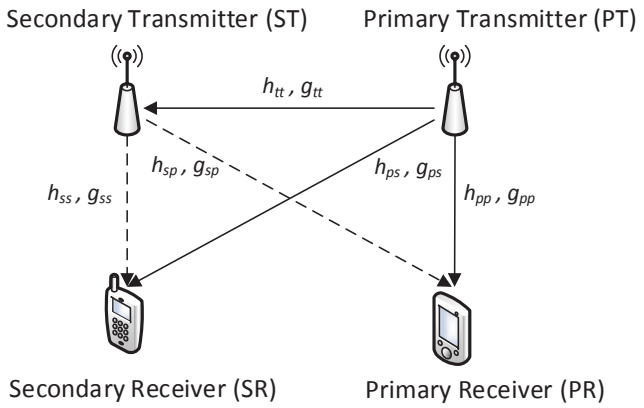


Fig. 1. System model ('h' means channel coefficient, and 'g' means channel gain).

II. System Model

We consider a cognitive system, consisting of one pair of PT and PR, and one pair of ST and SR, as depicted in Fig. 1. The spectrum is licensed to the primary system. The PT has a stable power supply, which transmits data with a fixed transmit power P_p. The ST is powered by harvested energy from RF signals transmitted by the PT. The ST opportunistically gains spectrum access opportunities in an overlay mode, i.e., when the link from the PT to the PR is not good enough, the ST could help forward the PT's signal and send its own signal to the SR as well by using NOMA.

The system is time slotted, and each time slot has a unit length. The channel coefficients between PT and PR, PT and SR, ST and PR, and ST and SR are denoted as h_{pp}, h_{ps}, h_{sp}, and h_{ss}, respectively. In subscript of the channel coefficients, the first symbol p or s means the primary or secondary transmitter, and the second symbol p or s means the primary or secondary receiver. In addition, the channel coefficient between PT and ST is denoted as h_{tt}. Accordingly, the channel gains (square of channel coefficient magnitude) of those channels are denoted as g_{pp}, g_{ps}, g_{sp}, g_{ss}, and g_{tt}, respectively. Block fading is assumed, which means that all the channel gains keep unchanged in each time slot and may change independently from slot to slot.

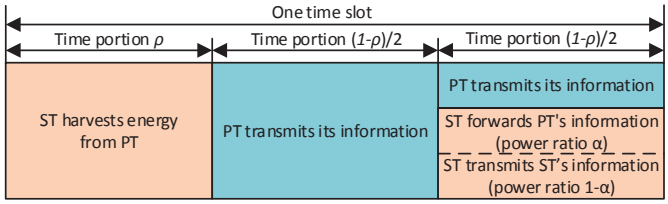


Fig. 2. The slotted structure of the energy harvesting cognitive NOMA system.

Basically, in a particular time slot, the system works in one of two transmission modes:

- Direct primary transmission mode
- Cooperative transmission mode.

A. Direct Primary Transmission Mode

We assume that the PT and ST always have enough information to be transmitted to the PR and SR, respectively. The PT has a target throughput[1], denoted as γ^T, for each time slot. At a slot, the achievable throughput of the direct link from the PT to the PR is expressed as $\log(1+\frac{P_p g_{pp}}{\sigma^2})$, where σ^2 is variance of the background additive white Gaussian noise (AWGN).

If the achievable throughput is not less than the PT's target throughput γ^T (which equivalently means that the channel gain of the direct primary link PT→PR satisfies $g_{pp} \geq (2^{\gamma^T}-1)\sigma^2/P_p$), then the system works in *direct primary transmission mode*. In this mode, the PT transmits data to the PR during the whole slot, whereas the ST is not allowed to access the spectrum. Thus, the ST harvests energy during the whole slot, which yields the battery energy at the end of the slot as $\max\{E_0+\eta P_p g_{tt}-E_c, 0\}$,[2] where E_0 is the ST's battery energy at the beginning of the slot, $\eta \in (0,1)$ is the energy conversion efficiency, and E_c means the energy expenditure in a time slot due to circuit operation and channel estimation.

B. Cooperative Transmission Mode

If the achievable throughput $\log(1+\frac{P_p g_{pp}}{\sigma^2})$ of the direct primary link PT→PR is lower than the PT's target throughput γ^T (which equivalently means that $g_{pp} < (2^{\gamma^T}-1)\sigma^2/P_p$), the ST is requested to help to forward the PT's message to PR by using decode-and-forward (DF) relaying. The ST can simultaneously send its own message to the SR based on downlink NOMA. In other words, the system works in the *cooperative transmission mode*.

In this mode, the slot is partitioned into three phases, as shown in Fig. 2. The first phase has length $\rho \in [0,1]$, while the second and third phases both have length $(1-\rho)/2$, where the value of ρ is a designed parameter, to be optimized hereafter.

[1]In this paper, "throughput" is defined as the amount of information bits that can be transmitted in a target slot.

[2]This amount of energy is also the battery energy at the beginning of the next slot.

1) The First Phase: During the first phase, the PT transmits wireless RF signals with power P_p, from which the ST harvests energy.[3] Then the ST's battery energy, after harvesting, has the form of $E_h = E_0+\eta\rho P_p g_{tt}$. If E_h is less than E_c, then the ST is not able to help the PT, and thus, the system has to work in the direct primary transmission mode. Therefore, for the system to work in the cooperative transmission mode, we should have $E_h \geq E_c$, which leads to $\rho \geq \max\{\frac{E_c-E_0}{\eta P_p g_{tt}}, 0\}$.

2) The Second Phase: During the second phase, the PT transmits its information signal x_p with power P_p, which is received by the PR and ST, as well as the SR.

At the PR, the received signal in the second phase of the slot is represented by $\sqrt{P_p}h_{pp}x_p + n_{pr}$, in which n_{pr} is the AWGN at the PR.

The received signal at the ST in the second phase of the slot is written as $\sqrt{P_p}h_{tt}x_p+n_{st}$, in which n_{st} is the AWGN at the ST. The achievable information rate for the transmission from the PT to the ST in the second phase of the slot is expressed as $R_{tt} = \frac{1-\rho}{2}\log(1+\frac{P_p g_{tt}}{\sigma^2})$. As the ST needs to decode the PT's signal, the achievable information rate R_{tt} should be not less than the target throughput γ^T of the primary system. Thus, we should have constraint $R_{tt} \geq \gamma^T$, based on which we have constraint $\rho \in \mathcal{A}_1 \triangleq \left[\max\{\frac{E_c-E_0}{\eta P_p g_{tt}}, 0\}, 1-\frac{2\gamma^T}{\log(1+\frac{P_p g_{tt}}{\sigma^2})}\right]$.

At the SR, the received signal in the second phase of the slot is represented by $\sqrt{P_p}h_{ps}x_p + n_{sr}$, in which n_{sr} is the AWGN at the SR.

3) The Third Phase: During the third phase, the PT transmits a copy of the primary signal x_p with power P_p. The ST applies downlink NOMA to transmit a superimposed signal consisting of the PT's signal x_p and the ST's own signal x_s, by using the harvested energy in a greedy manner, i.e., it uses up all the available energy stored in battery in this phase.[4] In specific, the ST transmits $x_{st} = \sqrt{\alpha P_e}x_p + \sqrt{(1-\alpha)P_e}x_s$, where α, a parameter to be optimized, is the power ratio for x_p, $(1-\alpha)$ is the power ratio for x_s, and P_e is the transmit power of ST, shown as

$$P_e = \frac{2(E_h-E_c)}{1-\rho} = \frac{2(\eta\rho P_p g_{tt}+E_0-E_c)}{1-\rho}. \quad (1)$$

Accordingly, the received signal at the PR in the third phase is $\sqrt{P_p}h_{pp}x_p + h_{sp}x_{st} + n_{pr}$. As the PR receives the PT's signal x_p in the second phase (from the PT) and the third phase (from the PT and ST), the PR employs maximal ratio combining (MRC) to combine the received PT's signal portions, which yields the overall throughput of the PT's signal x_p (also called the throughput of the primary system) in the time slot as

$$R_{pp} = \frac{1-\rho}{2}\log\left(1+\frac{P_p g_{pp}}{\sigma^2}+\frac{P_p g_{pp}+\alpha P_e g_{sp}}{(1-\alpha)P_e g_{sp}+\sigma^2}\right). \quad (2)$$

[3]Note that in the first phase, the wireless RF signals can actually be information signals of the PT for the PR. In other words, in the first phase, from the PT's signals, the ST tries to harvest energy while the PR tries to decode information. This portion of information for the PR will be investigated in Section III-F.

[4]Note that our method can be straightforwardly extended to the case when there exists a limit for the energy level that the ST can use in the third phase of the slot (i.e., there is some energy left at the end of the slot).

In the third phase, the received signal at the SR is $y_{ss} = \sqrt{P_p}h_{ps}x_p + h_{ss}x_{st} + n_{sr}$. In the received signal y_{ss}, the PT's signal portion x_p is interference to the SR's own signal x_s. Thus, similar to NOMA works in the literature, the SR can use SIC, i.e., it first decodes the PT's signal x_p, removes x_p from y_{ss}, and then decodes its own signal x_s.

When the SR decodes the PT's signal portion x_p, the achievable information rate is given as

$$R_{ps} = \frac{1-\rho}{2}\log\left(1 + \frac{P_p g_{ps}}{\sigma^2} + \frac{P_p g_{ps} + \alpha P_e g_{ss}}{(1-\alpha)P_e g_{ss} + \sigma^2}\right), \quad (3)$$

where MRC is used to combine the received PT's signal portions in the second and third phases. When performing SIC at the SR, ρ and α should be set up such that $R_{ps} \geq \gamma^T$, which means that the SR can successfully decode x_p, and perform SIC to remove x_p in y_{ss}. Subsequently, the achievable throughput of the secondary signal x_s (also called the secondary system's throughput) can be written as

$$R_{ss} = \frac{1-\rho}{2}\log\left(1 + \frac{(1-\alpha)P_e g_{ss}}{\sigma^2}\right). \quad (4)$$

Our objective is to maximize the secondary system's throughput while guaranteeing that the throughput of the primary system is not smaller than γ^T. Thus we formulate the following optimization problem.

Problem P1:

$$\max_{\alpha,\rho} \quad R_{ss} = \frac{1-\rho}{2}\log\left(1 + \frac{(1-\alpha)P_e g_{ss}}{\sigma^2}\right) \quad (5a)$$

$$\text{s.t.} \quad \rho \in \mathcal{A}_1; \min\{R_{pp}, R_{ps}\} \geq \gamma^T; 0 \leq \alpha \leq 1. \quad (5b)$$

In Problem P1, to perform SIC at the SR, we have constraint $R_{ps} \geq \gamma^T$. If the SR does not perform SIC, interference from the PT's signal is not cancelled, which harms the secondary throughput. On the other hand, if the SR does not perform SIC, the constraint $R_{ps} \geq \gamma^T$ can be removed, and we can have a larger feasible region of α and ρ, which benefits the secondary throughput. It is not clear whether the overall effect of not performing SIC is beneficial or harmful. Thus, we should also investigate an optimization problem, in which the SR does not perform SIC.

When the SR does not perform SIC, the achievable throughput of the secondary system is

$$R_{ss}^{w/o} = \frac{1-\rho}{2}\log\left(1 + \frac{(1-\alpha)P_e g_{ss}}{\alpha P_e g_{ss} + P_p g_{ps} + \sigma^2}\right), \quad (6)$$

in which superscript $(\cdot)^{w/o}$ stands for "without performing SIC." Accordingly, the following optimization problem can be formulated.

Problem P2:

$$\max_{\alpha,\rho} \quad R_{ss}^{w/o} = \frac{1-\rho}{2}\log\left(1 + \frac{(1-\alpha)P_e g_{ss}}{\alpha P_e g_{ss} + P_p g_{ps} + \sigma^2}\right) \quad (7a)$$

$$\text{s.t.} \quad \rho \in \mathcal{A}_1; R_{pp} \geq \gamma^T; 0 \leq \alpha \leq 1. \quad (7b)$$

As a summary, the optimal solution of the system is the better one between the optimal solutions of Problems P1 and P2, which has larger secondary throughput.[5]

III. Optimal Solution of Problem P1

Problem P1 is non-convex since the objective and the constraint functions are not jointly concave. Generally, it is hard to solve such a problem. In this work, we provide an efficient method to solve our Problem P1.

By careful inspection, there exists the following useful lemma.

Lemma 1: When the optimality of Problem P1 is achieved, constraint $\min\{R_{pp}, R_{ps}\} \geq \gamma^T$ should be active, i.e., we should have $\min\{R_{pp}, R_{ps}\} = \gamma^T$.

Proof: We use proof by contradiction. Suppose that when the optimality of Problem P1 is achieved, we have $\min\{R_{pp}, R_{ps}\} > \gamma^T$.

From (2) and (3), we know that R_{pp} and R_{ps} both are increasing functions of α. Furthermore, when $\alpha = 0$, we have $\min\{R_{pp}, R_{ps}\}|_{\alpha=0} \leq R_{pp}|_{\alpha=0} = \frac{1-\rho}{2}\log\left(1 + \frac{P_p g_{pp}}{\sigma^2} + \frac{P_p g_{pp}}{P_e g_{sp} + \sigma^2}\right) < \gamma^T$, in which the last inequality comes from the fact that

$$\begin{aligned}
&\frac{1-\rho}{2}\log\left(1 + \frac{P_p g_{pp}}{\sigma^2} + \frac{P_p g_{pp}}{P_e g_{sp} + \sigma^2}\right) \\
&\quad \leq \frac{1-\rho}{2}\log(1 + \frac{2P_p g_{pp}}{\sigma^2}) \leq \frac{1}{2}\log(1 + \frac{P_p g_{pp}}{\sigma^2})^2 \\
&\quad = \log\left(1 + \frac{P_p g_{pp}}{\sigma^2}\right) < \gamma^T. \quad (8)
\end{aligned}$$

In (8), the last inequality comes from the fact that the system works in the cooperative transmission mode when the achievable throughput $\log(1 + \frac{P_p g_{pp}}{\sigma^2})$ of the direct primary link from the PT to the PR is less than the PT's target throughput γ^T.

Thus, from the optimality point of Problem P1, we can decrease the value of α such that we still have $\min\{R_{pp}, R_{ps}\} \geq \gamma^T$ but we have a larger objective function R_{ss} (noting that R_{ss} is a decreasing function of α according to (4)). This is a contradiction.

This completes the proof. ■

Lemma 1 indicates that, at optimality of Problem P1, either R_{pp} or R_{ps} should be equal to γ^T.

If $R_{pp} = \gamma^T$, then from (2), we can see that α can be expressed by a function of ρ as

$$\alpha = F_p(\rho) \triangleq \frac{\mu_p(P_e g_{sp} + \sigma^2) - P_p g_{pp}}{(\mu_p + 1)P_e g_{sp}} \quad (9)$$

with

$$\mu_p \triangleq 2^{\frac{2\gamma^T}{1-\rho}} - \frac{P_p g_{pp}}{\sigma^2} - 1. \quad (10)$$

Note that by using (8), we can see $\mu_p > 0$.

When α is expressed as a function of ρ as in (9), the constraint $0 \leq \alpha \leq 1$ should be satisfied. Note that $\alpha > 0$ is satisfied automatically, since from (8), we have $\frac{1-\rho}{2}\log(1 + \frac{2P_p g_{pp}}{\sigma^2}) < \gamma^T$, based on which we have $\mu_p\sigma^2 - P_p g_{pp} > 0$,

[5] If both Problems P1 and P2 are infeasible, which means the cooperative transmission mode cannot make the throughput of the primary system be at least γ^T, then the system will work in direct primary transmission mode.

yielding $\alpha > 0$. Thus, only $\alpha \leq 1$ is considered here, which is equivalent to

$$\rho \in \mathcal{A}_2 \triangleq \left\{\rho | \rho \geq 0 \text{ and } P_e g_{sp} - \mu_p \sigma^2 + P_p g_{pp} \geq 0\right\}. \quad (11)$$

Lemma 2: Set $\mathcal{A}_2$ is a closed interval of ρ, with closed-form starting and ending points.

Proof: The proof is similar to the proof in Appendix A in [1] (conference version of this paper), and thus, is omitted due to space limit. ∎

If $R_{ps} = \gamma^T$, then from (3), we can see that α can be expressed as a function of ρ, as

$$\alpha = F_s(\rho) \triangleq \frac{\mu_s(P_e g_{ss} + \sigma^2) - P_p g_{ps}}{(\mu_s + 1) P_e g_{ss}} \quad (12)$$

with $\mu_s \triangleq 2^{\frac{2\gamma^T}{1-\rho}} - \frac{P_p g_{ps}}{\sigma^2} - 1$.

When α is expressed as a function of ρ as in (12), the constraint $0 \leq \alpha \leq 1$ is equivalent to

$$\rho \in \mathcal{A}_3 \triangleq \left\{\rho | \rho \geq 0 \text{ and } \mu_s(P_e g_{ss} + \sigma^2) - P_p g_{ps} \geq 0\right\} \cap \left\{\rho | \rho \geq 0 \text{ and } P_e g_{ss} - \mu_s \sigma^2 + P_p g_{ps} \geq 0\right\}. \quad (13)$$

In $\mathcal{A}_3$, set $\{\rho | \rho \geq 0 \text{ and } \mu_s(P_e g_{ss} + \sigma^2) - P_p g_{ps} \geq 0\}$ comes from $\alpha \geq 0$, which is a closed interval, since $\mu_s(P_e g_{ss} + \sigma^2) - P_p g_{ps}$ is an increasing function of ρ. Similar to set $\mathcal{A}_2$, set $\{\rho | \rho \geq 0 \text{ and } P_e g_{ss} - \mu_s \sigma^2 + P_p g_{ps} \geq 0\}$ is also a closed interval. Thus, $\mathcal{A}_3$ is a closed interval.

Afterwards, by comparing (2) and (3), we have the following four cases for Problem P1.

A. Case 1: When $g_{pp} \leq g_{ps}$ and $g_{sp} \leq g_{ss}$

For a specific value of ρ, R_{pp} in (2) and R_{ps} in (3) can be viewed as functions of α.

1) Intersections of Curves R_{pp} vs. α and R_{ps} vs. α over $\alpha \in [0, 1]$: By considering two curves: R_{pp} vs. α and R_{ps} vs. α over $\alpha \in [0, 1]$, we have the following lemma.

Lemma 3: For a specific value of ρ, the two curves R_{pp} vs. α and R_{ps} vs. α over $\alpha \in [0, 1]$ have up to two intersections.

Proof: For a given ρ, from (2) and (3), if we set $R_{pp} = R_{ps}$, we can obtain

$$L_1 P_e^2 (1-\alpha)^2 + \left[L_1 \left(\frac{\sigma^2}{g_{ss}} + \frac{\sigma^2}{g_{sp}}\right) + \frac{\frac{P_p g_{ps}}{g_{ss}} - \frac{P_p g_{pp}}{g_{sp}}}{\frac{\sigma^2}{g_{ss}} - \frac{\sigma^2}{g_{sp}}} + 1 \right] P_e (1-\alpha) + 2L_1 \frac{\sigma^4}{g_{ss} g_{sp}} - P_e = 0, \quad (14)$$

where $L_1 = \frac{\frac{P_p g_{ps}}{\sigma^2} - \frac{P_p g_{pp}}{\sigma^2}}{\frac{\sigma^2}{g_{ss}} - \frac{\sigma^2}{g_{sp}}} < 0$. Considering the left hand-side of (14) as a function of α, the two roots of (14) are

$$\text{ROOT}_1 = 1 - \frac{-L_2 + \sqrt{L_2^2 - 4L_1 L_3}}{2L_1 P_e}, \quad (15)$$

$$\text{ROOT}_2 = 1 - \frac{-L_2 - \sqrt{L_2^2 - 4L_1 L_3}}{2L_1 P_e}, \quad (16)$$

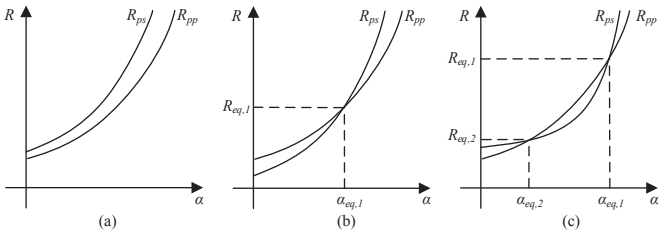


Fig. 3. An illustration of two curves R_{pp} vs. α and R_{ps} vs. α (for a specific ρ) in Case 1, in which the two curves have up to two intersections.

in which $L_2 = L_1(\frac{\sigma^2}{g_{ss}} + \frac{\sigma^2}{g_{sp}}) + \frac{\frac{P_p g_{ps}}{g_{ss}} - \frac{P_p g_{pp}}{g_{sp}}}{\frac{\sigma^2}{g_{ss}} - \frac{\sigma^2}{g_{sp}}} + 1$ and $L_3 = 2L_1 \frac{\sigma^4}{g_{ss} g_{sp}} - P_e < 0$. Note that when $L_2 < 0$, both ROOT_1 and ROOT_2 are larger than 1 and infeasible.

To make ROOT_1 shown in (15) feasible, we should have $0 \leq \text{ROOT}_1 \leq 1$, which equivalently means that $\rho \in \mathcal{A}_4 \triangleq \{\rho | 0 \leq \rho \leq 1, L_2 \geq 0, L_2^2 - 4L_1 L_3 \geq 0, 2L_1 P_e + L_2 - \sqrt{L_2^2 - 4L_1 L_3} \leq 0\}$.[6] The set $\mathcal{A}_4$ is a closed interval of ρ.

To make ROOT_2 shown in (16) feasible, we should have $0 \leq \text{ROOT}_2 \leq 1$, which equivalently means $\rho \in \mathcal{A}_5 \triangleq \{\rho | 0 \leq \rho \leq 1, L_2 \geq 0, L_2^2 - 4L_1 L_3 \geq 0, 2L_1 P_e + L_2 + \sqrt{L_2^2 - 4L_1 L_3} \leq 0\}$. The set $\mathcal{A}_5$ is an interval of ρ, and is a subset of $\mathcal{A}_4$.

Overall, for any specific ρ, if $\rho \in \bar{\mathcal{A}}_4 \triangleq [0,\ 1] \backslash \mathcal{A}_4$, then the two curves R_{pp} vs. α and R_{ps} vs. α do not have intersection; if $\rho \in \mathcal{A}_4 \backslash \mathcal{A}_5$, then the two curves have one intersection denoted as $(\alpha_{eq,1}, R_{eq,1})$, with $\alpha_{eq,1} = \text{ROOT}_1$; if $\rho \in \mathcal{A}_5$, then the two curves have two intersections denoted as $(\alpha_{eq,1}, R_{eq,1})$ and $(\alpha_{eq,2}, R_{eq,2})$, with $\alpha_{eq,1} = \text{ROOT}_1$ and $\alpha_{eq,2} = \text{ROOT}_2$. An illustration is given in Fig. 3.

This completes the proof. ∎

Below we show some features of $R_{eq,1}$ and $R_{eq,2}$.

$R_{eq,1}$ is expressed as

$$R_{eq,1} = R_{pp}|_{\alpha=\text{ROOT}_1} = \frac{1-\rho}{2} \log \left[1 + \frac{P_p g_{pp}}{\sigma^2} + \frac{P_p \frac{g_{ps}}{g_{ss}} - P_p \frac{g_{pp}}{g_{sp}}}{\frac{\sigma^2}{g_{ss}} - \frac{\sigma^2}{g_{sp}}} + L_1 \left(\frac{-L_2 + \sqrt{L_2^2 - 4L_1 L_3}}{2L_1} + \frac{\sigma^2}{g_{ss}} \right) \right]. \quad (17)$$

The second order derivative of $R_{eq,1}$ is given by

$$\frac{\mathrm{d}^2 R_{eq,1}}{\mathrm{d}\rho^2} = \frac{1}{(\ln 2) H_1(\rho)} \frac{L_1}{\sqrt{L_2^2 - 4L_1 L_3}} \left(\frac{1-\rho}{2} \frac{\mathrm{d}^2 P_e}{\mathrm{d}\rho^2} - \frac{\mathrm{d}P_e}{\mathrm{d}\rho} \right) - \frac{1-\rho}{2(\ln 2)\left(H_1(\rho)\right)^2} \frac{L_1^2}{L_2^2 - 4L_1 L_3} \left(\frac{\mathrm{d}P_e}{\mathrm{d}\rho}\right)^2 - \frac{1-\rho}{(\ln 2) H_1(\rho)} \frac{L_1^2}{\left(L_2^2 - 4L_1 L_3\right)^{\frac{3}{2}}} \left(\frac{\mathrm{d}P_e}{\mathrm{d}\rho}\right)^2, \quad (18)$$

[6] L_2 is not a function of ρ. By including "$L_2 \geq 0$" in the expression of $\mathcal{A}_4$, we mean that $\mathcal{A}_4$ will be a null set if $L_2 < 0$.

where

$$H_1(\rho)=1+\frac{P_p g_{pp}}{\sigma^2}+\frac{P_p\frac{g_{ps}}{g_{ss}}-P_p\frac{g_{pp}}{g_{sp}}}{\frac{\sigma^2}{g_{ss}}-\frac{\sigma^2}{g_{sp}}} + L_1\Big(\frac{-L_2+\sqrt{L_2^2-4L_1L_3}}{2L_1}+\frac{\sigma^2}{g_{ss}}\Big),$$

and

$$\begin{cases} \frac{\mathrm{d}P_e}{\mathrm{d}\rho} &= \frac{2(\eta P_p g_{tt}+E_0-E_c)}{(1-\rho)^2}, \\ \frac{\mathrm{d}^2P_e}{\mathrm{d}\rho^2} &= \frac{4(\eta P_p g_{tt}+E_0-E_c)}{(1-\rho)^3}. \end{cases} \tag{19}$$

From (18), since $\frac{1-\rho}{2}\frac{\mathrm{d}^2P_e}{\mathrm{d}\rho^2}-\frac{\mathrm{d}P_e}{\mathrm{d}\rho}=0$, $\frac{\mathrm{d}^2P_e}{\mathrm{d}\rho^2}>0$ and $\frac{\mathrm{d}P_e}{\mathrm{d}\rho}>0$, one can obtain that $\frac{\mathrm{d}^2R_{eq,1}}{\mathrm{d}\rho^2}<0$, which indicates that $R_{eq,1}$ is a concave function of ρ.

$R_{eq,2}$ is expressed as

$$\begin{aligned} R_{eq,2} &= R_{pp}|_{\alpha=\text{ROOT}_2} \\ &= \frac{1-\rho}{2}\log\Bigg[1+\frac{P_p g_{pp}}{\sigma^2}+\frac{P_p\frac{g_{ps}}{g_{ss}}-P_p\frac{g_{pp}}{g_{sp}}}{\frac{\sigma^2}{g_{ss}}-\frac{\sigma^2}{g_{sp}}} \\ &\quad + L_1\left(\frac{-L_2-\sqrt{L_2^2-4L_1L_3}}{2L_1}+\frac{\sigma^2}{g_{ss}}\right)\Bigg]. \end{aligned} \tag{20}$$

The second order derivative of $R_{eq,2}$ is given by

$$\begin{aligned} \frac{\mathrm{d}^2R_{eq,2}}{\mathrm{d}\rho^2} &= \frac{-1}{(\ln 2)H_2(\rho)}\frac{L_1}{\sqrt{L_2^2-4L_1L_3}}\Big(\frac{1-\rho}{2}\frac{\mathrm{d}^2P_e}{\mathrm{d}\rho^2}-\frac{\mathrm{d}P_e}{\mathrm{d}\rho}\Big) \\ &\quad - \frac{1-\rho}{2(\ln 2)\big(H_2(\rho)\big)^2}\frac{L_1^2}{L_2^2-4L_1L_3}\big(\frac{\mathrm{d}P_e}{\mathrm{d}\rho}\big)^2 \\ &\quad + \frac{1-\rho}{(\ln 2)H_2(\rho)}\frac{L_1^2}{\big(L_2^2-4L_1L_3\big)^{\frac{3}{2}}}\big(\frac{\mathrm{d}P_e}{\mathrm{d}\rho}\big)^2, \end{aligned} \tag{21}$$

where

$$H_2(\rho)=1+\frac{P_p g_{pp}}{\sigma^2}+\frac{P_p\frac{g_{ps}}{g_{ss}}-P_p\frac{g_{pp}}{g_{sp}}}{\frac{\sigma^2}{g_{ss}}-\frac{\sigma^2}{g_{sp}}} + L_1\Big(\frac{-L_2-\sqrt{L_2^2-4L_1L_3}}{2L_1}+\frac{\sigma^2}{g_{ss}}\Big).$$

Recall that intersection $(\alpha_{eq,2}, R_{eq,2})$ exists only when $\rho\in\mathcal{A}_5$. By observing (21), it can be seen that 1) when $\rho\in\mathcal{A}_6\triangleq\mathcal{A}_5\cap\{\rho|H_2(\rho)-\frac{1}{2}\sqrt{L_2^2-4L_1L_3}\geq 0\}$, we have $\frac{\mathrm{d}^2R_{eq,2}}{\mathrm{d}\rho^2}\geq 0$, which indicates that $R_{eq,2}$ is a convex function of ρ over $\mathcal{A}_6$; 2) when $\rho\in\mathcal{A}_7\triangleq\mathcal{A}_5\cap\{\rho|H_2(\rho)-\frac{1}{2}\sqrt{L_2^2-4L_1L_3}<0\}$, we have $\frac{\mathrm{d}^2R_{eq,2}}{\mathrm{d}\rho^2}<0$, which indicates that $R_{eq,2}$ is a concave function of ρ over $\mathcal{A}_7$. Note that $\mathcal{A}_6$ and $\mathcal{A}_7$ both are closed intervals of ρ.

2) Problem P1 with Case 1: From Lemma 3, when $\rho\in\bar{\mathcal{A}}_4$, the two curves R_{pp} vs. α and R_{ps} vs. α do not have intersection, which means that we always have $R_{pp}<R_{ps}$, as shown in Fig. 3(a). Thus, at optimality of Problem P1, we should have $R_{pp}=\gamma^T$. Then Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_p(\rho)} \tag{22a}$$

$$\text{s.t.} \quad \rho\in\mathcal{A}_1\cap\mathcal{A}_2\cap\bar{\mathcal{A}}_4, \tag{22b}$$

with $R_{ss}|_{\alpha=F_p(\rho)}=\frac{1-\rho}{2}\log\big(1+\frac{(P_e g_{sp}-\mu_p\sigma^2+P_p g_{pp})g_{ss}}{(\mu_p+1)g_{sp}\sigma^2}\big)$.

When $\rho\in\mathcal{A}_4\backslash\mathcal{A}_5$, the two curves have one intersection at $(\alpha_{eq,1}, R_{eq,1})$. It is interesting to point out that we further have two scenarios, as follows.

- If $R_{eq,1}\geq\gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp},R_{ps}\}=\gamma^T$), we have $R_{ps}\leq R_{pp}$, and thus, we have $R_{ps}=\gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_s(\rho)} \tag{23a}$$

$$\text{s.t.} \quad \rho\in\mathcal{A}_1\cap\mathcal{A}_3\cap[(\mathcal{A}_4\backslash\mathcal{A}_5)\cap\mathcal{A}_8] \tag{23b}$$

  in which $R_{ss}|_{\alpha=F_s(\rho)}=\frac{1-\rho}{2}\log\big(1+\frac{P_e g_{ss}-\mu_s\sigma^2+P_p g_{ps}}{(\mu_s+1)\sigma^2}\big)$ and $\mathcal{A}_8\triangleq\{\rho|R_{eq,1}\geq\gamma^T\}$. As aforementioned, $R_{eq,1}$ is a concave function of ρ. Thus, $\mathcal{A}_8$ is a closed interval of ρ.
- If $R_{eq,1}\leq\gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp},R_{ps}\}=\gamma^T$), we have $R_{pp}\leq R_{ps}$, and thus, we have $R_{pp}=\gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_p(\rho)} \tag{24a}$$

$$\text{s.t.} \quad \rho\in\mathcal{A}_1\cap\mathcal{A}_2\cap[(\mathcal{A}_4\backslash\mathcal{A}_5)\cap\bar{\mathcal{A}}_8] \tag{24b}$$

  in which $\bar{\mathcal{A}}_8=[0,\ 1]\backslash\mathcal{A}_8$. As $\mathcal{A}_8$ is a closed interval, $\bar{\mathcal{A}}_8$ is the union of two closed intervals.

When $\rho\in\mathcal{A}_5$, the two curves have two intersections at $(\alpha_{eq,1}, R_{eq,1})$ and $(\alpha_{eq,2}, R_{eq,2})$. It is also interesting to point out that we further have two scenarios, as follows.

- If $R_{eq,1}\leq\gamma^T$ or $R_{eq,2}\geq\gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp},R_{ps}\}=\gamma^T$), we have $R_{pp}\leq R_{ps}$, and thus, we have $R_{pp}=\gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_p(\rho)} \tag{25a}$$

$$\text{s.t.} \quad \rho\in\mathcal{A}_1\cap\mathcal{A}_2\cap\mathcal{A}_5\cap[\bar{\mathcal{A}}_8\cup\mathcal{A}_9] \tag{25b}$$

  in which $\mathcal{A}_9=\{\rho|R_{eq,2}\geq\gamma^T \text{ and } \rho\in\mathcal{A}_6\}\cup\{\rho|R_{eq,2}\geq\gamma^T \text{ and } \rho\in\mathcal{A}_7\}$. As $R_{eq,2}$ is convex over $\mathcal{A}_6$ and concave over $\mathcal{A}_7$, $\mathcal{A}_9$ is the union of three closed intervals.
- If $R_{eq,1}\geq\gamma^T$ and $R_{eq,2}\leq\gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp},R_{ps}\}=\gamma^T$), we have $R_{ps}\leq R_{pp}$, and thus, we have $R_{ps}=\gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_s(\rho)} \tag{26a}$$

$$\text{s.t.} \quad \rho\in\mathcal{A}_1\cap\mathcal{A}_3\cap\mathcal{A}_5\cap\mathcal{A}_8\cap\bar{\mathcal{A}}_9 \tag{26b}$$

  in which $\bar{\mathcal{A}}_9=\{\rho|R_{eq,2}\leq\gamma^T \text{ and } \rho\in\mathcal{A}_6\}\cup\{\rho|R_{eq,2}\leq\gamma^T \text{ and } \rho\in\mathcal{A}_7\}$, being the union of three closed intervals.

As a summary, for Case 1, the maximal objective function of Problem P1 is the largest one among the maximal objective functions of Problem (22), Problem (23), Problem (24), Problem (25) and Problem (26).

B. Case 2: When $g_{pp} > g_{ps}$ and $g_{sp} > g_{ss}$

Similar to Lemma 3, when $\rho \in \bar{\mathcal{A}}_4$, the two curves do not have intersection, which means that we always have $R_{ps} \leq R_{pp}$. Thus, at optimality of Problem P1, we should have $R_{ps} = \gamma^T$. Then Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_s(\rho)} \tag{27a}$$
$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_3 \cap \bar{\mathcal{A}}_4. \tag{27b}$$

When $\rho \in \mathcal{A}_4 \backslash \mathcal{A}_5$, the two curves have one intersection at $(\alpha_{eq,1}, R_{eq,1})$. We further have two scenarios, as follows.

- If $R_{eq,1} \geq \gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp}, R_{ps}\} = \gamma^T$), we have $R_{pp} \leq R_{ps}$, and thus, we have $R_{pp} = \gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_p(\rho)} \tag{28a}$$
$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_2 \cap [(\mathcal{A}_4 \backslash \mathcal{A}_5) \cap \mathcal{A}_8]. \tag{28b}$$

- If $R_{eq,1} \leq \gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp}, R_{ps}\} = \gamma^T$), we have $R_{ps} \leq R_{pp}$, and thus, we have $R_{ps} = \gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_s(\rho)} \tag{29a}$$
$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_3 \cap [(\mathcal{A}_4 \backslash \mathcal{A}_5) \cap \bar{\mathcal{A}}_8]. \tag{29b}$$

When $\rho \in \mathcal{A}_5$, the two curves have two intersections at $(\alpha_{eq,1}, R_{eq,1})$ and $(\alpha_{eq,2}, R_{eq,2})$. We further have two scenarios, as follows.

- If $R_{eq,1} \leq \gamma^T$ or $R_{eq,2} \geq \gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp}, R_{ps}\} = \gamma^T$), we have $R_{ps} \leq R_{pp}$, and thus, we have $R_{ps} = \gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_s(\rho)} \tag{30a}$$
$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_3 \cap \mathcal{A}_5 \cap [\bar{\mathcal{A}}_8 \cup \mathcal{A}_9]. \tag{30b}$$

- If $R_{eq,1} \geq \gamma^T$ and $R_{eq,2} \leq \gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp}, R_{ps}\} = \gamma^T$), we have $R_{pp} \leq R_{ps}$, and thus, we have $R_{pp} = \gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_p(\rho)} \tag{31a}$$
$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_5 \cap \mathcal{A}_8 \cap \bar{\mathcal{A}}_9. \tag{31b}$$

As a summary, for Case 2, the maximal objective function of Problem P1 is the largest one among the maximal objective functions of Problem (27), Problem (28), Problem (29), Problem (30) and Problem (31).

C. Case 3: When $g_{pp} \leq g_{ps}$ and $g_{sp} > g_{ss}$

Similar to the previous two cases, we also try to decide which one between R_{pp} and R_{ps} should be equal to γ^T.

Lemma 4: For a specific value of ρ, the two curves R_{pp} vs. α and R_{ps} vs. α over $\alpha \in [0,1]$ have at most one intersection at $(\alpha_{eq,1}, R_{eq,1})$, with $R_{eq,1}$ being a concave function of ρ.

Proof: The proof is similar to Lemma 3. The major difference is $L_1 > 0$ and $L_2 > 0$, which makes ROOT_2 in (16) larger than 1 and infeasible, yielding at most one intersection. ■

Note that we have $R_{pp}|_{\alpha=0} = \frac{1-\rho}{2}\log(1 + \frac{P_p g_{pp}}{\sigma^2} + \frac{P_p g_{pp}}{P_e g_{sp}+\sigma^2}) < R_{ps}|_{\alpha=0} = \frac{1-\rho}{2}\log(1 + \frac{P_p g_{ps}}{\sigma^2} + \frac{P_p g_{ps}}{P_e g_{ss}+\sigma^2})$. Thus, the two curves R_{pp} vs. α and R_{ps} vs. α have one intersection over $\alpha \in [0,1]$ if and only if $R_{pp}|_{\alpha=1} \geq R_{ps}|_{\alpha=1}$, which equivalently means that $\rho \in \mathcal{A}_4 = \left[\max\left\{1 - \frac{\eta P_p g_{tt} + E_0 - E_c}{\eta P_p g_{tt} + \frac{P_p g_{ps} - P_p g_{pp}}{g_{sp} - g_{ss}}}, 0\right\}, \ 1\right]$.

Similar to Lemma 3, when $\rho \in \bar{\mathcal{A}}_4 \triangleq [0, \ 1] \backslash \mathcal{A}_4$, the two curves do not have intersection, which means that we always have $R_{pp} < R_{ps}$. Thus, at optimality of Problem P1, we should have $R_{pp} = \gamma^T$. Then Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_p(\rho)} \tag{32a}$$
$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_2 \cap \bar{\mathcal{A}}_4. \tag{32b}$$

When $\rho \in \mathcal{A}_4$, the two curves have one intersection at $(\alpha_{eq,1}, R_{eq,1})$. We further have two scenarios, as follows.

- If $R_{eq,1} \geq \gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp}, R_{ps}\} = \gamma^T$), we have $R_{pp} \leq R_{ps}$, and thus, we have $R_{pp} = \gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_p(\rho)} \tag{33a}$$
$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_4 \cap \mathcal{A}_8. \tag{33b}$$

- If $R_{eq,1} < \gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp}, R_{ps}\} = \gamma^T$), we have $R_{ps} < R_{pp}$, and thus, we have $R_{ps} = \gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_s(\rho)} \tag{34a}$$
$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \bar{\mathcal{A}}_8. \tag{34b}$$

As a summary, for Case 3, the maximal objective function of Problem P1 is the largest one among the maximal objective functions of Problem (32), Problem (33), and Problem (34).

D. Case 4: $g_{pp} > g_{ps}$ and $g_{sp} \leq g_{ss}$

Similar to Case 3, we can also prove that Lemma 4 holds in Case 4. Different from Case 3, here we have $R_{pp}|_{\alpha=0} = \frac{1-\rho}{2}\log(1 + \frac{P_p g_{pp}}{\sigma^2} + \frac{P_p g_{pp}}{P_e g_{sp}+\sigma^2}) > R_{ps}|_{\alpha=0} = \frac{1-\rho}{2}\log(1 + \frac{P_p g_{ps}}{\sigma^2} + \frac{P_p g_{ps}}{P_e g_{ss}+\sigma^2})$. Thus, the two curves R_{pp} vs. α and R_{ps} vs. α have one intersection over $\alpha \in [0,1]$ if and only if $R_{pp}|_{\alpha=1} \leq R_{ps}|_{\alpha=1}$, which equivalently means that $\rho \in \mathcal{A}_4$.[7]

When $\rho \in \bar{\mathcal{A}}_4$, the two curves do not have intersection, which means that we always have $R_{pp} > R_{ps}$. Thus, at

[7]In Case 3, $R_{pp}|_{\alpha=1} \geq R_{ps}|_{\alpha=1}$ is equivalent to $\rho \in \mathcal{A}_4$, while in Case 4, $R_{pp}|_{\alpha=1} \leq R_{ps}|_{\alpha=1}$ is equivalent to $\rho \in \mathcal{A}_4$. This is because we have $g_{pp} \leq g_{ps}$ and $g_{sp} > g_{ss}$ in Case 3, while we have $g_{pp} > g_{ps}$ and $g_{sp} \leq g_{ss}$ in Case 4.

optimality of Problem P1, we should have $R_{ps} = \gamma^T$. Then Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_s(\rho)} \tag{35a}$$

$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_3 \cap \bar{\mathcal{A}}_4. \tag{35b}$$

When $\rho \in \mathcal{A}_4$, the two curves have one intersection at $(\alpha_{eq,1}, R_{eq,1})$. We have two scenarios as follows.

- If $R_{eq,1} \geq \gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp}, R_{ps}\} = \gamma^T$), we have $R_{ps} \leq R_{pp}$, and thus, we have $R_{ps} = \gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_s(\rho)} \tag{36a}$$

$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_3 \cap \mathcal{A}_4 \cap \mathcal{A}_8. \tag{36b}$$

- If $R_{eq,1} < \gamma^T$, then at optimality of Problem P1 (i.e., when $\min\{R_{pp}, R_{ps}\} = \gamma^T$), we have $R_{pp} < R_{ps}$, and thus, we have $R_{pp} = \gamma^T$ at optimality of Problem P1. Accordingly, Problem P1 is equivalent to

$$\max_{\rho} \quad R_{ss}|_{\alpha=F_p(\rho)} \tag{37a}$$

$$\text{s.t.} \quad \rho \in \mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_4 \cap \bar{\mathcal{A}}_8. \tag{37b}$$

As a summary, for Case 4, the maximal objective function of Problem P1 is the largest one among the maximal objective functions of Problem (35), Problem (36), and Problem (37).

In Cases 1∼4, all the equivalent problems are in the format of $\max_{\rho} \ R_{ss}|_{\alpha=F_p(\rho)}$ or $\max_{\rho} \ R_{ss}|_{\alpha=F_s(\rho)}$ under a constraint that ρ is within a closed interval (i.e., such as equivalent problems (23), (28), (32) and (35)) or within a union of multiple closed intervals (i.e., such as equivalent problems (25), (30), (34) and (37)). We have two observations:

- If ρ is within a union of multiple closed intervals, we can first get the optimal solution over each interval and pick up the best optimal solution.
- $R_{ss}|_{\alpha=F_p(\rho)}$ and $R_{ss}|_{\alpha=F_s(\rho)}$ can be expressed in a unified form as

$$R_{ss,n} = \frac{1-\rho}{2} \log\left(1 + \frac{(P_e g_{sn} - \mu_n \sigma^2 + P_p g_{pn}) g_{ss}}{(\mu_n + 1) g_{sn} \sigma^2}\right) \tag{38}$$

with $n \in \{p, s\}$. Here we have $R_{ss}|_{\alpha=F_n(\rho)} = R_{ss,n}$.

Therefore, in what follows, we focus on maximizing $R_{ss,n}$, $n \in \{p, s\}$ over $\rho \in \mathcal{B}$ with $\mathcal{B}$ being a closed interval.

E. Solving $\max_{\rho}$ $R_{ss,n}$ s.t. $\rho \in \mathcal{B}$

The objective function $R_{ss,n}$ is not a concave function. However, it is a quasiconcave function [27], as theoretically shown in the following theorem.

Theorem 1: The objective function $R_{ss,n}$ is quasiconcave with respect to $\rho \in \mathcal{B}$.

Proof: See Appendix A. ∎

As shown in [27], super-level sets of a quasiconcave function can be represented by inequalities of concave functions. As reference [27] does not provide methods to find the inequalities of concave functions, here we develop a method to find the inequalities of concave functions that can represent super-level sets of our quasiconcave function $R_{ss,n}$. As a result, we have the following theorem.

Theorem 2: For any $t \geq 0$, inequality $R_{ss,n} \geq t$ and inequality $\xi_{t,n} \geq 0$ are equivalent, in which $\xi_{t,n}$ is given as

$$\xi_{t,n} = (\lambda_n - 2^{\frac{2t}{1-\rho}})(1-\rho)(\mu_n + 1), \tag{39}$$

where

$$\lambda_n = 1 + \frac{(P_e g_{sn} - \mu_n \sigma^2 + P_p g_{pn}) g_{ss}}{(\mu_n + 1) g_{sn} \sigma^2}. \tag{40}$$

Proof: It is readily checked that $R_{ss,n} \geq t$ is equivalent to $\lambda_n - 2^{\frac{2t}{1-\rho}} \geq 0$.

Since $1-\rho$ and $\mu_n + 1$ are larger than 0, multiplying two positive numbers to $\lambda_n - 2^{\frac{2t}{1-\rho}}$ will not change its sign. Then we can see that $R_{ss,n} \geq t$ is equivalent to $\xi_{t,n} \geq 0$.

This completes the proof. ∎

From Theorem 2, identifying the maximized $R_{ss,n}$ in $\rho \in \mathcal{B}$ is equivalent to finding the maximal possible value of t such that there exists $\rho \in \mathcal{B}$ that makes $R_{ss,n} \geq t$, which is further equivalent to finding the maximal possible value of t such that there exists $\rho \in \mathcal{B}$ that makes $\xi_{t,n} \geq 0$. For a value of t, if there exists $\rho \in \mathcal{B}$ that makes $\xi_{t,n} \geq 0$, then we say that the t value is *feasible*. Thus, $\max_{\rho} \ R_{ss,n}$ s.t. $\rho \in \mathcal{B}$ is equivalent to finding the maximal feasible t value, which can be done by using a bisection search over $t \in [0, \log(1 + \frac{P_e g_{ss}}{\sigma^2})]$ (noting that from (4) and (6), it can be seen that an upper bound of $R_{ss,n}$ is $\log(1 + \frac{P_e g_{ss}}{\sigma^2})$).

In the bisection search over t, we need to decide whether or not a checked t value is feasible. For this purpose, the following theorem, which gives a feature of $\xi_{t,n}$, is helpful.

Theorem 3: For a given nonnegative value of t, the function $\xi_{t,n}$ is concave with respect to $\rho \in \mathcal{B}$.

Proof: The proof is similar to the proof in Appendix C in [1] (conference version of this paper), and thus, is omitted due to space limit. ∎

In the bisection search over t, if a checked t value is feasible, this equivalently means that for the checked t value, the maximal value of $\xi_{t,n}$ over $\rho \in \mathcal{B}$ is nonnegative. Thus, from Theorem 3, we can use a bisection search over $\rho \in \mathcal{B}$, to find the maximal value of $\xi_{t,n}$.[8] If any searched ρ makes $\xi_{t,n} \geq 0$, then the checked t value is feasible, and we can terminate the bisection search of ρ. If the found maximal value of $\xi_{t,n}$ is negative, then the checked t value is infeasible. Thus, we have two levels of bisection search, and we call the bisection search over t as outer bisection search, and call the bisection search over ρ as inner bisection search.

Generally, each inner bisection search may need to be done over $\rho \in \mathcal{B}$. Next we reduce the number of iterations in inner bisection search.

For a t value, define $\mathcal{R}_{t,n} = \{\rho | \xi_{t,n} \geq 0, \rho \in \mathcal{B}\}$. So $\mathcal{R}_{t,n}$ is a set of ρ that makes $\xi_{t,n} \geq 0$. Consider two t values: t_1 and t_2, with $t_1 < t_2$. We have

$$\mathcal{R}_{t_2,n} \overset{\text{(i)}}{=} \{\rho | R_{ss,n} \geq t_2\} \overset{\text{(ii)}}{\subseteq} \{\rho | R_{ss,n} \geq t_1\} \overset{\text{(iii)}}{=} \mathcal{R}_{t_1,n}, \tag{41}$$

[8] In the bisection search, we try to find a value of ρ that makes $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho} = 0$ (which means the maximal value of $\xi_{t,n}$ is achieved over $\rho \in \mathcal{B}$). Note that other methods can also be used here to find the maximal value of $\xi_{t,n}$ over $\rho \in \mathcal{B}$, such as gradient descent method and Newton's method.

in which steps (i) and (iii) are from Theorem 2, and step (ii) is from $t_1 < t_2$.

From (41), we can see that, if we know that t_1 is feasible, then when deciding whether or not t_2 is feasible, we only need to search ρ over $\mathcal{R}_{t_1,n}$, or over an interval of ρ that includes $\mathcal{R}_{t_1,n}$ as a subset. Denote $\mathcal{F} \triangleq [f_1, f_2]$ as the interval of ρ over which the inner bisection search is performed. So $\mathcal{F}$ is initially set to be $\mathcal{B}$. In the outer bisection search, for a checked t value (say $t^\dagger$), if $t^\dagger$ is feasible, we get an updated $\mathcal{F}$ (which is a subset of the previous $\mathcal{F}$, and includes $\mathcal{R}_{t^\dagger,n}$ as a subset). Then, in the outer bisection search, when we check feasibility of higher t values,[9] we only need to search ρ over the updated interval $\mathcal{F}$ (rather than over $\mathcal{B}$) in the inner bisection search, referred to as *feasible region shrinking*.

Based on this observation, we have Algorithm 1 for our inner bisection search. In Algorithm 1, the expression of $\xi_{t,n}$ is given in (39), while expression of $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}$ is given as

$$\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho} = (\frac{g_{ss}}{g_{sn}} - 1)\big(\mu_n - (1-\rho)\frac{\mathrm{d}\mu_n}{\mathrm{d}\rho}\big) + \frac{2\eta P_p g_{tt} g_{ss}}{\sigma^2} - 1 - \frac{P_p g_{pn} g_{ss}}{g_{sn}\sigma^2} - \Big((\mu_n+1)\big(\frac{2(\ln 2)t}{1-\rho} - 1\big) + (1-\rho)\frac{\mathrm{d}\mu_n}{\mathrm{d}\rho}\Big)2^{\frac{2t}{1-\rho}}. \tag{42}$$

Algorithm 1: Inner Bisection Search

Input: value of t to be checked for feasibility, and two end points of the interval $\mathcal{F}$: f_1 and f_2

Output: Feasibility of value t, updated f_1 and f_2

1 **if** $\xi_{t,n}|_{\rho=f_1} \geq 0$ *or* $\xi_{t,n}|_{\rho=f_2} \geq 0$ **then**
2 $\quad$ t is feasible. Terminate the algorithm;
3 **if** $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}|_{\rho=f_1} \times \frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}|_{\rho=f_2} \geq 0$ **then**
4 $\quad$ t is infeasible. Terminate the algorithm;
5 $f_1^{\text{BS}} \leftarrow f_1, f_2^{\text{BS}} \leftarrow f_2$;
6 $f_1^{\text{new}} \leftarrow f_1, f_2^{\text{new}} \leftarrow f_2$;
7 **if** $|f_2^{BS} - f_1^{BS}| < \epsilon_\rho$ **then**
8 $\quad$ t is infeasible. Terminate the algorithm;
9 $f_{\text{mid}} \leftarrow (f_1^{\text{BS}} + f_2^{\text{BS}})/2$;
10 **if** $\xi_{t,n}|_{\rho=f_{mid}} \geq 0$ **then**
11 $\quad$ t is feasible; $f_1 \leftarrow f_1^{\text{new}}, f_2 \leftarrow f_2^{\text{new}}$; Terminate the algorithm;
12 **if** $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}|_{\rho=f_{mid}} = 0$ **then**
13 $\quad$ t is infeasible. Terminate the algorithm;
14 **if** $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}|_{\rho=f_{mid}} > 0$ **then**
15 $\quad$ $f_1^{\text{BS}} \leftarrow f_{\text{mid}}$; $f_1^{\text{new}} \leftarrow f_{\text{mid}}$;Go to Step 7;
16 **if** $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}|_{\rho=f_{mid}} < 0$ **then**
17 $\quad$ $f_2^{\text{BS}} \leftarrow f_{\text{mid}}$; $f_2^{\text{new}} \leftarrow f_{\text{mid}}$;Go to Step 7;

In Steps 1 and 2 of Algorithm 1, we check whether $\rho = f_1$ or $\rho = f_2$ makes $\xi_{t,n}$ nonnegative. If yes, then the checked t value is feasible and we terminate the algorithm. If we proceed to Step 3, then we know that $\xi_{t,n}|_{\rho=f_1} < 0$ and $\xi_{t,n}|_{\rho=f_2} < 0$. In Step 3, we check whether $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}|_{\rho=f_1}$ and $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}|_{\rho=f_2}$ are both nonpositive or both nonnegative. If they are both nonpositive or both nonnegative, then $\xi_{t,n}$ is a decreasing or increasing function with respect to $\rho \in \mathcal{F}$, and thus, we can conclude that the checked t value is infeasible. So when we proceed to Step 5, we should have $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}|_{\rho=f_1} > 0$ and $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}|_{\rho=f_2} < 0$. Then we try to search the maximal point of $\xi_{t,n}$ over $\rho \in [f_1, f_2]$, by using inner bisection search of $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho}$ until $\frac{\mathrm{d}\xi_{t,n}}{\mathrm{d}\rho} = 0$. In the inner bisection search, $[f_1^{\text{BS}}, f_2^{\text{BS}}]$ represents the subinterval after bisecting the original interval of ρ, and $[f_1^{\text{new}}, f_2^{\text{new}}]$ represents the updated interval $\mathcal{F}$. In the inner bisection search, if a searched ρ value makes $\xi_{t,n} \geq 0$, then we know that the checked t value is feasible, and we update $\mathcal{F}$ (we can see that the updated $\mathcal{F}$ satisfies $\mathcal{R}_{t,n} \subseteq \mathcal{F}$), and terminate the algorithm (Steps 10–11). If the subinterval $[f_1^{\text{BS}}, f_2^{\text{BS}}]$ is sufficiently small (i.e., less than a threshold value ϵ_ρ) and no searched ρ value makes $\xi_{t,n} \geq 0$, then we know the checked t value is infeasible, and we do not update $\mathcal{F}$ (Steps 7–8).

The detailed algorithm for the outer bisection search is straightforward, and thus, is omitted here, for presentational simplicity.

Complexity: According to [28], the computational complexity of a bisection search is $O(\log(\frac{1}{\epsilon}))$, where ϵ is the pre-defined tolerance for convergence. Thus, the complexity of the proposed two-level bisection search is expressed as $O\big(\log(\frac{1}{\epsilon_t})\log(\frac{1}{\epsilon_\rho})\big)$, in which ϵ_t and ϵ_ρ are pre-defined convergence tolerance for the outer bisection search over t and the inner bisection search over ρ.

Impact of parameters ϵ_t and ϵ_ρ: Denote t^* as the maximal feasible t value, and denote ρ^* as the corresponding ρ that achieves t^*. Denote $\hat{t}$ as the t value found by our proposed algorithm.

For the outer bisection search over t, when it converges, we get a region of t, denoted as $[l, u]$ with $u - l < \epsilon_t$. Then $\hat{t} = l$ is the t value found by our algorithm.

If the inner bisection search (which does feasibility check for a specific t value) is always accurate, then when the outer bisection search converges, $t = l$ is feasible, $t = u$ is infeasible, and t^* falls within $[l, u]$. Thus, the gap between t^* and $\hat{t}(= l)$ is less than ϵ_t.

However, when the inner bisection search checks the feasibility of a value, say $t^\dagger$, close to t^*, it may not be accurate. It is possible that the inner bisection search may claim that $t^\dagger$ is infeasible, but actually $t^\dagger$ is feasible. If this happens, then when the outer bisection search converges at a region of t denoted as $[l, u]$, actually $t = l$ and $t = u$ are both feasible, and we have $l < u < t^*$ (in other words, t^* does not fall within $[l, u]$). Recall that our objective function $R_{ss,n}$ is a quasiconcave function, i.e., when the first order derivative is 0, the second order derivative is nonpositive. Thus, we can treat the quasiconcave function $R_{ss,n}$ as a concave function at the neighborhood of ρ^*. Based on this, $|t^* - u|$ is less than $\Big|\frac{\mathrm{d}R_{ss,n}}{\mathrm{d}\rho}|_{\rho=\rho^*-\epsilon_\rho}\Big|\epsilon_\rho$. Accordingly, the gap between t^* and $\hat{t}(= l)$ is less than $\Delta \triangleq \epsilon_t + \Big|\frac{\mathrm{d}R_{ss,n}}{\mathrm{d}\rho}|_{\rho=\rho^*-\epsilon_\rho}\Big|\epsilon_\rho$.

For $R_{ss,n}$, its first-order derivative at $\rho = \rho^*$ is zero, i.e.,

[9]Note that since $t^\dagger$ is feasible, we do not need to check t values lower than $t^\dagger$ in the outer bisection search.

$\frac{\mathrm{d}R_{ss,n}}{\mathrm{d}\rho}|_{\rho=\rho^*}=0$. Since $R_{ss,n}$ can be viewed as a concave function at the neighborhood of ρ^* and ϵ_ρ is small, we can see that $\left|\frac{\mathrm{d}R_{ss,n}}{\mathrm{d}\rho}|_{\rho=\rho^*-\epsilon_\rho}\right|$ is close to zero. Thus, Δ is small for small values of ϵ_t and ϵ_ρ.

F. Further Discussion: When Primary Information is Sent during the First Phase

In the first phase, the wireless RF signals can actually be information signals of the PT for the PR. In other words, in the first phase, from the PT's signals, the ST tries to harvest energy while the PR tries to decode information. Thus, the primary system has an additional throughput expressed as $\rho\log(1+\frac{P_p g_{pp}}{\sigma^2})$. It also means that the target throughput of the primary system during the second and third phase is $\gamma^T-\rho\log(1+\frac{P_p g_{pp}}{\sigma^2})$. Therefore, in Problem P1, we should replace γ^T with $\gamma^T-\rho\log(1+\frac{P_p g_{pp}}{\sigma^2})$, and accordingly, we have the following revised Problem P1:

Revised Problem P1:

$$\max_{\alpha,\rho} \quad R_{ss}=\frac{1-\rho}{2}\log\left(1+\frac{(1-\alpha)P_e g_{ss}}{\sigma^2}\right) \tag{43a}$$

$$\text{s.t.} \quad \rho\in\tilde{\mathcal{A}}_1; 0\le\alpha\le 1 \tag{43b}$$

$$\min\{R_{pp},R_{ps}\}\ge\gamma^T-\rho\log(1+\frac{P_p g_{pp}}{\sigma^2}) \tag{43c}$$

in which

$$\tilde{\mathcal{A}}_1 \triangleq \left[\max\{\frac{E_c-E_0}{\eta P_p g_{tt}},0\}, 1-\frac{2\left(\gamma^T-\log(1+\frac{P_p g_{pp}}{\sigma^2})\right)}{\log(1+\frac{P_p g_{tt}}{\sigma^2})-2\log(1+\frac{P_p g_{pp}}{\sigma^2})}\right].$$

Similar to Lemma 1, we should have either $R_{pp}=\gamma^T-\rho\log(1+\frac{P_p g_{pp}}{\sigma^2})$ or $R_{ps}=\gamma^T-\rho\log(1+\frac{P_p g_{pp}}{\sigma^2})$ at optimality of the revised Problem P1. Accordingly, when $R_{pn}=\gamma^T-\rho\log(1+\frac{P_p g_{pp}}{\sigma^2})$, $n\in\{p,s\}$, we can see that α can be expressed in terms of ρ as $\alpha=\tilde{F}_n(\rho)\triangleq\frac{\tilde{\mu}_n(P_e g_{sn}+\sigma^2)-P_p g_{pn}}{(\tilde{\mu}_n+1)P_e g_{sn}}$ with $\tilde{\mu}_n\triangleq(1+\frac{P_p g_{pp}}{\sigma^2})^2 2^{\frac{2\tilde{\gamma}^T}{1-\rho}}-\frac{P_p g_{pn}}{\sigma^2}-1$, in which $\tilde{\gamma}^T=\gamma^T-\log(1+\frac{P_p g_{pp}}{\sigma^2})$.

Afterwards, for each of the four cases defined in Section III-A∼III-D, the revised Problem P1 is equivalent to maximizing

$$R_{ss}|_{\alpha=\tilde{F}_n(\rho)}=\frac{1-\rho}{2}\log\left(1+\frac{(P_e g_{sn}-\tilde{\mu}_n\sigma^2+P_p g_{pn})g_{sn}}{(\tilde{\mu}_n+1)g_{sn}\sigma^2}\right) \tag{44}$$

over closed intervals.

Then similar to Theorem 1, we can still prove that $R_{ss}|_{\alpha=\tilde{F}_n(\rho)}$ is quasiconcave over each closed interval, and thus, a two-level bisection search can be used to find the optimal solution.

For presentation simplicity, in the sequel, we consider that the PT does not send information to the PR during the first phase of each slot in the cooperative transmission mode.

IV. Optimal Solution for Problem P2

Similar to Problem P1, when Problem P2 achieves the optimality, the constraint $R_{pp}\ge\gamma^T$ should take equality, which means $\alpha=F_p(\rho)$, as given in (9). Accordingly, Problem P2 is equivalent to

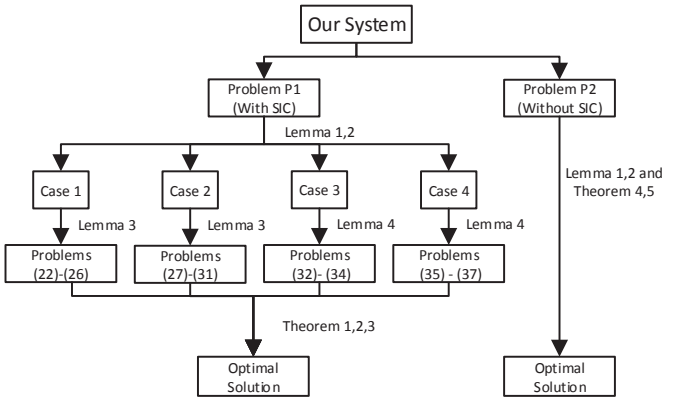


Fig. 4. Flow chart of the procedure for finding optimal solution of the considered system.

$$\max_{\rho} \quad R_{ss}^{w/o}|_{\alpha=F_p(\rho)} \tag{45a}$$

$$\text{s.t.} \quad \rho\in\mathcal{A}_1\cap\mathcal{A}_2 \tag{45b}$$

with $R_{ss}^{w/o}|_{\alpha=F_p(\rho)}=\frac{1-\rho}{2}\log\lambda^{w/o}$, where

$$\lambda^{w/o}=1+(P_e g_{sp}-\mu_p\sigma^2+P_p g_{pp})g_{ss}\Big\{\big[\mu_p(P_e g_{sp}+\sigma^2)-P_p g_{pp}\big]g_{ss}+(P_p g_{ps}+\sigma^2)(\mu_p+1)g_{sp}\Big\}^{-1}. \tag{46}$$

We have the following theorem for the objective function of Problem (45).

Theorem 4: The objective function $R_{ss}^{w/o}|_{\alpha=F_p(\rho)}$ is quasiconcave with respect to $\rho\in[0,1]$.

Proof: See Appendix B. ∎

Since $R_{ss}^{w/o}|_{\alpha=F_p(\rho)}$ is quasiconcave, similar to Theorem 2, for any $t\ge 0$, inequality $R_{ss}^{w/o}|_{\alpha=F_p(\rho)}\ge t$ is equivalent to $\xi_t^{w/o}\ge 0$, with

$$\xi_t^{w/o}\triangleq(\lambda^{w/o}-2^{\frac{2t}{1-\rho}})(1-\rho)\Big\{\big[\mu_p(P_e g_{sp}+\sigma^2)-P_p g_{pp}\big]g_{ss}+(P_p g_{ps}+\sigma^2)(\mu_p+1)g_{sp}\Big\}. \tag{47}$$

The following theorem gives a feature of $\xi_t^{w/o}$.

Theorem 5: For a given nonnegative value t, the function $\xi_t^{w/o}$ is concave with respect to $\rho\in[0,1]$.

Proof: See Appendix C. ∎

Therefore, a two-level bisection search similar to that in Section III-E can be used to find the optimal solution for Problem P2. The details are omitted here.

Overall, Fig. 4 shows the procedure for finding optimal solution of the considered system.

Remark: Interestingly, we have an observation that the maximal secondary throughput of Problem P1 may not guarantee to be larger than that of Problem P2. The reason is as follows. For specific α and ρ, the objective function value of Problem P1 is indeed larger than that of Problem P2. However, compared to Problem P2, Problem P1 has one more constraint $R_{ps}\ge\gamma^T$, which makes the feasible region of Problem P1

be a subset of the feasible region of Problem P2. Thus, in a larger feasible region, it is possible that the maximal secondary throughput of Problem P2 is larger than that of Problem P1.

A numerical example is also given here, which has the following parameter setting as: $g_{tt} = 7.15 \times 10^{-3}$, $g_{pp} = 2.69 \times 10^{-5}$, $g_{ps} = 2.65 \times 10^{-6}$, $g_{sp} = 1.1 \times 10^{-1}$, $g_{ss} = 2.88 \times 10^{-2}$, $P_p = 15$ dBm, $E_0 = 3.41 \times 10^{-5}$ J, $E_c = 1.5 \times 10^{-6}$ J, $\sigma^2 = -25$ dBm, and $\gamma^T = 0.5$ bps/Hz. In this specific example, the optimal solution of Problem P2 is 0.1448 bps/Hz, which is larger than the optimal solution of Problem P1, 0.0321 bps/Hz.

V. Numerical Results

We use Matlab simulation to evaluate the performance of our proposed algorithm[10]. Similar to [16], the channel power gain g_i is further represented as $g_i = \frac{\tilde{g}_i}{1+d_i^{\kappa_i}}$, where $i \in \mathcal{M} = \{tt, pp, ps, sp, ss\}$, $\tilde{g}_i$ is exponentially distributed with parameter 1, d_i is the distance of link i, and κ_i is the path loss exponent of link i. The distance values are: $d_{pp} = 12$ m, $d_{tt} = d_{ps} = d_{sp} = d_{ss} = 8$ m. The background noise variance is $\sigma^2 = -25$dBm and the energy conversion efficiency is $\eta = 0.5$. The energy expenditure in a time slot due to circuit operation and channel estimation is $E_c = 1.5 \times 10^{-6}$ J. In our simulation, a minimum throughput requirement 0.15 bps/Hz is set up at the ST. If the ST cannot achieve this throughput at a slot, our system works in direct primary transmission mode.

First we evaluate the chance of ST to access the channel. Three events, D_1, D_2 and D_3, are considered. Specifically, event D_1 is defined as $g_{pp} < (2^{\gamma^T} - 1)\sigma^2/P_p$, which means that the PT needs help. Event D_2 is defined as $g_{tt} \geq (2^{2\gamma^T} - 1)\sigma^2/P_p$, which means that it is possible to set up a ρ such that $R_{tt} \geq \gamma^T$ (i.e., the ST is able to decode information from the PT in the second phase). Event D_3 is defined as the event that the PT needs help (i.e., $g_{pp} < (2^{\gamma^T} - 1)\sigma^2/P_p$) and at least one of Problems P1 and P2 is feasible[11]. Clearly, when event D_3 happens, events D_1 and D_2 should also happen. And the probability of event D_3 is exactly the probability of the ST to access the channel. By setting $\gamma^T = 0.25$ bps/Hz and 0.5 bps/Hz, $\kappa_{pp} = 4$, $\kappa_i = 2$, $i \in \mathcal{M}\backslash\{pp\}$ and varying P_p from 0 dBm to 30 dBm, the occurrence probabilities of three events, $P(D_1), P(D_2)$, and $P(D_3)$, are shown in Fig. 5. Probabilities $P(D_1)$ and $P(D_2)$ are decreasing and increasing functions, respectively, of P_p, which is intuitive. Moreover, we get higher $P(D_1)$ and lower $P(D_2)$ through increasing γ^T from 0.25 bps/Hz to 0.5 bps/Hz. This is because a higher target rate γ^T means that the PT needs more help and the ST needs better channel gain g_{tt} to decode information from the PT. At low P_p, $P(D_1)$ is close to 1 (i.e., the PT almost always needs help from the ST), and thus, the curve of $P(D_3)$ follows the trend of the curve of $P(D_2)$, i.e., increases when P_p increases or γ^T decreases. At high P_p, $P(D_2)$ is close to 1, and thus, the curve of $P(D_3)$ follows the trend of the curve of $P(D_1)$, i.e., decreases when P_p increases or γ^T decreases.

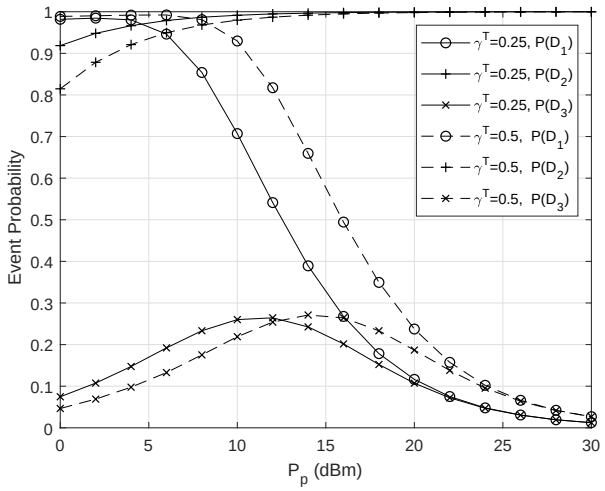


Fig. 5. The probabilities of three events D_1, D_2, and D_3.

Next we show by how much chance the performance of Problem P2 is better than that of Problem P1 (i.e., the maximal secondary throughput without SIC is higher than that with SIC). We set $\gamma^T = 0.1, 0.25, 0.5, 0.75$, and 1 bps/Hz, and $P_p = 0$ dBm and 10 dBm. For each (γ^T, P_p) pair, we run simulations for 10^6 time slots. In Table I, the numbers after slash are numbers of time slots in which both Problems P1 and P2 are feasible, while the numbers before slash are numbers of time slots when the maximal secondary throughput of Problem P2 is larger than that of Problem P1. It can be seen that, in most time slots when both problems are feasible, Problem P1 has better performance. However, it is still possible (with a small probability) that Problem P2 performs better. Therefore, when both problems are feasible, if we directly take the solution of Problem P1, we have a very large chance to get the overall optimal solution, to be verified below.

We compare our proposed algorithm with its two variants: an `SIC-if-possible` algorithm that takes the solution of Problem P1 when both Problem P1 and Problem P2 are feasible, and a `Never-SIC` algorithm that always takes the solution of Problem P2 even if both Problem P1 and Problem P2 are feasible. Fig. 6 shows secondary throughput performance of our proposed algorithm and the two variants for $\gamma^T = 0.25$ bps/Hz and 0.5 bps/Hz. It can be seen that the throughput of our proposed algorithm has similar trend as the channel access probability (the curve of $P(D_3)$ as shown in Fig. 5). By comparing Fig. 5 and Fig. 6, when P_p increases beyond 20 dBm, the channel access probability and throughput of the secondary system in the proposed algorithm both decrease, but the decrease rate of the channel access probability is higher, explained as follows. When P_p is high, the channel access probability is low, which means the ST has more chance to accumulate energy. Thus, at a slot, when the system works in cooperative transmission mode, the energy level of the ST is high, leading to high throughput

[10]In the sequel, by "our proposed algorithm," we mean the algorithm that takes the better solution of Problem P1 and Problem P2 if both problems are feasible, and takes the solution of Problem P2 if only Problem P2 is feasible. Note that if Problem P2 is infeasible, then Problem P1 should also be infeasible.

[11]Note that in cooperative transmission mode, when Problem P1 is feasible, Problem P2 is always feasible. Thus, "at least one of Problems P1 and P2 is feasible" is equivalent to "Problem P2 is feasible."

TABLE I
THE NUMBER OF TIME SLOTS IN WHICH THE MAXIMAL SECONDARY THROUGHPUT OF PROBLEM P2 IS LARGER THAN THAT OF PROBLEM P1, AND THE NUMBER OF TIME SLOTS WHEN BOTH PROBLEMS ARE FEASIBLE.

	$P_p = 0$ dBm	$P_p = 10$ dBm
$\gamma^T = 0.1$	134/111690	20/219865
$\gamma^T = 0.25$	150/74603	79/260055
$\gamma^T = 0.5$	57/46421	89/218943
$\gamma^T = 0.75$	5/31032	25/171037
$\gamma^T = 1$	0/21085	1/132606

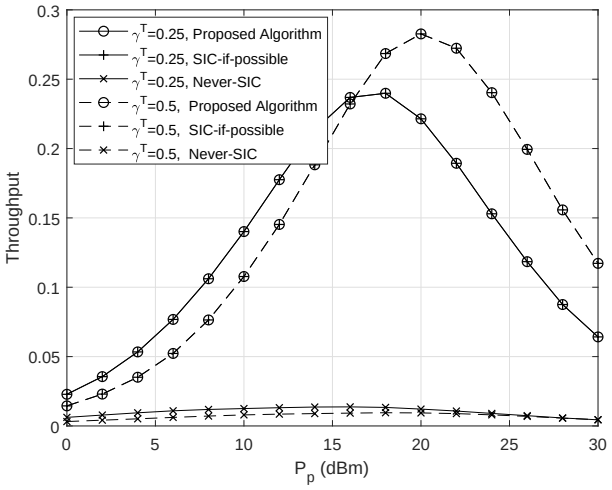


Fig. 6. Secondary throughput of the proposed algorithm, the `SIC-if-possible` algorithm, and the `Never-SIC` algorithm.

at the slot. Thus, when P_p increases beyond 20 dBm, the decrease rate of secondary system throughput is not as high as that of the channel access probability. From Fig. 6, the `SIC-if-possible` algorithm achieves almost the same performance as that of our proposed algorithm. This verifies our statement that when both Problem P1 and Problem P2 are feasible, if we directly take the solution of Problem P1, we have a very large chance to get the overall optimal solution. From Fig. 6, it can also be seen that there is a gap between the performance of the `Never-SIC` algorithm and that of our proposed algorithm. This is because it is with a very large probability (for example, more than 90% based on simulation results in Table I) that the case with SIC performs better than the case without SIC.

Remark: The above observations do not mean that our Problem P2 is useless. This is because it is possible that Problem P1 is infeasible but Problem P2 is feasible, in which scenario we have to take the solution of Problem P2.

In our proposed algorithm, we use two levels of bisection search. In the inner bisection search, we keep shrinking the feasible region as shown in Algorithm 1.[12] To show the benefit of this, we compare two scenarios: our proposed algorithm is implemented with and without shrinking the feasible region in inner bisection search. Recall that the bisecting in the inner bisection search may not be implemented in Algorithm 1 (for example, when $\xi_{t,n}|_{\rho=f_1} \geq 0$ or $\xi_{t,n}|_{\rho=f_2} \geq 0$, Algorithm 1 terminates at Step 2, and thus, the bisecting is not implemented). Thus, in the comparison, we only consider the time slots when the bisecting in the inner bisection search is carried out (e.g., Algorithm 1 proceeds to Step 5). The average number of inner-bisecting iterations in a two-level bisection search algorithm is shown in Fig. 7, with the tolerance of convergence for bisection search being $\epsilon_t = \epsilon_\rho = 10^{-3}$ and $\epsilon_t = \epsilon_\rho = 10^{-4}$. It can be seen that shrinking feasible region in inner bisection search reduces the number of iterations in the inner bisection search by around 10%.

[12]Algorithm 1 is for the case with performing SIC. For the case without performing SIC, we also keep shrinking the feasible region in the inner bisection search.

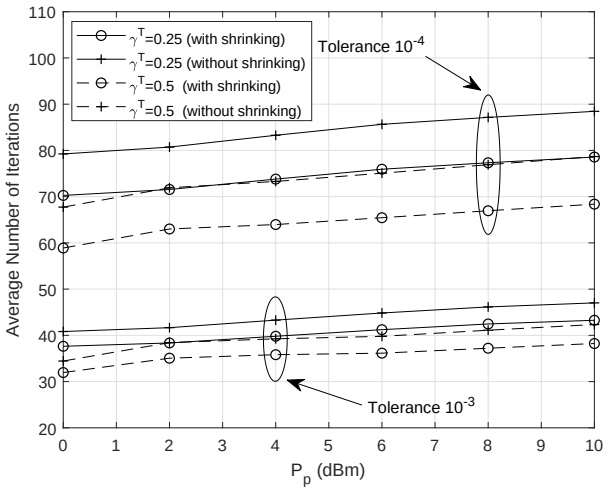


Fig. 7. The average number of inner-bisecting iterations with and without shrinking.

Now we try to compare our proposed algorithm with other algorithms. Since no existing work in the literature considers overlay cognitive NOMA enhanced with TS-based SWIPT, here we compare with two algorithms: an OMA algorithm that partitions one time slot into one harvesting phase and three equal-length information transmission phases for links PT→ST&PR, ST→PR, and ST→SR, respectively, and an equal power allocation (EPA) algorithm that allocates equal amount of energy for transmitting the PT's signal and the ST's signal in the third phase of a time slot. Fig. 8 shows secondary throughput of our proposed algorithm, and the OMA and EPA algorithms. Clearly, our proposed algorithm outperforms the EPA and OMA algorithms in terms of higher secondary throughput. Moreover, performance gap of our algorithm with the EPA and OMA algorithms shrink at high P_p, explained as follows. At high P_p, the probability that the PT needs help is low (as observed in Fig. 5). So at time slots when the PT does not need help, the ST will only accumulate energy. Thus, at a time slot when the PT needs help, the ST has a large chance to have high energy, leading to high signal-to-noise ratio (SNR) at the SR. The throughput is a logarithm function of the SNR. The logarithm function is a concave function, i.e., at high SNR, increase of SNR does not lead to much increase in throughput. Thus, the throughput gap of our algorithm with the EPA and OMA algorithms shrink at high P_p.

Distance plays a critical role in SWIPT-based systems, since

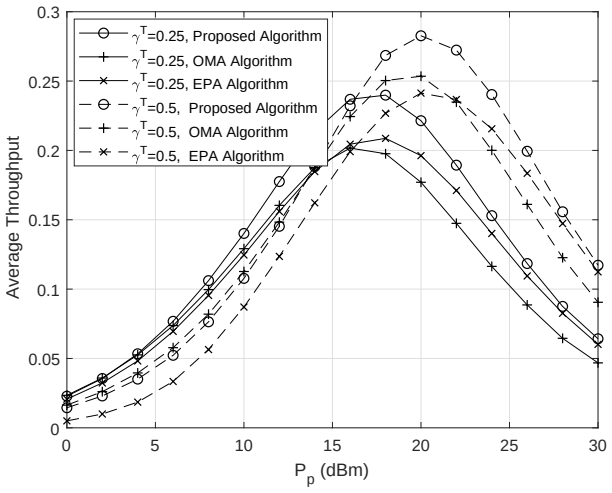


Fig. 8. Secondary throughput of our proposed algorithm, the OMA and the EPA algorithms.

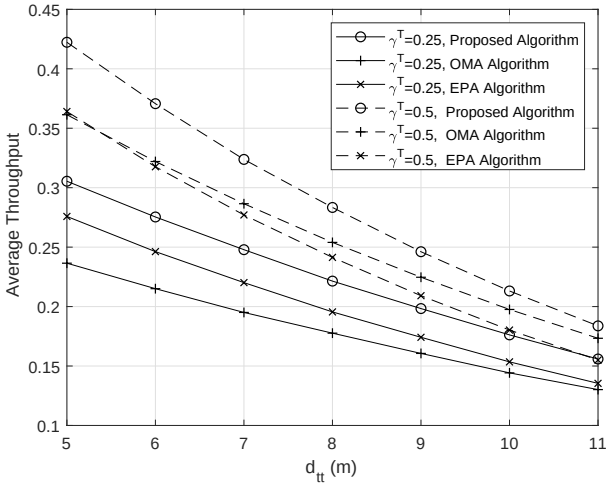


Fig. 9. The impact of distance on secondary throughput of our proposed algorithm, the OMA and the EPA algorithms.

the amount of harvested energy is largely determined by the distance from the RF transmitter. For the network topology considered in our simulation, we move the ST along the line segment of PT–ST, and fix the locations of other nodes. The secondary throughput is evaluated and illustrated in Fig. 9 for $P_p = 20$ dBm. It can be seen that the secondary throughout decreases with the increase of d_{tt} (distance from the PT to the ST), due to the decay of harvested energy amount at the ST.

VI. Conclusion

We have investigated the secondary throughput maximization problem for an overlay cognitive NOMA network aided by SWIPT. For the research problems with and without SIC at the SR, we have transformed the problems to equivalent problems, the objective functions of which are proved to be quasiconcave. Optimal solutions for the equivalent problems have been found by two-level bisection search, and a method has been developed to reduce the number of iterations in the inner bisection search. Interestingly, the non-SIC case is possible to achieve a better secondary performance than the SIC case. Since this happens with a small probability as shown in our simulation results, if the ST just picks up the solution of the SIC case for a slot when both cases are feasible, it has a very large chance to achieve the optimal performance at the slot. By this method, the ST does not have to solve two optimization problems at the slot. On the other hand, at a slot when the non-SIC case is feasible but the SIC case is infeasible, the ST needs to take the solution of the non-SIC case.

Our numerical evaluation leads to the following observations. 1) When the transmit power of the PT increases, the ST's channel access probability and throughput first increase and then decease. The first increase is because when the PT's transmit power increases, the ST has a higher chance to decode the PT's signal and a higher chance to harvest more energy. The subsequent decrease of the ST's channel access probability and throughput is because a high transmit power of the PT increases the probability that the PT does not need help from the ST. 2) Our proposed method can reduce the number of inner-bisection-search iterations by around 10%. 3) The distance between the PT and ST has a big effect on the secondary throughput.

Appendix A
The proof of Theorem 1

Define $\lambda_n = 1 + \frac{(P_e g_{sn} - \mu_n \sigma^2 + P_p g_{pn}) g_{ss}}{(\mu_n+1) g_{sn} \sigma^2}$. The first and second order derivatives of objective function $R_{ss,n}$ are derived as

$$\begin{cases} \frac{\mathrm{d}R_{ss,n}}{\mathrm{d}\rho} = \frac{-1}{2}\log(\lambda_n) + \frac{1-\rho}{2(\ln 2)\lambda_n}\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho}, \\ \frac{\mathrm{d}^2 R_{ss,n}}{\mathrm{d}\rho^2} = \frac{-1}{(\ln 2)\lambda_n}\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho} + \frac{1-\rho}{2(\ln 2)\lambda_n}\big(\frac{-1}{\lambda_n}(\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho})^2 + \frac{\mathrm{d}^2\lambda_n}{\mathrm{d}\rho^2}\big). \end{cases} \tag{48}$$

The expression of λ_n can be rewritten as $\lambda_n = 1 + \frac{\lambda_a}{\lambda_b}$, where $\lambda_a \triangleq (P_e g_{sn} - \mu_n \sigma^2 + P_p g_{pn}) g_{ss} \geq 0$ and $\lambda_b \triangleq (\mu_n + 1) g_{sn} \sigma^2 > 0$.[13] Then the first and second order derivatives of λ_n are given by

$$\begin{cases} \frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho} = \frac{1}{\lambda_b}\big(\frac{\mathrm{d}\lambda_a}{\mathrm{d}\rho} - \frac{\lambda_a}{\lambda_b}\frac{\mathrm{d}\lambda_b}{\mathrm{d}\rho}\big), \\ \frac{\mathrm{d}^2\lambda_n}{\mathrm{d}\rho^2} = \frac{-2\frac{\mathrm{d}\lambda_b}{\mathrm{d}\rho}\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho}}{\lambda_b} + \frac{\lambda_b\frac{\mathrm{d}^2\lambda_a}{\mathrm{d}\rho^2} - \lambda_a\frac{\mathrm{d}^2\lambda_b}{\mathrm{d}\rho^2}}{\lambda_b^2}, \end{cases} \tag{49}$$

where

$$\begin{cases} \frac{\mathrm{d}\lambda_a}{\mathrm{d}\rho} &= (g_{sn}\frac{\mathrm{d}P_e}{\mathrm{d}\rho} - \sigma^2\frac{\mathrm{d}\mu_n}{\mathrm{d}\rho})g_{ss}, \\ \frac{\mathrm{d}^2\lambda_a}{\mathrm{d}\rho^2} &= (g_{sn}\frac{\mathrm{d}^2P_e}{\mathrm{d}\rho^2} - \sigma^2\frac{\mathrm{d}^2\mu_n}{\mathrm{d}\rho^2})g_{ss}, \\ \frac{\mathrm{d}\lambda_b}{\mathrm{d}\rho} &= g_{sn}\sigma^2\frac{\mathrm{d}\mu_n}{\mathrm{d}\rho}, \\ \frac{\mathrm{d}^2\lambda_b}{\mathrm{d}\rho^2} &= g_{sn}\sigma^2\frac{\mathrm{d}^2\mu_n}{\mathrm{d}\rho^2}, \\ \frac{\mathrm{d}\mu_n}{\mathrm{d}\rho} &= \frac{2(\ln 2)\gamma^T}{(1-\rho)^2}2^{\frac{2\gamma^T}{1-\rho}}, \\ \frac{\mathrm{d}^2\mu_n}{\mathrm{d}\rho^2} &= \big(\frac{2(\ln 2)\gamma^T}{(1-\rho)^2}\big)^2 2^{\frac{2\gamma^T}{1-\rho}} + \frac{4(\ln 2)\gamma^T}{(1-\rho)^3}2^{\frac{2\gamma^T}{1-\rho}}, \end{cases} \tag{50}$$

with $\frac{\mathrm{d}P_e}{\mathrm{d}\rho}$ and $\frac{\mathrm{d}^2P_e}{\mathrm{d}\rho^2}$ given in (19).

[13] From (4), an equivalent form of λ_n is $\lambda_n = 1 + \frac{(1-\alpha)P_e g_{ss}}{\sigma^2}$, which is always not less than 1. Since $\lambda_n = 1 + \frac{\lambda_a}{\lambda_b}$ and $\lambda_b \triangleq (\mu_n + 1) g_{sp} \sigma^2 > 0$ (noting that for $\rho \in \mathcal{B}$, we always have $\mu_n > 0$), we have $\lambda_a \geq 0$.

According to [27, page 101], for a function over an interval, if its second order derivative is nonpositive when its first order derivative is 0, then the function is quasiconcave over the interval. Thus, next, we will show $R_{ss,n}$ is strictly quasiconcave, that is, $\frac{\mathrm{d}^2 R_{ss,n}}{\mathrm{d}\rho^2} < 0$ when $\frac{\mathrm{d}R_{ss,n}}{\mathrm{d}\rho} = 0$. So in the rest of this proof, we only consider what happens when $\frac{\mathrm{d}R_{ss,n}}{\mathrm{d}\rho} = 0$.

Based on the first equation of (48), $\frac{\mathrm{d}R_{ss,n}}{\mathrm{d}\rho} = 0$ is equivalent to

$$-\lambda_n \log(\lambda_n) + \frac{1-\rho}{(\ln 2)} \frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho} = 0. \tag{51}$$

As λ_n is always not less than 1, $\lambda_n \log(\lambda_n)$ is nonnegative. Together with (51), we have $\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho} \geq 0$. From the second equation of (49), we have $\frac{\mathrm{d}^2\lambda_n}{\mathrm{d}\rho^2} = \frac{-2\frac{\mathrm{d}\lambda_b}{\mathrm{d}\rho}\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho}}{\lambda_b} + \omega \leq \omega$, where $\omega = \frac{\lambda_b \frac{\mathrm{d}^2\lambda_a}{\mathrm{d}\rho^2} - \lambda_a \frac{\mathrm{d}^2\lambda_b}{\mathrm{d}\rho^2}}{\lambda_b^2}$. Here the inequality is because $\frac{\mathrm{d}\lambda_b}{\mathrm{d}\rho} > 0$ (from (50)), $\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho} \geq 0$, and $\lambda_b > 0$ (from definition of λ_b).

Using $\frac{\mathrm{d}^2\lambda_n}{\mathrm{d}\rho^2} \leq \omega$,, we can obtain the following for the second order derivative of $R_{ss,n}$

$$\begin{aligned} \frac{\mathrm{d}^2 R_{ss,n}}{\mathrm{d}\rho^2} &= \frac{-(1-\rho)(\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho})^2}{2(\ln 2)(\lambda_n)^2} + \frac{(1-\rho)\frac{\mathrm{d}^2\lambda_n}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho}}{2(\ln 2)\lambda_n} \\ &\leq \frac{-(1-\rho)(\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho})^2}{2(\ln 2)(\lambda_n)^2} + \frac{(1-\rho)\omega - 2\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho}}{2(\ln 2)\lambda_n}, \end{aligned} \tag{52}$$

in which the first equality is from (48). Apparently, for the right hand side of the inequality in (52), the first term is nonpositive. Next we show that the second term is negative. The numerator of the second term can be rewritten as

$$\begin{aligned} (1-\rho)\omega - 2\frac{\mathrm{d}\lambda_n}{\mathrm{d}\rho} = \; & \frac{\left((1-\rho)\frac{\mathrm{d}^2\lambda_a}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\lambda_a}{\mathrm{d}\rho}\right)\lambda_b}{\lambda_b^2} \\ & - \frac{\left((1-\rho)\frac{\mathrm{d}^2\lambda_b}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\lambda_b}{\mathrm{d}\rho}\right)\lambda_a}{\lambda_b^2}. \end{aligned} \tag{53}$$

For the two terms on the right hand-side of (53), we have

$$\begin{cases} (1-\rho)\frac{\mathrm{d}^2\lambda_a}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\lambda_a}{\mathrm{d}\rho} & = (1-\rho)(g_{sn}\frac{\mathrm{d}^2 P_e}{\mathrm{d}\rho^2} - \sigma^2\frac{\mathrm{d}^2\mu_n}{\mathrm{d}\rho^2})g_{ss} \\ & \quad -2(g_{sn}\frac{\mathrm{d}P_e}{\mathrm{d}\rho} - \sigma^2\frac{\mathrm{d}\mu_n}{\mathrm{d}\rho})g_{ss} \\ & = -\big((1-\rho)\frac{\mathrm{d}^2\mu_n}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\mu_n}{\mathrm{d}\rho}\big)\sigma^2 g_{ss} \\ & \overset{\text{(iv)}}{<} 0 \\ (1-\rho)\frac{\mathrm{d}^2\lambda_b}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\lambda_b}{\mathrm{d}\rho} & = \big((1-\rho)\frac{\mathrm{d}^2\mu_n}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\mu_n}{\mathrm{d}\rho}\big)g_{sn}\sigma^2 \\ & \overset{\text{(v)}}{>} 0, \end{cases} \tag{54}$$

in which steps (iv) and (v) come from the fact that

$$(1-\rho)\frac{\mathrm{d}^2\mu_n}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\mu_n}{\mathrm{d}\rho} \overset{\text{(vi)}}{=} \frac{\big(2(\ln 2)\gamma^T\big)^2}{(1-\rho)^3} 2^{\frac{2\gamma^T}{1-\rho}} > 0 \tag{55}$$

with step (vi) from (50).

Thus, the numerator of the second term on the right hand-side of the inequality in (52) is negative, and accordingly, we have $\frac{\mathrm{d}^2 R_{ss,n}}{\mathrm{d}\rho^2} < 0$ when $\frac{\mathrm{d}R_{ss,n}}{\mathrm{d}\rho} = 0$. Therefore, $R_{ss,n}$ is a quasiconcave function. This completes the proof.

Appendix B
The Proof of Theorem 4

Denote $\lambda^{w/o} = 1 + \frac{\lambda_a|_{n=p}}{\lambda_c}$, where $\lambda_a|_{n=p}$ and its first and second order derivatives are given in Appendix A, and λ_c and its first and second order derivatives are expressed as

$$\begin{cases} \lambda_c & = [\mu_p(P_e g_{sp} + \sigma^2) - P_p g_{pp}]g_{ss} \\ & \quad + (P_p g_{ps} + \sigma^2)(\mu_p + 1)g_{sp}, \\ \frac{\mathrm{d}\lambda_c}{\mathrm{d}\rho} & = (P_e g_{sp} + \sigma^2)g_{ss}\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho} + \mu_p g_{sp} g_{ss}\frac{\mathrm{d}P_e}{\mathrm{d}\rho} \\ & \quad + (P_p g_{ps} + \sigma^2)g_{sp}\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho}, \\ \frac{\mathrm{d}^2\lambda_c}{\mathrm{d}\rho^2} & = (P_e g_{sp} + \sigma^2)g_{ss}\frac{\mathrm{d}^2\mu_p}{\mathrm{d}\rho^2} + 2g_{sp}g_{ss}\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho}\frac{\mathrm{d}P_e}{\mathrm{d}\rho} \\ & \quad + \mu_p g_{sp} g_{ss}\frac{\mathrm{d}^2 P_e}{\mathrm{d}\rho^2} + (P_p g_{ps} + \sigma^2)g_{sp}\frac{\mathrm{d}^2\mu_p}{\mathrm{d}\rho^2}, \end{cases} \tag{56}$$

with $\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho}$ and $\frac{\mathrm{d}^2\mu_p}{\mathrm{d}\rho^2}$ given in (50), and $\frac{\mathrm{d}P_e}{\mathrm{d}\rho}$ and $\frac{\mathrm{d}^2 P_e}{\mathrm{d}\rho^2}$ given in (19).

The proof for quasiconcavity of $R_{ss}^{w/o}$ is similar to that in Appendix A if we replace λ_b in Appendix A by λ_c. The only difference is that we need to prove $(1-\rho)\frac{\mathrm{d}^2\lambda_c}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\lambda_c}{\mathrm{d}\rho} > 0$ instead of the second equation of (54) in Appendix A. Base on (56), we have

$$\begin{aligned} (1-\rho)\frac{\mathrm{d}^2\lambda_c}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\lambda_c}{\mathrm{d}\rho} & \overset{\text{(vii)}}{=} \big((1-\rho)\frac{\mathrm{d}^2\mu_p}{\mathrm{d}\rho^2} - 2\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho}\big) \\ & \quad \times \big((P_e g_{sp} + \sigma^2)g_{ss} + (P_p g_{ps} + \sigma^2)g_{sp}\big) \\ & \quad + 2(1-\rho)g_{sp}g_{ss}\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho}\frac{\mathrm{d}P_e}{\mathrm{d}\rho} \overset{\text{(viii)}}{>} 0, \end{aligned} \tag{57}$$

in which step (vii) uses (56) and $\frac{\mathrm{d}^2 P_e}{\mathrm{d}\rho^2} = \frac{2}{1-\rho}\frac{\mathrm{d}P_e}{\mathrm{d}\rho}$, and step (viii) uses (55), $\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho} > 0$ from (50), and $\frac{\mathrm{d}P_e}{\mathrm{d}\rho} > 0$ from (19).

Appendix C
Proof of Theorem 5

$\xi_t^{w/o}$ given in (47) can be rewritten as

$$\begin{aligned} \xi_t^{w/o} = \; & (1-\rho)(\mu_p + 1)(P_e g_{ss} + P_p g_{ps} + \sigma^2)g_{sp} - \\ & (1-\rho)\big\{[\mu_p(P_e g_{sp} + \sigma^2) - P_p g_{pp}]g_{ss} + \\ & (P_p g_{ps} + \sigma^2)(\mu_p + 1)g_{sp}\big\}2^{\frac{2t}{1-\rho}}. \end{aligned}$$

Defining $\chi \triangleq 2(\eta P_p g_{tt} + E_0 - E_c)$ and $\psi \triangleq 2\eta P_p g_{tt}$, P_e given in (1) can be represented as

$$P_e = \frac{1}{(1-\rho)}\big(\chi - (1-\rho)\psi\big). \tag{58}$$

Using (58) and applying some math manipulations, $\xi_t^{w/o}$ can be expressed as

$$\xi_t^{w/o} = \phi_a + \phi_b - \phi_c - \phi_d - \phi_e, \tag{59}$$

in which

$$\begin{cases} \phi_a & = \big(\chi - (1-\rho)\psi\big)(2^{\frac{2\gamma^T}{1-\rho}} - \frac{P_p g_{pp}}{\sigma^2})g_{ss}g_{sp}, \\ \phi_b & = (1-\rho)(2^{\frac{2\gamma^T}{1-\rho}} - \frac{P_p g_{pp}}{\sigma^2})(P_p g_{ps} + \sigma^2)g_{sp}, \\ \phi_c & = \big(\chi - (1-\rho)\psi\big)(2^{\frac{2\gamma^T}{1-\rho}} - \frac{P_p g_{pp}}{\sigma^2} - 1)g_{ss}g_{sp}2^{\frac{2t}{1-\rho}}, \\ \phi_d & = (1-\rho)(2^{\frac{2\gamma^T}{1-\rho}} - \frac{2P_p g_{pp}}{\sigma^2} - 1)g_{ss}\sigma^2 2^{\frac{2t}{1-\rho}}, \\ \phi_e & = (1-\rho)(2^{\frac{2\gamma^T}{1-\rho}} - \frac{P_p g_{pp}}{\sigma^2})(P_p g_{ps} + \sigma^2)g_{sp}2^{\frac{2t}{1-\rho}}. \end{cases} \tag{60}$$

The first order derivative of $\xi_t^{w/o}$ is expressed by

$$\frac{\mathrm{d}\xi_t^{w/o}}{\mathrm{d}\rho} = \frac{\mathrm{d}\phi_a}{\mathrm{d}\rho} + \frac{\mathrm{d}\phi_b}{\mathrm{d}\rho} - \frac{\mathrm{d}\phi_c}{\mathrm{d}\rho} - \frac{\mathrm{d}\phi_d}{\mathrm{d}\rho} - \frac{\mathrm{d}\phi_e}{\mathrm{d}\rho}, \tag{61}$$

where

$$\begin{cases} \frac{\mathrm{d}\phi_a}{\mathrm{d}\rho} = \big(\psi(\mu_p+1) + (1-\rho)P_e\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho}\big)g_{ss}g_{sp}, \\ \frac{\mathrm{d}\phi_b}{\mathrm{d}\rho} = \big(-\mu_p - 1 + (1-\rho)\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho}\big)(P_pg_{ps}+\sigma^2)g_{sp}, \\ \frac{\mathrm{d}\phi_c}{\mathrm{d}\rho} = \Big(\psi\mu_p + (1-\rho)P_e\big(\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho} + \mu_p\frac{2(\ln 2)t}{(1-\rho)^2}\big)\Big)g_{ss}g_{sp}2^{\frac{2t}{1-\rho}}, \\ \frac{\mathrm{d}\phi_d}{\mathrm{d}\rho} = \Big((1-\rho)\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho} + (\mu_p - \frac{P_pg_{pp}}{\sigma^2})\big(\frac{2(\ln 2)t}{1-\rho}-1\big)\Big) \\ \qquad \times g_{ss}\sigma^2 2^{\frac{2t}{1-\rho}}, \\ \frac{\mathrm{d}\phi_e}{\mathrm{d}\rho} = \Big((1-\rho)\frac{\mathrm{d}\mu_p}{\mathrm{d}\rho} + (\mu_p+1)\big(\frac{2(\ln 2)t}{1-\rho}-1\big)\Big)(P_pg_{ps} \\ \qquad + \sigma^2)g_{sp}2^{\frac{2t}{1-\rho}}. \end{cases} \tag{62}$$

Define two functions $G(y)$ and $K(y)$ over $y \geq 0$ as $G(y) = \frac{\big(2(\ln 2)y\big)^2}{(1-\rho)^3}$ and $K(y) = \frac{4(\ln 2)y}{(1-\rho)^3} + \frac{\big(2(\ln 2)y\big)^2}{(1-\rho)^4}$, respectively. We have

$$G(\gamma^T+t) = G(\gamma^T) + G(t) + \frac{8(\ln 2)^2\gamma^T t}{(1-\rho)^3} \geq G(\gamma^T) + G(t), \tag{63}$$

$$K(\gamma^T+t) = K(\gamma^T) + K(t) + \frac{8(\ln 2)^2\gamma^T t}{(1-\rho)^4} \geq K(\gamma^T) + K(t). \tag{64}$$

Then using $G(y)$ and $K(y)$, the second order derivatives of $(1-\rho)2^{\frac{2y}{1-\rho}}$ and $2^{\frac{2y}{1-\rho}}$ are written as $\frac{\mathrm{d}^2((1-\rho)2^{\frac{2y}{1-\rho}})}{\mathrm{d}\rho^2} = G(y)2^{\frac{2y}{1-\rho}}$ and $\frac{\mathrm{d}^2(2^{\frac{2y}{1-\rho}})}{\mathrm{d}\rho^2} = K(y)2^{\frac{2y}{1-\rho}}$. Using this, we have

$$\begin{cases} \frac{\mathrm{d}^2\phi_a}{\mathrm{d}\rho^2} &= \big(\chi K(\gamma^T) - \psi G(\gamma^T)\big)g_{ss}g_{sp}2^{\frac{2\gamma^T}{1-\rho}}, \\ \frac{\mathrm{d}^2\phi_b}{\mathrm{d}\rho^2} &= G(\gamma^T)(P_pg_{ps}+\sigma^2)g_{sp}2^{\frac{2\gamma^T}{1-\rho}}, \\ \frac{\mathrm{d}^2\phi_c}{\mathrm{d}\rho^2} &= \Big(\big(\chi K(\gamma^T+t) - \psi G(\gamma^T+t)\big)2^{\frac{2(\gamma^T+t)}{1-\rho}} \\ &\quad -\big(\chi K(t) - \psi G(t)\big)\big(\frac{P_pg_{pp}}{\sigma^2}+1\big)2^{\frac{2t}{1-\rho}}\Big)g_{ss}g_{sp}, \\ \frac{\mathrm{d}^2\phi_d}{\mathrm{d}\rho^2} &= G(\gamma^T+t)g_{ss}\sigma^2 2^{\frac{2(\gamma^T+t)}{1-\rho}} \\ &\quad -G(t)\big(\frac{2P_pg_{pp}}{\sigma^2}+1\big)g_{ss}\sigma^2 2^{\frac{2t}{1-\rho}}, \\ \frac{\mathrm{d}^2\phi_e}{\mathrm{d}\rho^2} &= G(\gamma^T+t)(P_pg_{ps}+\sigma^2)g_{sp}2^{\frac{2(\gamma^T+t)}{1-\rho}} \\ &\quad -G(t)\frac{P_pg_{pp}}{\sigma^2}(P_pg_{ps}+\sigma^2)g_{sp}2^{\frac{2t}{1-\rho}}. \end{cases} \tag{65}$$

In the expressions for $\frac{\mathrm{d}^2\phi_a}{\mathrm{d}\rho^2}$ and $\frac{\mathrm{d}^2\phi_c}{\mathrm{d}\rho^2}$, both of them contain the item $\chi K(y) - \psi G(y)$, where y takes the value γ^T, t, or (γ^T+t). By checking $\chi K(y) - \psi G(y)$, it can be represented as the following

$$\begin{aligned} \chi K(y) - \psi G(y) &= \chi\big(\tfrac{4(\ln 2)y}{(1-\rho)^3} + \tfrac{\big(2(\ln 2)y\big)^2}{(1-\rho)^4}\big) - \psi\tfrac{\big(2(\ln 2)y\big)^2}{(1-\rho)^3} \\ &= \tfrac{4(\ln 2)y}{(1-\rho)^3}\big(\chi + (\ln 2)yP_e\big), \end{aligned} \tag{66}$$

in which the last equality uses (58).

Then $\frac{\mathrm{d}^2\phi_a}{\mathrm{d}\rho^2} - \frac{\mathrm{d}^2\phi_c}{\mathrm{d}\rho^2}$ is expressed as

$$\begin{aligned} &\tfrac{\mathrm{d}^2\phi_a}{\mathrm{d}\rho^2} - \tfrac{\mathrm{d}^2\phi_c}{\mathrm{d}\rho^2} \\ &= \big(\chi K(\gamma^T) - \psi G(\gamma^T)\big)g_{ss}g_{sp}2^{\frac{2\gamma^T}{1-\rho}} \\ &\quad - \Big(\big(\chi K(\gamma^T+t) - \psi G(\gamma^T+t)\big)2^{\frac{2(\gamma^T+t)}{1-\rho}} \\ &\quad - \big(\chi K(t) - \psi G(t)\big)\big(\tfrac{P_pg_{pp}}{\sigma^2}+1\big)2^{\frac{2t}{1-\rho}}\Big)g_{ss}g_{sp} \\ &= -\Big(\tfrac{4(\ln 2)\gamma^T}{(1-\rho)^3}\big(\chi + (\ln 2)\gamma^T P_e\big)(2^{\frac{2t}{1-\rho}}-1)2^{\frac{2\gamma^T}{1-\rho}} \\ &\quad + \tfrac{4(\ln 2)t}{(1-\rho)^3}\big(\chi + (\ln 2)tP_e\big)\mu_p 2^{\frac{2t}{1-\rho}} \\ &\quad + \tfrac{8(\ln 2)^2\gamma^T t}{(1-\rho)^3}P_e 2^{\frac{2(\gamma^T+t)}{1-\rho}}\Big)g_{ss}g_{sp} \leq 0, \end{aligned} \tag{67}$$

in which the first equality is from (65), and the second equality is from the first equality of (63), the first equality of (64), (10), (66), and (58).

Moreover, we have

$$\begin{aligned} \tfrac{\mathrm{d}^2\phi_b}{\mathrm{d}\rho^2} - \tfrac{\mathrm{d}^2\phi_e}{\mathrm{d}\rho^2} &\overset{\text{(ix)}}{=} -\Big(G(\gamma^T+t)2^{\frac{2(\gamma^T+t)}{1-\rho}} - G(\gamma^T)2^{\frac{2\gamma^T}{1-\rho}} \\ &\quad -G(t)\tfrac{P_pg_{pp}}{\sigma^2}2^{\frac{2t}{1-\rho}}\Big)(P_pg_{ps}+\sigma^2)g_{sp} \\ &\overset{\text{(x)}}{\leq} -\Big(G(\gamma^T)(2^{\frac{2t}{1-\rho}}-1)2^{\frac{2\gamma^T}{1-\rho}} \\ &\quad +G(t)(\mu_p+1)2^{\frac{2t}{1-\rho}}\Big)(P_pg_{ps}+\sigma^2)g_{sp} \leq 0, \end{aligned} \tag{68}$$

in which (ix) is from (65), and (x) is from $G(\gamma^T+t) \geq G(\gamma^T) + G(t)$ and (10).

From (65) and $G(\gamma^T+t) \geq G(\gamma^T) + G(t)$, we have

$$\begin{aligned} \tfrac{\mathrm{d}^2\phi_d}{\mathrm{d}\rho^2} &\geq \big(G(\gamma^T) + G(t)\big)g_{ss}\sigma^2 2^{\frac{2(\gamma^T+t)}{1-\rho}} \\ &\quad - G(t)\big(\tfrac{2P_pg_{pp}}{\sigma^2}+1\big)g_{ss}\sigma^2 2^{\frac{2t}{1-\rho}} \\ &= G(\gamma^T)g_{ss}\sigma^2 2^{\frac{2(\gamma^T+t)}{1-\rho}} + G(t)(\mu_p - \tfrac{P_pg_{pp}}{\sigma^2})g_{ss}\sigma^2 2^{\frac{2t}{1-\rho}} > 0. \end{aligned} \tag{69}$$

Combining (59), (67), (68), and (69), it can be seen that $\frac{\mathrm{d}^2\xi_t^{w/o}}{\mathrm{d}\rho^2} < 0$, namely, $\xi_t^{w/o}$ is a strictly concave function.

REFERENCES

[1] F. Li, H. Jiang, R. Fan, and P. Tan, "Optimal cooperative strategy in energy harvesting cognitive radio networks," in *IEEE 86th Vehicular Technology Conference (VTC2017-Fall)*, Sept. 2017, pp. 1–6.

[2] S. Haykin, "Cognitive radio: brain-empowered wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 2, pp. 201–220, Feb. 2005.

[3] 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.

[4] Y. C. Liang, Y. Zeng, E. C. Y. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.

[5] R. Fan and H. Jiang, "Optimal multi-channel cooperative sensing in cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 9, no. 3, pp. 1128–1138, Mar. 2010.

[6] R. Fan, H. Jiang, Q. Guo, and Z. Zhang, "Joint optimal cooperative sensing and resource allocation in multichannel cognitive radio networks," *IEEE Trans. Veh. Technol.*, vol. 60, no. 2, pp. 722–729, Feb. 2011.

[7] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE VTC2013-Spring*, June 2013, pp. 1–5.

[8] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C. L. I, and H. V. Poor, "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, X. Lei, G. K. Karagiannidis, R. Schober, J. Yuan, and V. K. Bhargava, "A survey on non-orthogonal multiple access for 5G networks: Research challenges and future trends," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2181–2195, Oct. 2017.

[10] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[11] 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.

[12] L. Yang, Q. Ni, L. Lv, J. Chen, X. Xue, H. Zhang, and H. Jiang, "Cooperative non-orthogonal layered multicast multiple access for heterogeneous networks," *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1148-1165, Feb. 2019.

[13] Z. Ding, P. Fan, and H. V. Poor, "Impact of user pairing on 5G nonorthogonal multiple-access downlink transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6010–6023, Aug. 2016.

[14] Y. Zhang, Q. Yang, T. Zheng, H. Wang, Y. Ju, and Y. Meng, "Energy efficiency optimization in cognitive radio inspired non-orthogonal multiple access," in *Proc. 2016 IEEE PIMRC*, Sep. 2016, pp. 1–6.

[15] W. Liang, Z. Ding, Y. Li, and L. Song, "User pairing for downlink non-orthogonal multiple access networks using matching algorithm," *IEEE Trans. Commun.*, vol. 65, no. 12, pp. 5319–5332, Dec. 2017.

[16] 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. 10 152–10 157, Dec. 2016.

[17] M. Mohammadi, B. K. Chalise, A. Hakimi, Z. Mobini, H. A. Suraweera, and Z. Ding, "Beamforming design and power allocation for full-duplex non-orthogonal multiple access cognitive relaying," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 5952–5965, Dec. 2018.

[18] 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, Jun. 2017.

[19] L. Lv, L. Yang, H. Jiang, T. H. Luan, and J. Chen, "When NOMA meets multiuser cognitive radio: Opportunistic cooperation and user scheduling," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 6679–6684, Jul. 2018.

[20] L. Lv, J. Chen, Q. Ni, Z. Ding, and H. Jiang, "Cognitive non-orthogonal multiple access with cooperative relaying: A new wireless frontier for 5G spectrum sharing," *IEEE Commun. Mag.*, vol. 56, no. 4, pp. 188–195, Apr. 2018.

[21] S. Ulukus, A. Yener, E. Erkip, O. Simeone, M. Zorzi, P. Grover, and K. Huang, "Energy harvesting wireless communications: A review of recent advances," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 3, pp. 360–381, Mar. 2015.

[22] 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, no. 2, pp. 757–789, 2nd Quart. 2015.

[23] L. R. Varshney, "Transporting information and energy simultaneously," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 1612–1616.

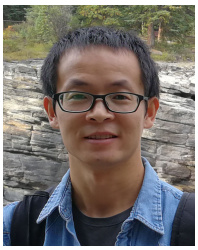
[24] X. Zhou, R. Zhang, and C. K. Ho, "Wireless information and power transfer: Architecture design and rate-energy tradeoff," *IEEE Trans. Commun.*, vol. 61, no. 11, pp. 4754–4767, Nov. 2013.

[25] D. Wang and S. Men, "Secure energy efficiency for noma based cognitive radio networks with nonlinear energy harvesting," *IEEE Access*, vol. 6, pp. 62 707–62 716, Oct. 2018.

[26] 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, Sep. 2017.

[27] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

[28] J. F. Traub and A. G. Werschulz, *Complexity and Information*. Cambridge, U.K.: Cambridge Univ. Press, 1998.



Fudong Li received the B.Sc. and M.Sc. degrees in communication engineering from Shandong University, Jinan, China, in 2009 and 2012, respectively. Since 2015, he has been working towards his Ph.D. degree at the University of Alberta, Edmonton, Alberta, Canada. His research interests include resource allocation in wireless systems, non-orthogonal multiple access, and energy harvesting.

Hai Jiang (SM'15) received the Ph.D. degree in electrical engineering from the University of Waterloo, Waterloo, Ontario, Canada, in 2006. He is currently a Professor at the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, Alberta, Canada. His research interests include radio resource management, cognitive radio networking, and cooperative communications.

Rongfei Fan received the Ph. D degree in electrical engineering from the University of Alberta, Edmonton, Alberta, Canada, in 2012. Since 2013, he has been a faculty member at the Beijing Institute of Technology, Beijing, China, where he is currently an Assistant Professor in the School of Information and Electronics. His research interests include cognitive radio, resource allocation in wireless networks, energy harvesting and statistical signal processing.

Peng Tan (M'04) received the Ph.D. degree in electrical engineering from the University of Alberta, Edmonton, Alberta, Canada, in 2006. He is currently with TELUS Communications Inc., where he is leading the technology and architecture development for 5G and immersive media experience. His research interests include 4G and 5G wireless technologies, as well as content delivery over wireless and IP networks.