

Wireless Communications and Mobile Computing

Green Communication and Networking

Lead Guest Editor: Yongpeng Wu

Guest Editors: Fuhui Zhou, Zan Li, Shunqing Zhang, Zheng Chu,
and Wolfgang H. Gerstaecker



Green Communication and Networking

Wireless Communications and Mobile Computing

Green Communication and Networking

Lead Guest Editor: Yongpeng Wu

Guest Editors: Fuhui Zhou, Zan Li, Shunqing Zhang, Zheng Chu,
and Wolfgang H. Gerstacker



Copyright © 2018 Hindawi. All rights reserved.

This is a special issue published in “Wireless Communications and Mobile Computing.” All articles are open access articles distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Editorial Board

- Javier Aguiar, Spain
Wessam Ajib, Canada
Muhammad Alam, China
Eva Antonino-Daviu, Spain
Shlomi Arnon, Israel
Leyre Azpilicueta, Mexico
Paolo Barsocchi, Italy
Alessandro Bazzi, Italy
Zdenek Becvar, Czech Republic
Francesco Benedetto, Italy
Olivier Berder, France
Ana M. Bernardos, Spain
Mauro Biagi, Italy
Dario Bruneo, Italy
Jun Cai, Canada
Zhipeng Cai, USA
Claudia Campolo, Italy
Gerardo Canfora, Italy
Rolando Carrasco, UK
Vicente Casares-Giner, Spain
Luis Castedo, Spain
Ioannis Chatzigiannakis, Greece
Lin Chen, France
Yu Chen, USA
Hui Cheng, UK
Ernestina Cianca, Italy
Riccardo Colella, Italy
Mario Collotta, Italy
Massimo Condoluci, Sweden
Bernard Cousin, France
Telmo Reis Cunha, Portugal
Igor Curcio, Finland
Laurie Cuthbert, Macau
Donatella Darsena, Italy
Pham Tien Dat, Japan
André de Almeida, Brazil
Antonio De Domenico, France
Antonio de la Oliva, Spain
Gianluca De Marco, Italy
Luca De Nardis, Italy
Liang Dong, USA
Mohammed El-Hajjar, UK
Oscar Esparza, Spain
Maria Fazio, Italy
- Mauro Femminella, Italy
Manuel Fernandez-Veiga, Spain
Gianluigi Ferrari, Italy
Ilario Filippini, Italy
Jesus Fontecha, Spain
Luca Foschini, Italy
A. G. Fragkiadakis, Greece
Sabrina Gaito, Italy
Óscar García, Spain
Manuel García Sánchez, Spain
L. J. García Villalba, Spain
José A. García-Naya, Spain
Miguel Garcia-Pineda, Spain
A.-J. García-Sánchez, Spain
Piedad Garrido, Spain
Vincent Gauthier, France
Carlo Giannelli, Italy
Carles Gomez, Spain
Juan A. Gomez-Pulido, Spain
Ke Guan, China
Antonio Guerrieri, Italy
Daojing He, China
Paul Honeine, France
Sergio Ilarri, Spain
Antonio Jara, Switzerland
Xiaohong Jiang, Japan
Minho Jo, Republic of Korea
Shigeru Kashiara, Japan
Dimitrios Katsaros, Greece
Minseok Kim, Japan
Mario Kolberg, UK
Nikos Komninos, UK
Juan A. L. Riquelme, Spain
Pavlos I. Lazaridis, UK
Tuan Anh Le, UK
Xianfu Lei, China
Hoa Le-Minh, UK
Jaime Lloret, Spain
Miguel López-Benítez, UK
Martín López-Nores, Spain
Javier D. S. Lorente, Spain
Tony T. Luo, Singapore
Maode Ma, Singapore
Imadeldin Mahgoub, USA
- Pietro Manzoni, Spain
Álvaro Marco, Spain
Gustavo Marfia, Italy
Francisco J. Martinez, Spain
Davide Mattera, Italy
Michael McGuire, Canada
Nathalie Mitton, France
Klaus Moessner, UK
Antonella Molinaro, Italy
Simone Morosi, Italy
Kumudu S. Munasinghe, Australia
Enrico Natalizio, France
Keivan Navaie, UK
Thomas Newe, Ireland
Wing Kwan Ng, Australia
Tuan M. Nguyen, Vietnam
Petros Nicopolitidis, Greece
Giovanni Pau, Italy
Rafael Pérez-Jiménez, Spain
Matteo Petracca, Italy
Nada Y. Philip, UK
Marco Picone, Italy
Daniele Pinchera, Italy
Giuseppe Piro, Italy
Vicent Pla, Spain
Javier Prieto, Spain
Rüdiger C. Prys, Germany
Junaid Qadir, Pakistan
Sujan Rajbhandari, UK
Rajib Rana, Australia
Luca Reggiani, Italy
Daniel G. Reina, Spain
Abusayeed Saifullah, USA
Jose Santa, Spain
Stefano Savazzi, Italy
Hans Schotten, Germany
Patrick Seeling, USA
Muhammad Z. Shakir, UK
Mohammad Shojafar, Italy
Giovanni Stea, Italy
Enrique Stevens-Navarro, Mexico
Zhou Su, Japan
Luis Suarez, Russia
Ville Syrjälä, Finland



Hwee Pink Tan, Singapore
Pierre-Martin Tardif, Canada
Mauro Tortonesi, Italy
Federico Tramarin, Italy
Reza Monir Vaghefi, USA

Juan F. Valenzuela-Valdés, Spain
Aline C. Viana, France
Enrico M. Vitucci, Italy
Honggang Wang, USA
Jie Yang, USA

Sherali Zeadally, USA
Jie Zhang, UK
Meiling Zhu, UK

Contents

Green Communication and Networking

Yongpeng Wu , Fuhui Zhou, Zan Li , Shunqing Zhang, Zheng Chu, and Wolfgang H. Gerstacker
Editorial (3 pages), Article ID 1921353, Volume 2018 (2018)

Prolonging the Lifetime of Wireless Sensor Networks: A Review of Current Techniques

Felicia Engmann , Ferdinand Apietu Katsriku , Jamal-Deen Abdulai , Kofi Sarpong Adu-Manu,
and Frank Kataka Banaseka
Review Article (23 pages), Article ID 8035065, Volume 2018 (2018)

Low Cost and High Efficiency Hybrid Architecture Massive MIMO Systems Based on DFT Processing

Weiqiang Tan , Guixian Xu, Elisabeth De Carvalho, Mu Zhou , Lisheng Fan , and Chunguo Li 
Research Article (11 pages), Article ID 7597290, Volume 2018 (2018)

AN-Aided Transmit Beamforming Design for Secured Cognitive Radio Networks with SWIPT

Weili Ge , Zhengyu Zhu , Zhongyong Wang , and Zhengdao Yuan
Research Article (13 pages), Article ID 6956313, Volume 2018 (2018)

A Context-Aware Location Differential Perturbation Scheme for Privacy-Aware Users in Mobile Environment

Xuejun Zhang , Haiyan Huang, Shan Huang, Qian Chen, Tao Ju, and Xiaogang Du 
Research Article (15 pages), Article ID 9173519, Volume 2018 (2018)

Joint Range-Doppler-Angle Estimation for OFDM-Based RadCom System via Tensor Decomposition

Bo Kong , Yuhao Wang , Xiaohua Deng, and Dong Qin 
Research Article (12 pages), Article ID 2708416, Volume 2018 (2018)

Performance and Power Consumption Analysis of IEEE802.11ah for Smart Grid

Zhe Zheng , Wenpeng Cui, Lei Qiao, and Jinghong Guo
Research Article (8 pages), Article ID 5286560, Volume 2018 (2018)

Energy Efficiency Maximization of Dynamic CoMP-JT Algorithm in Dense Small Cell Networks

Xuefei Peng, Jiandong Li , and Yifei Xu
Research Article (7 pages), Article ID 8572489, Volume 2018 (2018)

Resource Allocation for Green Cognitive Radios: Energy Efficiency Maximization

Zhou Yang, Wenqian Jiang , and Gang Li
Research Article (16 pages), Article ID 1327030, Volume 2018 (2018)

Power-Splitting Scheme for Nonlinear Energy Harvesting AF Relaying with Direct Link

Xiaobo Bai , Jingfeng Shao, Jiangang Tian, and Liqin Shi 
Research Article (8 pages), Article ID 7906957, Volume 2018 (2018)

Security-Reliability Tradeoff Analysis in Multisource Multirelay Cooperative Networks with Multiple Cochannel Interferers

Weilong Hu , Jiangbo Si , and Hongyan Li
Research Article (12 pages), Article ID 2379427, Volume 2018 (2018)

Dynamic Power Splitting Strategy for SWIPT Based Two-Way Multiplicative AF Relay Networks with Nonlinear Energy Harvesting Model

Tianci Wang , Guangyue Lu, Yinghui Ye , and Yuan Ren
Research Article (9 pages), Article ID 1802063, Volume 2018 (2018)

Energy-Efficient Incentives Resource Allocation Scheme in Cooperative Communication System

Zi Yan Liu , Pan Mao, Li Feng, and Shi Mei Liu
Research Article (8 pages), Article ID 5452120, Volume 2018 (2018)

Energy Efficiency Maximized Resource Allocation for Opportunistic Relay-Aided OFDMA Downlink with Subcarrier Pairing

Tao Wang , Chao Ma, Yanzan Sun, Shunqing Zhang, and Yating Wu
Research Article (10 pages), Article ID 9046847, Volume 2018 (2018)

Average SEP of AF Relaying in Nakagami- m Fading Environments

Dong Qin , Yuhao Wang , and Tianqing Zhou 
Research Article (7 pages), Article ID 6581827, Volume 2018 (2018)

Editorial

Green Communication and Networking

Yongpeng Wu ¹, **Fuhui Zhou**,² **Zan Li** ³, **Shunqing Zhang**,⁴
Zheng Chu,⁵ and **Wolfgang H. Gerstacker**⁶

¹Shanghai Key Laboratory of Navigation and Location Based Services, Shanghai Jiao Tong University, Minhang, 200240, China

²The School of Information Engineering, Nanchang University, 330031, China

³The Integrated Service Networks Lab of Xidian University, Xi'an, 710071, China

⁴The Shanghai Institute for Advanced Communication and Data Science, Shanghai University, Shanghai, 200444, China

⁵The 5G Innovation Centre, Institute for Communication Systems, University of Surrey, Guildford GU2 7XH, UK

⁶The Institute for Digital Communications of Friedrich-Alexander University, Erlangen, 91054, Germany

Correspondence should be addressed to Yongpeng Wu; yongpeng.wu2016@gmail.com

Received 13 August 2018; Accepted 13 August 2018; Published 2 September 2018

Copyright © 2018 Yongpeng Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Energy crisis and rising concerns on Greenhouse Gas (GHG) emissions have always been crucial issues faced by the development of wireless communication techniques. As a more energy-efficient network architecture, green communications and networks (GCN) have recently attracted significant attention from academia and industry. In particular, the newly designed GCN not only can alleviate the greenhouse effect and decrease the operational expenditure but also can attain sustainable development due to the descending independence on fossil fuel and the exploitation of renewable energy resources. In order to enable the technical and economical GCNs, several emerging techniques have been proposed including energy-efficient and energy harvesting techniques. Although these emerging techniques have drawn considerable attention and have been studied recently, there are still many open theoretical and practical problems to be addressed. Specifically, most of the existing works have focused on optimizing a single objective of GCNs, such as energy efficiency. Since there are multiple conflicting objectives in GCNs, e.g., spectral efficiency and energy efficiency, multiobjective strategies are required to be in order to achieve a good tradeoff among the conflicting objectives. Moreover, since considered nonorthogonal multiple access techniques have advantages in energy efficiency and massive connectivity, how to apply nonorthogonal multiple access techniques into GCNs needs to be further investigated. Furthermore, the conventional linear energy harvesting model is ideal in

practice. How to design GCNs under practical nonlinear energy harvesting models is required to be focused.

This special issue aims to provide a comprehensive overview of the state of the art in theory and practice for realizing GCNs, which will bring together researchers from academia, industry, and governmental agencies to promote the research and development needed to address the major challenges that pertain to this cutting-edge research topic.

- (i) Multiobjective resource allocation strategies for GCNs
- (ii) Multiobjective optimization theory for GCNs
- (iii) Multiobjective energy-efficient techniques for GCNs
- (iv) Energy-efficient cooperative techniques for GCNs
- (v) Energy-efficient nonorthogonal multiple access techniques for GCNs
- (vi) Practical energy harvesting models for GCNs
- (vii) Resource optimization for GCNs under practical nonlinear energy harvesting models
- (viii) Physical layer security techniques for GCNs
- (ix) Multiantenna techniques for GCNs

Twenty-five papers were submitted for this special issue. Our distinguished reviewers from respective research fields narrowed the field to fourteen papers which were finally

accepted. The following is a short summary of the findings of each of these papers.

T. Wang et al. studied an energy efficiency maximization for an orthogonal frequency division multiple access downlink network aided by a relay station with subcarrier pairing. A resource allocation scheme was proposed to maximize the energy efficiency. A resource allocation algorithm was proposed to solve the formulated problem. Simulation results have shown the impact of minimum required rate of the network on the energy efficiency of the network.

Z. Y. Liu et al. have proposed an energy-efficient incentives resource allocation scheme for a cooperative communication system. The energy efficiency was maximized by the proposed scheme. Moreover, an improved Intelligent Water Drops algorithm combined with Genetic Algorithm. Simulation results demonstrate that the proposed resource allocation scheme can effectively improve the energy efficiency.

Z. Yang et al. have studied a green cognitive radio network. The energy efficiency maximized problems were studied in delay-insensitive green cognitive radio and delay-sensitive green cognitive radio. Two algorithms based on the proposed resource allocation strategies were proposed to solve the formulated problems. Simulation results have shown that the maximum energy efficiency of the secondary user achieved under the average transmit power constraint is higher than that achieved under the peak transmit power constraint.

X. Peng et al. have proposed a joint user association and power allocation to maximize the energy efficiency of dense small cell networks. A dynamic coordinated multipoint joint transmission algorithm was proposed to improve energy efficiency. The proposed solution has a much lower computational complexity. Simulation results have shown that the proposed solution has a better performance.

W. Tan et al. have studied the hybrid architecture of multiuser massive MIMO systems. The digital domain utilized the zero-forcing precoding scheme and the analog domain used discrete Fourier transform processing to significantly reduce hardware cost and energy consumption. The analytical expressions on the total achievable spectral efficiency and energy efficiency were derived. The total achievable energy efficiency of hybrid architectures with discrete Fourier transform processing outperforms other massive MIMO architectures.

T. Wang et al. studied an energy-constrained two-way multiplicative amplify-and-forward relay network with simultaneous wireless information and power transfer. A practical nonlinear energy harvesting model was considered. The outage throughput was maximized by the proposed resource allocation scheme. Simulation results have shown the superiority of proposed the dynamic power splitting strategy in terms of the outage throughput.

X. Bai et al. designed a power-splitting scheme in an amplify-and-forward energy-constrained relay system with simultaneous wireless information and power transfer in the presence of a direct link between the source and the destination. A practical nonlinear energy harvesting (EH) model was considered. The authors have proposed a resource

allocation scheme for maximizing the system capacity. Simulation results have demonstrated that a higher system capacity can be achieved when the PS scheme is optimized based on nonlinear EH models instead of the linear EH model

W. Ge et al. have studied multiple-input single-output secured cognitive radio networks relying on simultaneous wireless information and power transfer. An AN-aided transmit beamforming was optimized to improve the security of the studied networks. The transmit power of the information signal was minimized subject to the secrecy rate constraint, the harvested energy constraint, and the total transmit power. Simulation results have validated the performance of the proposed algorithms.

F. Engmann et al. have presented a survey paper for the current techniques for prolonging the lifetime of wireless sensor networks. This paper presented the state of the art in the energy management schemes, such as energy harvesting, energy transfer, and energy conservation methods. The remaining challenges and the open issues for future research work were discussed.

B. Kong et al. studied radar and communication systems with orthogonal frequency division multiplexing. This system are promising in the next generation wireless communication due to its high energy efficiency and spectral efficiency. The authors have proposed a joint Range-Doppler-Angle Estimation scheme. A parameterized rectification method was proposed to incorporate the inherent structures of the factor matrices. The numerical experiments have demonstrated superior performance of the proposed algorithm compared with the existing methods.

W. Hu et al. studied a cooperative relaying wireless communication networks. The physical layer security of a multisource multirelay cooperative relaying communication network was investigated by considering the influence of cochannel interference from a security-reliability tradeoff perspective. Theoretical and simulation results have shown that a better security-reliability tradeoff performance can be achieved by increasing the number of sources, relays, and cochannel interferers.

D. Qin et al. studied an exact average symbol error probability for amplify and forward relaying in independent Nakagami-m fading environments. The closed-form symbol error probability was expressed in the form of Lauricella multivariate hypergeometric function. Four modulation modes were studied. Simulation results have verified the theoretical analysis.

Z. Zheng et al. have presented performance and power consumption analysis of IEEE802.11ah for smart grid. The authors have simulated the new added highly robust 1MHz bandwidth and Modulation Coding Scheme 10 in the 802.11ah physical layer and analyzed the coverage range and energy saving performance of 802.11ah based on the simulation results. The analysis has shown that the 802.11ah at 1 MHz MCS 10 mode can obtain 2.5-3 dB gain.

X. Zhang et al. proposed a context-aware location differential perturbation scheme for privacy-aware users in mobile environment. It can enhance the user's location privacy without requiring a trusted third party. This improves the operational efficiency of the system. The proposed scheme

can provide strong privacy guarantees with a bounded accuracy loss while improving retrieval accuracy.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this article.

Yongpeng Wu
Fuhui Zhou
Zan Li
Shunqing Zhang
Zheng Chu
Wolfgang H. Gerstaecker

Review Article

Prolonging the Lifetime of Wireless Sensor Networks: A Review of Current Techniques

Felicia Engmann ¹, **Ferdinand Apietu Katsriku** ², **Jamal-Deen Abdulai** ²,
Kofi Sarpong Adu-Manu² and **Frank Kataka Banaseka**²

¹*School of Technology, Ghana Institute of Management and Public Administration, Ghana*

²*Department of Computer Science, University of Ghana, Ghana*

Correspondence should be addressed to Felicia Engmann; fnaengmann@st.ug.edu.gh

Received 9 March 2018; Revised 10 July 2018; Accepted 17 July 2018; Published 15 August 2018

Academic Editor: Zheng Chu

Copyright © 2018 Felicia Engmann et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There has been an increase in research interest in wireless sensor networks (WSNs) as a result of the potential for their widespread use in many different areas like home automation, security, environmental monitoring, and many more. Despite the successes gained, the widespread adoption of WSNs particularly in remote and inaccessible places where their use is most beneficial is hampered by the major challenge of limited energy, being in most instances battery powered. To prolong the lifetime for these energy hungry sensor nodes, energy management schemes have been proposed in the literature to keep the sensor nodes alive making the network more operational and efficient. Currently, emphasis has been placed on energy harvesting, energy transfer, and energy conservation methods as the primary means of maintaining the network lifetime. These energy management techniques are designed to balance the energy in the overall network. The current review presents the state of the art in the energy management schemes, the remaining challenges, and the open issues for future research work.

1. Introduction

Energy efficiency has become a major theme in wireless sensor network (WSN) research. The interest in energy efficiency may be attributed to limitations imposed by the batteries used to power such devices. These batteries are usually the main source of power for these devices and are characterized by a limited lifespan, after which they are recharged or discarded. WSNs form the backbone of ubiquitous computing applications such as military surveillance, disaster, environmental, structural, health and security, and wildlife and habitat monitoring as well as precision agriculture. Deployment of sensor nodes is usually in inaccessible environments, and with limited battery capacity their lifetime is usually an issue of major concern. Several techniques have been proposed in the literature to increase the lifetime of sensor nodes as well as the sensor networks [1–6]. In recent times, long lasting sensor nodes that may never die have been proposed [7–9].

Several definitions have been proposed for the lifetime of a sensor network; however, a generally accepted definition is when the network degrades to a point when it is no longer

able to perform its intended function [10]. This could be when any of the following events occur: when the first sensor node dies or when a number or percentage of the nodes die or when the network is partitioned such that there is no communication between the subnetworks or when coverage is lost [10–12].

To help extend the lifetime of sensor nodes and networks, energy conservation methods are usually employed. In this, an effort is made to reduce the energy consumed by the unit. The authors in [3] broadly categorized energy conservation schemes under the three main headings: duty cycling, data driven, and mobility driven techniques. Duty cycling is aimed at reducing idle listening when the node's radio waits in vain for frames and overhearing when nodes stay active listening to uninterested frames. Data driven techniques use some parameters of the data themselves to make decisions to reduce energy consumption during communication while mobility schemes consider the mobility of the sink or relay nodes as a factor affecting the energy consumed in the network.

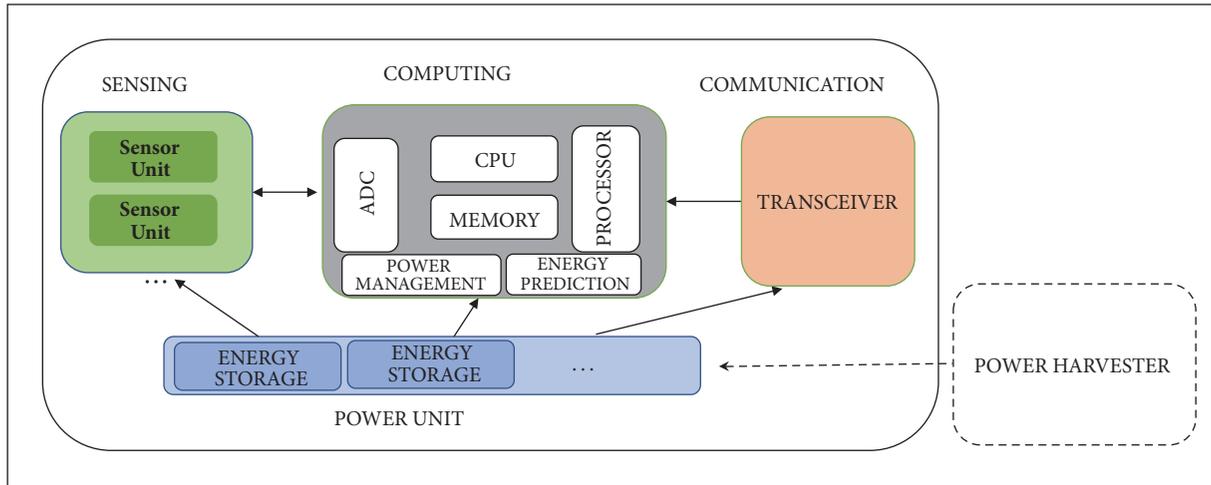


FIGURE 1: A typical architecture of a wireless sensor node. Adapted from [64].

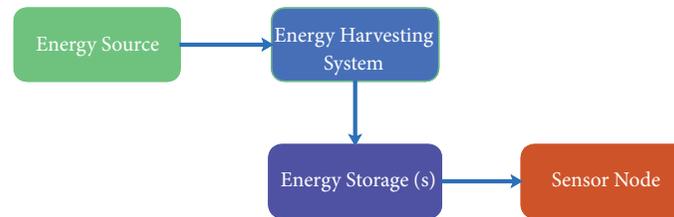


FIGURE 2: Ambient energy harvesting to store and use. Adapted from [37].



FIGURE 3: Ambient energy harvesting for direct use. Adapted from [37].

In Figure 1 the architecture of a typical wireless sensor node is shown. Each component of the sensor node as seen in MicaZ mote is presented. As may be observed, a typical node will consist of four major components, a sensing unit, processing unit, communications unit, and a power unit. Of the different components, the communications unit, which involves data transfer (involving both transmission and reception), expends a significantly higher proportion of the energy available [13]. This is represented in Figure 11.

Typical energy conservation techniques simply seek to prolong the lifespan of the network by reducing the energy used and do not typically require the introduction of new sources of energy. To increase the energy available to the sensor nodes, energy harvesting techniques have been proposed [14, 15]. A key limitation of this technique is that energy sources may not always be available and hence there is the need to store the harvested energy using rechargeable batteries or low-powered supercapacitors as in Figure 2 although in some cases the energy is utilized directly by the nodes as shown in Figure 3.

Another recent technique employed for prolonging the lifetime of sensor nodes and the network is transferring

energy from energy rich node to the energy deficient nodes. Energy transfer may be done wirelessly from a specialized energy harvesting node or an energy resourced node to an energy hungry node in the same network. The energy transfer may be continuous or on-demand but is limited by the cost of charge and discharge losses associated with it. Several approaches have been proposed in the literature to provide reliable energy transfer to increase the network lifetime [9, 16–22].

The method of ensuring that nodes have enough energy to function in the network by maintaining appropriate energy levels and transferring energy from an energy resourced node to an energy hungry node in the same network is referred to as energy balancing. The use of energy balancing approach to extend the life time of the networks may involve the use of any of the following schemes or a combination of them: energy conservation, energy harvesting, or wireless energy transfer. Energy conservation is the sparing use of energy in sensor nodes to allow sensor networks to be able to function as required [3]. It usually involves minimizing the communication cost in nodes [3, 15] since the radio is known to be the greatest consumer of the available energy [12, 23, 24]. It may also be achieved by developing energy efficient routing protocols, clustering approaches, sleep/wake-up optimization (duty cycling), and in some cases mobility [15, 25, 26].

Energy management schemes are the techniques designed for the efficient use of energy in a network [23] and in some instances for efficient use of harvested energy

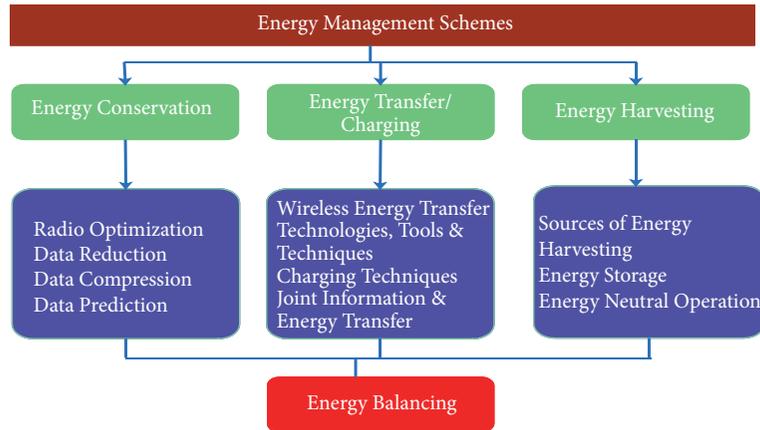


FIGURE 4: Energy management schemes in WSN.

[15]. Although some of the proposed energy management techniques assume that data acquisition through sensing consumes less energy than data transmission [3, 23], this may not be so for all applications [27, 28] especially in the case of energy hungry sensors, e.g., gas sensors. Most of these techniques are used to prolong the lifetime of sensor nodes by either reducing energy consumption or replenishing the consumed energy in battery powered nodes or low-powered capacitors. In this paper, we attempt to categorize the proposed energy management techniques into energy conservation mechanisms, energy harvesting, and energy transfer/wireless charging mechanisms. In some applications, [29, 30] sensing may consume significant percentage of the energy available. The broad categorizations of energy management used in this paper are energy conservation mechanisms, energy harvesting, and energy transfer/wireless charging mechanisms as presented in Figure 4. A holistic approach to achieving energy balancing in a network not only must be limited to energy harvesting and transfer but also should include energy conservation.

In this paper we present the state of the art in energy management schemes (i.e., energy harvesting, energy conservation, and energy transfer) and present the techniques used for harvesting, transferring, and conserving energy in WSNs. We discuss management schemes related to radio optimization, data reduction, aggregation, compression and prediction, and wireless transfer technologies and techniques. We present the concept of energy balancing in wireless sensor network when energy harvesting, transfer, and conservation are used efficiently. We discuss limitations in existing simulators and emulators that are designed for modeling WSN applications. Finally, we discuss the challenges and future research directions for energy management schemes.

The rest of the paper is organized as follows. In Section 2, we describe the different energy harvesting sources available and approaches. In Section 3, we describe the different wireless energy transfer techniques and technologies and present the various simulation and emulation tools for modeling recent WSNs applications. Section 4 provides current techniques for energy conservation. In Section 5 current energy balancing schemes are presented. In Section 6, we present

the challenges related to energy management schemes and provide future research directions in developing wireless sensor networks that consider the trio (i.e., harvesting, transferring, and conserving) energy management scheme. Finally, Section 7 concludes the paper.

2. Energy Harvesting

Energy harvesting approaches scavenge for energy from the external environment such as wind, vibrations, solar, acoustic, and thermal. The techniques used in energy harvesting convert energy from the environment into electrical energy that can be used in wireless sensing nodes/devices. In wireless sensor networks, energy harvesting can be used to overcome the challenge of energy depletion that causes shorter lifetime of the nodes in the network and in other cases of the black hole problem [31]. To realize the promised benefits of energy harvesting, concerted effort is required on the part of researchers to address some outstanding issues. Energy harvesting does not guarantee immortal nodes and continuous operation due to the uncontrollable energy sources, making them unpredictable and difficult to model. The constant unavailability of energy harvesting sources is discussed in [15, 32, 33]; hence a buffer is proposed to store energy for later use, using a battery-less sensor node and low-powered capacitors to act as buffers [5, 34], as shown in Figures 2 and 3. An example is solar energy which is not available for harvesting at night due to the absence of the sun [35]. Table 1 gives specifications of some commercially available solar energy harvesting units for use in sensor nodes. In energy harvesting, nodes in the network may be attached with special devices for scavenging energy from the ambient environment for conversion into electric energy. In the case of solar energy, the size of the panel is directly proportional to the amount of energy converted through the photovoltaic technique [34]. This poses a challenge when the energy harvesting device becomes larger than the sensor node. Special energy harvesting devices may therefore be provided in the network to scavenge energy and then wirelessly transfer them to nodes. Powercast technology [36] harvests energy from intentional, anticipated, and known

TABLE 1: Specifications of solar energy harvesting sensor nodes [37].

Node	Solar Panel Power (mW)	Solar Panel Size (in ²)	Energy Availability (mWh/day)	Storage Type	Battery Type	Battery Capacity (mAh)	Sensor Node Used	MPPT Usage
Heliomote	190	3.75 * 2.5	1140	Battery	Ni-MH	1800	Mica2	No
HydroWatch	276	2.3 * 2.3	139	Battery	Ni-MH	2500	TelosB	Yes
Flechl	-	4.56 * 3.35	2100	Battery	Ni-MH	2500	NA	NO
Everlast	450	2.25 * 3.75	2700	Supercap (100F)	NA	NA	NA	Yes
SolarBiscuit	150	2 * 2	900	Supercap (1F)	NA	NA	NA	NO
Sunflower	4 PIN Photodiodes 20 mW	NA	100	Supercap (0.2F)	NA	NA	NA	NO
AmbiMax	400	3.75 * 2.5	1200	Supercap (two 22F) & Battery	Li-poly	200	TelosB	NO
Prometheus	130	3.23 * 1.45	780	Supercap (two 22F) & Battery	Li-poly	200	TelosB	NO

sources using the Powerharvester Receivers. Powerharvester Receivers are designed for 50 standard antennas on the 902 928 MHz frequency band.

2.1. Sources of Energy Harvesting. The source from which energy is harvested in a sensor network is a valuable resource since it determines the amount of energy available to the network and the rate of conversion from the source to electrical energy. Energy harvested may be classified under ambient sources, which are sources available in the surrounding environment and human sources [37]. Ambient sources of energy discussed include solar, vibration, thermal, and radio frequency.

2.2. Solar Energy Harvesting. Solar energy is an affordable and clean source of energy given its abundance in the environment. The harvesting of solar energy through photovoltaic effect is seen as the likely choice for sensor nodes with energy harvesting [15, 35, 38]. Even with its abundance there are times of the day when solar will not be available; hence there is a need for energy storage that balances the energy stored with the consumption rate of the sensor node. In [32], an energy neutral operation is employed when solar energy was the only source of energy and the sensor node has no battery. Solar energy is obtained when a solar cell receives sunlight with appropriate energy. The amount of energy derived from a typical solar system is dependent on the amount of illumination and the surface area of the solar cell with power conversion efficiencies of 15% to 25% on crystalline silicon PV cells [34]. The other known PV cells are the monocrystalline, polycrystalline, and thin-film based [34]. Table 1 is a summary of some commercial solar harvesting tools from [37] and their specifications making them useful in WSNs.

2.3. Vibration Energy Harvesting. Vibrational energy may be obtained through activities that produce sufficient vibrations like subways, industrial machinery, and vehicles. Amount of energy harvested is approximated in 100-W range using mechanical-to-electrical energy generators (MEEG) that use piezoelectric (ferroelectrics) and magnetostrictive materials, and electrostatic or electromagnetic mechanisms to harvest energy [39, 40]. The harvested energy is directly proportional to the size of the MEEG used. In sensor networks where the smaller size of the node is a requirement, vibration may not be the best choice.

2.4. Thermal Energy Harvesting. Thermal energy is based on the existence of a temperature difference within an environment. The amount of energy obtainable is determined by the Carnot cycle as

$$\frac{(T_h - T_l)}{T_h} = \frac{\Delta T}{T_h} \quad (1)$$

where T_h and T_l are the maximum and minimum temperatures of the thermodynamic cycle.

Efficiency values up to 17% have been achieved for small temperature gradients based on the Carnot cycle [41]. Thermal energy harvesting has found application in many areas

including use in devices attached to the body and implantable devices such as pacemakers for the heart. It is possible to envisage their use in other monitoring applications where a temperature difference exists. A thermal energy harvester capable of achieving an output of 100 μW was reported in [42].

2.5. Radio Frequencies Energy Harvesting. Given the large number of radio transmitters available in any urban environment, harvesting energy from this source is very appealing. Those devices capable of using harvested RF energy will have very limited power requirements. In addition, they must be in close proximity to the energy source or have a very large antenna for collecting the energy. The basic principle of operation is for the antennas to receive RF energy from the atmosphere and convert them to electrical signals as shown in Figure 5. The matching circuit is made up of capacitor and inductor components and is used to maximize RF energy in the circuit. The voltage multiplier is made up of diodes and capacitors and the resulting energy is stored in either supercapacitors or rechargeable batteries. The conversion of RF signals to DC energy is dependent on the source of the power, antenna gains, and distance between source and receiver nodes and the energy conversion rate [15], given that the power density of a receiving antenna is

$$P = \frac{E^2}{Z_o} \quad (2)$$

where E is the electric field and Z_o is the radiation resistance of free space. Assuming $Z_o = 377$ ohms and an E value of 0.5 V/m we obtain a power density value of 0.13 $\mu W/cm^2$. Electric field values larger than 1 V/m are extremely rare. Progress in the use of RF sources will require advancement in power requirements of wireless sensor nodes.

Some of the RF technologies that exist but are not optimized for WSN use include Bluetooth, Wifi technology (IEEE 802.11a/b/h/g), and Ultra-Wideband (UWB IEEE 802.15.3). UWB has greater ratio of velocity with lower power consumption as compared to Wifi and Bluetooth but is limited to short range communications. Others that are being developed for WSN use include Wavenis by Coronis Systems, Wibree by Nokia, and Zigbee which is widely used by most WSN systems. RF power harvesting shown in Figure 5, convert RF energy emitted by RF sources such as TV signals and wireless radio networks.

2.6. Energy Storage. The use of harvested energy in WSN has a limitation since it is not always available. There is often a need to store the harvested energy for later use. Sensor nodes equipped for energy harvesting either have attached storage devices to store the harvested energy for later use as in Figure 3 or may not have storage devices but directly use the harvested energy in the node as in Figure 2. The storage devices could be either batteries (rechargeable and nonrechargeable) or supercapacitors. To replenish energy levels in WSN, rechargeable batteries and supercapacitors are used. Batteries have limited recharge cycles [5], and hence to prolong the lifetime of nodes energy conservation techniques

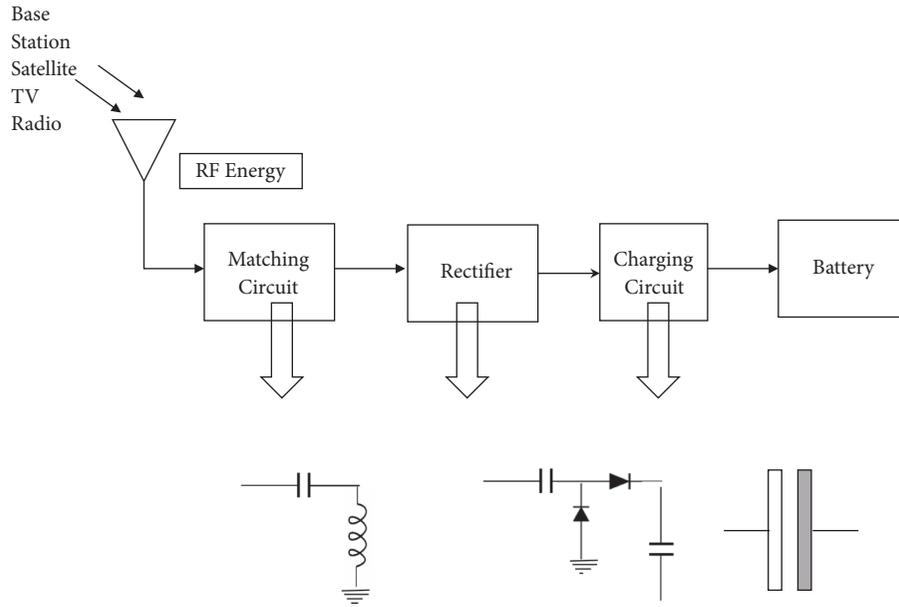


FIGURE 5: RF energy harvesting system.

must be employed together with energy harvesting. The use of supercapacitors [5, 43] is an alternative to batteries and may be repowered by energy harvesting. Supercapacitors may be recharged with a recharge cycle of half a million years with a 10-year functioning lifetime before the energy stored is reduced by 20% [44]. Supercapacitors have become more useful because of the high density of energy stored and their smaller size which is appropriate for WSN nodes.

2.7. Batteries. The use of batteries in wireless sensor nodes is to act as sources of energy, but with their limited capacity, energy management has become an important research area. Replacing batteries when their capacity is depleted is inconvenient in most applications of WSN. Energy harvesting gives opportunity for nodes to receive energy either from the ambient environment or from intentional sources [15]. This energy can be stored in batteries and hence the batteries must have the ability to be recharged. The amount of harvested energy is usually less than needed to charge a battery and hence must be stored to accumulate for intermittent use. Earlier power harvesting used traditional electrolytic capacitors as energy storage [45, 46] but with their limited energy density, their output energy per discharge cycle is very limited.

2.8. Rechargeable Batteries. Rechargeable batteries are preferred in WSN due to their high energy density [47, 48]. A 40 mA nickel metal hydride (NiMH) rechargeable battery could be charged from zero state to maximum state within an hour under the vibration from a typical vibrating machine. Rechargeable batteries include nickel cadmium (NiCad), NiMH, and lithium ion/polymer (lithium) rechargeable batteries. NiCad has memory effect not suitable for shallow charging, but energy harvesting is usually slow charging. NiMH and lithium rechargeable batteries were discussed in

[49]. The energy densities of the NiMH are typically 60-80 Wh/kg while that of the lithium rechargeable batteries could be as high as 120-140 Wh/kg. NiMH batteries are rated for 300-500 cycles while lithium batteries are rated for 500-1000 cycles, but their lifetime decreases with frequent charge/discharge cycles. Even with the higher cycle efficiency of the lithium rechargeable batteries, they are limited with shorter lifetime. The electrolyte decays causing an increase in the internal resistance. This causes the stored energy being unable to be discharged.

2.9. Capacitors. Supercapacitors are alternatives to using rechargeable batteries. Traditional electrolytic capacitors due to their low energy density are usually not encouraged in WSN. Current research proposes the use of supercapacitors. Supercapacitors are 10-100 times higher in energy density than traditional electrolytic capacitors. Supercapacitors usually have an energy density of 1-10 Wh/kg high enough for applications in WSN. They are mostly preferred in energy harvesting systems due to the higher cycles, usually rated as high as 100,000 cycles [49]. Table 1 presents some energy harvesting nodes using rechargeable batteries and supercapacitors.

2.10. Charge and Discharge Efficiency. Rechargeable and supercapacitors may not have efficiency 100%. The Coulombic efficiency or charge/discharge efficiency is defined as

$$n = \frac{\text{Discharge}_{\text{energy}}}{\text{charge}_{\text{energy}}} \quad (3)$$

The types of rechargeable batteries mostly used in WSN [47, 50] are the cadmium, NiCd, nickel metal hydride (NiMH), lithium based (Li+), and sealed lead acid (SLA). SLA is not often used because even with low energy densities it also

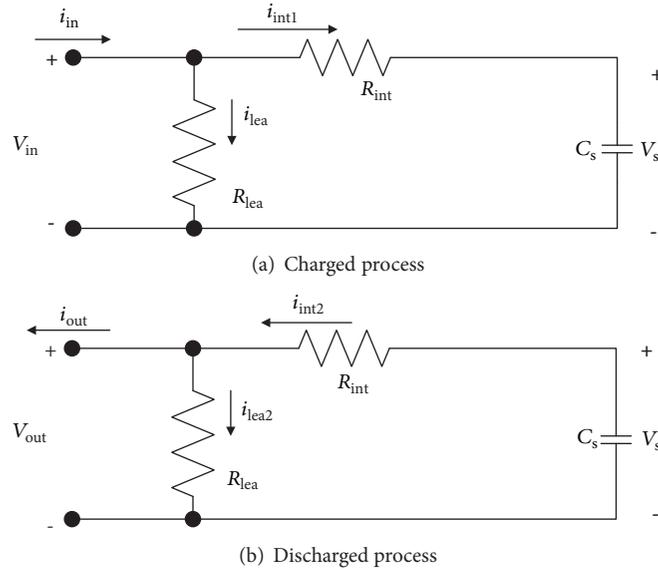


FIGURE 6: Charge and discharge process in rechargeable batteries.

has shallow discharge cycles which causes temporary capacity losses, also known as memory effect. The NiMH and Li+ batteries are mostly used, but Li+ is more efficient. They have lower rate of discharge and longer lifetime cycle. But they are also more expensive and have complicated charging circuits, degrading faster when subjected to deep discharge cycles. NiMH batteries also degrade to 80% of their charging capacities after repeating 100% discharges, degrading to 500 cycles. Supercapacitors store charges but self-discharge at a much faster rate than batteries to as much as 5.9% on a day. They have lower energy-to-density ratios of about 5 Wh/kg with high charge and discharge efficiency of 97-98%. The charge and discharge efficiencies of frequently used supercapacitors, Li+ and NiMH capacitors, are 95,92 and 65%, respectively [49]. In Figure 6 there is a diagram that shows the process of charge and discharge in a typical rechargeable battery. The NiMH rechargeable battery has the least rate of decrease and has the largest leakage loss in comparison to Li+ and supercapacitors.

2.11. Energy Neutral Operation (ENO). In energy harvesting wireless sensor network (EH-WSN), energy neutral operation (ENO) aims at achieving the desired network performance that can be supported by the energy harvested from the required energy sources (i.e., solar, vibration, and RF) and the network-wide operations (i.e., routing, clustering, and duty cycling) over longer periods of time [32]. The implementation of WSN in environmental monitoring applications requiring uninterrupted operations has become common in recent years. In such systems, power is constantly supplied to the sensor nodes for efficient performance. The implementation of ENO generally improves the network lifetime indefinitely [51, 52]. Several researchers have proposed different schemes to achieve energy neutrality in the network [51–53].

In [51], energy neutral operation was achieved by harvesting solar energy to improve the systems performance. The

authors focused mainly on the solar available periods and designed a dynamic power management scheme that allows the system to be operational for longer periods of time. To prove the performance of their approach, trace-driven simulations were performed based on real-world data collected over 11-year period. This scheme for power management over a long-term period achieved better performance compared to other similar schemes. In [53], a real-time demonstration of vibration energy harvester was adapted to improve the network through energy neutrality. The authors measured different parameters such as data transmission and reception since these components consume a great deal of energy when the network is operational. Network-wide operations such as routing, clustering, and the node duty cycling allow wireless sensor nodes to maintain their energy levels through energy neutral operation [52, 54]. The uncertainty in the amount of harvested energy results in a more challenging protocol design and energy prediction models for such networks. In most EH-WSN, designers and developers seek to maximize the overall network performance while meeting ENO. In order to overcome these challenges, it is relevant to compensate the energy harvesting systems by providing energy chargers and transfer power from these charges to power the sensor nodes.

3. Wireless Energy Transfer

Wireless energy transfer also known as wireless power transfer is the ability to transfer electrical energy from a source storage to some destination storage without any plugs or wires [55]. In 1900, Nikola Tesla experimented the wireless transfer of power from device to another without contact with large electric fields. These large electric fields diminished the energy transfer efficiency and coupled with the size of large antennas required to make these transfers feasible [8] Tesla's invention was abandoned. Due to the pervasive

use of portable devices, wireless power transfer or wireless charging (these terms being used interchangeably in this paper) has reemerged with much acceptance, already having commercial use in applications, for example, the electric toothbrush and mobile phone wireless charging like in Apple iPhone, Samsung Qi, etc. Wireless power transfer has been achieved in applications such as RFID and medical implants using nearfield coupling. In 2007, Witricity was reintroduced by [56] who reported of powering a 60 W bulb from 2 meters with 40% efficiency using strongly coupled magnetic resonance [57]. Application areas include the electric vehicle charging applications [58] medical sensors, implantable devices and consumer electronics, and power transfer in concrete [59].

In [36], the transfer of RF energy (between 850 and 950 MHz, central frequency of 915 MHz) broadcasts radio waves in the 915 MHz ISM band and a receiver tuning into the same frequency harvest RF power. The work [14] reports that 45 mW of energy is harvested within 10 cm of the RF transmitter with 1% efficiency. Earlier research in wireless power transfer considered transfer over single hops until [22] demonstrated the possibility of transferring energy over multihops. Their method gives room for possibility of neighboring nodes to charge energy deprived nodes that may be outside the charging range of energy charging devices like energy transmitters [22, 60]. Initial power transfer approaches had limitations in usage due to requirements such as close contact, continuous line of sight, and accurate alignment in charging direction. The work [56] experimentally demonstrated the transfer of power from magnetic inductive coil to another magnetic inductive coil that are in resonance. Resonance is achieved by the interplay between distributed inductance from a transmitting coil and the distributed capacitance from the receiving coil. Strongly coupled magnetic resonance between the coils enables the transfer of power between the coils, and this is not affected by obstructive interfaces; it is nearly omnidirectional and not limited by line of sight. The work [56] suggested that the receiving coil could be smaller for portable devices without decreasing the efficiency of transfer.

3.1. Wireless Energy Transfer Technologies. The broad categorization of wireless power transfer technologies is inductive coupling, electromagnetic radiation, and magnetic resonant coupling.

3.1.1. Inductive Coupling. Inductive coupling is the near field wireless transmission of electrical energy from a primary coil to a secondary coil. It is generated when an alternating current in a primary coil from a source generates a varying magnetic field that induces a terminal voltage of a secondary coil at a receiver. In inductive coupling, the size of coil is directly proportional to the amount of energy generated. Its charging efficiency is reduced over short distances. Its simplicity and ease of use have led to several commercial applications including electric toothbrushes, charging pads for mobile phones or laptop and medical implants and RFID tags.

The addition of a parallel capacitor to the secondary coil to form a resonant circuit at the operating frequency increases the voltage received. The Wireless Power Consortium in 2010 approved the first wireless charging standard (Qi) for low-power inductive charging. In an application where robot swarms were powered, resonance was applied on the receiving coil but not on the transmitting coil, to minimize performance variations from the interactions of the robots [61]. It is from inductive coupling that other wireless power transfer methods like resonant magnetic coupling were derived where resonance is applied to both the transmitting and receiving coils, and power transfer is done with little radiated losses. Inductive coupling operates at a frequency of 13.56 and 135 MHz. The transmission range is less than 1 meter and works best when the transmitter and receiver are in close contact (0 cm giving the highest power transmission) and have accurate alignment in the charging direction. These limitations make inductive coupling not desirable in WSN.

3.1.2. Magnetic Resonant Coupling. First presented by [56], magnetic resonant coupling works on the principle of magnetic resonant coils where coils on the same resonance frequency are strongly coupled through nonradiative magnetic resonance. Energy is transferred from a source coil to a receiver coil on the same resonance frequency with little losses to external off-resonance objects. The coils could be made small enough to fit into portable devices such as sensor nodes without decreasing efficiency. Experimental results from charging a 60-W light bulb at 2 m in [56] reported 40% power transfer efficiency. Challenges in magnetic resonance coupling include orientation and interference, with the maximum charging distance of 2 m achieved only when the transmitting and receiving coils are aligned coaxially. A 45-degree rotation of the coaxial alignment reduces the coupling factor and when charging multiple devices mutual coupling between the receiving coil and other objects may cause interference. Cannon [62] demonstrated power transfer from a single resonant transmitting coil to multiple resonant receivers provided they meet these two conditions: (1) coils on the receiver must remain in the uniform magnetic field generated by the transmitting coil; (2) mutual inductances between the receiving coils must have negligible effect on the resonant interaction. This means receiving coils must be far enough from each other that their interactions with the transmitting coil are decoupled. Designing a network of mobile nodes that use magnetic resonant coupling is therefore a challenge due to the second limitation. Given the limited distances allowed in resonance coupling and the coupling interference of nodes, new research challenges are opened in power transfer in WSN.

3.1.3. Electromagnetic Radiation. Electromagnetic radiation or EM radiation emits energy from the transmitting antenna of a source to the receiving antenna through EM waves. The electromagnetic spectrum can contain regions of ambient energy levels of low and high regions and the efficiency of conversion depends on the part of the spectrum. Classifications of EM radiations are omnidirectional and

TABLE 2: Wireless power transfer technologies [8].

Wireless Power Transfer Technologies	Strengths	Weakness	Examples
Inductive Coupling	Simple, high power transfer efficiency in centimeters	Short Charging distance, requiring accurate alignment in charging direction.	Electric toothbrush, charging pad for cell phones and laptops.
Electromagnetic	Omnidirectional	Rapid drop of power transfer efficiency over distances, ultra-low-power reception	Charging WSN for environmental monitoring (temperature, moisture, light)
	Unidirectional	Effective power transmission over long distances	Requiring LOS and complicated tracking mechanisms.
Magnetic Resonant Coupling	High efficiency over several meters under omnidirection. Not requiring LOS and insensitive to weather conditions	High Efficiency only within several meters	Charging mobile devices, electric vehicles, implantable devices and WSNs.

unidirectional. Omnidirectional radiation transmits broadcast EM waves in an assigned ISM band and a receiver in the same frequency harvests the radio power. Unidirectional radiation on the other hand transmits from one source to a receiving antenna in an assigned band. Omnidirectional EM waves dissipate over distances and in a paper by [8] the power transfer efficiency was 1.5% with a receiver at 30 cm. EM radiations with omnidirectional antennas can be used in low-power sensor nodes with low sensing activities to prevent hazards to humans. To achieve high power transmission in unidirectional antennas, microwave beams transmitted on microwave frequency of 2.45 and 5.8 GHz is used. Laser-beamed systems can be used for unidirectional power transfer under the visible or near infrared frequency spectrum. Unidirectional radiation is not suitable for wireless sensor networks because they require line of sight and has complicated tracking mechanisms. Omnidirectional radiations are used in applications where either the location of nodes is unknown a priori or nodes are mobile. Powercast technology is a commercially developed device that uses the EM waves to transfer radio frequency (RF) power from a source to receiver(s).

A summary of the wireless energy transfer technologies is presented in Table 2.

3.2. Energy Transfer Tools and Techniques. Energy transfer may be achieved using stationed energy sources that transfer energy to nodes in the network [16] or by the use of mobile chargers [8, 16, 63]. Mobile chargers have been widely used in the literature. Tools for energy transfer include mobile chargers, charging vehicles or robots, energy transmitters, and the sensor nodes themselves.

3.2.1. Mobile Chargers. The use of mobile chargers has been used by several authors when the energy of the nodes runs low [8, 69, 70] using either human manned chargers or robots. Mobile charging vehicles have also been used [8, 55, 70]. A mobile charging vehicle carries a battery

charging station that is assumed to have enough power to charge several nodes. The mobile charging vehicles may have power harvesters attached that scavenge energy from the environment and hence ensure continuous power supply to the battery station [9, 71, 72]. They could also have attached rechargeable batteries such that, after a cycle of charging nodes in the network, the mobile charging vehicles return to a stationed power source to be replenished with its energy [55, 69]. The works [19, 20] studied the optimization of the vacation time of the wireless charging vehicles (WCV) over the cycle time. The use of multiple mobile chargers has also been proposed in [69] where a scheduling algorithm is used to charge sensor nodes in a large network. Single mobile chargers may not carry enough energy to recharge energy node in the network if it is not equipped with energy harvesting.

The deployment of multiple wireless chargers has been studied [63, 69, 70] and different methods are proposed. Of such methods is triangular deployment [8] of deploying multiple mobile chargers such that a charger is placed at each vertex of the triangle. The side length of the triangle yields the minimum number of nodes for covering a plane. The method accounts for the fading of recharge signals in space unlike in binary disks. The infrastructure used even though fixed could have mobile nodes, and it is argued that having mobile nodes as opposed to stationary nodes could have the advantage of having fewer chargers. The work [73] proposed a hierarchical charging structure of multiple chargers in a network. Two groups of mobile chargers are formed: the lower mobile chargers that charge the sensor nodes in the network and the higher Special Chargers to charge the mobile chargers.

Some approaches of wireless charging include the use of specialized nodes called energy transmitters [74] to harvest energy from the environment and then transfer to ordinary nodes in the network. Electromagnetic induction, inductive coupling, and RF energy transfer techniques have been discussed and proposed in literature [14]. With all these

approaches being discussed, there is still a need to use energy as a scarce resource in the network to prolong the lifetime. Near field coupling is applied in RFID tags and medical implant. The efficiency of near field coupling techniques decays with distance at a rate of $1/r^6$ [75].

3.2.2. Energy Transmitters. Energy transmitters used in research assume that the nodes have some attached antennas for receiving and transmitting energy from some device to sensor node. Powercast technology [36] that introduced the TX91501 Powercast Transmitter uses the 915-MHz ISM band to transmit radio waves for power and data. The TX91501 transmitter can broadcast power and data over 12 meters to Powercast receivers. The Powercast receivers convert RF energy into DC for devices that are batteryless or wirelessly trickle charge batteries. In [16] they realized the high broadcast range and power of the energy transmitters of as much as 3 Watts introduces interference in data communication. A model for concurrent energy and data created three regions in a network: (i) the wireless charging region within which data can be transferred directly from a transmitter to neighboring nodes; (ii) communication region where nodes can communicate with the ET but cannot be charged; and (iii) interference range where nodes receive interference from the ETs energy transmission and therefore data communication is affected during energy transfer of the ETs.

3.2.3. Monitoring Techniques. Monitoring approaches provide means to check the energy consumption levels in sensor nodes and to report the amount of energy remaining in the node. This method of monitoring energy levels gives the energy source storage nodes the ability to know which nodes to charge and the schedule for charging. Some methods and techniques employed in research for monitoring energy levels are as discussed below. In [72], three charging schemes are proposed. In two of the proposed schemes, nodes monitor their energy levels and inform mobile chargers when their energy levels go below predefined threshold. A mobile charger patrols the shortest path computed and charges nodes within the threshold. In the third scheme, a function creates a charging scheme based on the nodes residual energy and its distance to a charger for chargers that charge more than 1 sensor node.

In [71], a three-tier architecture of stationary sensors, a mobile charger, and an energy station is proposed. The energy station computes a charging sequence for the network when sensor nodes periodically send information about their energy level state to the energy station. Qi-Ferry, a mobile charger used in [76], describes the tour of the QiF as it goes through the network to charge nodes below some distance threshold. The aim of QiF is to maximize the number of sensors charges in a tour while reducing the amount of energy spent by the charger in a tour. Qi-Ferry monitors its energy level and iteratively removes tour stop from the tour until all nodes are sufficiently charged.

In [55] an optimal path is formulated for a mobile charger to periodically charge all nodes in the network, and this optimal path is the shortest Hamiltonian cycle. The method is to maximize the ratio of the time the mobile

station spends at home recharging itself (also known as the vacation time) to the time spent in a cycle while charging nodes. The assumption is that the mobile charger is charged enough that it may not be depleted in a cycle. But this may not always be so, especially in instances where the rate of consumption of some nodes due the location-specific channel behavior of some nodes that may change. In [32], distributed energy harvesting routing is proposed where battery-less nodes using energy directly harvested from the ambient environment can predict the amount of harvested energy and based on the rate of consumption of the node, the real-time residual energy is used to create a distributed protocol. This protocol is limited and may not be practical for most energy harvesting sources whose continuous availability in the environment may not be predicted and amount harvested may not be modeled.

3.2.4. Charging Techniques. Techniques for charging sensor nodes are dependent on the source of energy being harvested and transferred and the antennas used. RF energy charging is different from the conventional constant voltage charging. This is because the RF power received for recharging a capacitor is constant for an RF source that transmits constant power to a fixed distance [65]. Some of the charging techniques for RF energy are discussed in [65] as multiple antenna transmissions, distributed beamforming of multiple antennas, and cooperative relay and protocol based optimizations. This technique is described using the illustration in Figure 7.

In [22] three charging techniques are proposed based on Witricity and investigate the possibility of transferring energy efficiently over multiple hops using long-time resonant electromagnetic states with localized slowly evanescent field patterns. The techniques are as follows:

- (i) The store and forward technique assumes that nodes are equipped with rechargeable batteries. The main power source which is assumed to be stationary charges neighboring nodes till their batteries are fully charged or the source reaches an energy threshold of 50%. The energy may then be transferred to nodes in the neighbors next hop and stored in batteries and then transferred to the next hop. It is assumed that nodes are attached with antennas for both transmitting and receiving energy and size and signal interference is not a problem.
- (ii) The direct flow technique works on the principle that a single node can couple with multiple nodes simultaneously. Nodes couple with the previous and the next nodes in their path from the source to the destination nodes, receiving energy without storing in batteries, directly transmitting to the next node. Charge and discharge losses associated with energy transfer are not incurred except at the last node. Coupling coefficient of a pair of nodes does not affect that of the next pair of nodes since the coupling coefficient depends on the radius of the coils and not the distance between nodes.
- (iii) The hybrid technique uses a combination of the store and forward technique and the virtual circuit

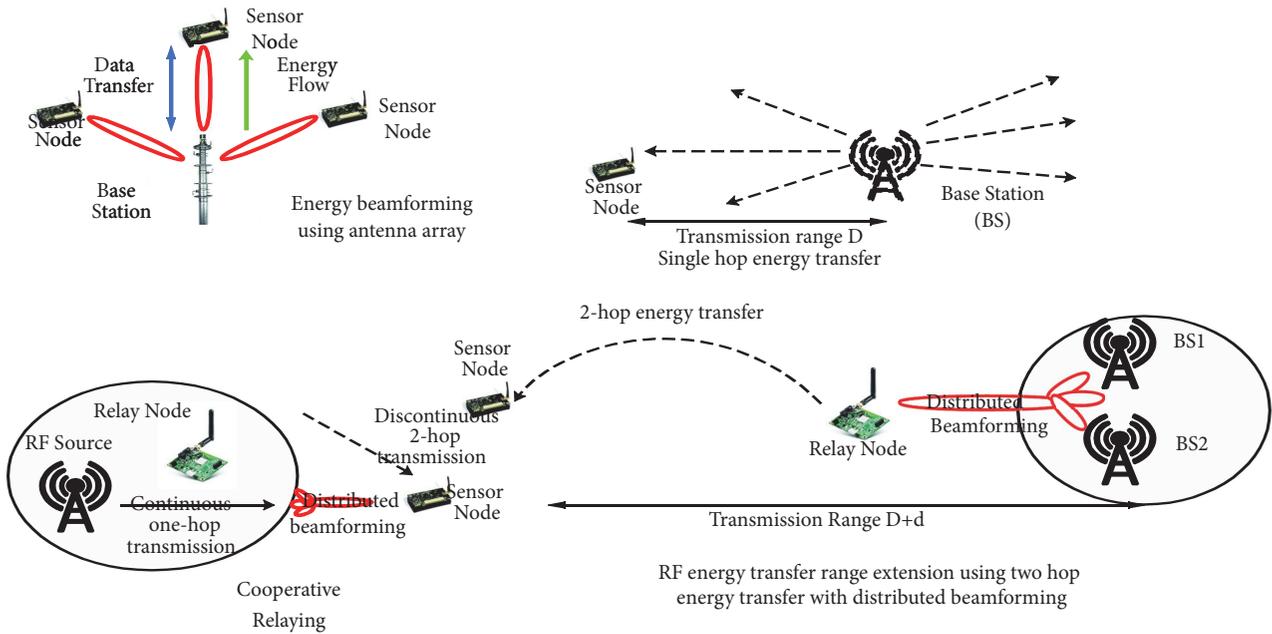


FIGURE 7: Beamforming techniques for RF energy harvesting [65].

technique. It uses a two-step approach. To transfer energy to the N th node, where M th node comes before the N th node in terms of distance from the energy transmitter, the direct flow is used to transfer m -hops to the M th node where it is stored; then through another direct flow technique the energy stored in the M th node is transferred to the N th node. The advantage is to be able to transfer to several hops while overcoming the challenges of the charge and discharge losses. The limitation of the technique that could transfer energy to as much as 20 hops is that real life test of the technique has not been made to ascertain the feasibility. Real nodes having both transmitter and receiver antennas are yet to be produced and tested.

3.3. Joint Wireless Information and Energy Transfer. RF signals carry both energy and information and with the widespread use and application in sensor networks and IoT, research into JWJET has attracted significant attention. Optimal transmission strategies and performance limitation of JWJET have been studied under perfect full channel state information at the transmitter (CSIT) [77], considering the downlink of a cellular system with single base station and multiple stations, cooperative relay system in [78], and broadcasting system in [79]. Transmission relies on the CSIT but acquisition of the full CSIT incurred large overheads [80]. Partial CSIT has been considered in [81, 82] with robust beamforming schemes. In [81], MISO downlink broadcasting channel with three nodes and a single MISO uplink channel in [82] are considered. In [83], a two-user MIMO interference channel in which a receiver either decodes incoming messages or harvests RF energy to operate forever is studied. A transmission strategy that achieves maximum energy beamforming and minimum leakage beamforming for a

rate-energy tradeoff region is achieved. A new transmission strategy that satisfies the condition of signal-to-leakage-and-energy-ratio maximum beamforming is also proposed. A general k -user MIMO interference channel is explored in [84] where three scenarios are investigated: (i) multiple energy harvesting receiver and a single ID receiver; (ii) multiple IDs and single energy harvesting receiver; (iii) multiple IDs and multiple energy harvesting receivers, where IDs are devices for information decoding and energy harvesters are for receiving RF energy from the ambient environment. An energy beamforming scheme requires partial CSIT and reduces the feedback overhead in a two-user MIMO and k -user MIMO IFC using Geodesic information/energy beamforming strategies. A summary of beamforming schemes is presented in Figure 5. The work [65] discusses the RF harvesting efficiency prevalent in low RF-to-DC conversion efficiency and receiver sensitivity, with new communications techniques enhancing the usability of RF energy harvesting.

3.4. Simulation and Emulation Tools

3.4.1. Castalia Simulator. Castalia [85] is an open source simulator built on top of OMNeT++, and it is optimized for testing distributed algorithms and protocols and features an accurate channel/radio model, radio behavior, and other aspects of communication. It has parameters for clock drift, sensor bias and node energy consumption, CPU energy consumption, memory usage, CPU time, and some implementation for MAC and routing protocols. It provides a first-order analysis of algorithms and protocols before their implementation of node platforms.

3.4.2. EKHO. EKHO [66] is an emulation tool that records and emulates energy harvesting conditions from diverse sources in the ambient environment such as solar, thermal,

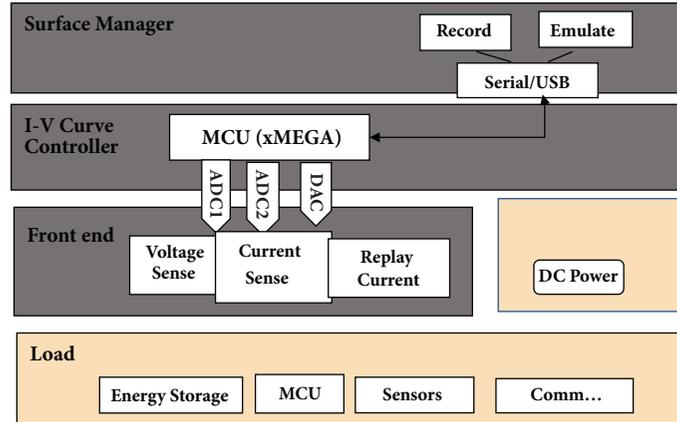


FIGURE 8: EKHO, a simulator for energy harvesting [66].

RF, and kinetic. The energy harvesting environmental conditions are recorded and stored in a digital format which could be replayed through analog front-ends serving as energy source. EKHO is presented in Figure 8.

3.4.3. OMNeT++. OMNeT++ [86] is a discrete event simulator based on C++ for modeling communications networks, multiprocessors, and distributed and parallel systems. It was developed as an open source tool that could be used for educational, academic, and research oriented applications, to bridge the gap between open source research oriented simulators like the NS-2 and expensive commercial simulators like the OPNET. It is available on Linux, MAC OS/X, and Windows. OMNeT provides basic machinery for users to write simulations and consists of modules that communicate by message passing. It has two major simulation model frameworks: the mobility framework and the INET framework. OMNeT++, unlike NS-2 and NS-3 which are network simulators, is a simulation platform upon which researchers could build their own simulation frameworks. It does not have a framework for energy modeling.

3.4.4. PASES. Power Aware Simulator for Embedded Systems (PASES) [67] is a SystemC based framework that is a combination of an event-driven simulation engine and a hardware, applications, and network models composer. It is a simulation and design space exploration framework that is used for power consumption analysis of WSNs application, communication, and platform layers. It gives performance and energy analysis of WSN hardware platforms and provides a gap between pure network oriented WSN tools and tools for architecture specific simulation environment. It supports Platform Based Design methodology and provides power analysis for different platforms by defining abstraction layers for the application, communication, hardware, power supply, and sensing modules of the network and node. PASES provides these abstractions: the software layer provides users with the application layer (AL) to define application functionality.

The communication layer (CL) could be tweaked to meet network requirements such as throughput, latency, and

energy efficiency. The Architecture Resource Layer (ARL) provides the service layer (SL) that provides APIs for composing different applications. The Resource Behavioral Layer defines hardware resources for target architectures and the Resource Annotation Layer specifies energy-performance details. At the energy level, the Energy Source Layer collects energy sources for sensor nodes. The Energy Source Layer provides energy source models such as batteries and super-capacitors and may include models for energy harvesting. But the energy harvesting models are not complete. PASES is presented in Figure 9.

3.4.5. COOJA. The Contiki OS Java or COOJA simulator [87] is a Contiki sensor node operating system and usually integrated with MSPSim to form the COOJA/MSPSim. COOJA allows simultaneous cross-level simulation at the application (network level), operating system, and machine code level. COOJA combines the elevated level behavior of a node to the low-level sensor node hardware in a single simulation. COOJA supports adding and using different radio mediums. It allows for the flexible additions and replacements of its parts including the radio medium, the hardware node, and plug-ins for input/output. With all the cross-level support provided in COOJA, it does not have an energy model and the energy parameters of nodes may not be properly analyzed during simulations. COOJA has a Visualizer.java class that could be extended to provide GUI and has support for memory and radio model simulations but has relatively low efficiency with increasing number of nodes hence not scalable.

3.4.6. Network Simulator 2 (NS-2). NS-2 or Network Simulator 2 is a discrete event simulator based on the Object Oriented Programming (OOP) and consists of two languages: C++ and Object Oriented Tool Command Language (OTcL) bound together by TcLCL. Codes written in OTcL will be visualized using NAM and XGRAPH with optional python bindings. NS2 has support for protocols such as 802.11, 802.16 and 802.15.4 but is limited with support for sensing. Parameters such as energy model, packet formats, and MAC protocols are different from those used in real

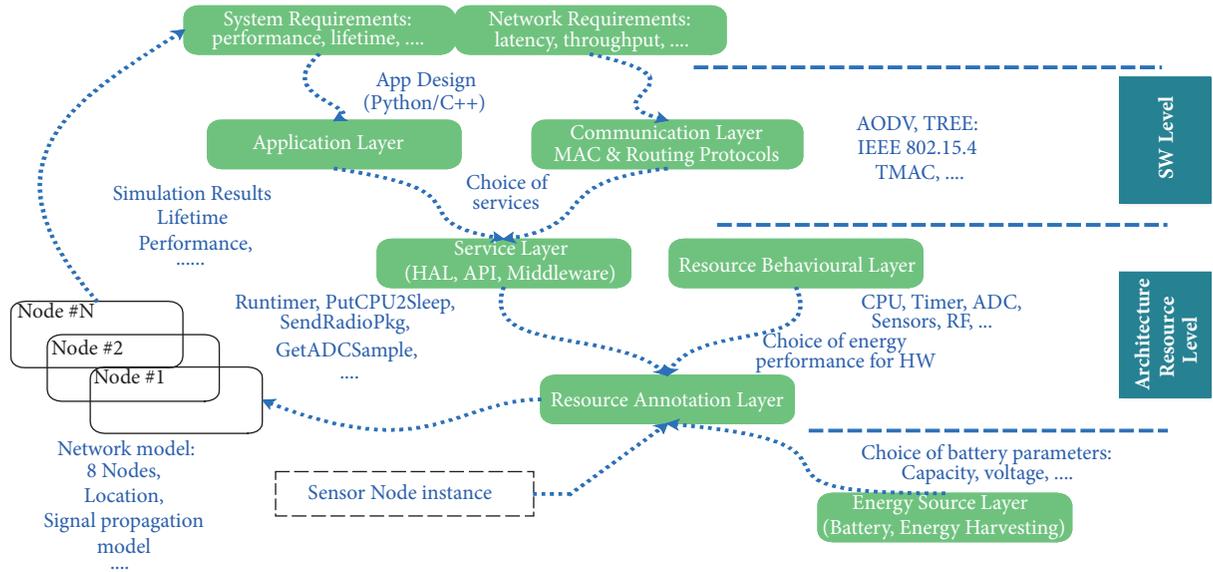


FIGURE 9: Design space exploration methodology of PASES [67].

sensor network nodes. NS-2 has parameters for residual energy but does not give models for keeping track of energy consumed by the different components and does not have a model for energy harvesting.

3.4.7. Network Simulator 3 (NS-3). NS-3 [88] is not considered an extension of NS-2 but is an entirely new simulator written in C++ with optional python bindings. Energy model in NS-3 consists of two major components: energy source and the device energy model. The energy source is an abstract base class that provides interface for updating/recording total energy consumption on a node, keeping track of remaining energy, decreasing energy, and when the energy is completely depleted. The device energy model monitors the state of the device to calculate its energy consumption. It provides an interface for updating the residual energy in the energy source and gives notification from the energy source when energy is completely depleted and maintains a record of the total energy consumed by the device. NS-3 provides energy model for Wifi Radio with states *IDLE*, *CCA BUSY*, *TX*, *RX*, and *SWITCHING*. Developers may extend on the models in NS-3 to model different scenarios that may not be present in current releases.

NS-3 allows for the definition of new energy sources that incorporate the contributions of an energy harvester, with the addition of an energy harvester component with existing energy source as well as the possibility of evaluating the interaction between energy sources and the different energy harvesting models. The work [68] provided an extension of the current energy models in NS-3 introducing the concept of energy harvesting. Two energy harvesting models are as follows: the basic energy harvester, providing time-varying, uniformly distributed amount of energy and the energy harvester that recharges the energy source. A model is for a supercapacitor energy source and a device energy model is for energy consumed by a sensor node. A model for an energy predictor was introduced that is supposed to predict

the amount of energy that will be available in the future based on information from the basic energy source and energy harvester. An extended diagram of NS-3 with modules for energy harvesting is presented in Figure 10.

All the above tools provide some support for energy modeling and even some simulators like the PAWiS, WSNNet, OPNET, and Qualnet not mentioned above provide energy modeling but not completely. Support for energy modeling is still an open research especially when multihop energy transfer is considered in WSN.

4. Energy Conservation

Energy conservation methods are concerned with reducing energy consumption of the nodes. To conserve energy, the major components in a sensor node that consume energy must be controlled. The lifetime of a sensor node, which the lifetime of the network is dependent on, is an indication of how much energy is consumed and the amount of energy available for use [11].

Definition of a network frequently used in literature is of n-of-n such that

$$T_n^m = \min T_v \tag{4}$$

where T_v is the lifetime of node v [11]. The lifetime of the node, which is a function of energy consumed and energy available for use in the network, depends on the activities of various components, being the sensing, processing, radio, and power supply units, with typical energy consumption of the various units of the node presented in Figure 11.

The sensing component consists of the sensors with Analog-to-Digital (ADC) converters for collecting data from the environment that are then fed into the processing unit. The processing component manages the node by performing internal computations and aggregation of data with other nodes in the network and has a storage unit/memory included working as a temporary buffer. The transceiver

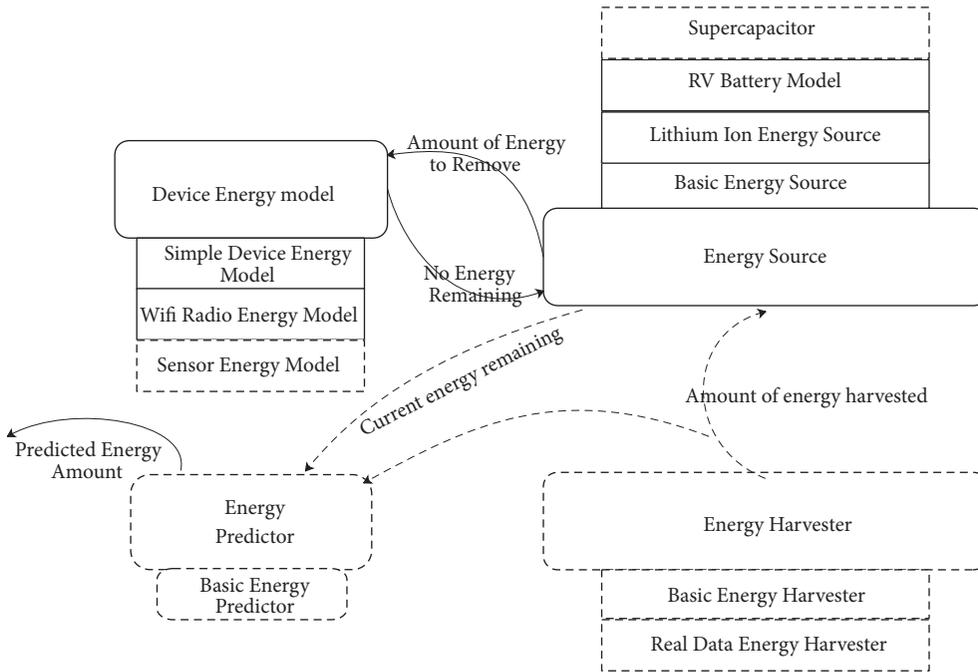


FIGURE 10: NS3 extension with energy harvesting model and basic energy model [68].

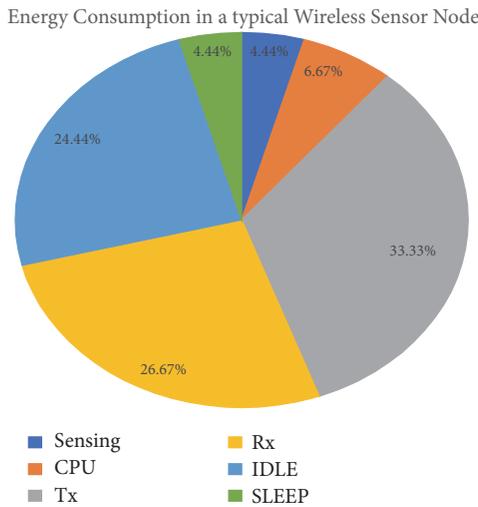


FIGURE 11: Energy consumption of the components in a sensor node.

also known as the radio unit connects the node to the network. The power unit consists of the battery or low-powered capacitors and serves as the source of energy. They may also be supported with power scavenging units for energy harvesting. The control of the energy consumption of the various components of a sensor node has led to the generation of some methods in energy conservation which may be classified as radio optimization techniques, data reduction, and efficient routing techniques. In general, energy conservation methods focus on networking and sensing. Networking comprises the management of the sensor nodes and the design of the network protocols while sensing is based on the techniques to reduce the frequency of sensing.

Energy conservation methods provide means of reducing energy consumption by the different components of the sensor node as shown in Figure 1. Of the different components, data communication expends the maximum energy available [13], where communication involves both transmission and reception.

The sensing component consists of the sensors with Analog-to-Digital (ADC) converters for collecting data from the environment that are then fed into the processing unit. The processing component manages the node by performing internal computations and aggregation of data with other nodes in the network and has a storage unit/memory included working as a temporary buffer. The transceiver, also known as the radio unit, connects the node to the network. The power unit consists of the battery or low-powered capacitors and serves as the source of energy. They may also be supported with power scavenging units for energy harvesting. The control of the energy consumption of the various components of a sensor node has led to the generation of some methods in energy conservation which may be classified as radio optimization techniques, data reduction, and efficient routing techniques. In general, energy conservation methods focus on networking and sensing. Networking comprises the management of the sensor nodes and the design of the network protocols while sensing is based on the techniques to reduce the frequency of sensing.

4.1. Radio Optimization Technique. The radio module is responsible for wireless communication and is the component that consumes significant amounts of energy in the network. To optimize the radio, techniques used include cooperative communication schemes, sleep/wake-up schemes, duty cycling, and radio optimization parameters such as radio

coding and modulation techniques, power transmission, and antenna direction. The radio transceiver is one component that consumes much energy since it is used in data communication. Energy conservation methods focus more on data transmission since more energy is expended from the node during data transmission than data processing/computation. Energy consumed during sensing may be considered in energy conservation since the energy consumed may be comparable to or even greater than communication [27, 28, 89] in some applications and cannot be ignored. Radio optimization techniques provide means of mitigating the energy consumption of sensor nodes due to wireless communication. Radio optimization techniques considered include SISO, MIMO, cooperative communication schemes, sleep wake-up schemes, and Transmission Power Control.

4.2. Single-Input Single-Output (SISO). Single-Input Single-Output (SISO) refers to the direct transmission of data from single nodes to a base station usually through a single hop transmission. Challenges of SISO include data congestion, collisions, and loss of energy when the distance between node and base station is big.

4.3. Multiple Input Multiple Output (MIMO). Multiple Input Multiple Output (MIMO) systems assume that multiple antennas on nodes transmit data to multiple receivers. The application of MIMO spreads the power to transmit among different antennas on nodes in the network to achieve power gains. This increases the bandwidth for high data rates and bit-error-rate performance requirements [74, 90]. MIMO has challenges in WSN due to the limited physical size of typical sensor nodes that cannot support multiple antennas and the energy consumed by the circuit energy of the transmitter and receiver in the system. When the number of antennas increase, the circuit energy consumed by multiple antennas increases [74]. To mitigate the limitations of MIMO, cooperative MIMO technologies are constructed that minimize the energy consumed in transmission especially in long range transmission where the benefits of MIMO outweigh SISO [90, 91]. SISO has efficient energy consumption for short range transmissions but still requires approaches to minimize circuit energy consumption. The reader could read papers [74, 92] for further benefits of MIMO.

4.4. Cooperative Communication Schemes. Cooperative communication schemes provide means of communication by allowing the terminals/antennas in a multiuser environment to collaborate in communicating in a sensor network, using the broadcast nature of wireless networks. Single antennas in a multiple user environment collaborate to share their antennas to form a virtual multiple antenna transmission, thereby gaining the benefits of MIMO systems while overcoming their challenges, such as size, cost, hardware, and deployment limitations [76, 93]. Wireless nodes in cooperative communication systems act as transmitters and also cooperative agents for other users. Two ways of cooperation are introduced in [76] called the relay cooperation and node cooperation. Relay cooperation is when extra relay nodes help to transmit data from source to destination. This is

achieved by relay nodes receiving data from sources and then transmitting to destinations using some cooperation protocols (amplify-and-forward, decode-and-forward, and compress-and-forward). In node cooperative systems, nodes cooperate by either communication terminal using combined processing or coordinating the strategies for communication at the terminals. Challenges in the implementation of cooperative communication systems include, but are not limited to, cooperation assignment and hand-off, network interference, transmitting and receiving requirements of wireless systems, and the loss of rate to the cooperating mobile system [93].

4.5. Sleep/Wake-Up Schemes. Sleep/wake-up schemes adapt the node to the activities of the network to conserve energy by putting the radio to sleep, to minimize idle listening (idle sensing of the channel). Duty cycle is defined as the ratio of time nodes that are active during their lifetime and the sum of the times when the node is on and asleep [94], which means nodes alternate between sleep and wake-up times. The radio transceiver of nodes is made to sleep when there is no communication and wakes up when data transmission is required. The alternating of sleep and wake-up periods is referred to as duty cycling. The downside of this technique is that data generated during the vacation period (when node is sleeping) may be lost. Another technique that is data driven senses the channel until some data is generated; then it wakes up for transmission. Unnecessary data may be transmitted to the sink increasing energy consumption and could also be too energy consuming if the data sensing is not negligible.

To optimize the sleep period, event-driven systems adapt selective and incremental wake-up scheme, where low-powered sensors continuously monitor the system, until some event trigger is received; then nodes are triggered for high-quality detection and quality sensing. Another method puts all nodes to sleep but they wake up when there is a demand by another node to communicate. This means nodes are active only for a minimum time during communication. No sensing of the network is required and it is appropriate for applications where sensing consumes much energy and periods of data communication are known a priori. Thirdly, there are methods where all nodes sleep and wake up at the same time according to a wake-up schedule. These methods are appropriate for data gathering applications where aggregation may be required but more collisions will be introduced in such networks as all nodes wake up at the same time. Asynchronous methods allow nodes to wake-up independently with overlapping wake-up periods with the neighbors. Such networks require active periods of sensing or nodes will have to wake up frequently when sender sends long preamble or receiver remains active for longer periods. All will require huge energy cost to the network and hence this is not an efficient energy conservation technique, but good for QoS purposes [3].

4.6. Transmission Power Control. The aim of Transmission Power Control (TPC) approaches is to dynamically adjust the transmission power of the radio to maintain an effective communication link between pairs of nodes while not transmitting at full power capacity [95, 96]. Factors such

as distance and link quality affect the transmission power within a transmitter-receiver pair. A survey of TPCs by [95] investigated existing approaches of protocol development that were based on single-hop communication in WSN. An Adaptive Transmission Power Control (ATPC) [96] was proposed that builds a model for communication where neighboring nodes create a correlation between transmission power and link quality. A feedback-based TPC algorithm is employed to dynamically maintain individual link quality, creating a pairwise adjustment for ATPC that saves energy with online control and is robust to environmental changes. A TPC method for SCADA systems is used for industrial control of their remote stations and a central site [97]. Using a fuzzy based algorithm, a minimum number of transmission paths are maintained between the sink and source nodes while maintaining minimum multihops. The effect of different Transmission Power Control protocols on the lifetime of WSN is studied in [98] when power levels and strategies for transmission power assignment are discretized. The bandwidth of TPCs and the granularity of the power control of the link-level affect the energy consumed.

4.7. Data Reduction. Data reduction techniques in WSN reduce the amount of data that is transmitted to the destination, usually the sink, thereby reducing the number of transmissions. These techniques reduce the bandwidth needed to send data as it traverses the network from source to destination (usually the sink). Some techniques used are data aggregation, compression, and prediction. Others are network coding and efficient routing.

4.8. Aggregation. Aggregation techniques fuse data as it traverses the network from one node to the other to the sink. Its main aim is to aggregate data in an efficient manner to increase the network lifetime [99]. Since near nodes share similar data by spatial correlation, energy is wasted when the same data value is routed from multiple sources to the sink. Transmitting 1 KB of data over 100 m expends energy as much as executing 300 million instructions on a typical processor with 100MIPS [100]. In-network data aggregation can be broadly categorized as Address-Centric (AC) and Data Centric (DC). It reduces medium access contention and the number of transmitted packets and minimizes packet transmission delays. Aggregation in the network can be done via data aggregation tree (DAT for flat networks) and by clustering for hierarchical networks. Some key points in data aggregation are as follows:

- (i) Nodes sense data values on the entire network and route to neighbor nodes.
- (ii) Sensor nodes can receive different versions of the same message from different nodes in the network.
- (iii) Data is combined from diverse sources and routes to mitigate redundancy.
- (iv) Intermediate nodes must access the content of packets to be aggregated.
- (v) Nodes must wait for a predefined waiting time (WT).

A survey of data aggregation algorithms is presented in [101] and analyzes different solutions against performance metrics such as data latency and accuracy. In data aggregation algorithms, there is usually latency and accuracy based on the application area. Energy efficiency, aggregation freshness, and collision avoidance are some performance metrics used in data aggregation.

4.9. Data Compression. These are techniques that reduce the size of sensed data before transmission. This reduces the amount of energy consumed in processing and transmitting data in individual nodes in the network, reducing the size of bandwidth used. A basic assumption in compression is that the amount of energy consumed compressing a bit of data into b , such that $a < b$, must be smaller than the amount consumed in transmitting $a-b$ string of data [102]. The work [103] presented a survey of mechanisms for data compression. The assumption used in data compression is that multiples of energy consumed per 480 addition instructions are consumed for every bit of data transmitted by radio. If more than 1 bit of data is taken from sensed data by data compression, total power consumed by transmitting that data will be significantly reduced. Data compression techniques in WSN take into consideration the size of the compression algorithm and the processing speed (that of a typical WSN node is 128MB and 4 MHz, respectively). Examples of these compression techniques are coding by ordering, pipelined in-network compression, low-complexity video compression, and distributed compression [102].

4.10. Data Prediction. Prediction is a term given to the process of inferring missing values in a dataset based on statistical or empirical probability or the estimation of future values on some historical data. A prediction method is a function with two inputs, the set of observed values and the set of parameters. A model created for prediction is deterministic and obtained from the observed values, but one could have several prediction models from the same prediction method or algorithm [104].

Prediction methods require additional information about the observed data which may be known to the user before deployment which can be applied to the statistical data. The additional information may support assumptions made in predictions that determine the feasibility of the model. This feature of prediction models makes them more reliable compared to machine learning techniques that use fewer assumptions of the data in exchange for the time to adjust parameters and adapt to the observed data set. This does not give the user the opportunity to see the prediction accuracy before testing with real data. The work [104] groups prediction schemes under two main headings: single prediction schemes and dual prediction schemes. Single prediction schemes are made at one point in the network which could either be closer to the sensor nodes or close to the data collection point. In this, sensor nodes may sense data but based on the reliability of predictions of the sensed data predict changes to the amount of data measured and transmitted. The advantage is that each device may decide to adapt itself based on the predictions or not without a

need to synchronize with neighbors of their decisions without incurring any overhead cost of communications. This could eventually reduce the quality of the information derived from the cluster heads since most data may not be transmitted from the sensor nodes. With the autonomy of cluster heads coupled with the spatiotemporal correlation of sensor nodes placed near each other, probabilistic models could be generated with good distributions and confidence levels that may be used to predict measurements thereby reducing transmissions.

Applications of this models are used in adaptive sampling [105, 106], clustering [107], and data compression [108, 109]. Dual prediction schemes on the other hand make predictions of the cluster heads and in sensor nodes. When sensor nodes measure values outside the threshold of the prediction models, the value from the sensor nodes is transmitted to the cluster heads which then sends to the sensor nodes the correct value. Frequent transmissions and therefore energy consumption are hereby reduced. The aim of dual prediction schemes is to mitigate the number of transmissions without compromising on the quality of measurements made by the systems and hence a tradeoff between the number of transmissions for new prediction model distribution and the reliability of the channel is always made [109, 110].

5. Energy Balancing

Energy balancing techniques that distribute the amount of energy in the network such that nodes have equal amount of energy and have prolonged lifetime have been discussed in this paper. They comprise data reduction schemes that reduce the amount of data that is delivered to the sink node and balancing schemes that optimize the distribution of energy available to the node and energy consumption of the nodes in the network.

Balancing schemes proposed in recent research discussed means of distributing and managing the energy in a sensor node [23]. Clustering schemes have been used to balance energy in the network, and they were processes where nodes are grouped together with a coordinator, usually known as the cluster head, that perform specialized functions such as data fusion and aggregation, and communicate this data from its clusters to the base station. Most published clustering approaches form groups of nodes and allow these nodes to select a cluster head based on some criteria. The cluster head selection can be randomized [111] or based on degree of connectivity [1]. Some approaches include the residual energy of the node as a criterion for cluster head selection [7, 112].

The possibility of unbalanced energy consumption in the network was due to the different consumption rates of energy of nodes and their distance from the base station, causing some clusters to be of high energy while others are of low energy, in a situation known as the black hole problem [31]. To solve this problem, the unequal clustering approaches [113] have been proposed. A round-robin method causes cluster heads to be rotated among the nodes in the network of homogeneous nodes (beginning the network formation with the same energy level) and have the same capabilities. The assumption is that due to different transmission and reception rates of data of the individual nodes energy depletion

will not be the same throughout the network. Rotating cluster heads changes the topology of the network at each round of rotation and imposes change over overheads [114, 115].

All cluster heads in the network must be notified of the change while cluster heads change their routing tables and scheduling strategy. Some methods include the addition of high energy specialized nodes in the network to balance the load in various locations of the sink node to balance energy consumption. In [116], high energy nodes called gateways are proposed which form equal sizes of clusters in the network based on the cost of communication and the load on the gateways. These gateways act as cluster heads and perform energy consuming tasks like data fusion and organization of nodes for special tasks. The method solves the problem of extra overhead incurred by frequent reclustering on nodes since this task is performed by the gateways. The addition of specialized nodes comes at an extra cost and the optimized number added in a network must be considered. Other methods balance the energy consumption in the network by forming clusters of unequal sizes [7, 117, 118]. In these approaches, the size of clusters increases as one approaches the base station. The assumption is that, for nodes further away from the base station, multihopping data through relay cluster is more energy efficient than directly since the amount of energy required to transmit data from one node to another is directly proportional to the distance between the two nodes. Cluster heads aggregate data from their clusters and relay data from other cluster heads to the base station. This means cluster heads closer to the base station will be depleted of their energy faster than nodes on the peripheral.

This paper includes energy harvesting schemes augmented with energy transfer technologies to distribute energy available in the network fairly such that nodes are not depleted of their energy below some threshold when they are no more operational. Energy conservation techniques are also included to ensure efficient use of energy by the sensor node with minimal consumption.

6. Challenges and Future Research Directions

In energy transfer in wireless sensor networks, some researchers have attempted to solve the distance related energy transfer issues [8, 34, 55, 63]. Despite the attempts to resolve these issues, there remains a great deal of work in this area. In this section, we present some challenges in energy harvesting and energy transfer and propose likely future works.

6.1. Challenges

Cost of Experimentation Using Testbeds. Research in WSN requires comprehensive evaluations process that could be verified and reproduced. Most studies done over the years have evaluated theoretical analysis and simulations lacking experimental evaluations. Over the years research into the transfer of energy from node to node or from an energy transmitter to nodes by either single hop or multihop has been proposed as a solution to making sensor networks immortal. Most of such research is based on the modeling of the

networks and the charging scenarios. Creating real testbed experiments for energy transfer is still ongoing research with a little breakthrough. The work [56] performed simulation and temporary transfer of energy over magnetically coupled resonance coils of 1 m diameter or less with 60% efficiency. This shows a direct transfer without storing or retransmission of the power. Powercast Technology [36] has the Powercast energy transmitter that provides an EIRP of 3 W for 5 W DC input but does not generate continuous RF output. The work [22] proposed a solution for multihop energy transfer using theory and simulations to investigate the phenomenon of slowly evanescent field patterns that can transfer energy efficiently. Their results proved the transfer of energy over 20 hops but lack testbed tests. The work [119] demonstrated multihop RF energy transfer within two hops using the MICA2 mote operating on the supercapacitors on the Powercast P1110 Evaluation Board. The setup was made of A HAMEG RF synthesizer HM8135, an intermediate node which is made up of the P1110 Evaluation Board, and a modified MICA2 mote powered from the 50mF supercapacitor on the board and has a 6.1 dBi antenna for transmitting energy in the form of data packets to the end node (P1110 Evaluation Board). This setup is not automated and requires reconfigurations each time the nodes position or topology changes. It is also not scalable and is limited to two hops with modified sensor nodes. For efficient energy transfer, cross-layer support for energy transfer comprising the MAC, link, and application layers is critical for the implementation of wireless energy transfer in WSN. There is lack of hardware designed to support energy transfer and the lack of optimal energy-aware routing protocols that consider the concurrent transfer of energy and data in a network.

For charging nodes in an entire network, specialized devices such as mobile chargers or robots have been designed, with shortest path algorithms and optimal paths developed for easy charging, but these specific nodes increase the overall cost of energy transfer on test beds. The introduction of specialized nodes for energy harvesting and transfer called energy transmitters also increases the cost of implementation. Sensor nodes due to their usage and places of deployment are expected to be smaller in size; this becomes a challenge when antennas for energy transmission and reception must both be attached to common nodes for receiving and transmitting energy.

Designing Energy Transfer Models in Current Simulators. Simulation tools of wireless sensor networks currently lack features that support useful energy models for energy harvesting from renewable and sustainable energy sources [34]. There is a need to either develop energy models in existing simulation tools for energy harvesting and monitoring or develop energy modeling simulators for wireless sensor networks. NS-2 energy model comprises the radio energy model parameters and allows a user to set the initial energy on the mote but does not have models for energy harvesting or the ability to transfer energy from node to node. NS-3 has an energy model which has models for the device energy model, the fundamental energy source (which is usually the Li-ion battery but can allow for modifications with

other energy sources like supercapacitors) and the energy harvesting model. The energy harvesting model increases the energy stored in the energy source and could be modeled as some other energy harvesting source [68] like in solar energy harvesters in [120]. Other simulators like Castalia, EKHO, and COOJA do not have models for energy modeling. Castalia does not provide battery or energy modeling and therefore does not support lifetime estimation simulations. It also does not have postprocessing tools for GUI support [86]. PASES [67] is a design space framework that was created with power awareness for the different components of the sensor node. It has the energy level that introduces the Energy Source Layer that analyzes the power consumption of the hardware components and models for energy sources like batteries and supercapacitors and energy harvesters. It has model for the device energy model but does not give a user the flexibility to customize protocols at the MAC and network layer; files are in XML and Python and do not support other low-level languages like C and C++ which is used in most simulators and test beds. Interfacing PASES with testbed tools like what is done in COOJA is not possible and PASES does not also support increasing number of sensor nodes being added at runtime.

Prediction Models for Energy Harvesting. The unpredictability of energy due to the continuous supply of energy harvesting sources to predictive models for energy harvesting that depend on the residual of energy in the network based on the availability of the energy source is difficult due to the unpredictability of energy sources. Routing protocol design is a challenge due to the fluctuation of the available energy in the network which affects which nodes will be awake at every point in time to receive broadcast packets. This makes broadcasting not suitable for WSN with energy harvesting [5]. The use of energy-aware duty cycling algorithms becomes a challenge if the energy on the nodes is dependent on the harvested energy. This could create erratic sleep/wake-up cycles since the residual energy may not be known a priori [121].

Inductive Coupling. Inductive coupling has been revised as a technique for wireless energy transfer to handle its limitation issues such as alignment and distance that are critical to their deployment and implementation. A strongly coupled inductive resonance coupling technique is introduced by [56] and has become a viable means of energy transfer. The challenge is the health implications of inductive resonance in the human environment due to constant exposure to radiations and the discomfort when used in humans [122]. Inductive coupling works within few centimeters, and therefore the distance of operation is limited. Scalability of transmission using inductive coupling is still a challenge since nodes must be tuned to avoid interference due to mutual coupling effect [123]. Since inductive coupling requires alignment of nodes, it is a challenge to design nodes with mobility.

6.2. Future Research Directions. Current research on modeling energy transfer is showing positive directions for single hop and multihop energy transfer. There is the need to

develop or improve existing simulation tools to support energy transfer.

Single Energy Transmitters. The possibility of single energy transmitters transferring energy to multiple receivers simultaneously with multihop energy transfer with minimum charge and discharge losses is still open for research. Current research proposes using multiple transmitters to simultaneously charge nodes in large networks. But due to problem of signal interference with multiple chargers, there is a limit to the number that could be used in a network. The possibility of using a single energy transmitter that continuously receives energy through energy transfer could be a huge solution if multihop energy transfer is explored.

Extending Recent Simulation and Emulation Tools. Current simulation and emulation tools on the market (both commercial and open source) lack the full capability to test and evaluate new energy harvesting and transfer applications and protocols in WSN. For example, the current simulators and emulators used for modeling WSN applications such as Network Simulator 3 (NS-3) [88], Castalia Simulator [85], EKHO [66], OMNeT++ [86], COOJA [87], and PASES [67] do not reflect the network behavior accurately. Most of these simulators do not include models for energy transfer which makes performance evaluation difficult. Network Simulator three (NS-3) was recently extended to include two energy harvesting models [68]. The first model (i.e., the basic energy harvester) uses a generic random variable to provide energy to the harvester. The second model recharges the energy source with datasets based on real values obtained from solar panel. Development of simulation and emulation environments that support all aspects of energy harvesting and transfer is still an open research and will be a valuable tool to evaluate the proposed energy transfer techniques.

Experimental Demonstration of Multihop Energy Transfer. In WSN, the available energy harvested from the energy source may have practical implementation or simulation of multihop energy transfer that reduces the charge and discharge losses with better transfer efficiencies is still an open research. For charging nodes in an entire network, specialized devices such as mobile chargers or robots have been designed, where shortest path algorithms and optimal paths are developed for easy charging, but these specialized nodes increase the overall cost of energy transfer on test beds. The introduction of specialized nodes for energy harvesting and transfer called energy transmitters also increases the cost of implementation. Sensor nodes due to their usage and places of deployment are expected to be smaller in size; this becomes a challenge when antennas for energy transmission and reception must both be attached to ordinary nodes for receiving and transmitting energy. The need to design special nodes that could have contained both transmitter and receiver antennas on small sized nodes will enable the efficient implementation of energy transfer. There is also the need to develop optimal energy routing protocols to support multihop energy transfer that have the MAC and link layer support.

7. Conclusions

The role wireless sensor networks play in monitoring human activities in the last decade cannot be underestimated. Over the years, the introduction of energy management schemes that seek to prolong the lifetime of the sensor node and the overall network has been proposed, but the amount of energy required by the sensors to be operational all the time remains a challenge. In this paper, we have provided the trio energy management scheme that when fully implemented will keep the network alive forever. We first discussed the broad categorization of energy harvesting technologies and techniques and followed the discussion with the current energy transfer techniques and finally the approaches for conserving energy. Although there is an extensive work on each of these management schemes, there are still several other challenges that need to be addressed by the research community for effective implementation of the trio schemes.

Conflicts of Interest

The authors declare there are no conflicts of interest regarding this paper publication.

References

- [1] S. Soro and W. B. Heinzelman, "Cluster head election techniques for coverage preservation in wireless sensor networks," *Ad Hoc Networks*, vol. 7, no. 5, pp. 955–972, 2009.
- [2] C. Wang, J. Li, Y. Yang, and F. Ye, "Combining solar energy harvesting with wireless charging for hybrid wireless sensor networks," *IEEE Transactions on Mobile Computing*, vol. 17, no. 3, pp. 560–576, 2017.
- [3] G. Anastasi, M. Conti, M. Di Francesco, and A. Passarella, "Energy conservation in wireless sensor networks: a survey," *Ad Hoc Networks*, vol. 7, no. 3, pp. 537–568, 2009.
- [4] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer Networks*, vol. 38, no. 4, pp. 393–422, 2002.
- [5] W. K. G. Seah, Z. A. Eu, and H.-P. Tan, "Wireless sensor networks powered by ambient energy harvesting (WSN-HEAP)—survey and challenges," in *Proceedings of the 1st International Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology (Wireless VITAE '09)*, pp. 1–5, Aalborg, Denmark, May 2009.
- [6] L. J. Chien, M. Drieger, P. Sebastian, and L. H. Hiung, "A simple solar energy harvester for wireless sensor networks," in *Proceedings of the 6th International Conference on Intelligent and Advanced Systems (ICIAS '16)*, pp. 1–6, August 2016.
- [7] S. Soro and W. B. Heinzelman, "Prolonging the lifetime of wireless sensor networks via unequal clustering," in *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS '05)*, pp. 236–243, Washington, DC, USA, April 2005.
- [8] L. Xie, Y. Shi, Y. T. Hou, and A. Lou, "Wireless power transfer and applications to sensor networks," *IEEE Wireless Communications Magazine*, vol. 20, no. 4, pp. 140–145, 2013.
- [9] T. Rault, A. Bouabdallah, and Y. Challal, "Energy efficiency in wireless sensor networks: a top-down survey," *Computer Networks*, vol. 67, pp. 104–122, 2014.

- [10] Y. Chen and Q. Zhao, "On the lifetime of wireless sensor networks," *IEEE Communications Letters*, vol. 9, no. 11, pp. 976–978, 2005.
- [11] I. Dietrich and F. Dressler, "On the lifetime of wireless sensor networks," *ACM Transactions on Sensor Networks*, vol. 5, no. 1, article 5, 2009.
- [12] J. Ren, Y. Zhang, K. Zhang, A. Liu, J. Chen, and X. S. Shen, "Lifetime and Energy Hole Evolution Analysis in Data-Gathering Wireless Sensor Networks," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 788–800, 2016.
- [13] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Communications Magazine*, vol. 40, no. 8, pp. 102–105, 2002.
- [14] T. Soyata, L. Copeland, and W. Heinzelman, "RF energy harvesting for embedded systems: a survey of tradeoffs and methodology," *IEEE Circuits and Systems Magazine*, vol. 16, no. 1, pp. 22–57, 2016.
- [15] F. K. Shaikh and S. Zeadally, "Energy harvesting in wireless sensor networks: a comprehensive review," *Renewable & Sustainable Energy Reviews*, vol. 55, pp. 1041–1054, 2016.
- [16] M. Y. Naderi, K. R. Chowdhury, S. Basagni, W. Heinzelman, S. De, and S. Jana, "Experimental study of concurrent data and wireless energy transfer for sensor networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '14)*, pp. 2543–2549, December 2014.
- [17] R. J. M. Vullers, R. V. Schaijk, H. J. Visser, J. Penders, and C. Hoof, "Energy harvesting for autonomous wireless sensor networks," *IEEE Journal of Solid-State Circuits*, vol. 2, no. 2, pp. 29–38, 2010.
- [18] R. Du, C. Fischione, and M. Xiao, "Joint node deployment and wireless energy transfer scheduling for immortal sensor networks," in *Proceedings of the 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt '17)*, pp. 1–8, Paris, France, May 2017.
- [19] L. Xie, Y. Shi, Y. T. Hou, and H. D. Sherali, "Making sensor networks immortal: an energy-renewal approach with wireless power transfer," *IEEE/ACM Transactions on Networking*, vol. 20, no. 6, pp. 1748–1761, 2012.
- [20] L. Xie, Y. Shi, Y. T. Hou, W. Lou, H. D. Sherali, and S. F. Midkiff, "On renewable sensor networks with wireless energy transfer: the multi-node case," in *Proceedings of the 9th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '12)*, pp. 10–18, June 2012.
- [21] C. Zhu, K. Liu, C. Yu, R. Ma, and H. Cheng, "Simulation and experimental analysis on wireless energy transfer based on magnetic resonances," in *Proceedings of the Vehicle Power and Propulsion Conference (VPPC '08)*, pp. 1–4, 2008.
- [22] M. K. Watfa, H. AlHassanieh, and S. Selman, "Multi-hop wireless energy transfer in WSNs," *IEEE Communications Letters*, vol. 15, no. 12, pp. 1275–1277, 2011.
- [23] A. A. Babayo, M. H. Anisi, and I. Ali, "A Review on energy management schemes in energy harvesting wireless sensor networks," *Renewable & Sustainable Energy Reviews*, vol. 76, pp. 1176–1184, 2017.
- [24] R. Doost-Mohammady and K. R. Chowdhury, "Transforming healthcare and medical telemetry through cognitive radio networks," *IEEE Wireless Communications Magazine*, vol. 19, no. 4, pp. 67–73, 2012.
- [25] A. Khelil, F. K. Shaikh, P. Szczytowski, B. Ayari, and N. Suri, "Map-based design for autonomic wireless sensor networks," in *Autonomic Communication*, pp. 309–326, Springer, 2009.
- [26] M. di Francesco, S. K. Das, and G. Anastasi, "Data collection in wireless sensor networks with mobile elements: a survey," *ACM Transactions on Sensor Networks*, vol. 8, no. 1, article 7, 2011.
- [27] C. Alippi, G. Anastasi, M. Di Francesco, and M. Roveri, "Energy management in wireless sensor networks with energy-hungry sensors," *IEEE Instrumentation & Measurement Magazine*, vol. 12, no. 2, pp. 16–23, 2009.
- [28] M. A. Razzaque and S. Dobson, "Energy-efficient sensing in wireless sensor networks using compressed sensing," *Sensors*, vol. 14, no. 2, pp. 2822–2859, 2014.
- [29] V. Raghunathan, S. Ganeriwal, and M. Srivastava, "Emerging techniques for long lived wireless sensor networks," *IEEE Communications Magazine*, vol. 44, no. 4, pp. 108–114, 2006.
- [30] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.
- [31] E. Ever, R. Luchmun, L. Mostarda, A. Navarra, and P. Shah, "UHEED: an unequal clustering algorithm for wireless sensor networks," in *Proceedings of the 1st International Conference on Sensor Networks (SENSORNETS '12)*, pp. 185–193, February 2012.
- [32] A. Kansal, J. Hsu, S. Zahedi, and M. B. Srivastava, "Power management in energy harvesting sensor networks," *ACM Transactions on Embedded Computing Systems*, vol. 6, no. 4, article 32, 2007.
- [33] E. D. Dunlop, L. Wald, and M. Suri, *Solar Energy Resource Management for Electricity Generation from Local Level to Global Scale*, Nova Science Publishers Inc., 2006.
- [34] F. Akhtar and M. H. Rehmani, "Energy replenishment using renewable and traditional energy resources for sustainable wireless sensor networks: a review," *Renewable & Sustainable Energy Reviews*, vol. 45, pp. 769–784, 2015.
- [35] P. T. V. Bhuvaneshwari, R. Balakumar, V. Vaidehi, and P. Balamuralidhar, "Solar energy harvesting for wireless sensor networks," in *Proceedings of the 1st International Conference on Computational Intelligence, Communication Systems and Networks (CICSYN '09)*, pp. 57–61, July 2009.
- [36] P. Technologies, "Documentation," *Powercast, Wireless Power for a Wireless World*, 2018, <http://www.powercastco.com/documentation/>.
- [37] K. S. Adu-Manu, N. Adam, C. Tapparelo, H. Ayatollahi, and W. Heinzelman, "Energy-harvesting wireless sensor networks (EH-WSNs): a review," *ACM Transactions on Sensor Networks (TOSN)*, vol. 14, no. 2, p. 10, 2018.
- [38] T. J. Kazmierski and S. Beeby, *Energy Harvesting Systems*, Springer, 2014.
- [39] C. Ó. Mathúna, T. O'Donnell, R. V. Martinez-Catala, J. Rohan, and B. O'Flynn, "Energy scavenging for long-term deployable wireless sensor networks," *Talanta*, vol. 75, no. 3, pp. 613–623, 2008.
- [40] S. Roundy, P. K. Wright, and J. M. Rabaey, "Energy scavenging for wireless sensor networks," *Norwell*, 2003.
- [41] D. M. Rowe, "Conversion efficiency and figure-of-merit," in *CRC Handbook of Thermoelectrics*, p. 31, CRC Press, 1995.
- [42] I. Stark, "Invitedtalk: Thermal energy harvesting with thermo life," in *Proceedings of the International Workshop Wearable and Implantable Body Sensor Networks (BSN '06)*, pp. 19–22, 2006.
- [43] A. S. Aricò, P. Bruce, B. Scrosati, J.-M. Tarascon, and W. van Schalkwijk, "Nanostructured materials for advanced energy conversion and storage devices," *Nature Materials*, vol. 4, pp. 366–377, 2005.

- [44] F. I. Simjee and P. H. Chou, "Efficient charging of supercapacitors for extended lifetime of wireless sensor nodes," *IEEE Transactions on Power Electronics*, vol. 23, no. 3, pp. 1526–1536, 2008.
- [45] T. H. Ng and W. H. Liao, "Sensitivity analysis and energy harvesting for a self-powered piezoelectric sensor," *Journal of Intelligent Material Systems and Structures*, vol. 16, no. 10, pp. 785–797, 2005.
- [46] N. S. Shenck and J. A. Paradiso, "Energy scavenging with shoe-mounted piezoelectrics," *IEEE Micro*, vol. 21, no. 3, pp. 30–42, 2001.
- [47] G. K. Ottman, H. F. Hofmann, A. C. Bhatt, and G. A. Lesieutre, "Adaptive piezoelectric energy harvesting circuit for wireless remote power supply," *IEEE Transactions on Power Electronics*, vol. 17, no. 5, pp. 669–676, 2002.
- [48] H. A. Sodano, D. J. Inman, and G. Park, "Comparison of piezoelectric energy harvesting devices for recharging batteries," *Journal of Intelligent Material Systems and Structures*, vol. 16, no. 10, pp. 799–807, 2005.
- [49] M. J. Guan and W. H. Liao, "Characteristics of energy storage devices in piezoelectric energy harvesting systems," *Journal of Intelligent Material Systems and Structures*, vol. 19, no. 6, pp. 671–680, 2008.
- [50] V. Raghunathan, A. Kansal, J. Hsu, J. Friedman, and M. Srivastava, "Design considerations for solar energy harvesting wireless embedded systems," in *Proceedings of the 4th International Symposium on Information Processing in Sensor Networks (IPSN '05)*, pp. 457–462, April 2005.
- [51] B. Buchli, F. Sutton, J. Beutel, and L. Thiele, "Dynamic power management for long-term energy neutral operation of solar energy harvesting systems," in *Proceedings of the 12th ACM Conference*, pp. 31–45, November 2014.
- [52] S. Peng, T. Wang, and C. P. Low, "Energy neutral clustering for energy harvesting wireless sensors networks," *Ad Hoc Networks*, vol. 28, pp. 1–16, 2015.
- [53] S. Baghaee, S. Chamanian, H. Ulsan, O. Zorlu, E. Uysal-Biyikoglu, and H. Kulah, "Demonstration of energy-neutral operation on a WSN testbed using vibration energy harvesting," in *Proceedings of the 20th European Wireless Conference (EW '14)*, pp. 47–52, May 2014.
- [54] C. M. Vigorito, D. Ganesan, and A. G. Barto, "Adaptive control of duty cycling in energy-harvesting wireless sensor networks," in *Proceedings of the 4th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks (SECON '07)*, pp. 21–30, IEEE, San Diego, Calif, USA, June 2007.
- [55] Y. Shi, L. Xie, Y. T. Hou, and H. D. Sherali, "On renewable sensor networks with wireless energy transfer," in *Proceedings of the IEEE INFOCOM '11*, pp. 1350–1358, Shanghai, China, April 2011.
- [56] A. Kurs, A. Karalis, R. Moffatt, J. D. Joannopoulos, P. Fisher, and M. Soljacic, "Wireless power transfer via strongly coupled magnetic resonances," *Science*, vol. 317, no. 5834, pp. 83–86, 2007.
- [57] X. Wei, Z. Wang, and H. Dai, "A critical review of wireless power transfer via strongly coupled magnetic resonances," *Energies*, vol. 7, no. 7, pp. 4316–4341, 2014.
- [58] F. Musavi and W. Eberle, "Overview of wireless power transfer technologies for electric vehicle battery charging," *IET Power Electronics*, vol. 7, no. 1, pp. 60–66, 2014.
- [59] O. Jonah and S. V. Georgakopoulos, "Wireless power transfer in concrete via strongly coupled magnetic resonance," *IEEE Transactions on Antennas and Propagation*, vol. 61, no. 3, pp. 1378–1384, 2013.
- [60] F. Engmann, J.-D. Abdulai, and J. Q. Azasoo, "Enhancing the reliability of wsn through wireless energy transfer," in *Proceedings of the International Conference on Computational Science and Its Applications*, pp. 610–618, Springer, 2016.
- [61] J. Li, K. Li, and W. Zhu, "Improving sensing coverage of wireless sensor networks by employing mobile robots," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO '07)*, pp. 899–903, December 2007.
- [62] B. L. Cannon and D. D. Stancil, "Magnetic resonant coupling as a potential means for wireless power transfer to multiple small receivers," *IEEE Transactions on Power Electronics*, vol. 24, no. 7, pp. 1819–1825, 2009.
- [63] M. Erol-Kantarci and H. T. Mouftah, "Suresense: sustainable wireless rechargeable sensor networks for the smart grid," *IEEE Wireless Communications Magazine*, vol. 19, no. 3, pp. 30–36, 2012.
- [64] K. S. Adu-Manu, C. Tapparello, W. Heinzelman, F. A. Katsriku, and J.-D. Abdulai, "Water quality monitoring using wireless sensor networks: Current trends and future research directions," *ACM Transactions on Sensor Networks*, vol. 13, no. 1, 2017.
- [65] D. Mishra, S. De, S. Jana, S. Basagni, K. Chowdhury, and W. Heinzelman, "Smart RF energy harvesting communications: Challenges and opportunities," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 70–78, 2015.
- [66] J. Hester, T. Scott, and J. Sorber, "Ekho: realistic and repeatable experimentation for tiny energy-harvesting sensors," in *Proceedings of the 12th ACM Conference on Embedded Networked Sensor Systems (SenSys '14)*, pp. 1–15, New York, NY, USA, November 2014.
- [67] I. Minakov and R. Passerone, "PASES: an energy-aware design space exploration framework for wireless sensor networks," *Journal of Systems Architecture*, vol. 59, no. 8, pp. 626–642, 2013.
- [68] C. Tapparello, H. Ayatollahi, and W. Heinzelman, "Energy harvesting framework for network simulator 3 (ns-3)," in *Proceedings of the 2nd International Workshop on Energy Neutral Sensing Systems (ENSsys '14)*, pp. 37–42, Memphis, TN, USA, November 2014.
- [69] H. P. Dai, X. B. Wu, L. J. Xu, G. Chen, and S. Lin, "Using minimum mobile chargers to keep large-scale wireless rechargeable sensor networks running forever," in *Proceedings of the IEEE 22nd International Conference on Computer Communication and Networks (ICCCN '13)*, pp. 1–7, August 2013.
- [70] W. Xu, W. Liang, X. Lin, and G. Mao, "Efficient scheduling of multiple mobile chargers for wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 9, pp. 7670–7683, 2016.
- [71] Y. Peng, Z. Li, W. Zhang, and D. Qiao, "Prolonging sensor network lifetime through wireless charging," in *Proceedings of the 31st IEEE Real-Time Systems Symposium (RTSS '10)*, pp. 129–139, IEEE, December 2010.
- [72] W. Yao, M. Li, and M.-Y. Wu, "Inductive charging with multiple charger nodes in wireless sensor networks," in *Proceedings of the Asia-Pacific Web Conference*, pp. 262–270, Springer, 2006.
- [73] A. Madhja, S. Nikolettseas, and T. P. Raptis, "Hierarchical, collaborative wireless energy transfer in sensor networks with multiple Mobile Chargers," *Computer Networks*, vol. 97, pp. 98–112, 2016.
- [74] H. Ayatollahi, C. Tapparello, and W. Heinzelman, "Transmitter-receiver energy efficiency: a trade-off in MIMO wireless sensor

- networks,” in *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC '15)*, pp. 1476–1481, March 2015.
- [75] A. B. Constantine, “Microstrip Antennas,” in *Antenna Theory: Analysis and Design*, John Wiley & Sons, 3rd edition, 2005.
- [76] Q. Li, R. Q. Hu, Y. Qian, and G. Wu, “Cooperative communications for wireless networks: techniques and applications in LTE-advanced systems,” *IEEE Wireless Communications Magazine*, vol. 19, no. 2, pp. 22–29, 2012.
- [77] K. Huang and E. Larsson, “Simultaneous information and power transfer for broadband wireless systems,” *IEEE Transactions on Signal Processing*, vol. 61, no. 23, pp. 5972–5986, 2013.
- [78] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, “Relaying protocols for wireless energy harvesting and information processing,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3622–3636, 2013.
- [79] R. Zhang and C. K. Ho, “MIMO broadcasting for simultaneous wireless information and power transfer,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 1989–2001, 2013.
- [80] J. Park and B. Clerckx, “Joint wireless information and energy transfer with reduced feedback in MIMO interference channels,” *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 8, pp. 1563–1577, 2015.
- [81] Z. Xiang and M. Tao, “Robust beamforming for wireless information and power transmission,” *IEEE Wireless Communications Letters*, vol. 1, no. 4, pp. 372–375, 2012.
- [82] X. Chen, C. Yuen, and Z. Zhang, “Wireless energy and information transfer tradeoff for limited-feedback multiantenna systems with energy beamforming,” *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, pp. 407–412, 2014.
- [83] J. Park and B. Clerckx, “Joint wireless information and energy transfer in a two-user MIMO interference channel,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 8, pp. 4210–4221, 2013.
- [84] J. Park and B. Clerckx, “Joint wireless information and energy transfer in a K-user MIMO interference channel,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 10, pp. 5781–5796, 2014.
- [85] Castalia, “Castalia wireless sensor network simulator,” 2016.
- [86] F. Chen, I. Dietrich, R. German, and F. Dressler, “An energy model for simulation studies of wireless sensor networks using OMNeT++,” *Praxis der Informationsverarbeitung und Kommunikation*, vol. 32, no. 2, pp. 133–138, 2009.
- [87] H. Sundani, H. Li, V. Devabhaktuni, M. Alam, and P. Bhattacharya, “Wireless sensor network simulators a survey and comparisons,” *International Journal of Computer Networks*, vol. 2, no. 5, pp. 249–265, 2011.
- [88] H. Wu, S. Nabar, and R. Poovendran, “An energy framework for the network simulator 3 (ns-3),” in *Proceedings of the 4th International ICST Conference on Simulation Tools and Techniques*, Barcelona, Spain, March 2011.
- [89] M. K. Stojcev, M. R. Kosanovic, and L. R. Golubovic, “Power management and energy harvesting techniques for wireless sensor nodes,” in *Proceedings of the 9th International Conference on Telecommunications in Modern Satellite, Cable, and Broadcasting Services (TELSIKS '09)*, pp. 65–72, October 2009.
- [90] S. Cui, A. J. Goldsmith, and A. Bahai, “Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks,” *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1089–1098, 2004.
- [91] S. Hussain, A. Azim, and J. H. Park, “Energy efficient virtual MIMO communication for wireless sensor networks,” *Telecommunication Systems*, vol. 42, no. 1-2, pp. 139–149, 2009.
- [92] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, “An overview of massive MIMO: benefits and challenges,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 742–758, 2014.
- [93] A. Nosratinia, T. E. Hunter, and A. Hedayat, “Cooperative communication in wireless networks,” *IEEE Communications Magazine*, vol. 42, no. 10, pp. 74–80, 2004.
- [94] S. Basagni, M. Y. Naderi, C. Petrioli, and D. Spenza, “Wireless sensor networks with energy harvesting,” *Mobile Ad Hoc Networking: Cutting Edge Directions*, pp. 701–736, 2013.
- [95] I. Khemapech, A. Miller, and I. Duncan, “A survey of transmission power control in wireless sensor networks,” in *Proceedings of the PGNNet*, pp. 15–20, 2007.
- [96] S. Lin, F. Miao, J. Zhang et al., “ATPC: Adaptive transmission power control for wireless sensor networks,” *ACM Transactions on Sensor Networks*, vol. 12, no. 1, 2016.
- [97] K. V. Kumar and E. Baburaj, “Energy efficient transmission power control in sensor nodes of WSN SCADA systems using cognitive fuzzy systems,” *International Journal of Applied Engineering Research*, vol. 11, no. 2, pp. 1478–1484, 2016.
- [98] H. Cotuk, K. Bicakci, B. Tavli, and E. Uzun, “The impact of transmission power control strategies on lifetime of wireless sensor networks,” *IEEE Transactions on Computers*, vol. 63, no. 11, pp. 2866–2879, 2014.
- [99] A. Tripathi, S. Gupta, and B. Chourasiya, “Survey on data aggregation techniques for wireless sensor networks,” *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, no. 7, pp. 7366–7371, 2014.
- [100] L. C. Zhong, R. Shah, C. Guo, and J. Rabaey, “An ultra-low power and distributed access protocol for broadband wireless sensor networks,” *IEEE Broadband Wireless Summit*, vol. 3, 2001.
- [101] M. Bagaa, Y. Challal, A. Ksentini, A. Derhab, and N. Badache, “Data aggregation scheduling algorithms in wireless sensor networks: solutions and challenges,” *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1339–1368, 2014.
- [102] N. Kimura and S. Latifi, “A survey on data compression in wireless sensor networks,” in *Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC '05)*, vol. 2, pp. 8–13, April 2005.
- [103] T. Srisooksai, K. Keamarungsi, P. Lamsrichan, and K. Araki, “Practical data compression in wireless sensor networks: a survey,” *Journal of Network and Computer Applications*, vol. 35, no. 1, pp. 37–59, 2012.
- [104] G. M. Dias, B. Bellalta, and S. Oechsner, “A survey about prediction-based data reduction in wireless sensor networks,” *ACM Computing Surveys*, vol. 49, no. 3, 2016.
- [105] C. Alippi, G. Anastasi, M. Di Francesco, and M. Roveri, “An adaptive sampling algorithm for effective energy management in wireless sensor networks with energy-hungry sensors,” *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 2, pp. 335–344, 2010.
- [106] B. Srbinovski, M. Magno, B. O’Flynn, V. Pakrashi, and E. Popovici, “Energy aware adaptive sampling algorithm for energy harvesting wireless sensor networks,” in *Proceedings of the Sensors Applications Symposium (SAS '15)*, pp. 1–6, 2015.
- [107] Y. Yin, F. Liu, X. Zhou, and Q. Li, “An efficient data compression model based on spatial clustering and principal component analysis in wireless sensor networks,” *Sensors*, vol. 15, no. 8, pp. 19443–19465, 2015.

- [108] X. Cao, S. Madria, and T. Hara, "Efficient Z-order encoding based multi-modal data compression in WSNs," in *Proceedings of the 37th IEEE International Conference on Distributed Computing Systems (ICDCS '17)*, pp. 2185–2192, June 2017.
- [109] M. Wu, L. Tan, and N. Xiong, "Data prediction, compression, and recovery in clustered wireless sensor networks for environmental monitoring applications," *Information Sciences*, vol. 329, pp. 800–818, 2016.
- [110] B. Stojkoska, D. Solev, and D. Davcev, "Data prediction in WSN using variable step size LMS algorithm," in *Proceedings of the 5th International Conference on Sensor Technologies and Applications (SENSORCOMM '11)*, pp. 191–196, August 2011.
- [111] W. R. Heinzelman, A. Chandrakasan, and H. Balakrishnan, "Energy-efficient communication protocol for wireless micro-sensor networks," in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences (HICSS '00)*, p. 10, January 2000.
- [112] M. Ye, C. Li, G. Chen, and J. Wu, "EECS: an energy efficient clustering scheme in wireless sensor networks," in *Proceedings of the 24th IEEE International Performance, Computing, and Communications Conference (IPCCC '05)*, pp. 535–540, IEEE, April 2005.
- [113] G. V. Selvi and R. Manoharan, "A survey of energy efficient unequal clustering algorithms for wireless sensor networks," *International Journal of Computer Applications*, vol. 79, no. 1, 2013.
- [114] R. Sharma, G. Jain, and S. Gupta, "Enhanced Cluster-head selection using round robin technique in WSN," in *Proceedings of the International Conference on Communication Networks (ICCN '15)*, pp. 37–42, Gwalior, India, November 2015.
- [115] A. A. Abbasi and M. Younis, "A survey on clustering algorithms for wireless sensor networks," *Computer Communications*, vol. 30, no. 14-15, pp. 2826–2841, 2007.
- [116] C. Alippi, R. Camplani, C. Galperti, and M. Roveri, "A robust, adaptive, solar-powered WSN framework for aquatic environmental monitoring," *IEEE Sensors Journal*, vol. 11, no. 1, pp. 45–55, 2011.
- [117] G. Chen, C. Li, M. Ye, and J. Wu, "An unequal cluster-based routing protocol in wireless sensor networks," *Wireless Networks*, vol. 15, no. 2, pp. 193–207, 2009.
- [118] H. Bagci and A. Yazici, "An energy aware fuzzy approach to unequal clustering in wireless sensor networks," *Applied Soft Computing*, vol. 13, no. 4, pp. 1741–1749, 2013.
- [119] K. Kaushik, D. Mishra, S. De et al., "Experimental demonstration of multi-hop RF energy transfer," in *Proceedings of the IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC '13)*, pp. 538–542, September 2013.
- [120] G. Benigno, O. Briante, and G. Ruggieri, "A sun energy harvester model for the network simulator 3 (ns-3)," in *Proceedings of the 12th Annual IEEE International Conference on Sensing, Communication, and Networking - Workshops (SECON Workshops '15)*, pp. 1–6, Seattle, WA, USA, June 2015.
- [121] R. C. Carrano, D. Passos, L. C. S. Magalhaes, and C. V. N. Albuquerque, "Survey and taxonomy of duty cycling mechanisms in wireless sensor networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 181–194, 2014.
- [122] I. Mayordomo, T. Drager, P. Spies, J. Bernhard, and A. Pflaum, "An overview of technical challenges and advances of inductive wireless power transmission," in *Proceedings of the IEEE*, vol. 101, no. 6, pp. 1302–1311, 2013.
- [123] S. Bi, C. K. Ho, and R. Zhang, "Wireless powered communication: opportunities and challenges," *IEEE Communications Magazine*, vol. 53, no. 4, pp. 117–125, 2015.

Research Article

Low Cost and High Efficiency Hybrid Architecture Massive MIMO Systems Based on DFT Processing

Weiqiang Tan ¹, Guixian Xu,² Elisabeth De Carvalho,² Mu Zhou ³,
Lisheng Fan ¹ and Chunguo Li ⁴

¹School of Computer Science and Educational Software, Guangzhou University, Guangzhou 510006, China

²Department of Electronic Systems, Aalborg University, 9400 Aalborg, Denmark

³Chongqing Key Lab of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

⁴National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China

Correspondence should be addressed to Lisheng Fan; lsfan@gzhu.edu.cn

Received 24 April 2018; Revised 10 July 2018; Accepted 2 August 2018; Published 14 August 2018

Academic Editor: Zheng Chu

Copyright © 2018 Weiqiang Tan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Low cost and high efficiency, defined as energy efficiency (EE) and spectral efficiency (SE), have raised more and more attention in the fifth generation (5G) communication systems due to steadily rising hardware cost, energy consumption, and mobile traffic. This paper studies the hybrid architecture of multiuser massive MIMO systems, where the digital domain utilizes the zero-forcing (ZF) precoding scheme and the analog domain uses discrete Fourier transform (DFT) processing that significantly reduces hardware cost and energy consumption. We derive analytical expressions on the total achievable SE and EE, as well as offering insight into some engineering parameters in the system performance. Our aim is to achieve low cost and high efficiency massive MIMO system, with constraints on the overall transmit power, the number of users, and the number of radio frequency (RF) chains. Results exhibit that the total achievable SE of the hybrid architectures with DFT processing is inferior to the full digital architectures and hybrid architectures with the ideal phase shifters, but the performance attenuation can be compensated by providing the more input SNR and higher number of RF chains. Moreover, we find that the total achievable EE of hybrid architectures with DFT processing outperforms other massive MIMO architectures that include a full digital implementation, ideal phase shifters, and a switched network.

1. Introduction

As environmental issue has attracted more and more concern from public, green radio communication has emerged as the most important concept in the fifth generation (5G) communication systems [1–4]. Along with improved traffic rate, transmission reliability, and quality of service, the upcoming 5G communication systems aim at improving the energy efficiency without compromising on the user experience [5–8]. Massive multiple-input multiple-output (MIMO), viewed as one of the important techniques for 5G wireless communication, has attracted a great deal of research interest in current years. For a conventional massive multiple-input multiple-output (MIMO) system with fully digital architecture, every antenna element requires to deploy

a radio frequency (RF) chain that is equipped with a high-resolution analog-to-digital converter (ADC). The energy consumption of the ADCs grows exponentially with the number of ADC quantization bits and also increases with carrier frequency bandwidth and signal sampling rate. However, massive MIMO systems employing large-scale number antennas and high-resolution ADCs lead to unaffordable energy consumption and hardware cost [9–12].

Hybrid architectures massive MIMO systems, which consist of analog processing and digital precoding, are one such alternative with promising techniques in 5G communication systems [13–15]. Such architectures not only achieve high total achievable SE, but also keep low cost and energy consumption. There has been a lot of work devoted to developing analog beamforming design. For example, the

authors in [16, 17] proposed a novel algorithm for point-to-point MIMO systems, which aimed to reduce the computational complexity of system and maximize the total achievable spectral efficiency (SE) on different subcarriers. To handle the severe energy consumption and hardware cost, the authors in [18] proposed user scheduling algorithms and resource allocation schemes, which aimed to maximize the total achievable energy efficiency (EE) on hybrid massive system. However, all the aforementioned studies are based on the ideal phase shifters at the analog domain, which has a potential to enhance the complexity of the system, especially for a huge quantity of data. In order to simplify the complexity of system, the work of [19] proposed hybrid architecture with a switches network, where the ideal phase shifters are replaced by a switches network. Results reported that the switches network is a simpler solution but suffers from the significantly performance loss of the system.

Fortunately, the work of [20] found that using discrete Fourier transform (DFT) processing at analog domain has similar nature with electromagnetic lens and outperforms the fully connected phase shifters, which has low power consumption and is easy to implement. In light of these benefits, the DFT processing at the analog domain has gained great attention [21]. Results in [22] demonstrated that the DFT processing is carried out by the Butler circuit, which has low cost and high energy efficiency. Most importantly, the authors in [23] reported that the DFT processing is treated as the most promising technique used for signal transfer, which has been recommended by various industrial ports. Meanwhile, the authors in [24] investigated hybrid massive MIMO system based on DFT processing and further obtained the bounds on the achievable SE of such system, where the zero-forcing (ZF) precoding is adopted by the baseband processing. These studies confirm that the hybrid massive MIMO system based on DFT processing provides advantages for massive MIMO systems. However, the aforementioned work of [23, 24] did not derive an exact expression on the achievable SE and did not include analytical results on the achievable EE for hybrid massive MIMO systems.

In order to compensate for this gap in the study of achievable SE and EE, this paper presents a novel hybrid analog/digital architecture for massive MIMO systems, where analog processing is implemented using DFT processing. Moreover, we further investigate the EE and SE of the hybrid architecture massive MIMO system, which aims to effectively reduce the power consumption and hardware cost without compromising on the achievable SE. Our main contributions are as follows:

- (i) We present a novel hybrid architecture massive system based on the DFT processing and derive exact and approximated expressions on the total achievable SE of system. According to the derived analytical result, we further offer insight into the impact of several engineering parameters in the system performance, which constraints on the overall transmit power, the number of users, and the number of RF chains.

- (ii) By establishing the realistic power consumption model and employing the derived approximate expression, we evaluate the total EE for hybrid architectures with DFT processing. Closed-form expressions for the optimal the number of RF chains and the SNR by maximizing the achievable EE are derived.
- (iii) Results exhibit that the total achievable SE of the hybrid architectures with DFT precessing is inferior to the full digital architectures and hybrid architectures with the ideal phase shifters, but the performance loss can be compensated by offering more RF chains and more SNR. We also find that hybrid architectures with DFT precessing exhibit much better total achievable EE than the other hybrid architectures massive MIMO systems.

The reminder of the paper is listed as follows. Section 2 presents the signal model and hybrid architectures massive MIMO system with DFT precessing. In Section 3, we analyze the system's performance that includes the achievable SE and EE. Based on the derived theoretical results, we attain the several engineering insights by considering the special cases. Section 4 provides the numerical results to confirm the obtained analytical results. Section 5 outlines the whole paper.

Notations. Matrix and vector are used by upper and lower case boldface, $\det(\cdot)$, $\text{diag}(\cdot)$, and $(\cdot)!$ stand for a matrix determinant, diagonal matrix, and factorial operation, respectively, $[\mathbf{G}]_{k,k}$ denotes the k -th diagonal elements of \mathbf{G} , $[\cdot]$ accounts for rounding to the nearest integer, and \mathbf{I}_K and $\mathcal{C}^{n \times n}$ account for an identity matrix of $K \times K$ dimension and an complex Gaussian matrix of $n \times n$ dimension, respectively.

2. System Model

We firstly present the signal model associated with a hybrid analog/digital architecture in massive MIMO system, where the analog domain utilizes the low cost DFT processing, and then we describe channel model that consists of the small and large fading.

2.1. Signal Model. As depicted in Figure 1, we present a hybrid architecture of a multiuser massive MIMO system, where the BS comprises N_t transmit antennas simultaneously serving M users. Each user has a single-antenna element and the N_t transmit antennas are connected to N_s RF chains via a DFT processing with parameters restricted as $M \ll N_t$ and $M \leq N_s \leq N_t$. Assume that \mathbf{x} denotes a transmitted signal vector of $M \times 1$ dimension for all users with $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}_M$. When the input signal is the Gaussian symbol [25–27], the received signal after analog DFT processing technique and digital linear precoding scheme can be given by

$$\mathbf{y} = \sqrt{P}\mathbf{G}^H\mathbf{F}\mathbf{W}\mathbf{x} + \mathbf{n}, \quad (1)$$

in which P stands for the input SNR of the system, \mathbf{W} represents the digital baseband precoder matrix, \mathbf{F} denotes the analog processing matrix, \mathbf{n} denotes the additive white Gaussian noise (AWGN), and \mathbf{G} accounts for the channel

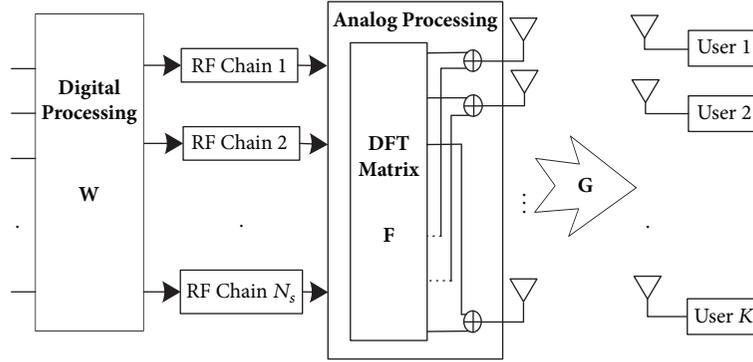


FIGURE 1: Illustration of hybrid architecture massive MIMO system, in which the analog domain adopts DFT processing.

matrix that includes the large-fading coefficient and fast-fading coefficient, which can be given by

$$\mathbf{G} = \mathbf{H}\mathbf{D}^{1/2}, \quad (2)$$

where \mathbf{H} accounts for the small-scale fading matrix, whose entries follow a complex Gaussian distribution, and \mathbf{D} denotes large-scale fading matrix that diagonal entries are given by $[\mathbf{D}]_{k,k} = \beta_k$ and nondiagonal entries equal zero. In practical, large-scale fading coefficient β_k contains the shadowing fading and path loss of the user k , which is modeled as $\beta_k = z_k r_k^{-\gamma}$, where z_k and γ account for a log-normal random variable of the user k and the decay exponent of path loss, respectively, and r_k accounts for the distance between the BS and the user k .

2.2. DFT Processing Implemented. In the previous works about hybrid massive MIMO systems [13, 28], phase shifters or a network of switches requires the ideal CSI to be available at the BS. To overcome this drawback brought by phase shifters, this paper investigates a hybrid massive MIMO system as showed in Figure 1, in which the analog domain is implemented by using DFT processing. For hybrid architecture with DFT processing, all transmitted signals via a partial DFT matrix preprocessing before digital baseband precoding, in which the partial DFT matrix is derived by uniformly randomly selecting the rows of a DFT matrix. Therefore, the partial matrix \mathbf{F} is constructed by

$$\mathbf{F} = \frac{1}{\sqrt{N_t}} \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & \cdots & w^{(N_s-1)} \\ 1 & w^2 & w^4 & \cdots & w^{2(N_s-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & w^{(N_t-1)} & w^{2(N_t-1)} & \cdots & w^{(N_t-1)(N_s-1)} \end{bmatrix}, \quad (3)$$

where $w = e^{-2\pi i/N_t}$.

Remark 1. From energy efficiency and economic perspective, the DFT processing is appealing for analog domain in massive MIMO systems since it has low energy consumption, less attenuation, and high performance, while the DFT processing

is implemented by adopting a field programmable gate array (FPGA) [29] or an integrated Butler circuit [30], which is easy to implement and efficiently reduces hardware complexity. In addition, from the actual practical implementation, the DFT processing able to embrace all propagation signal from K users and the DFT processing is implemented offline, which makes sure all RF chains always be activated.

3. Performance Analysis of System

In this section, our main task aims to derive the exact expression on the achievable SE. By using the obtained theoretical result, we offer the insight of engineering parameters on the achievable SE. And then the total achievable EE is investigated by considering the energy consumption of such a system. Finally a brief discussion is provided on how to achieve the EE maximization.

3.1. Hybrid Massive MIMO System. A hybrid architecture massive MIMO system with DFT processing is displayed in Figure 1, which is split into DFT processing and digital precoding processing successively. Before DFT processing at the analog domain and digital precoding at the baseband domain, an equivalent channel matrix \mathbf{G}_{eq} is equivalent to a composite matrix composed of the channel matrix and partial DFT matrix, i.e., $\mathbf{G}_{eq} = \mathbf{G}^H \mathbf{F}$. Thus, the digital precoding matrices are based on \mathbf{G}_{eq} . In addition, we make an assumption that the accurate and instantaneous CSI is available. In practical setups, the CSI is efficiently derived by channel estimation [31] and shared with the users using limited feedback techniques such as given in [32]. The ZF precoding scheme is able to completely eliminate interference among users, which achieves the performance of the non-linear precoding schemes, for example, dirty paper coding (DPC) [33]. Thus, we would like to employ zero-forcing (ZF) precoding at the digital domain; \mathbf{W} can be written as

$$\mathbf{W} = \mathbf{G}_{eq}^H (\mathbf{G}_{eq} \mathbf{G}_{eq}^H)^{-1} \mathbf{\Xi}, \quad (4)$$

where $\mathbf{\Xi}$ accounts for the normalization matrix that satisfies the constraint condition, i.e., $[\mathbf{\Xi}]_{k,k} = 1/\sqrt{[\mathbf{W}^H \mathbf{W}]_{k,k}}$, which makes sure the precoding power constraint is satisfied. When the input symbol is a Gaussian random variable, the

achievable SE of the user k with a ZF precoding scheme is expressed as

$$R_k = \mathbb{E} \left\{ \log_2 \left(1 + \frac{1}{\left[(\mathbf{G}_{eq} \mathbf{G}_{eq}^H)^{-1} \right]_{k,k}} \right) \right\}. \quad (5)$$

Considering all the users, the total achievable ergodic SE of system is identical to a sum of the achievable ergodic SE per user, which is calculated as

$$R_{\text{sum}} = \sum_{k=1}^M R_k. \quad (6)$$

According to (5), we see that the expectation is carried out via ergodic of the equivalent channel matrix. In order to derive the exact expression of the achievable SE, we have to calculate the probability density function (p.d.f) of the SINR term. In the following section, we shall present the detailed process about the achievable SE.

3.2. Spectral Efficiency Analysis

Theorem 2. Consider hybrid architecture massive MIMO systems with DFT processing; the exact expression on the achievable ergodic SE of the user k with ZF precoding scheme is calculated as

$$R_k = \log_2(e) e^{1/P\beta_k} \sum_{h=1}^{N_s-M+1} E_h \left(\frac{1}{P\beta_k} \right), \quad (7)$$

where $E_h(\cdot)$ denotes the exponential integral function of order h . For real nonzero values of x and the integer h , the exponential integral $E_h(x)$ is defined as

$$E_h(x) = \int_1^{\infty} \frac{e^{-xt}}{t^h} dt. \quad (8)$$

Proof. See Appendix A. \square

From Theorem 2, it is shown that the achievable SE of the user k is related to the number of RF chains, the input SNR, and the number of users. However, it is hard to get an intuition about the impact of the system parameters on the achievable SE, since it invokes the exponential integral function. Therefore, we try to calculate an approximate expression on the achievable SE, which enables us to directly observe the general trends of the SE as a function of the most determinant parameters.

Corollary 3. Considering a hybrid architecture massive MIMO system based on DFT processing, an approximate expression on the achievable SE can be given by

$$\bar{R}_k = \log_2(1 + P\beta_k(N_s - M + 1)). \quad (9)$$

Proof. To begin with, we invoke the results in [34] and directly obtain the following approximate expression:

$$e^{1/x} \sum_{n=1}^{N-1} E_n \left(\frac{1}{x} \right) \approx \log(1 + x(N-1)). \quad (10)$$

According to the above approximate equation, we have the following result:

$$\begin{aligned} e^{1/P\beta_k} \sum_{h=1}^{N_s-M+1} E_h \left(\frac{1}{P\beta_k} \right) \\ \approx \log(1 + P\beta_k(N_s - M + 1)). \end{aligned} \quad (11)$$

Substituting (11) into (8), we can derive the desired result. \square

With the derived tractable expression in (9) at hand, we obtain clear insights into the SE behavior with respect to different system parameters. From (9), we observe that \bar{R}_k is concerned with the large-fading coefficient, the input SNR, the number of RF chains, and the number of users. It can be seen that \bar{R}_k is a monotonically increasing function of the number of RF chains and the input SNR. Additionally, assuming a fixed number of BS antennas and the input SNR, \bar{R}_k is a monotonically decreasing function of the number of users. However, as the number of users grows, the total achievable SE increases since it is the multiplication of the achievable ergodic per rate by M . Besides, the total achievable SE is a concave function with respect to the number of users M , and there exists a unique globally optimal number of users maximizing the total achievable SE according to convex optimization theory. In the following corollary, we keep the other parameters being fixed and concentrate on finding the best number of users to achieve the total achievable SE maximization.

Corollary 4. For hybrid architecture massive MIMO systems based on DFT processing, there exists a unique globally optimal number of users M^{opt} that achieve the total achievable SE maximization, which is

$$M^{\text{opt}} = \left\lfloor \frac{\Delta}{P\beta} \left(1 - \frac{1}{W_0(\Delta \exp(1))} \right) \right\rfloor, \quad (12)$$

where $\Delta = 1 + P\beta(N_s + 1)$ and $W_0(\cdot)$ represents the Lambert function.

Proof. See Appendix B. \square

Corollary 4 reveals an important conclusion that the achievable SE will not linearly increase with the number of users, but it is a convex function. This implies that the optimal number of users M^{opt} in (12) achieves the total achievable SE maximization. Additionally, it worth noting that Corollary 4 typically shows us a noninteger value with respect to M^{opt} ; however, the quasi-concavity of the solution in (12) indicates that M^{opt} is obtained at one of the two closest integers.

3.3. Energy Efficiency Analysis. Green radio communication has emerged as a new prominent indicator in the design of 5G communication systems [35, 36]. Thus, the total achievable EE has become a primary concern, which can be viewed as the total achievable SE divided by the total energy consumption, which is established as

$$\eta_E = \frac{BR_{\text{sum}}}{P_{\text{total}}}, \quad (13)$$

where B stands for the available bandwidth, whose unit is MHz, P_{total} accounts for the overall energy consumption, and R_{sum} was defined as the total achievable SE of system, which is provided in (6).

We now present a practical energy consumption study of the system. According to the prior works of [37–39], the main energy consumption comes in several varieties: circuit energy P_C , signal processing energy P_{SP} , and signal transmission energy P_T . Therefore, the total energy consumption in our system can then be calculated as

$$P_{\text{total}} = P_C + P_{\text{SP}} + P_T. \quad (14)$$

Next, we shall introduce a detailed explanation of each energy consumption part in the following.

- (1) Circuit energy: the circuit loss of system mainly comes from three parts: which includes the LNA module, DFT processor, and one RF chain. Thus, the circuit energy is calculated as

$$P_C = N_t P_{\text{LAN}} + P_{\text{DFT}} + N_s P_{\text{RF}}, \quad (15)$$

where P_{LAN} accounts for the energy consumption from a LNA, P_{DFT} denotes the maximum energy consumption from DFT processor at the analog domain, and P_{RF} is the energy consumption for a single RF chain.

- (2) Signal processing energy: the total energy consumption mainly comes from the signal processing, which is given by [39]

$$P_{\text{SP}} = K P_{\text{cd}} + \left(\frac{2M^2 N_t + 2MN_t}{LT} + \frac{2M^3}{3LT} \right) + \frac{N_t M}{L}, \quad (16)$$

where $P_{\text{cd}} = P_{\text{code}} + P_{\text{dec}}$, P_{code} and P_{dec} stand for the energy consumption caused by the coding

and decoding symbols, respectively. The second term can be understood as the energy computation of precoding due to the matrix inversion of the ZF scheme. The values of T , L , and U stand for the coherence time, the computational efficiency, and the number of blocks per sec, respectively. The third term can be understood as the energy computation of information symbols during signal transmission.

- (3) Signal transmission energy: the energy consumption mainly comes from the loss of signal transmission in the circuits and transmission, which is function of the input SNR and is equivalent to $P_T = P/\eta$, where η accounts for the effective coefficient of the input SNR.

Under the above consideration about the energy consumption, we shall derive a closed-form expression of total achievable EE. By plugging (14) into (13), the total achievable EE of system is expressed as

$$\eta_E = \frac{B \log_2 (1 + P\beta_k (N_s - M + 1))}{N_t I_1 + N_s P_{\text{RF}} + M (P_{\text{cd}} + 2M^2/3LT) + P/\eta + P_{\text{DFT}}}, \quad (17)$$

where $I_1 = P_{\text{LAN}} + M/L + 2M^2/LT$. Based on the above optimization problem, it is shown that the total achievable EE in (17) is related to other fixed energy consumption, the number of BS antennas, the number of RF chains, the input SNR, and the number of users. Maintaining the number of BS antennas, the number of users, the input SNR, and the number of RF chains constant, we will focus on finding the best input SNR that achieves the total achievable EE maximization.

Corollary 5. *Keeping the number of BS antennas, the number of RF chains, the number of users, and the other parameters ($B, I_1, L, T, \beta_k, \eta, P_{\text{RF}}, P_{\text{cd}}, P_{\text{DFT}}$) constant, the optimal SNR is calculated as (18), which can be found below.*

$$p^{\text{opt}} = \frac{\exp \left\{ W_0 \left(\eta \beta_k (N_s - M + 1) - 1 / \left(N_t I_1 + N_s P_{\text{RF}} + M (P_{\text{cd}} + 2M^2/3LT) + P_{\text{DFT}} \right) + 1 \right) \right\} - 1}{\beta_k (N_s - M + 1)} \quad (18)$$

Proof. To complete the proof, we consider the optimization function $g(x)$ as follows:

$$\max g(x) = \frac{f \log_2 (a + bx)}{c + dx} \quad (19)$$

with constants $a, c \geq 0$ and $b, d, f > 0$. With the aid of the conclusion provided in [40], we infer that the objective function is strictly quasi-concave and there exists an optimal x that maximizes $g(x)$; the optimal x can be given by

$$x^{\text{opt}} = \frac{e^{W_0((bc-ad)/de)+1} - a}{b}, \quad (20)$$

where e is the natural number. By capitalizing on the above result and keep the system parameters constant except the SNR, we arrive at the desired result. \square

Corollary 5 provides an explicit guideline on how much SNR should be offered at the BS to maximize the EE. But, it is worth noting that the optimal SNR involves the Lambert function $W(x)$, which is a strictly increasing formula in accordance with the random variance x [41, 42]. In the following, we shall present the numerical result to verify the derived analytical result.

Corollary 6. *Keeping the number of BS antennas, the SNR, the number of users, and the other parameters*

($B, I_1, N_t, L, T, \beta_k, \eta, P_{RF}, P_{cd}, P_{DFT}$) as constant, the number of RF chains is calculated as

$$N_s^{opt} = \left\lceil \left(\frac{1}{P\beta_k} \right) \cdot \exp \left(W_0 \left(\frac{P\beta_k (2M^3/3LT + P/\eta + N_t I_1 + I_2)}{P_{RF} e} \right) + 1 \right) + (M - 1) \right\rceil \quad (21)$$

where $I_2 = MP_{cd} + P_{DFT} - P_{RF} (1 - M)$.

Proof. Using the similar method, the desired result is directly obtained by directly utilizing the results provided in (20) and setting the other parameters being constant except the number of RF chains. \square

4. Numerical Results

In this section, we provide the Monte-Carlo simulations to confirm the theoretical results on the achievable SE, which have been derived in Section 3, and then we compare the achievable SE for full digital architectures and hybrid architectures by considering realistic power consumption. In all simulations, the cell radius is 1000 meters, the guard zone ring region is set to 100 meters, the number of users is set to eight, the bandwidth of carrier frequency is $B = 10$ MHz, the computational efficiency is $L = 12.8$, the coherence time is set to $T = 32$ ms, the decay exponent of large-scale fading is given by $\gamma = 2.1$, and the standard deviation of large-scale fading is $\sigma_{shad} = 4.9$ dB. The large-fading coefficients β_k ($k = 1, \dots, 8$) are randomly created as follows: $\{3.14, 0.09, 0.154, 5.38, 3.29, 0.16, 11.93, 1.82\} \times 10^{-3}$. Unless specified otherwise, the above simulation parameters are used. All the numerical results are averaged over 10,000 independent channel realizations.

Figure 2 depicts the total achievable SE versus the input SNR, where all curves include the exact expression of Theorem 2, the approximate expression of Corollary 3, and the simulated results. We can see that the curves for the simulated results, the approximate expression, and exact expression coincide perfectly in the entire SNR regime. This observation confirms the derived analytical expressions in Theorem 2 and Corollary 3. Meanwhile, Figure 2 also provides the total achievable SE for the full digital architectures [43], as well as the hybrid architectures with the ideal phase shifters and switch network [19]. It is shown that the total achievable SE of full digital architectures always outperforms the one of hybrid architectures, regardless of ideal phase shifters, switch network, and DFT precoding. However, the total achievable SE of the hybrid architectures and DFT precoding is superior to that of the hybrid architectures with the ideal phase shifters, but it is inferior to that of the hybrid architectures with the ideal phase shifters. As expected, there is a significant performance gap that is caused by inaccurate CSI and randomly DFT selection. According to (9), we know

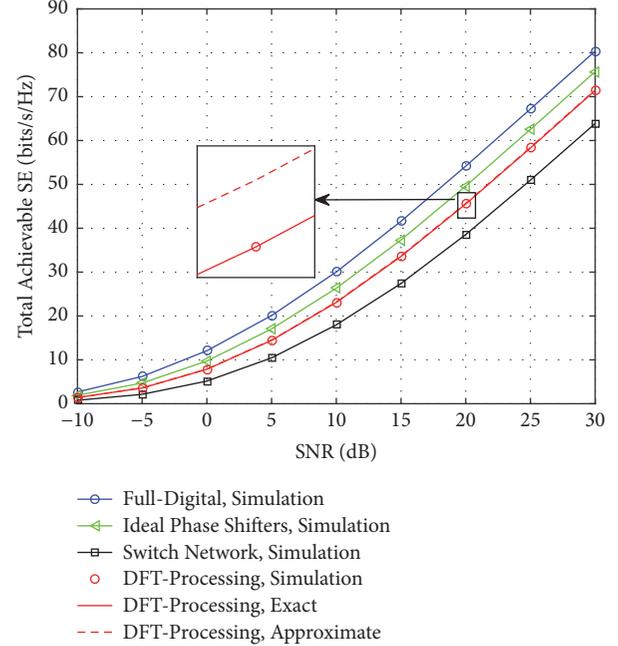


FIGURE 2: Total achievable SE versus the input SNR for different architectures ($N_t = 100$, $M = 8$, and $N_s = 50$).

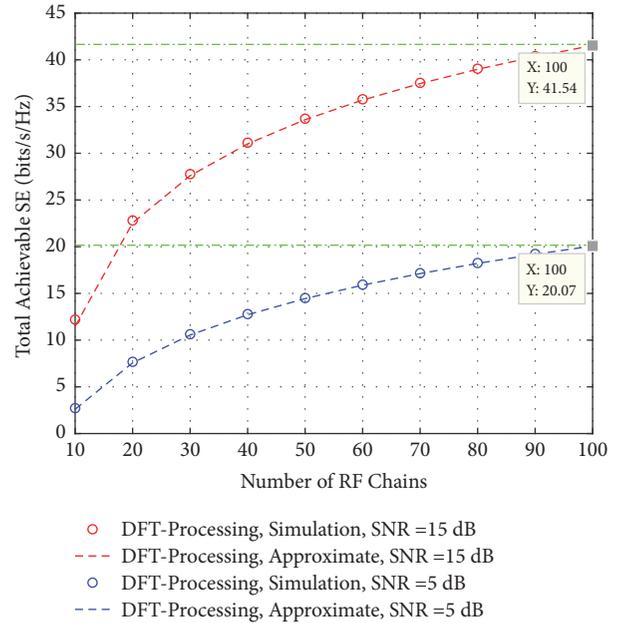


FIGURE 3: Total achievable SE versus the number of RF chains ($N_t = 100$, $M = 8$, and SNR = 5 dB, 15 dB).

that the performance gap can be compensated for offering the more input SNR and the more number of RF chains.

In Figure 3, we depict the total achievable SE versus the number of RF chains. As expected, the total achievable SE increases with the number of RF chains, which keeps pace with theoretical results in Corollary 3. For comparison, the total achievable SE for the full digital architectures is

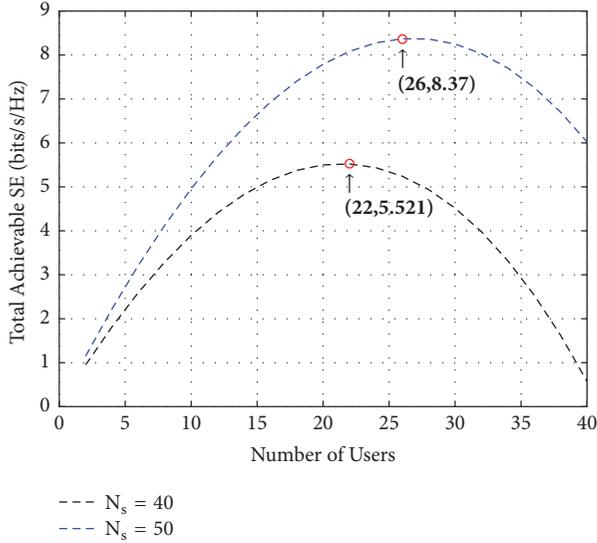


FIGURE 4: Total achievable SE with DFT processing versus the number of users ($N_t = 128$, SNR = 10 dB, and $N_s = 40, 50$).

also provided using [43, Table I], respectively. We can see that, with fixed number of users, number of BS antennas, and SNR (15 dB and 5 dB), the total achievable SE for full digital architectures maintains a constant value and is 41.54 and 20.07 bit/s/Hz, respectively, which keeps constant. This is because each antenna is connected to the RF chain. In addition, only when the number of available RF chains equals the number of BS antennas ($N_s = N_t = 100$), the difference between hybrid architectures and full digital architectures tends to zero. Moreover, we observe that the total achievable SE with larger SNR regime always outperforms that one of the smaller SNR regime for the reason that the larger transmitted power boosts the total achievable SE.

Figure 4 depicts the total achievable SE varying with the number of users, in which the curves only showed the hybrid architecture with DFT processing. To guarantee fairness among users, we assume that the large-scale fading of each user is set to be identical, namely, $\beta_k = 1 \times 10^{-3}, \forall M$. Moreover, given the number of BS antennas and the input SNR ($N_t = 128$ and SNR = 10 dB), we observe that the total achievable SE increases first and then decreases as the number of users increases from 2 to 40. Meanwhile, we depict the total achievable SE for different numbers of RF chains ($N_s = 40, 50$); no matter what the number of RF chains is utilized, a globally optimal number of users always exists that achieves the total achievable SE optimum, which keeps pace with theoretical result in Corollary 4. In addition, keeping the parameters N_s , P , and β constant and plugging these parameters into (B.1) and (12), the best number of users and corresponding total achievable SE of system is given by (22, 5.521) and (26, 8.370), respectively, which further support our theoretical result in Corollary 4.

In Figures 5 and 6, we compare the total achievable EE for a hybrid massive MIMO system based on DFT processing, ideal phase shifters, and switch network, as well as massive MIMO systems with full digital architecture. Before depicting

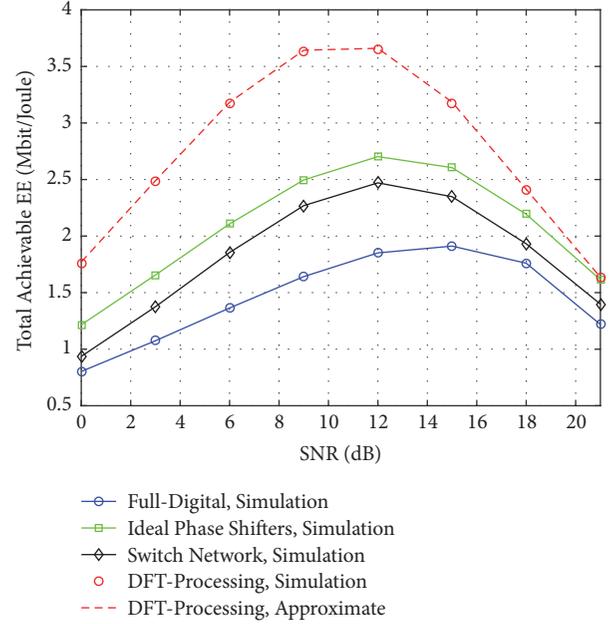


FIGURE 5: Total achievable EE versus SNR for different architectures ($N_t = 128$, $N_s = 50$, and $M = 8$).

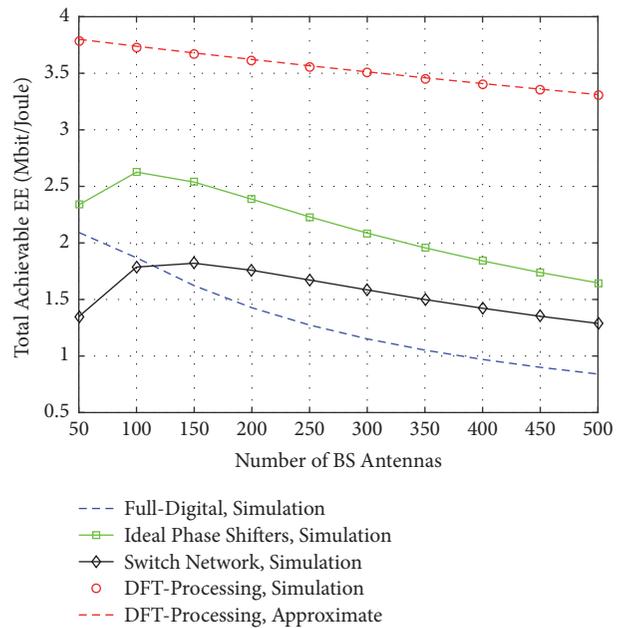


FIGURE 6: Total achievable EE versus the number of BS antennas for architectures (SNR = 10 dB, $N_s = 50$, and $M = 8$).

the total achievable EE, we shall present the energy consumption for full digital architecture, hybrid architectures with ideal phase shifters, and switch network hybrid architectures because the energy consumption of hybrid architectures with DFT processing has been provided in Section 3. With the help of the energy consumption models that have been provided in [19, 37, 44], we know that, for full digital architecture, hybrid architectures with ideal phase shifters, and switch network hybrid architectures, the energy consumption of the

signal processing part and the signal transmission part is the same as the hybrid architectures with DFT processing, and only the energy consumption of the circuit part has changed drastically. For full digital architecture, the energy consumption of the circuit part is provided by [44]

$$P_{\text{Full}} = N_t P_{\text{BS}} + P_{\text{SYS}} + MP_{\text{UE}}, \quad (22)$$

where $P_{\text{BS}} = 1$ Watt, $P_{\text{SYS}} = 2$ Watt, and $P_{\text{UE}} = 0.3$ Watt. For a hybrid architecture massive MIMO system based on ideal phase shifters and switch network, the energy consumption of circuit part is provided by [19]

$$P_{**} = N_t (K + 1) P_{\text{LAN}} + N_t MP_{**} + MP_{\text{RF}}, \quad (23)$$

where the subscript ** represents two different cases of the ideal phase shifter and switch network. The detailed energy consumption is provided by $P_{\text{LAN}} = 20 \times 10^{-3}$ Watt, $P_{\text{PS}} = 30 \times 10^{-3}$ Watt, $P_{\text{SW}} = 5 \times 10^{-3}$ Watt, $P_{\text{FR}} = 40 \times 10^{-3}$ Watt, $P_{\text{cod}} = 4$ Watt, and $P_{\text{dec}} = 500 \times 10^{-3}$ Watt, and assume that the efficiency of transmitter power is $\eta = 0.5$ at the BS. These above parameters are obtained from the results of [19, 37] for hybrid architectures.

Figure 5 displays the total achievable EE as a function of the input SNR. These curves are shown for hybrid architectures based on DFT processing, ideal phase shifters, and switch network. It is shown that the total achievable EE of hybrid architectures based on DFT processing is always superior to full digital architectures, as well as hybrid architectures ideal phase shifters and switch network. This is because the DFT processing has low cost and uses a few number of RF chains. However, the total achievable EE of the architectures with ideal phase shifters is better than that of the architectures with switch network for the reason that the degradation of the achievable SE with switch network as compared with the ideal phase shifters case is significant, especially for switch network case. Furthermore, the total achievable EE of hybrid architectures with DFT processing invariably outperforms the other in the input SNR regime, which indicates that the DFT processing is the best one, because the energy consumption of DFT processing is very small. In addition, it is shown that the total achievable EE first improves the input SNR and then reduces while the input SNR increases. Indeed, there exists the best input SNR that makes the total achievable EE of system maximization, which is consistent with analytical results in Corollary 5.

Figure 6 depicts the total achievable EE versus the number of BS antennas and includes full digital architecture and hybrid architectures with DFT processing, ideal phase shifters, and switch network. It is shown that the total achievable EE of hybrid architectures with DFT processing decreases by increasing the number of BS antennas. This is because keeping the number of RF chains and the SNR fixed, the total achievable SE becomes a constant, but the total energy consumption increases by increasing the number of BS antennas. Finally, this leads to the total achievable EE of hybrid architectures with DFT processing decreases. Obviously, the total achievable EE with DFT processing outperforms that of other hybrid architectures. For the hybrid architectures with ideal phase shifters and switch

network, it is shown that the best number of BS antennas exists that makes the total achievable EE optimum, which is consistent with the conclusion described in [45]. For full digital architectures, it is shown that as the number of BS antennas increases, the total achievable EE decreases. This is because each antenna needs to deploy a unique RF chain, which is augmented by increasing the number of BS antennas.

5. Conclusion

In this paper, low cost and high efficiency hybrid massive MIMO system has been studied, in which the analog domain uses DFT processing and the digital domain utilizes ZF precoding. We derived the exact expressions on the achievable SE. Results showed that compared with full digital architecture and hybrid architecture with ideal phase shifters, there is some degradation of achievable SE for hybrid architecture with DFT processing. Fortunately, the degraded achievable SE can be compensated by providing the more number of RF chains and the input SNR. Moreover, we also studied the total achievable EE and derived exact expression of the achievable EE. Compared to conventional massive MIMO system, regardless of full digital and hybrid architecture, results showcased that the total achievable EE of hybrid architectures with DFT processing invariably outperforms the full digital architectures and hybrid architectures with switch network and ideal phase shifters. We also found out that the achievable EE decreases as the number of antennas chains increases, while the best SNR exists that realizes the total achievable EE optimum. We believe that the hybrid massive MIMO systems with DFT processing have low cost and high efficiency for 5G communication system. In the future, we will incorporate some other wireless techniques such as [46–49] to further enhance the performance of the massive MIMO system.

Appendix

A. Proof of Theorem 2

To calculate the exact expression on the achievable SE in (5), we shall recall that the definition of $\mathbf{G}_{\text{eq}} = \mathbf{G}^H \mathbf{F}$ with $\mathbf{G} = \mathbf{H}\mathbf{D}^{1/2}$. By plugging these results into (8), the achievable ergodic SE in (8) is expressed as

$$R_k = \mathbb{E} \left\{ \log_2 \left(1 + \frac{P\beta_k}{[\mathbf{H}^H \mathbf{A}\mathbf{H}]_{k,k}^{-1}} \right) \right\}. \quad (\text{A.1})$$

For convenience, we firstly define an auxiliary matrix \mathbf{A} , which is identical to $\mathbf{A} = \mathbf{F}\mathbf{F}^H$. According to the properties of the DFT matrix, we easily know that the column vectors of the DFT matrix are mutually orthogonal, and then the auxiliary matrix \mathbf{A} is transformed by the eigen-decomposition, which can be written as

$$\mathbf{A} = \mathbf{F}\mathbf{F}^H = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H, \quad (\text{A.2})$$

where $\mathbf{\Lambda}$ is the diagonal matrix, whose entries are one or zero. Since the partial matrix \mathbf{F} is a $N_t \times N_s$ dimension of

DFT, by applying the eigen-decomposition $\mathbf{\Lambda}$ is a diagonal matrix having N_s unit singular values, whose diagonal entries can be expressed as $\mathbf{\Lambda} = \text{diag}\{1, \dots, 1, 0, \dots, 0\}$. By applying the unitary transformation, the matrix $\mathbf{H}^H \mathbf{A} \mathbf{H}$ has the same distribution form as $\mathbf{T}^H \mathbf{T}$, where \mathbf{T} reduces to a complex Gaussian matrix of $M \times N_s$ dimension and is equivalent to N_s columns of \mathbf{H} . Thus, (A.1) can be written as

$$R_k = \mathbb{E} \left\{ \log_2 \left(1 + \frac{P\beta_k}{[\mathbf{T}^H \mathbf{T}]_{k,k}^{-1}} \right) \right\}. \quad (\text{A.3})$$

We define

$$X_k = \frac{1}{[(\mathbf{T}^H \mathbf{T})]_{k,k}^{-1}}. \quad (\text{A.4})$$

Thus, (A.3) can be rewritten as

$$R_k = \mathbb{E} \{ \log_2 (1 + P\beta_k X_k) \}. \quad (\text{A.5})$$

We study the evaluation of R_k according to the following expression:

$$R_k = \log_2(e) \int_0^\infty \ln(1 + P\beta_k x_k) p(x_k) dx_k. \quad (\text{A.6})$$

According to random matrix theory, when the entries of small-scale fading \mathbf{T} are i.i.d. Rayleigh random variances [50–52], the p.d.f. of X_k is given by

$$p(x_k) = \frac{e^{-x_k}}{(N_s - M)!} x_k^{N_s - M}. \quad (\text{A.7})$$

Substituting (A.7) into (A.6) and employing the following integration identity:

$$\begin{aligned} & \int_0^\infty \ln(1 + a\lambda) \lambda^{q-1} e^{-b\lambda} dy \\ &= (q-1)! e^{b/a} b^{-q} \sum_{h=1}^q E_h \left(\frac{b}{a} \right). \end{aligned} \quad (\text{A.8})$$

By doing more basic mathematical manipulations, we finish the detailed proof.

B. Proof of Corollary 4

For fairness among users, we assume that the large-scale fading of each user is identical, namely, $\beta_k = \beta, \forall M$. According to the achievable SE of per user (9), the total achievable SE of system can be expressed as

$$\bar{R}_{\text{sum}} = M \log_2 (1 + P\beta (N_s - M + 1)). \quad (\text{B.1})$$

We differentiate the total achievable SE in (B.1) in terms of random variable M , the first-order partial derivative of $\bar{R}_{\text{sum}}(M)$ is calculated as

$$\begin{aligned} & \frac{\bar{R}_{\text{sum}}(M)}{\partial M} \\ &= \frac{(\ln(1 + P\beta (N_s - M + 1)) - P\beta M / (1 + P\beta (N_s - M + 1)))}{\ln 2}. \end{aligned} \quad (\text{B.2})$$

Since the second term is negative, it is hard to judge directly if the value of (B.2) is negative or positive, and we need to further solve the second order partial derivative of $\bar{R}_{\text{sum}}(M)$. The detailed computation and associated analysis of second-order partial can be found in the works [53–55]. The second-order partial derivative of $\bar{R}_{\text{sum}}(M)$ is calculated as

$$\begin{aligned} \frac{\bar{R}_{\text{sum}}^2(M)}{\partial M^2} &= \left(\frac{-P\beta M}{1 + P\beta (N_s - M + 1)} \right. \\ & \left. + \frac{-P\beta (N_s + P + 1)}{(1 + P\beta (N_s - M + 1))^2} \right). \end{aligned} \quad (\text{B.3})$$

We found that $\bar{R}_{\text{sum}}^2(M)/\partial M^2 < 0$ due to the value of numerator is a negative; this shows that the function $\bar{R}_{\text{sum}}(M)$ is concave with respect to M . According to convex optimization theory, it is easy to know that a unique globally optimal number of users always exist, which enables us to achieve the total achievable SE maximization. To check the optimal value, let the first-order partial derivative be equal to zero that holds

$$M^{\text{opt}} = \left\{ M \mid \frac{\partial \bar{R}_{\text{sum}}(M)}{\partial M} = 0 \right\}. \quad (\text{B.4})$$

Plugging (B.2) into (B.4), we attain

$$\ln(1 + P\beta (N_s - M + 1)) = \frac{P\beta}{1 + P\beta (N_s - M + 1)}. \quad (\text{B.5})$$

In order to solve the above equation in terms of random variable M , we observe that the equation is so sophisticated because it involves the logarithmic function. We start by letting $y = 1 + P\beta (N_s - M + 1)$; (B.5) can be simplified as

$$\begin{aligned} & (1 + P\beta (N_s + 1)) \exp(1) \\ &= \frac{1 + P\beta (N_s + 1)}{y} \cdot \exp\left(\frac{1 + P\beta (N_s + 1)}{y}\right). \end{aligned} \quad (\text{B.6})$$

By applying the properties of the Lambert function [41] that holds $x = W_0(x) \exp(W_0(x))$, we can attain the following equation:

$$y = \frac{1 + P\beta (N_s + 1)}{W_0(1 + P\beta (N_s + 1) \exp(1))}. \quad (\text{B.7})$$

Substituting y into (B.7) and playing some basic mathematical transformations, we finish the proof.

Data Availability

The authors derived the writing material from different journals as provided in the references. A MATLAB tool has been utilized to simulate the concept.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported in part by the Natural Science Foundation of Guangdong Province of China under Grant 2018A030310338, the National Natural Science Foundation of China under Grants 61671144, 61801132, and 61871139, the Project of Educational Commission of Guangdong Province of China under Grant 2017KQNCX155, the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2014A030306027, the Innovation Team Project of Guangdong Province University under Grant 2016KCXTD017, the Science and Technology Program of Guangzhou under Grant 201807010103, the Project of Fundamental Science and Frontier Technology Research Project of Chongqing under Grant cstc2017jcyjAX0380, the Guangdong Science and Technology Project under Grant 2017A050506058, the Major Research Project of Guangdong Provincial Department of Education under Grant 2016KZDXM035, the Innovation Group Project of Guangzhou under Grant 1201610010, and the Yangcheng Scholars Project of Guangzhou under Grant 1201561560.

References

- [1] F. Zhou, L. Fan, X. Lei, G. Luo, H. Zhang, and J. Zhao, "Edge Caching With Transmission Schedule for Multiuser Multirelay Networks," *IEEE Communications Letters*, vol. 22, no. 4, pp. 776–779, 2018.
- [2] W. Tan, W. Huang, X. Yang, W. Liu, and L. Fan, "Multiuser precoding scheme and achievable rate analysis for massive," *EURASIP Journal on Wireless Communications and Networking*, 2018.
- [3] B. Liu, Q. Zhu, W. Tan, and H. Zhu, "Congestion-Optimal WiFi Offloading with User Mobility Management in Smart Communications," *Wireless Communications and Mobile Computing*, vol. 2018, pp. 1–15, 2018.
- [4] H. Huang, J. Xia, X. Liu et al., "Switch-and-stay combining for energy harvesting relaying systems," *Physical Communication*, vol. 28, pp. 28–34, 2018.
- [5] F. Rusek, D. Persson, B. K. Lau et al., "Scaling up MIMO: opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.
- [6] F. Zhou, M. Du, Y. Wang, and G. Luo, "Joint source-channel coding for band-limited backhauls in coordinated multi-point systems," *IET Communications*, vol. 10, no. 13, pp. 1562–1570, 2016.
- [7] R. Zhao, Y. Yuan, L. Fan, and Y.-C. He, "Secrecy Performance Analysis of Cognitive Decode-and-Forward Relay Networks in Nakagami-m Fading Channels," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 549–563, 2017.
- [8] J. Yuan, S. Jin, W. Xu, W. Tan, M. Matthaiou, and K.-K. Wong, "User-Centric Networking for Dense C-RANs: High-SNR Capacity Analysis and Antenna Selection," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 5067–5080, 2017.
- [9] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: benefits and challenges," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 136–146, 2014.
- [10] J. Xia, F. Zhou, X. Lai et al., "Cache Aided Decode-and-Forward Relaying Networks: From the Spatial View," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 5963584, 9 pages, 2018.
- [11] X. Lai, J. Xia, M. Tang, H. Zhang, and J. Zhao, "Cache-aided multiuser cognitive relay networks with outdated channel state information," *IEEE Access*, vol. 6, pp. 21879–21887, 2018.
- [12] W. Tan, S. Jin, C. Wen, and T. Jiang, "Spectral efficiency of multi-user millimeter wave systems under single path with uniform rectangular arrays," *EURASIP Journal on Wireless Communications and Networking*, vol. 181, pp. 1–13, 2017.
- [13] H. Shuangfeng, I. Chih-Lin, X. Zhikun, and C. Rowell, "Large-scale antenna systems with hybrid analog and digital beamforming for millimeter wave 5G," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 186–194, 2015.
- [14] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.
- [15] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831–846, 2014.
- [16] Chanhong K., Taeyoung K., and Ji-Yun S., "Multi-beam transmission diversity with hybrid beamforming for MIMO-OFDM systems," in *Proceedings of the 2013 IEEE Globecom Workshops (GC Wkshps)*, pp. 61–65, Atlanta, GA, December 2013.
- [17] Z. Song, Z. Zhang, X. Liu, Y. Liu, and L. Fan, "Simultaneous cooperative spectrum sensing and wireless power transfer in multi-antenna cognitive radio," *Physical Communication*, vol. 29, pp. 78–85, 2018.
- [18] R. Zi, X. Ge, J. Thompson, C.-X. Wang, H. Wang, and T. Han, "Energy Efficiency Optimization of 5G Radio Frequency Chain Systems," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 758–771, 2016.
- [19] R. Mendez-Rial, C. Rusu, N. Gonzalez-Prelcic, A. Alkhateeb, and R. W. Heath, "Hybrid MIMO Architectures for Millimeter Wave Communications: Phase Shifters or Switches?" *IEEE Access*, vol. 4, pp. 247–267, 2016.
- [20] A. Garcia-Rodriguez, V. Venkateswaran, P. Rulikowski, and C. Masouros, "Hybrid Analog-Digital Precoding Revisited Under Realistic RF Modeling," *IEEE Wireless Communications Letters*, vol. 5, no. 5, pp. 528–531, 2016.
- [21] J. Butler and R. Lowe, "Beam forming matrix simplifies design of electronically scanned antennas," *Electron. Design*, pp. 170–173, 1961.
- [22] H. A. Darwish and I. Hartimo, "A close accord on DFT based frequency and phasor estimators for numerical relays," *Materials Research Bulletin*, pp. 29–36, 2001.
- [23] H. A. Darwish and M. Fikri, "Practical considerations for recursive DFT implementation in numerical relays," *IEEE Transactions on Power Delivery*, vol. 22, no. 1, pp. 42–49, 2007.
- [24] W. Tan, M. Matthaiou, S. Jin, and X. Li, "Spectral Efficiency of DFT-Based Processing Hybrid Architectures in Massive MIMO," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 586–589, 2017.
- [25] D. Deng, C. Li, L. Fan, X. Liu, and F. Zhou, "Impact of Antenna Selection on Physical-Layer Security of NOMA Networks," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 2390834, 11 pages, 2018.
- [26] F. Shi, L. Fan, X. Liu, Z. Na, and Y. Liu, "Probabilistic Caching Placement in the Presence of Multiple Eavesdroppers," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 2104162, 10 pages, 2018.

- [27] D. Deng, M. Yu, J. Xia, Z. Na, J. Zhao, and Q. Yang, "Wireless powered cooperative communications with direct links over correlated channels," *Physical Communication*, vol. 28, pp. 147–153, 2018.
- [28] L. Liang, W. Xu, and X. Dong, "Low-complexity hybrid precoding in massive multiuser MIMO systems," *IEEE Wireless Communications Letters*, vol. 3, no. 6, pp. 653–656, 2014.
- [29] S. Suh, A. Basu, C. Schlottmann, P. E. Hasler, and J. R. Barry, "Low-power discrete Fourier transform for OFDM: a programmable analog approach," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 58, no. 2, pp. 290–298, 2011.
- [30] A. F. Molisch, X. Zhang, S. Y. Kung, and J. Zhang, "DFT-based hybrid antenna selection schemes for spatially correlated MIMO channels," in *Proceedings of the 14th IEEE 2003 International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC2003*, pp. 1119–1123, China, September 2003.
- [31] H. Yang and T. L. Marzetta, "Performance of conjugate and zero-forcing beamforming in large-scale antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 172–179, 2013.
- [32] Y.-G. Lim, C.-B. Chae, and G. Caire, "Performance Analysis of Massive MIMO for Cell-Boundary Users," *IEEE Transactions on Wireless Communications*, vol. 14, no. 12, pp. 6827–6842, 2015.
- [33] M. Vu and A. Paulraj, "MIMO wireless linear precoding," *IEEE Signal Processing Magazine*, vol. 24, no. 5, pp. 86–105, 2007.
- [34] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, New York: Dover, 1974.
- [35] C. Li, Y. Li, K. Song, and L. Yang, "Energy efficient design for multiuser downlink energy and uplink information transfer in 5G," *Science China Information Sciences*, vol. 59, no. 2, pp. 1–8, 2016.
- [36] C. Li, K. Song, D. Wang, F.-C. Zheng, and L. Yang, "Optimal remote radio head selection for cloud radio access networks," *Science China Information Sciences*, vol. 59, no. 10, pp. 59–73, 2016.
- [37] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3059–3075, 2015.
- [38] J. Li, M. Wen, X. Jiang, and W. Duan, "Space-Time Multiple-Mode Orthogonal Frequency Division Multiplexing with Index Modulation," *IEEE Access*, vol. 5, pp. 23212–23222, 2017.
- [39] C. Kong, C. Zhong, M. Matthaiou, and Z. Zhang, "Performance of downlink massive MIMO in rician fading channels with ZF precoder," in *Proceedings of the IEEE International Conference on Communications, ICC 2015*, pp. 1776–1782, UK, June 2015.
- [40] E. Björnson, L. Sanguinetti, J. Hoydis, and M. Debbah, "Designing multi-user MIMO for energy efficiency: When is massive MIMO the answer?" in *Proc. IEEE WCNC*, pp. 56–64, 2014.
- [41] R. M. Corless, G. H. Gonnet, D. E. G. Hare, and D. E. Knuth, "On the Lambert W function," *Advances in Computational Mathem.*, *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, 1996.
- [42] Y. Lv, H. Wu, Y. Liu et al., "Quantitative research on the influence of particle size and filling thickness on aerogel glazing performance," *Energy and Buildings*, vol. 174, pp. 190–198, 2018.
- [43] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Transactions on Communications*, vol. 61, no. 4, pp. 1436–1449, 2013.
- [44] X. Gao, L. Dai, S. Han, I. Chih-Lin, and R. W. Heath, "Energy-efficient hybrid analog and digital precoding for MmWave MIMO systems with large antenna arrays," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 998–1009, 2016.
- [45] C. Sun, X. Gao, and Z. Ding, "Spectral efficiency and power allocation for mixed-ADC massive MIMO system," *IEEE Transactions on Signal Processing*, vol. 65, no. 11, pp. 2962–2974, 2017.
- [46] W. Zhou, D. Deng, J. Xia, and Z. Shao, "The Precoder Design with Covariance Feedback for Simultaneous Information and Energy Transmission Systems," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 8472186, 17 pages, 2018.
- [47] D. Deng, X. Li, L. Fan, W. Zhou, R. Qingyang Hu, and Z. Zhou, "Secrecy Analysis of Multiuser Untrusted Amplify-and-Forward Relay Networks," *Wireless Communications and Mobile Computing*, vol. 2017, Article ID 9580639, 11 pages, 2017.
- [48] X. Liu, X. Zhang, M. Jia, L. Fan, W. Lu, and X. Zhai, "5G-based green broadband communication system design with simultaneous wireless information and power transfer," *Physical Communication*, vol. 28, pp. 130–137, 2018.
- [49] L. Fan, R. Zhao, F.-K. Gong, N. Yang, and G. K. Karagiannidis, "Secure Multiple Amplify-and-Forward Relaying over Correlated Fading Channels," *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 2811–2820, 2017.
- [50] X. Wang, H. Zhang, L. Fan, and Y. Li, "Performance of Distributed Switch-and-Stay Combining for Cognitive Relay Networks with Primary Transceiver," *Wireless Personal Communications*, vol. 97, no. 2, pp. 3031–3042, 2017.
- [51] Z. Na, Y. Wang, X. Li et al., "Subcarrier allocation based Simultaneous Wireless Information and Power Transfer algorithm in 5G cooperative OFDM communication systems," *Physical Communication*, vol. 29, pp. 164–170, 2018.
- [52] M. Zhao, D. Deng, W. Zhou, and L. Fan, "Non-renewable energy efficiency optimization in energy harvesting relay-assisted system," *Physical Communication*, vol. 29, pp. 183–190, 2018.
- [53] Y. Liang, H. Wu, G. Huang, J. Yang, and H. Wang, "Thermal performance and service life of vacuum insulation panels with aerogel composite cores," *Energy and Buildings*, vol. 154, pp. 606–617, 2017.
- [54] Y. Jian, H. Wu, G. Huang, Y. Liang, and Y. Liao, "Modeling and coupling effect evaluation of thermal conductivity of ternary opacifier/fiber/aerogel composites for super-thermal insulation," *Materials Design*, vol. 133, no. 5, pp. 224–236, 2017.
- [55] J. Yang, H. Wu, M. Wang, and Y. Liang, "Prediction and optimization of radiative thermal properties of nano TiO₂ assembled fibrous insulations," *International Journal of Heat and Mass Transfer*, vol. 117, pp. 729–739, 2018.

Research Article

AN-Aided Transmit Beamforming Design for Secured Cognitive Radio Networks with SWIPT

Weili Ge ¹, Zhengyu Zhu ¹, Zhongyong Wang ¹ and Zhengdao Yuan²

¹School of Information Engineering, Zhengzhou University, Zhengzhou, China

²Zhengzhou Information Science and Technology Institute, Zhengzhou, China

Correspondence should be addressed to Zhengyu Zhu; zhuzhengyu6@gmail.com and Zhongyong Wang; zywangzuz@gmail.com

Received 6 April 2018; Revised 25 May 2018; Accepted 7 June 2018; Published 13 August 2018

Academic Editor: Zan Li

Copyright © 2018 Weili Ge et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We investigate multiple-input single-output secured cognitive radio networks relying on simultaneous wireless information and power transfer (SWIPT), where a multiantenna secondary transmitter sends confidential information to multiple single-antenna secondary users (SUs) in the presence of multiple single-antenna primary users (PUs) and multiple energy-harvesting receivers (ERs). In order to improve the security of secondary networks, we use the artificial noise (AN) to mask the transmit beamforming. Optimization design of AN-aided transmit beamforming is studied, where the transmit power of the information signal is minimized subject to the secrecy rate constraint, the harvested energy constraint, and the total transmit power. Based on a successive convex approximation (SCA) method, we propose an iterative algorithm which reformulates the original problem as a convex problem under the perfect channel state information (CSI) case. Also, we give the convergence of the SCA-based iterative algorithm. In addition, we extend the original problem to the imperfect CSI case with deterministic channel uncertainties. Then, we study the robust design problem for the case with norm-bounded channel errors. Also, a robust SCA-based iterative algorithm is proposed by adopting the \mathcal{S} -Procedure. Simulation results are presented to validate the performance of the proposed algorithms.

1. Introduction

The fifth-generation (5G) wireless technology is expected to satisfy an increasing demand for wireless device, such as high data service and radio coverage [1, 2]. However, the explosive increase of mobile terminal has resulted in severe scarcity of radio spectrum resources, which has become an outstanding problem [3–5]. As one of the most efficient ways to alleviate the problem of spectrum scarcity for green communications and networks, cognitive radio (CR) is a promising approach for heightening the spectrum utilization ratio [6, 7]. In CR networks, so long as the interference produced by the secondary transmitter (ST) is tolerable to each primary user (PU), a secondary user (SU) can employ the licensed spectrum of the primary system [8]. Although CR technology can significantly increase in the spectrum efficiency, the energy scarcity still give rise to a major bottleneck problem for the quality of service (QoS) and the long lifetime of wireless users [9].

Recently, a promising technique, simultaneous wireless information and power transfer (SWIPT), has been proposed to use the radio frequency- (RF-) enabled signal harvesting power in wireless networks, which can greatly help to solve the bottleneck problem of energy scarcity [10–18]. Compared to the conventional energy harvesting techniques, such as solar power and wind, there is an advantage for SWIPT in providing more stable and controllable energy to portable wireless devices [10]. For the SWIPT operation, the work [16] examined colocated receivers which employ a power splitter for energy harvesting (EH) and information decoding (ID). Therefore, it is really important to research the combining SWIPT with CR, which has the dual function of improving both the energy efficiency and spectrum efficiency, and it has attracted more attention in [19–28]. In SWIPT-based CR system, the energy efficiency optimization was considered in [19, 20]. Moreover, the SWIPT approach for CR schemes was investigated in cooperative relay networks [21], a thresholding-based antenna selection multiple-input

multiple-output (MIMO) systems [22], multiuser MISO system [23–25, 27], cooperative nonorthogonal multiple access networks [26], and sensing-Based wideband CR [28].

On the other hand, secrecy transmission has gained attentions in communication systems [29]. Unlike traditional cryptographic methods which are normally adopted in the network layer, physical layer security was developed from information theoretical aspects to improve the secrecy capacity of wireless transmission systems [30]. In the conventional SWIPT systems, since the energy-harvesting receivers (ERs) are normally assumed to be closer to the transmitter compared with information receivers (IRs), this results in a new information security issue. In such a situation, ERs have a possibility of eavesdropping the information sent to the IRs and thus can become potential eavesdroppers [31]. As a result, physical layer security has been recognized as an important issue for SWIPT systems [32]. Moreover, a few techniques have been proposed for multiple antenna secrecy systems to cause more interference to eavesdroppers [31, 33]. In SWIPT operation, artificial noise (AN) was embedded in the transmit beamforming signal to confuse the eavesdroppers and harvest power simultaneously [33].

In addition, due to the inherent characteristics of CR with SWIPT, ERs may illegitimately access the PU bands and change the radio environment. In this case, the legitimate SU is unable to use frequency bands of the PU. Thus, in order to satisfy secure communication and EH requirement, the security of CR SWIPT with ERs is also of great importance [34–38]. In [34], an outage-constrained secrecy rate maximization (SRM) problem has been investigated in an underlay MIMO CR network where the secondary transmitter (ST) provides SWIPT to all receivers. In order to guarantee secure communication and energy harvesting in MISO CR network with SWIPT, [35] studied a robust secure AN-aided beamforming and power splitting (PS) design under imperfect channel state information (CSI). For a CR network consisting of a PS-based to decode information and harvest energy simultaneously, [36] analyzed the behavior of the nodes by using game theoretic techniques and proved the existence of a unique Nash equilibrium (NE) strategy. Considering a system with ERs acting as potential eavesdroppers in CR-based SWIPT, [37] has derived a closed-form analytical expression for the exact secrecy outage probability. In a CR MIMO-SWIPT broadcast channel, [38] studied the SRM problem by designing secrecy AN-aided precoding.

Motivated by the above observations, in this paper, we study secrecy transmission over a MISO CR system with SWIPT, which consists of one primary transmitter-receiver pair (denoted as PT and PU), one multiantenna ST, multiple SUs, and multiple ERs (potentially eavesdropper). Under the perfect CSI case, we formulate a transmit power of the information signal minimization (TPISM) problem subject to the secrecy rate constraint for SUs and PU, the harvested energy constraint for ERs, and total transmit power constraint. We seek to jointly design strategies of secure beamforming and AN for MISO CR with SWIPT wiretap channels. Then, the framework is extended to robust designs for the imperfect CSI case by adopting deterministic CSI uncertainties. The main contributions for this paper are summarized as follows:

- (i) For the two different types of channel models, the secrecy rate constraint with linear fractional programming is equivalently converted into several constraints with exponential form by introducing exponential variables.
- (ii) For the perfect CSI case at ST and PT, unlike the conventional semidefinite relaxation (SDR), we propose a novel reformulation of the TPISM problem. A successive convex approximation- (SCA-) based iterative algorithm is proposed, where the nonconvex constraint is approximated as a convex one.
- (iii) For the imperfect CSI case, we use the norm-bounded channel uncertainty to model channel. First, by applying the triangle inequality, the original constraints can be transformed as the infinitely inequality constraints. Then, by employing \mathcal{S} -Procedure we convert these infinitely inequality constraints into finite linear matrix inequalities (LMIs). At last, by utilizing a SCA method, the recast TPISM problem is transformed to a semidefinite programming (SDP) problem, which can be directly solved to obtain a local optimal solution. A robust SCA-based iterative algorithm is proposed.

The rest of this paper is organized as follows: Section 2 presents the system model and problem formulation. In Section 3, the SCA-based iterative algorithm is proposed with perfect CSI case. Robust SCA-based iterative algorithm is developed with imperfect CSI case in Section 4. Section 5 illustrates the simulation results. Finally, we conclude the paper in Section 6.

Notation. Vectors and matrices are denoted by bold lowercase and uppercase letters, respectively. $(\cdot)^H$ represents the Hermitian transpose. For a vector \mathbf{x} , $\|\mathbf{x}\|$ indicates the Euclidean norm. $|\cdot|$ defines the absolute value of a complex scalar. $\mathbb{C}^{M \times L}$ and $\mathbb{H}^{M \times L}$ describe the space of $M \times L$ complex matrices and Hermitian matrices, respectively. For a matrix \mathbf{A} , $\mathbf{A} \geq \mathbf{0}$ means that \mathbf{A} is positive semidefinite, and $\text{tr}(\mathbf{A})$ indicate the trace, respectively. $\mathbb{E}\{\cdot\}$ describes the mathematical expectation. $\Re\{\cdot\}$ stands for the real part of a complex number. $[x]^+$ equals $\max\{x, 0\}$. \mathbf{I} denote the identity matrix with appropriate size.

2. System Model and Problem Formulation

In this section, we consider a MISO secured SWIPT cognitive radio networks, which consists of one PT, one PU, one multiantenna ST, K SU, and L ERs, as shown in Figure 1. We assume that the ST is equipped with N_S transmit antennas and each ER and each SU has one receive antenna. This system operates as follows: the primary network shares their frequency spectrum with the secondary network, whereas the confidential messages from the PT and the ST are intercepted by the eavesdroppers.

In order to achieve secure transmission, the ST employs transmit beamforming with AN, which acts as interference to

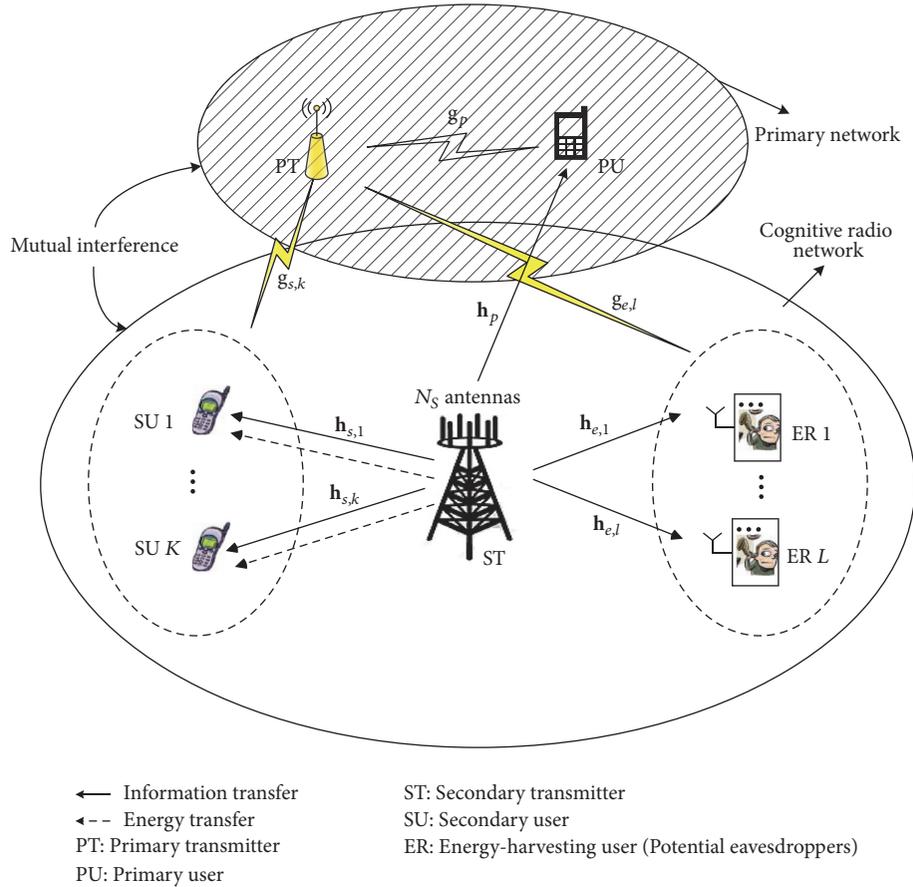


FIGURE 1: MISO Secured cognitive radio networks with SWIPT.

the ERs and provides energy to the SUs. The transmit signal vector \mathbf{x}_s from the ST can be written as

$$\mathbf{x}_s = \sum_{j=1}^K \mathbf{w}_j s_j + \mathbf{z}, \quad (1)$$

where $\mathbf{w}_j \in \mathbb{C}^{N_s}$ defines the transmit beamforming vector, s_j with $\mathbb{E}\{s_j^2\} = 1$ is the information-bearing signal intended for the SUs, and \mathbf{z} represents the energy-carrying AN.

We assume a frequency-flat slow-fading channel. Therefore, the received signal at the PU, the k th SU, and the l th ER can be given by, respectively,

$$y_p = \sqrt{P_p} g_p s_p + \mathbf{h}_p^H \mathbf{x}_s + n_p, \quad (2)$$

$$y_k = \sqrt{P_p} g_{s,k} s_p + \mathbf{h}_{s,k}^H \mathbf{x}_s + n_{s,k}, \quad k = 1, \dots, K, \quad (3)$$

$$y_{e,l} = \sqrt{P_p} g_{e,l} s_p + \mathbf{h}_{e,l}^H \mathbf{x}_s + n_{e,l}, \quad l = 1, \dots, L. \quad (4)$$

where $g_p \in \mathbb{C}$ and $\mathbf{h}_p \in \mathbb{C}^{N_s}$ are denoted by the channel between the PT and PU as well as that between the ST and PU, $g_{s,k} \in \mathbb{C}$ and $\mathbf{h}_{s,k} \in \mathbb{C}^{N_s}$ indicate the channel between the PT and the k th SU as well as that between the ST and the k th SU, $g_{e,l} \in \mathbb{C}$ and $\mathbf{h}_{e,l} \in \mathbb{C}^{N_s}$ are the channel between

the PT and the l th ER as well as that between the ST and the l th ER, s_p is confidential information-bearing signal for the PU from the PT satisfying $\mathbb{E}\{s_p^2\} = 1$, and P_p represents the transmitting power from the PT. In addition, $n_p \sim \mathcal{CN}(0, \sigma_p^2)$, $n_{s,k} \sim \mathcal{CN}(0, \sigma_{s,k}^2)$, and $n_{e,l} \sim \mathcal{CN}(0, \sigma_{e,l}^2)$ denote the complex Gaussian noise at the PU, the k th SU, and the l th ER, respectively.

Thus, the channel capacity of the k th SU can be written as

$$R_{s,k} = \log \left(1 + \frac{\mathbf{h}_{s,k}^H \mathbf{w}_k \mathbf{w}_k^H \mathbf{h}_{s,k}}{\mathbf{h}_{s,k}^H \left(\sum_{j \neq k} \mathbf{w}_j \mathbf{w}_j^H + \mathbf{Z} \right) \mathbf{h}_{s,k} + P_p |g_{s,k}|^2 + \sigma_{s,k}^2} \right). \quad (5)$$

where $\mathbf{Z} = \mathbf{z} \mathbf{z}^H$. According to [3, 25], it is assumed that the k th SU can successfully decode the information from the PT by exploiting the successive interference cancellation and, thus, (5) can be rewritten as

$$\tilde{R}_{s,k} = \log \left(1 + \frac{\mathbf{h}_{s,k}^H \mathbf{w}_k \mathbf{w}_k^H \mathbf{h}_{s,k}}{\mathbf{h}_{s,k}^H \left(\sum_{j \neq k} \mathbf{w}_j \mathbf{w}_j^H + \mathbf{Z} \right) \mathbf{h}_{s,k} + \sigma_{s,k}^2} \right). \quad (6)$$

Moreover, the channel capacity of the l th ER for decoding the desired signal of the k th SU can be represented as

$$R_{e,lk} = \log \left(1 + \frac{\mathbf{h}_{e,l}^H \mathbf{w}_k \mathbf{w}_k^H \mathbf{h}_{e,l}}{\sigma_{e,l}^2 + \mathbf{h}_{e,l}^H \left(\sum_{j \neq k} \mathbf{w}_j \mathbf{w}_j^H + \mathbf{Z} \right) \mathbf{h}_{e,l} + P_p |g_{e,l}|^2} \right). \quad (7)$$

Thus, the secrecy capacity of the k th SU can be written as

$$R_{su,k} = \left[\bar{R}_{s,k} - \max_l R_{e,lk} \right]^+, \quad \forall k, l. \quad (8)$$

In order to achieve the more reliable secrecy capacity (i.e., the minimum channel capacity at the k -th SU), (7) can be thus rewritten as

$$\bar{R}_{e,lk} = \log \left(1 + \frac{\mathbf{h}_{e,l}^H \mathbf{w}_k \mathbf{w}_k^H \mathbf{h}_{e,l}}{\sigma_{e,l}^2 + \mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l}} \right). \quad (9)$$

Then, we can obtain a lower bound of the secrecy capacity of the k th SU as follows:

$$R_{su,k}^{min} = \left[\bar{R}_{s,k} - \max_l \bar{R}_{e,lk} \right]^+, \quad \forall k, l. \quad (10)$$

Additionally, the channel capacity of the PU can be expressed as

$$R^{pu} = \log \left(1 + \frac{P_p |g_p|^2}{\mathbf{h}_p^H \left(\sum_{k=1}^K \mathbf{w}_k \mathbf{w}_k^H + \mathbf{Z} \right) \mathbf{h}_p + \sigma_p^2} \right). \quad (11)$$

The channel capacity of the l th ER for decoding the PU is given as

$$R_{e,l}^{pu} = \log \left(1 + \frac{P_p |g_{e,l}|^2}{\mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} + \sigma_{e,l}^2} \right). \quad (12)$$

Hence, the secrecy capacity of the PU under the considered worst case scenario can be written as

$$R_s^{pu} = \left[R^{pu} - \max_l R_{e,l}^{pu} \right]^+, \quad \forall l. \quad (13)$$

Moreover, the harvested power at the l th ER is expressed as, respectively,

$$E_l^e = \eta_{e,l} \left(\mathbf{h}_{e,l}^H \left(\sum_{j=1}^K \mathbf{w}_j \mathbf{w}_j^H + \mathbf{Z} \right) \mathbf{h}_{e,l} + P_p |g_{e,l}|^2 + \sigma_{e,l}^2 \right), \quad (14)$$

$\forall l,$

where $0 \leq \eta_{e,l} \leq 1$ denote the energy conversion efficiency at the l th ER.

In this paper, our aim is to minimize the transmit power of the information signal subject to the secrecy rate constraint

at the SUs, the harvested energy constraint at the ERs, and the total transmit power constraint. Based on the system models, the TPISM problem is formulated as

$$\min_{\mathbf{w}_k, \mathbf{Z}} \sum_{k=1}^K \|\mathbf{w}_k\|^2 \quad (15a)$$

$$\text{s.t.} \quad \min_k R_{su,k}^{min} \geq \bar{R}^{su}, \quad \forall k, \quad (15b)$$

$$R_s^{pu} \geq \bar{R}^{pu}, \quad \forall l, \quad (15c)$$

$$\min_l E_l^e \geq \bar{E}^e, \quad \forall l, \quad (15d)$$

$$\sum_{k=1}^K \|\mathbf{w}_k\|^2 + \text{tr}(\mathbf{Z}) \leq P, \quad \forall k, \quad (15e)$$

$$\mathbf{Z} \geq \mathbf{0}. \quad (15f)$$

where P is the total transmit power, \bar{R}^{su} and \bar{R}^{pu} denote the secrecy capacity requirement of the k th SU and the PU, and \bar{E}^e mean the harvested power requirement of the l th ER, respectively. The constraints (15b) and (15c) guarantee that the minimum secrecy rate should be achieved by the k th SU and the PU, respectively. The constraint (15d) guarantees that the minimum harvested power at the l th ER is no less than \bar{E}^e , respectively. The constraint (15e) limits the total transmit power of the ST. It is seen that problem (15a)–(15f) are nonconvex, difficult to be solved directly.

3. Proposed Algorithm with Perfect CSI

In this section, we investigate a joint design of the transmit beamforming and AN for systems under the assumption that the PT and ST can obtain perfect CSI. Problems (15a)–(15f) can be rewritten as

$$\min_{\mathbf{w}_k, \mathbf{Z}} \sum_{k=1}^K \|\mathbf{w}_k\|^2 \quad (16a)$$

$$\text{s.t.} \quad \log \left(1 + \frac{\mathbf{h}_{s,k}^H \mathbf{w}_k \mathbf{w}_k^H \mathbf{h}_{s,k}}{\mathbf{h}_{s,k}^H \left(\sum_{j \neq k} \mathbf{w}_j \mathbf{w}_j^H + \mathbf{Z} \right) \mathbf{h}_{s,k} + \sigma_{s,k}^2} \right) \quad (16b)$$

$$- \log \left(1 + \frac{\mathbf{h}_{e,l}^H \mathbf{w}_k \mathbf{w}_k^H \mathbf{h}_{e,l}}{\sigma_{e,l}^2 + \mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l}} \right) \geq \bar{R}^{su}, \quad \forall l, k,$$

$$\log \left(1 + \frac{P_p |g_p|^2}{\sum_{j=1}^K \mathbf{h}_p^H \mathbf{w}_j \mathbf{w}_j^H \mathbf{h}_p + \mathbf{h}_p^H \mathbf{Z} \mathbf{h}_p + \sigma_p^2} \right) \quad (16c)$$

$$- \log \left(1 + \frac{P_p |g_{e,l}|^2}{\mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} + \sigma_{e,l}^2} \right) \geq \bar{R}^{pu}, \quad \forall l, k,$$

$$\eta_{e,l} \left(\mathbf{h}_{e,l}^H \left(\sum_{j=1}^K \mathbf{w}_j \mathbf{w}_j^H + \mathbf{Z} \right) \mathbf{h}_{e,l} + P_p |g_{e,l}|^2 + \sigma_{e,l}^2 \right) \quad (16d)$$

$$\geq \bar{E}^e, \quad \forall l, \quad (16d)$$

$$(15e), (15f). \quad (16e)$$

In order to solve the nonconvex problems (16a)–(16e), we firstly define $\mathbf{W}_k \triangleq \mathbf{w}_k \mathbf{w}_k^H$ and denote $\mathbf{W}_Z = \sum_{j=1}^K \mathbf{w}_j \mathbf{w}_j^H + \mathbf{Z}$. Then, the constraints (16b) and (16c) can be reformulated as

$$\log \left(\frac{\mathbf{h}_{s,k}^H (\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z}) \mathbf{h}_{s,k} + \sigma_{s,k}^2 + \mathbf{h}_{s,k}^H \mathbf{W}_k \mathbf{h}_{s,k}}{\mathbf{h}_{s,k}^H (\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z}) \mathbf{h}_{s,k} + \sigma_{s,k}^2} \right) \quad (17a)$$

$$- \log \left(\frac{\sigma_{e,l}^2 + \mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} + \mathbf{h}_{e,l}^H \mathbf{W}_k \mathbf{h}_{e,l}}{\sigma_{e,l}^2 + \mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l}} \right) \geq \bar{R}^{su},$$

$$\log \left(\frac{\mathbf{h}_p^H \mathbf{W}_Z \mathbf{h}_p + \sigma_p^2 + P_p |g_p|^2}{\mathbf{h}_p^H \mathbf{W}_Z \mathbf{h}_p + \sigma_p^2} \right) \quad (17b)$$

$$- \log \left(\frac{\mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} + \sigma_{e,l}^2 + P_p |g_{e,l}|^2}{\mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} + \sigma_{e,l}^2} \right) \geq \bar{R}^{pu},$$

where they can be also rewritten as

$$\frac{(\mathbf{h}_{s,k}^H \mathbf{W}_Z \mathbf{h}_{s,k} + \sigma_{s,k}^2)(\sigma_{e,l}^2 + \mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l})}{(\mathbf{h}_{s,k}^H (\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z}) \mathbf{h}_{s,k} + \sigma_{s,k}^2)(\sigma_{e,l}^2 + \mathbf{h}_{e,l}^H (\mathbf{Z} + \mathbf{W}_k) \mathbf{h}_{e,l})} \quad (18a)$$

$$\geq 2^{\bar{R}^{su}}, \quad \forall l, k,$$

$$\frac{(\mathbf{h}_p^H \mathbf{W}_Z \mathbf{h}_p + \sigma_p^2 + P_p |g_p|^2)(\mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} + \sigma_{e,l}^2)}{(\mathbf{h}_p^H \mathbf{W}_Z \mathbf{h}_p + \sigma_p^2)(\mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} + \sigma_{e,l}^2 + P_p |g_{e,l}|^2)} \geq 2^{\bar{R}^{pu}}, \quad (18b)$$

$$\forall l, k.$$

For solving linear fractional programming (18a) and (18b), we can introduce the following exponential variables to equivalently convert. Then, we introduce slack variables x_k , y_l , t_k , $r_{l,k}$, x_p , u_l , t_p , and s_l . Equations (18a) and (18b) can be transformed as follows, respectively:

$$e^{x_k + y_l - t_k - r_{l,k}} \geq 2^{\bar{R}^{su}}, \quad (19a)$$

$$\mathbf{h}_{s,k}^H \mathbf{W}_Z \mathbf{h}_{s,k} + \sigma_{s,k}^2 \geq e^{x_k}, \quad (19b)$$

$$\sigma_{e,l}^2 + \mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} \geq e^{y_l}, \quad (19c)$$

$$\mathbf{h}_{s,k}^H \left(\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z} \right) \mathbf{h}_{s,k} + \sigma_{s,k}^2 \leq e^{t_k}, \quad (19d)$$

$$\sigma_{e,l}^2 + \mathbf{h}_{e,l}^H (\mathbf{Z} + \mathbf{W}_k) \mathbf{h}_{e,l} \leq e^{r_{l,k}}. \quad (19e)$$

$$e^{x_p + u_l - t_p - s_l} \geq 2^{\bar{R}^{pu}}, \quad (20a)$$

$$\mathbf{h}_p^H \mathbf{W}_Z \mathbf{h}_p + \sigma_p^2 + P_p |g_p|^2 \geq e^{x_p}, \quad (20b)$$

$$\mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} + \sigma_{e,l}^2 \geq e^{u_l}, \quad (20c)$$

$$\mathbf{h}_p^H \mathbf{W}_Z \mathbf{h}_p + \sigma_p^2 \leq e^{t_p}, \quad (20d)$$

$$\mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} + \sigma_{e,l}^2 + P_p |g_{e,l}|^2 \leq e^{s_l}. \quad (20e)$$

It is noted that the aforementioned constraints (19a), (19d), (19e), (20a), (20d), and (20e) are not still convex. Firstly, (19a) and (20a) can be reshaped respectively as the following convex constraints:

$$e^{-x_k - y_l + t_k + r_{l,k}} \leq 2^{-\bar{R}^{su}}, \quad \forall k, \forall l, \quad (21a)$$

$$e^{-x_p - u_l + t_p + s_l} \leq 2^{-\bar{R}^{pu}}, \quad \forall l. \quad (21b)$$

Secondly, an SCA method is used to jointly design the secure beamforming and AN matrix. Let us define $t_k(n)$, $r_{l,k}(n)$, $t_p(n)$, and $s_l(n)$ as the variables t_k , $r_{l,k}$, t_p , and s_l at the n th iteration for the SCA method. By adopting a Taylor series expansion $e^{x_i(n)}(x_i - x_i(n) + 1) \leq e^{x_i}$, we can convert the nonconvex constraints (19d), (19e), (20d), and (20e) to their corresponding convex approximations as

$$\mathbf{h}_{s,k}^H \left(\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z} \right) \mathbf{h}_{s,k} + \sigma_{s,k}^2 \quad (22a)$$

$$\leq e^{t_k(n)} (t_k - t_k(n) + 1),$$

$$\sigma_{e,l}^2 + \mathbf{h}_{e,l}^H (\mathbf{Z} + \mathbf{W}_k) \mathbf{h}_{e,l} \leq e^{r_{l,k}(n)} (r_{l,k} - r_{l,k}(n) + 1), \quad (22b)$$

$$\mathbf{h}_p^H \mathbf{W}_Z \mathbf{h}_p + \sigma_p^2 \leq e^{t_p(n)} (t_p - t_p(n) + 1), \quad (22c)$$

$$\mathbf{h}_{e,l}^H \mathbf{Z} \mathbf{h}_{e,l} + \sigma_{e,l}^2 + P_p |g_{e,l}|^2 \leq e^{s_l(n)} (s_l - s_l(n) + 1). \quad (22d)$$

At last, we consider the constraints (16d) and (15e), which can be converted as

$$\mathbf{h}_{e,l}^H \mathbf{W}_Z \mathbf{h}_{e,l} + P_p |g_{e,l}|^2 + \sigma_e^2 \geq \frac{\bar{E}^e}{\eta_{e,l}}, \quad \forall l, \quad (23a)$$

$$\sum_{k=1}^K \text{tr}(\mathbf{W}_k) + \text{tr}(\mathbf{Z}) \leq P, \quad \forall k. \quad (23b)$$

According to equations from (16a) to (23b), an SCA-based iterative algorithm is proposed. At the $(n+1)$ th iteration, by removing the nonconvex rank-one constraint $\text{rank}(\mathbf{W}_k) = 1$, $\forall k$, problems (16a)–(16e) can be thus reformed as

$$\begin{aligned} \min_{\Omega} \quad & \sum_{k=1}^K \text{tr}(\mathbf{W}_k) \\ \text{s.t.} \quad & (15f), (19b), (19c), (20b), (20c), (21a), (21b), (22a), (22b), (22c), (22d), (23a), (23b), (23c), \quad \mathbf{W}_k \geq \mathbf{0}, \\ & \Omega = \{ \mathbf{W}_k, \mathbf{Z}, x_k, y_l, t_k, r_{l,k}, x_p, u_l, t_p, s_l \}. \end{aligned} \quad (24)$$

Set $n = 0$ and initialize $\Psi(0) = \{t_k(0), r_{l,k}(0), t_p(0), s_l(0)\}$.
Repeat
 (i) Solve problem (24) with $\{t_k(n), r_{l,k}(n), t_p(n), s_l(n)\}$
 and denote a solution by $\{t_k(n+1), r_{l,k}(n+1), t_p(n+1), s_l(n+1)\}$.
 (ii) Update $n \leftarrow n + 1$.
Until Convergence

ALGORITHM 1: SCA-based iterative algorithm with perfect CSI.

For given $\Psi(n) = \{t_k(n), r_{l,k}(n), t_p(n), s_l(n)\}$ as the optimal solution obtained at the n th iteration, problem (24) is convex by removing the nonconvex $\text{rank}(\mathbf{W}_k) = 1$ constraint, which can be solved by using an CVX tools [39]. From SCA method, the convex approximation with the current solution is iteratively updated until the constraints (22a)–(22d) hold with equality, which implies that (16a)–(16e) can be optimally solved. The optimal solution obtained by the proposed SCA-based iterative algorithm at the n th iteration is assumed to be $(t_k(n), r_{l,k}(n), t_p(n), \text{ and } s_l(n))$, which can achieve a stable point until the SCA-based iterative algorithm converges [35], and this is summarized in Algorithm 1.

4. Proposed Algorithm with Imperfect CSI

Because of channel estimation and quantization errors, it may not be possible to have perfect CSI in practice. In this section, we extend the proposed algorithm to more practical scenarios with imperfect CSI. First, we introduce one scenario of the norm-bounded channel uncertainty and then provide the problem formulation based on the norm-bounded channel uncertainty. Moreover, we consider a joint robust design of transmit beamforming and AN.

4.1. Norm-Bounded Channel Uncertainty. Now, we adopt imperfect CSI based on the deterministic model [31, 35]. In particular, we assume that the actual channel vector $\mathbf{h}_{s,k}$ lies within a ball with the radius $\varepsilon_{s,k}$ around the estimated channel vector $\bar{\mathbf{h}}_{s,k}$ from the BS to the k th user, i.e.,

$$\begin{aligned} \mathbf{h}_{s,k} &\in \mathcal{H}_{s,k} = \{\bar{\mathbf{h}}_{s,k} + \Delta\mathbf{h}_{s,k} \mid \|\Delta\mathbf{h}_{s,k}\| \leq \varepsilon_{s,k}\}, \quad \forall l, \\ \mathbf{h}_{e,l} &\in \mathcal{H}_{e,l} = \{\bar{\mathbf{h}}_{e,l} + \Delta\mathbf{h}_{e,l} \mid \|\Delta\mathbf{h}_{e,l}\| \leq \varepsilon_{e,l}\}, \quad \forall l, \\ \mathbf{h}_p &\in \mathcal{H}_p = \{\bar{\mathbf{h}}_p + \Delta\mathbf{h}_p \mid \|\Delta\mathbf{h}_p\| \leq \varepsilon_p\}, \quad \forall l, \\ g_p &\in \widehat{\mathcal{G}}_p = \{\bar{g}_p + \Delta g_p \mid |\Delta g_p| \leq \gamma_p\}, \\ g_{s,k} &\in \widehat{\mathcal{G}}_{s,k} = \{\bar{g}_{s,k} + \Delta g_{s,k} \mid |\Delta g_{s,k}| \leq \gamma_{s,k}\}, \quad \forall k, \\ g_{e,l} &\in \widehat{\mathcal{G}}_{e,l} = \{\bar{g}_{e,l} + \Delta g_{e,l} \mid |\Delta g_{e,l}| \leq \gamma_{e,l}\}, \quad \forall l. \end{aligned} \quad (25)$$

where $\bar{\mathbf{h}}_{s,k}$, $\bar{\mathbf{h}}_{e,l}$, $\bar{\mathbf{h}}_p$, \bar{g}_p , $\bar{g}_{s,k}$, and $\bar{g}_{e,l}$ are the estimated channel available, the channel estimation errors $\Delta\mathbf{h}_{s,k}$, $\Delta\mathbf{h}_{e,l}$, $\Delta\mathbf{h}_p$, Δg_p , $\Delta g_{s,k}$, and $\Delta g_{e,l}$ are bounded by $\varepsilon_{s,k}$, $\varepsilon_{e,l}$, ε_p , γ_p , $\gamma_{s,k}$, and $\gamma_{e,l}$ respectively.

By taking the norm-bounded channel uncertainty model into account, the TPISM problem can be rewritten as

$$\min_{\mathbf{w}_k, \mathbf{Z}} \sum_{k=1}^K \|\mathbf{w}_k\|^2 \quad (26a)$$

$$\text{s.t. } \log \left(1 + \frac{(\bar{\mathbf{h}}_{s,k} + \Delta\mathbf{h}_{s,k})^H \mathbf{w}_k \mathbf{w}_k^H (\bar{\mathbf{h}}_{s,k} + \Delta\mathbf{h}_{s,k})}{(\bar{\mathbf{h}}_{s,k} + \Delta\mathbf{h}_{s,k})^H (\sum_{j \neq k} \mathbf{w}_j \mathbf{w}_j^H + \mathbf{Z}) (\bar{\mathbf{h}}_{s,k} + \Delta\mathbf{h}_{s,k}) + \sigma_{s,k}^2} \right) \quad (26b)$$

$$- \log \left(1 + \frac{(\bar{\mathbf{h}}_{e,l} + \Delta\mathbf{h}_{e,l})^H \mathbf{w}_k \mathbf{w}_k^H (\bar{\mathbf{h}}_{e,l} + \Delta\mathbf{h}_{e,l})}{\sigma_{e,l}^2 + (\bar{\mathbf{h}}_{e,l} + \Delta\mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta\mathbf{h}_{e,l})} \right) \geq \bar{R}^{su}, \quad \forall l, k,$$

$$\log \left(1 + \frac{P_p |\bar{g}_p + \Delta g_p|^2}{(\bar{\mathbf{h}}_p + \Delta\mathbf{h}_p)^H (\sum_{j=1}^K \mathbf{w}_j \mathbf{w}_j^H + \mathbf{Z}) (\bar{\mathbf{h}}_p + \Delta\mathbf{h}_p) + \sigma_p^2} \right) \quad (26c)$$

$$- \log \left(1 + \frac{P_p |\bar{g}_{e,l} + \Delta g_{e,l}|^2}{(\bar{\mathbf{h}}_{e,l} + \Delta\mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta\mathbf{h}_{e,l}) + \sigma_{e,l}^2} \right) \geq \bar{R}^{pu}, \quad \forall l, k,$$

$$\eta_{e,l} \left((\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H \mathbf{W}_Z (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) + P_p |\bar{g}_{e,l} + \Delta g_{e,l}|^2 + \sigma_e^2 \right) \geq \bar{E}^e, \quad \forall l, \quad (26d)$$

$$(15e), (15f). \quad (26e)$$

Utilizing a similar methodology in (17a)–(18b), the constraints (26b) and (26c) can be reformulated as

$$\frac{\left((\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k})^H \left(\sum_{j=1}^K \mathbf{W}_j + \mathbf{Z} \right) (\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k}) + \sigma_{s,k}^2 \right) \left(\sigma_{e,l}^2 + (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) \right)}{\left((\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k})^H \left(\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z} \right) (\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k}) + \sigma_{s,k}^2 \right) \left(\sigma_{e,l}^2 + (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H (\mathbf{Z} + \mathbf{W}_k) (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) \right)} \geq 2^{\bar{R}^{su}}, \quad (27a)$$

$$\frac{\left((\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p)^H \mathbf{W}_Z (\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p) + \sigma_p^2 + P_p |\bar{g}_p + \Delta g_p|^2 \right) \left((\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) + \sigma_{e,l}^2 \right)}{\left((\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p)^H \mathbf{W}_Z (\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p) + \sigma_p^2 \right) \left((\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) + \sigma_{e,l}^2 + P_p |\bar{g}_{e,l} + \Delta g_{e,l}|^2 \right)} \geq 2^{\bar{R}^{pu}}. \quad (27b)$$

Then, by introducing $\hat{x}_k, \hat{y}_l, \hat{t}_k, \hat{r}_{l,k}, \hat{x}_p, \hat{u}_l, \hat{t}_p$, and \hat{s}_l as slack variables, we equivalently convert the linear fractional programming (27a) and (27b) as follows, respectively:

$$e^{\hat{x}_k + \hat{y}_l - \hat{t}_k - \hat{r}_{l,k}} \geq 2^{\bar{R}^{su}}, \quad \forall k, \forall l, \quad (28a)$$

$$\begin{aligned} \min_{\|\Delta \mathbf{h}_{s,k}\| \leq \epsilon_{s,k}} & \left((\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k})^H \mathbf{W}_Z (\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k}) + \sigma_{s,k}^2 \right) \\ & \geq e^{\hat{x}_k}, \quad \forall k, \end{aligned} \quad (28b)$$

$$\begin{aligned} \min_{\|\Delta \mathbf{h}_{e,l}\| \leq \epsilon_{e,l}} & \sigma_{e,l}^2 + (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) \geq e^{\hat{y}_l}, \\ & \forall l, \end{aligned} \quad (28c)$$

$$\max_{\|\Delta \mathbf{h}_{s,k}\| \leq \epsilon_{s,k}} \left((\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k})^H \left(\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z} \right) \right) \quad (28d)$$

$$\cdot (\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k}) + \sigma_{s,k}^2 \leq e^{\hat{t}_k}, \quad \forall k,$$

$$\begin{aligned} \max_{\|\Delta \mathbf{h}_{e,l}\| \leq \epsilon_{e,l}} & \sigma_{e,l}^2 + (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H (\mathbf{Z} + \mathbf{W}_k) (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) \\ & \leq e^{\hat{r}_{l,k}}, \quad \forall k, \forall l. \end{aligned} \quad (28e)$$

$$e^{\hat{x}_p + \hat{u}_l - \hat{t}_p - \hat{s}_l} \geq 2^{\bar{R}^{pu}}, \quad \forall l, \quad (29a)$$

$$\begin{aligned} \min_{\|\Delta \mathbf{h}_p\| \leq \epsilon_p} & \left((\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p)^H \mathbf{W}_Z (\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p) + \sigma_p^2 \right) \\ & + P_p |\bar{g}_p + \Delta g_p|^2 \geq e^{\hat{x}_p}, \end{aligned} \quad (29b)$$

$$\begin{aligned} \min_{\|\Delta \mathbf{h}_{e,l}\| \leq \epsilon_{e,l}} & \left((\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) + \sigma_{e,l}^2 \right) \geq e^{\hat{u}_l}, \\ & \forall l, \end{aligned} \quad (29c)$$

$$\max_{\|\Delta \mathbf{h}_p\| \leq \epsilon_p} \left((\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p)^H \mathbf{W}_Z (\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p) + \sigma_p^2 \right) \leq e^{\hat{t}_p}, \quad (29d)$$

$$\begin{aligned} \max_{\|\Delta \mathbf{h}_{e,l}\| \leq \epsilon_{e,l}} & \left((\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) + \sigma_{e,l}^2 \right) \\ & + P_p |\bar{g}_{e,l} + \Delta g_{e,l}|^2 \leq e^{\hat{s}_l}, \quad \forall l. \end{aligned} \quad (29e)$$

By employing the slack variables (i.e., d_k, w_l, b_k , and $f_{l,k}$) for (28b)–(28e), respectively, (28a)–(28e) can be equivalently modified as

$$\min_{\|\Delta \mathbf{h}_{s,k}\| \leq \epsilon_{s,k}} \left((\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k})^H \mathbf{W}_Z (\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k}) \right) \geq d_k, \quad (30a)$$

$$\min_{\|\Delta \mathbf{h}_{e,l}\| \leq \epsilon_{e,l}} \left((\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) \right) \geq w_l, \quad (30b)$$

$$\max_{\|\Delta \mathbf{h}_{s,k}\| \leq \epsilon_{s,k}} \left((\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k})^H \left(\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z} \right) \right) \quad (30c)$$

$$\cdot (\bar{\mathbf{h}}_{s,k} + \Delta \mathbf{h}_{s,k}) \leq z_k,$$

$$\max_{\|\Delta \mathbf{h}_{e,l}\| \leq \epsilon_{e,l}} \left((\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H (\mathbf{Z} + \mathbf{W}_k) (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) \right) \quad (30d)$$

$$\leq f_{l,k},$$

$$d_k + \sigma_{s,k}^2 \geq e^{\hat{x}_k}, \quad (30e)$$

$$\sigma_{e,l}^2 + w_l \geq e^{\hat{y}_l}, \quad (30f)$$

$$z_k + \sigma_{s,k}^2 \leq e^{\hat{t}_k}, \quad (30g)$$

$$\sigma_{e,l}^2 + f_{l,k} \leq e^{\hat{r}_{l,k}}. \quad (30h)$$

For the worst-case based design in (26b)–(26d), the PT channel gains are upper-bounded and lower-bounded using the following triangle inequality properties:

$$\begin{aligned} |x + y|^2 &\leq (|x| + |y|)^2 = |x|^2 + |y|^2 + 2|x||y|, \\ |x + y|^2 &\geq (|x| - |y|)^2 = |x|^2 + |y|^2 - 2|x||y|. \end{aligned} \quad (31)$$

Applying (31), it follows that

$$\begin{aligned} g_{e,l}^{max} &\triangleq \max_{\Delta g_{e,l} \leq \gamma_{e,l}} |g_{e,l}|^2 \\ &= \max_{\Delta g_{e,l} \leq \gamma_{e,l}} \left(|\bar{g}_{e,l}|^2 + |\Delta g_{e,l}|^2 + 2|\bar{g}_{e,l}||\Delta g_{e,l}| \right) \\ &\leq |\bar{g}_{e,l}|^2 + \varepsilon_{e,l}^2 + 2\varepsilon_{e,l}|\bar{g}_{e,l}|, \\ g_{e,l}^{min} &\triangleq \min_{\Delta g_{e,l} \leq \gamma_{e,l}} |g_{e,l}|^2 \\ &= \min_{\Delta g_{e,l} \leq \gamma_{e,l}} \left(|\bar{g}_{e,l}|^2 + |\Delta g_{e,l}|^2 - 2|\bar{g}_{e,l}||\Delta g_{e,l}| \right) \\ &\geq |\bar{g}_{e,l}|^2 + \varepsilon_{e,l}^2 - 2\varepsilon_{e,l}|\bar{g}_{e,l}|, \\ g_p^{min} &\triangleq \min_{\Delta g_p \leq \gamma_p} |g_p|^2 \\ &= \min_{\Delta g_p \leq \gamma_p} \left(|\bar{g}_p|^2 + |\Delta g_p|^2 - 2|\bar{g}_p||\Delta g_p| \right) \\ &\geq |\bar{g}_p|^2 + \varepsilon_p^2 - 2\varepsilon_p|\bar{g}_p|, \\ g_{s,k}^{min} &\triangleq \min_{\Delta g_{s,k} \leq \gamma_{s,k}} |g_{s,k}|^2 \\ &= \min_{\Delta g_{s,k} \leq \gamma_{s,k}} \left(|\bar{g}_{s,k}|^2 + |\Delta g_{s,k}|^2 + 2|\bar{g}_{s,k}||\Delta g_{s,k}| \right) \\ &\geq |\bar{g}_{s,k}|^2 + \varepsilon_{s,k}^2 - 2\varepsilon_{s,k}|\bar{g}_{s,k}|. \end{aligned} \quad (32)$$

Similarly, we introduce the slack variables d_p , v_l , b_p , and q_l for (29b)–(29e). Then, by substituting the above results (32) into (29b) and (29e), respectively, we can equivalently express (29a)–(29e) as

$$\min_{\|\Delta \mathbf{h}_p\| \leq \varepsilon_p} (\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p)^H \mathbf{W}_Z (\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p) \geq d_p, \quad (33a)$$

$$\min_{\|\Delta \mathbf{h}_{e,l}\| \leq \varepsilon_{e,l}} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) \geq v_l, \quad (33b)$$

$$\max_{\|\Delta \mathbf{h}_p\| \leq \varepsilon_p} (\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p)^H \mathbf{W}_Z (\bar{\mathbf{h}}_p + \Delta \mathbf{h}_p) \leq b_p, \quad (33c)$$

$$\max_{\|\Delta \mathbf{h}_{e,l}\| \leq \varepsilon_{e,l}} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l})^H \mathbf{Z} (\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l}) \leq q_l, \quad (33d)$$

$$d_p + \sigma_p^2 + P_p g_p^{min} \geq e^{\hat{x}_p}, \quad (33e)$$

$$v_l + \sigma_{e,l}^2 \geq e^{\hat{u}_l}, \quad (33f)$$

$$b_p + \sigma_p^2 \leq e^{\hat{t}_p}, \quad (33g)$$

$$q_l + \sigma_{e,l}^2 + P_p g_{e,l}^{max} \leq e^{\hat{s}_l}. \quad (33h)$$

In order to make (26b)–(26d) more tractable, by applying \mathcal{S} -procedure [40], we can convert the infinitely inequality constraints (30a)–(30d) and (33a)–(33d) into finite linear matrix inequalities (LMIs). For completeness, the \mathcal{S} -procedure is presented in Lemma 1 in the following.

Lemma 1 (\mathcal{S} -procedure [40, Appendix B.2]). *Let a function $\mathbf{f}_m(\mathbf{x})$ with $\mathbf{x} \in \mathbb{C}^{N \times 1}$ ($m = 1, 2$) be defined as*

$$\mathbf{f}_m(\mathbf{x}) = \mathbf{x}^H \mathbf{A}_m \mathbf{x} + 2\text{Re} \{ \mathbf{b}_m^H \mathbf{x} \} + \mathbf{c}_m \quad (34)$$

where $\mathbf{A}_m \in \mathbb{H}^{N \times N}$, $\mathbf{b}_m \in \mathbb{C}^{N \times 1}$, and $\mathbf{c}_m \in \mathbb{R}^{N \times 1}$. Then, $\mathbf{f}_m(\mathbf{x}) \leq 0$ holds if and only if there exists $\theta \geq 0$ such that

$$\theta \begin{bmatrix} \mathbf{A}_1 & \mathbf{b}_1 \\ \mathbf{b}_1^H & \mathbf{c}_1 \end{bmatrix} - \begin{bmatrix} \mathbf{A}_2 & \mathbf{b}_2 \\ \mathbf{b}_2^H & \mathbf{c}_2 \end{bmatrix} \geq \mathbf{0}, \quad (35)$$

provided that there is a point $\hat{\mathbf{x}}$ which satisfies $\mathbf{f}_m(\hat{\mathbf{x}}) < 0$.

To employ the \mathcal{S} -procedure, we rewrite the constraint (30a) as

$$\begin{aligned} \Delta \mathbf{h}_{s,k}^H \mathbf{W}_Z \Delta \mathbf{h}_{s,k} + 2\Re \{ \bar{\mathbf{h}}_{s,k}^H \mathbf{W}_Z \Delta \mathbf{h}_{s,k} \} + \bar{\mathbf{h}}_{s,k}^H \mathbf{W}_Z \bar{\mathbf{h}}_{s,k} \\ \geq d_k. \end{aligned} \quad (36)$$

According to Lemma 1, by using a slack variable $\lambda_{s,k}$, (36) can be expressed as

$$\begin{bmatrix} \lambda_{s,k} \mathbf{I} + \mathbf{W}_Z & \mathbf{W}_Z \bar{\mathbf{h}}_{s,k} \\ \bar{\mathbf{h}}_{s,k}^H \mathbf{W}_Z & \bar{\mathbf{h}}_{s,k}^H \mathbf{W}_Z \bar{\mathbf{h}}_{s,k} - d_k - \lambda_{s,k} \varepsilon_{s,k}^2 \end{bmatrix} \geq \mathbf{0}. \quad (37)$$

Using the same approach for the constraints (30b)–(30d) and (33a)–(33d), we have

$$\begin{bmatrix} \lambda_{e,l} \mathbf{I} + \mathbf{Z} & \mathbf{Z} \bar{\mathbf{h}}_{e,l} \\ \bar{\mathbf{h}}_{e,l}^H \mathbf{Z} & \bar{\mathbf{h}}_{e,l}^H \mathbf{Z} \bar{\mathbf{h}}_{e,l} - w_l - \lambda_{e,l} \varepsilon_{e,l}^2 \end{bmatrix} \geq \mathbf{0}, \quad (38a)$$

$$\begin{bmatrix} \alpha_{s,k} \mathbf{I} - \sum_{j \neq k} \mathbf{W}_j - \mathbf{Z} & - \left(\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z} \right) \bar{\mathbf{h}}_{s,k} \\ -\bar{\mathbf{h}}_{s,k}^H \left(\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z} \right) & -\bar{\mathbf{h}}_{s,k}^H \left(\sum_{j \neq k} \mathbf{W}_j + \mathbf{Z} \right) \bar{\mathbf{h}}_{s,k} + z_k - \alpha_{s,k} \varepsilon_{s,k}^2 \end{bmatrix} \succeq \mathbf{0}, \quad (38b)$$

$$\begin{bmatrix} \alpha_{e,l} \mathbf{I} - \mathbf{Z} - \mathbf{W}_k & -(\mathbf{Z} + \mathbf{W}_k) \bar{\mathbf{h}}_{e,l} \\ -\bar{\mathbf{h}}_{e,l}^H (\mathbf{Z} + \mathbf{W}_k) & -\bar{\mathbf{h}}_{e,l}^H (\mathbf{Z} + \mathbf{W}_k) \bar{\mathbf{h}}_{e,l} + f_{l,k} - \alpha_{e,l} \varepsilon_{e,l}^2 \end{bmatrix} \succeq \mathbf{0}, \quad (38c)$$

$$\begin{bmatrix} \beta_p \mathbf{I} + \mathbf{W}_Z & \mathbf{W}_Z \bar{\mathbf{h}}_p \\ \bar{\mathbf{h}}_p^H \mathbf{W}_Z & \bar{\mathbf{h}}_p^H \mathbf{W}_Z \bar{\mathbf{h}}_p - d_p - \beta_p \varepsilon_p^2 \end{bmatrix} \succeq \mathbf{0}, \quad (39a)$$

$$\begin{bmatrix} \beta_{e,l} \mathbf{I} + \mathbf{Z} & \mathbf{Z} \bar{\mathbf{h}}_{e,l} \\ \bar{\mathbf{h}}_{e,l}^H \mathbf{Z} & \bar{\mathbf{h}}_{e,l}^H \mathbf{Z} \bar{\mathbf{h}}_{e,l} - \nu_l - \beta_{e,l} \varepsilon_{e,l}^2 \end{bmatrix} \succeq \mathbf{0}, \quad (39b)$$

$$\begin{bmatrix} \alpha_p \mathbf{I} - \mathbf{W}_Z & -\mathbf{W}_Z \bar{\mathbf{h}}_p \\ -\bar{\mathbf{h}}_p^H \mathbf{W}_Z & -\bar{\mathbf{h}}_p^H \mathbf{W}_Z \bar{\mathbf{h}}_p + b_p - \alpha_p \varepsilon_p^2 \end{bmatrix} \succeq \mathbf{0}, \quad (39c)$$

$$\begin{bmatrix} \nu_{e,l} \mathbf{I} - \mathbf{Z} & -\mathbf{Z} \bar{\mathbf{h}}_{e,l} \\ -\bar{\mathbf{h}}_{e,l}^H \mathbf{Z} & -\bar{\mathbf{h}}_{e,l}^H \mathbf{Z} \bar{\mathbf{h}}_{e,l} + q_l - \nu_{e,l} \varepsilon_{e,l}^2 \end{bmatrix} \succeq \mathbf{0}, \quad (39d)$$

where $\lambda_{e,l}$, $\alpha_{s,k}$, $\alpha_{e,l}$, β_p , $\beta_{e,l}$, α_p , and $\nu_{e,l}$ are slack variables.

By using the above results of Taylor series expansion in (22a)–(22d), we can transform the nonconvex constraints (30g), (30h), (33g), and (33h) into the corresponding convex forms, respectively, as

$$z_k + \sigma_{s,k}^2 \leq e^{\hat{t}_k(m)} (\hat{t}_k - \hat{t}_k(m) + 1), \quad (40a)$$

$$\sigma_{e,l}^2 + f_{l,k} \leq e^{\hat{r}_{l,k}(m)} (\hat{r}_{l,k} - \hat{r}_{l,k}(m) + 1), \quad (40b)$$

$$b_p + \sigma_p^2 \leq e^{\hat{t}_p(m)} (\hat{t}_p - \hat{t}_p(m) + 1), \quad (40c)$$

$$q_l + \sigma_{e,l}^2 + P_p g_{e,l}^{max} \leq e^{\hat{s}_l(m)} (\hat{s}_l - \hat{s}_l(m) + 1). \quad (40d)$$

where $\hat{t}_k(m)$, $\hat{r}_{l,k}(m)$, $\hat{t}_p(m)$, and $\hat{s}_l(m)$ are the variables \hat{t}_k , $\hat{r}_{l,k}$, \hat{t}_p , and \hat{s}_l at the n th iteration for the SCA method.

Here, we consider the EH constraint (30c). By use the inequalities (32) and (26d) can be rewritten as

$$\left(\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l} \right)^H \mathbf{W}_Z \left(\bar{\mathbf{h}}_{e,l} + \Delta \mathbf{h}_{e,l} \right) + P_p g_{e,l}^{min} + \sigma_e^2$$

$$\min_{\Phi} \sum_{k=1}^K \text{tr}(\mathbf{W}_k)$$

$$\text{s.t.} \quad (23b), (30e), (30f), (33e), (33f), (37), (38a), (38b), (38c), (39a), (39b), (39c), (39d), (40a), (40b), (40c), (40d), (42), (43a), (43b), \quad (44)$$

$$\Phi = \left\{ \mathbf{W}_k, \mathbf{Z}, \hat{x}_k, \hat{y}_l, \hat{t}_k, \hat{r}_{l,k}, \hat{x}_p, \hat{u}_l, \hat{t}_p, \hat{s}_l, d_k, w_l, z_k, f_{l,k}, d_p, \nu_l, b_p, q_l, \lambda_{s,k}, \lambda_{e,l}, \alpha_{s,k}, \alpha_{e,l}, \beta_p, \beta_{e,l}, \alpha_p, \nu_{e,l}, \zeta_{s,k}, \vartheta_{e,l} \right\}, \quad \mathbf{W}_k \succeq \mathbf{0}, \mathbf{Z} \succeq \mathbf{0}.$$

For given $\{\hat{t}_k(m), \hat{r}_{l,k}(m), \hat{t}_p(m), \hat{s}_l(m)\}$ at the m th iteration, problem (44) is a convex SDP form without the nonconvex

$$\geq \frac{\bar{E}^c}{\eta_{e,l}}, \quad \forall l.$$

(41)

Introducing slack variables $\vartheta_{e,l}$, we can transform (41) as

$$\begin{bmatrix} \vartheta_{e,l} \mathbf{I} + \mathbf{W}_Z & \mathbf{W}_Z \bar{\mathbf{h}}_{e,l} \\ \bar{\mathbf{h}}_{e,l}^H \mathbf{W}_Z & \bar{\mathbf{h}}_{e,l}^H \mathbf{W}_Z \bar{\mathbf{h}}_{e,l} + P_p g_{e,l}^{min} + \sigma_e^2 - \frac{\bar{E}^c}{\eta_{e,l}} - \vartheta_{e,l} \varepsilon_{e,l}^2 \end{bmatrix} \succeq \mathbf{0}, \quad (42)$$

Finally, similarly to (21a) and (21b), we need to recast the nonconvex constraints (28a) and (29a) as follows:

$$e^{-\hat{x}_k - \hat{y}_l + \hat{t}_k + \hat{r}_{l,k}} \leq 2^{-\bar{R}^{su}}, \quad \forall k, \forall l, \quad (43a)$$

$$e^{-\hat{x}_p - \hat{u}_l + \hat{t}_p + \hat{s}_l} \leq 2^{-\bar{R}^{pu}}, \quad \forall l. \quad (43b)$$

Combining all the results in (27a)–(34) and (36)–(43b), (26b)–(26d) can be reformulated as

rank-one constraint $\text{rank}(\mathbf{W}_k) = 1, \forall k$, which can be solved by convex optimization packages such as CVX [39],

Set $m = 0$ and initialize $\Psi(0) = \{\hat{t}_k(0), \hat{r}_{l,k}(0), \hat{t}_p(0), \hat{s}_i(0)\}$.

Repeat

(i) Solve problem (44) with $\{\hat{t}_k(m), \hat{r}_{l,k}(m), \hat{t}_p(m), \hat{s}_i(m)\}$
and denote a solution by $\{\hat{t}_k(m+1), \hat{r}_{l,k}(m+1), \hat{t}_p(m+1), \hat{s}_i(m+1)\}$.

(ii) Update $m \leftarrow m + 1$.

Until Convergence

ALGORITHM 2: Robust SCA-based iterative algorithm with imperfect CSI.

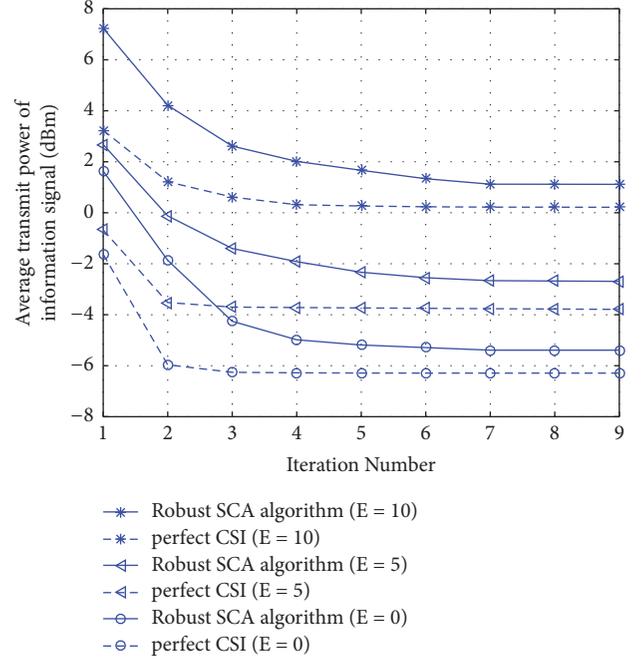
to update the solution for the $m + 1$ -th iteration until the algorithm converges. In addition, the proposed robust SCA-based iterative algorithm with imperfect CSI is summarized in Algorithm 2.

5. Simulation Results

In this section, we provide the simulation results to validate the performance of our proposed algorithms. We consider a system with two SUs and two ERs. The ST is equipped with six transmit antennas ($N_S = 6$). We define $d_{SU_k}^{ST} = 5$ m, $d_{ER_l}^{ST} = 3$ m, and $d_{PU}^{ST} = 10$ m as the distance between the ST and the k th SU, the l th ER, and the PU, respectively. Also, we fix $d_{SU_k}^{PT} = 4$ m, $d_{ER_l}^{PT} = 3$ m, and $d_{PU}^{PT} = 2$ m as the distance between the PT and the k th SU, the l th ER, and the PU, respectively, unless otherwise specified. Thus, the estimated channels $\bar{\mathbf{h}}_{s,k}$, $\bar{\mathbf{h}}_{e,l}$, $\bar{\mathbf{h}}_p$, $\bar{\mathbf{g}}_p$, $\bar{\mathbf{g}}_{s,k}$, and $\bar{\mathbf{g}}_{e,l}$ are respectively, modelled as $\bar{\mathbf{h}}_{s,k} = H(d_{SU_k}^{ST})\bar{\mathbf{h}}_I$, $\bar{\mathbf{h}}_{e,l} = H(d_{ER_l}^{ST})\bar{\mathbf{h}}_I$, $\bar{\mathbf{h}}_p = H(d_{PU}^{ST})\bar{\mathbf{h}}_I$, $\bar{\mathbf{g}}_{s,k} = H(d_{SU_k}^{ST})\bar{\mathbf{g}}_I$, $\bar{\mathbf{g}}_{e,l} = H(d_{ER_l}^{ST})\bar{\mathbf{g}}_I$, and $\bar{\mathbf{g}}_p = H(d_{PU}^{ST})\bar{\mathbf{g}}_I$, where $H(d) = (c/4\pi f_c)(1/d)^{\kappa/2}$, $\bar{\mathbf{h}}_I \sim \mathcal{CN}(0, \mathbf{I})$, and $\bar{\mathbf{g}}_I \sim \mathcal{CN}(0, 1)$, and we set the speed of light, the carrier frequency, and the path loss exponent as $c = 3 \times 10^8$ ms⁻¹, $f_c = 900$ MHz, and $\kappa = 2.7$, respectively. In addition, the noise power at the PU, SU, and ER is assumed to be $\sigma_p^2 = \sigma_{s,k}^2 = \sigma_{e,l}^2 = -90$ dBm. Also the additional noise power of all the SUs is $\delta_{a,k}^2 = -50$ dBm, $\forall k$. We fix the channel error bound for the deterministic model as $\varepsilon = \varepsilon_{s,k} = \varepsilon_{e,l} = \varepsilon_p$, $\forall k$, $\forall l$ and $\gamma = \gamma_p = \gamma_{s,k} = \gamma_{e,l}$, $\forall k$, $\forall l$. The EH efficiency coefficients are set to $\eta_{e,l} = \eta_{e,k} = 0.3$. In our simulations, we compare the following transmit designs: the perfect CSI case, robust SCA-based scheme, no-AN scheme which is obtained by setting $\mathbf{Z} = \mathbf{0}$, and the nonrobust scheme which assumes no uncertainty in the CSI.

Figure 2 illustrates the convergence performance of the proposed SCA-aided iterative algorithm with respect to (w.r.t.) iteration numbers. Here, we set $\bar{R}^{su} = \bar{R}^{pu} = 0.5$ bps/Hz, $E = \bar{E}^c = 1$ dBm, $P_p = 20$ dBm, and $\varepsilon = \gamma = 0.01$. It is easily seen from the plots that the convergence of all perfect CSI cases can be quickly achieved within 5 iterations. It is observed that robust SCA-based scheme converges slower than the perfect CSI case regardless of E . This is due to the fact that the number of variables in the robust SCA-based scheme is greater than the perfect CSI case.

Figure 3 compares the average transmit power of the information signal w.r.t. the target secrecy rate at SU with

FIGURE 2: Average transmit power of the information signal w.r.t. iteration numbers with various E .

$\bar{R}^{pu} = 0.5$ bps/Hz, $P_p = 20$ dBm and $E = 3$ dBm. It is observed that the performance gaps of perfect CSI scheme with $\varepsilon = 0.01$ and 0.1 over the robust SCA-aided iterative algorithm are 0.5 dB and 1.1 dB at all target secrecy rate region, respectively. Also, the transmit power for the no-AN scheme grows faster than the robust schemes as \bar{R}^{su} increases. We can see that, for $\varepsilon = 0.01$ and 0.1 , the robust SCA-aided iterative algorithm outperforms the no-AN scheme and nonrobust scheme. This is due to the help of the AN.

Moreover, the transmit power of the information signal performance is plotted in Figure 4 with $\bar{R}^{su} = 0.5$ bps/Hz, $P_p = 20$ dBm, and $\bar{E}^c = 5$ dBm. One can observe from Figure 4 that the average transmit power of the information signal increases as the target secrecy rate at PU becomes large. Also, the performance gap between the robust SCA-aided iterative algorithm and the no-AN scheme becomes large at high secrecy rate region. Also, there are 0.4 dB and 0.9 dB gaps between perfect CSI and robust SCA-aided iterative algorithm curves for $\varepsilon = 0.01$ and 0.1 , respectively.

Finally, in Figure 5, the transmit power of the information signal w.r.t. the target harvested power is illustrated with

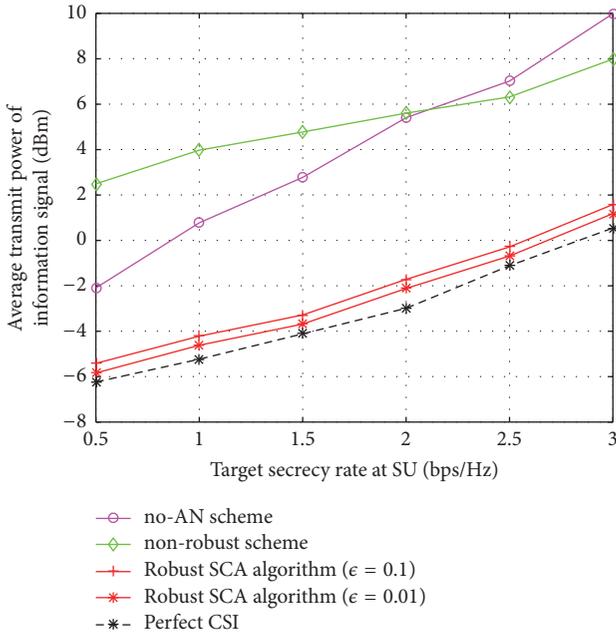


FIGURE 3: Average transmit power of the information signal w.r.t. the target secrecy rate at SU.

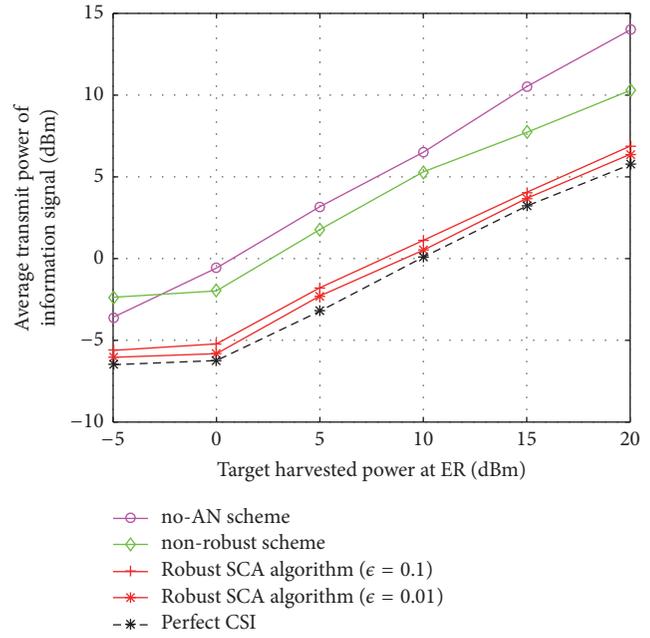


FIGURE 5: Average transmit power of the information signal w.r.t. the harvested power.

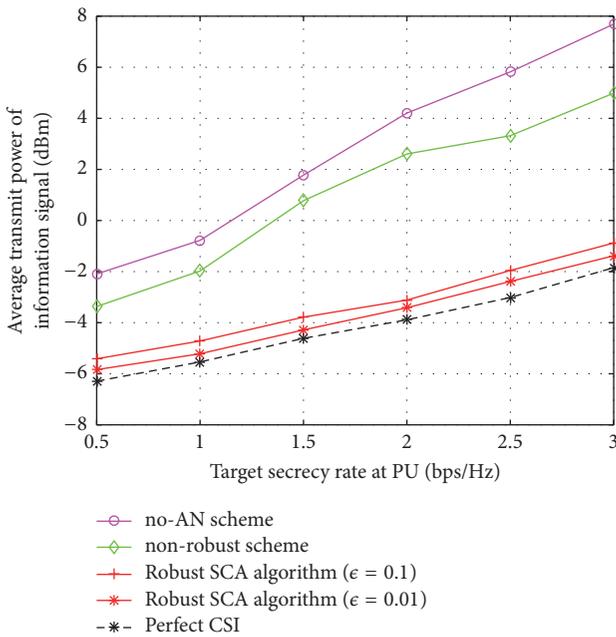


FIGURE 4: Average transmit power of the information signal w.r.t. the target secrecy rate at PU.

$\bar{R}^{su} = \bar{R}^{du} = 1$ bps/Hz and $P_p = 30$ dBm. We can see that the robust SCA-aided iterative algorithm with $\epsilon = 0.01$ outperforms the nonrobust scheme by 3.6 dB and the performance gains of the robust SCA-aided iterative algorithm over the no-AN scheme become larger as the target harvested power increases. Similarly, when $\bar{E}^e \leq 0$ dBm, the performance of the proposed algorithms changes slowly. This is due to the introduction of AN for the CR system.

6. Conclusion

In this paper, we have studied AN-aided secure beamforming designs for MISO Secured CR networks with SWIPT. Our aim is to minimize the transmit power of the information signal subject to the secrecy rate constraint, the harvested energy constraint, and total transmit power constraint. As the original problem is nonconvex, we recast the original problem as a convex form by using SCA method. Then, we have proposed an SCA-based iterative algorithm for the perfect CSI case. Moreover, we have extended the proposed algorithm to the norm-bounded channel uncertainty model. Finally, simulation results have been given to validate the performance of our proposed algorithms.

Data Availability

The original code of this paper cannot be made available.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Nature Science Foundation of China under Grants no. 61571402 and 61771431 and by the National Science and Technology Major Project under Grant no. 2017ZX03001001.

References

[1] S. Zhang, Q. Wu, S. Xu, and G. Y. Li, "Fundamental Green Tradeoffs: Progresses, Challenges, and Impacts on 5G

- Networks,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 33–56, 2017.
- [2] Z. Li, L. Guan, C. Li, and A. Radwan, “A Secure Intelligent Spectrum Control Strategy for Future THz Mobile Heterogeneous Networks,” *IEEE Communications Magazine*, vol. 56, no. 6, pp. 116–123, 2018.
 - [3] H. Zhang, F. Fang, J. Cheng, K. Long, W. Wang, and V. C. Leung, “Energy-Efficient Resource Allocation in NOMA Heterogeneous Networks,” *IEEE Wireless Communications Magazine*, vol. 25, no. 2, pp. 48–53, 2018.
 - [4] Y. Wu, R. Schober, D. W. Ng, C. Xiao, and G. Caire, “Secure massive MIMO transmission with an active eavesdropper,” *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 62, no. 7, pp. 3880–3900, 2016.
 - [5] D. Wang, N. Zhang, Z. Li, F. Gao, and X. Shen, “Leveraging High Order Cumulants for Spectrum Sensing and Power Recognition in Cognitive Radio Networks,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 2, pp. 1298–1310, 2018.
 - [6] S. Haykin, “Cognitive radio: brain-empowered wireless communications,” *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 2, pp. 201–220, 2005.
 - [7] A. Gupta and R. K. Jha, “A survey of 5G network: architecture and emerging technologies,” *IEEE Access*, vol. 3, pp. 1206–1232, 2015.
 - [8] Z. Li, S. Cheng, F. Gao, and Y.-C. Liang, “Sequential Detection for Cognitive Radio with Multiple Primary Transmit Power Levels,” *IEEE Transactions on Communications*, vol. 65, no. 7, pp. 2769–2780, 2017.
 - [9] Z. Li, D. Wang, P. Qi, and B. Hao, “Maximum eigenvalue based sensing and power recognition for multi-antenna cognitive radio system,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 10, pp. 8218–8229, 2015.
 - [10] R. Zhang and C. K. Ho, “MIMO broadcasting for simultaneous wireless information and power transfer,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 1989–2001, 2013.
 - [11] Z. Zhu, Z. Chu, Z. Wang, and I. Lee, “Outage Constrained Robust Beamforming for Secure Broadcasting Systems with Energy Harvesting,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 11, pp. 7610–7620, 2016.
 - [12] S. Kisseleff, I. F. Akyildiz, and W. H. Gerstacker, “Magnetic induction-based simultaneous wireless information and power transfer for single information and multiple power receivers,” *IEEE Transactions on Communications*, vol. 65, no. 3, pp. 1396–1410, 2017.
 - [13] Y. Dong, X. Ge, M. J. Hossain, J. Cheng, and V. C. M. Leung, “Proportional Fairness-Based Beamforming and Signal Splitting for MISO-SWIPT Systems,” *IEEE Communications Letters*, vol. 21, no. 5, pp. 1135–1138, 2017.
 - [14] Y. Dong, M. J. Hossain, and J. Cheng, “Joint power control and time switching for SWIPT systems with heterogeneous QoS requirements,” *IEEE Communications Letters*, vol. 20, no. 2, pp. 328–331, 2016.
 - [15] Z. Zhu, S. Huang, Z. Chu, F. Zhou, D. Zhang, and I. Lee, “Robust Designs of Beamforming and Power Splitting for Distributed Antenna Systems With Wireless Energy Harvesting,” *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–12, 2018.
 - [16] D. W. K. Ng and R. Schober, “Secure and Green SWIPT in Distributed Antenna Networks With Limited Backhaul Capacity,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 5082–5097, 2015.
 - [17] H. Zhang, J. Du, J. Cheng, K. Long, and V. C. Leung, “Incomplete CSI Based Resource Optimization in SWIPT Enabled Heterogeneous Networks: A Non-Cooperative Game Theoretic Approach,” *IEEE Transactions on Wireless Communications*, vol. 17, no. 3, pp. 1882–1892, 2018.
 - [18] Y. Lou, Y. Zheng, J. Cheng, and H. Zhao, “Performance of SWIPT-Based Differential AF Relaying Over Nakagami- m Fading Channels With Direct Link,” *IEEE Wireless Communications Letters*, vol. 7, no. 1, pp. 106–109, 2018.
 - [19] Z. Hu, N. Wei, and Z. Zhang, “Optimal Resource Allocation for Harvested Energy Maximization in Wideband Cognitive Radio Network with SWIPT,” *IEEE Access*, vol. 5, pp. 23383–23394, 2017.
 - [20] F. Benkhelifa and M. Alouini, “Prioritizing Data/Energy Thresholding-Based Antenna Switching for SWIPT-Enabled Secondary Receiver in Cognitive Radio Networks,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 4, pp. 782–800, 2017.
 - [21] J. Yan and Y. Liu, “A Dynamic SWIPT Approach for Cooperative Cognitive Radio Networks,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 12, pp. 11122–11136, 2017.
 - [22] F. Benkhelifa and M.-S. Alouini, “A thresholding-based antenna switching in MIMO cognitive radio networks with SWIPT-enabled secondary receiver,” in *Proceedings of the 2017 IEEE International Conference on Communications, ICC 2017*, France, May 2017.
 - [23] P. V. Tuan and I. Koo, “Optimal multiuser MISO beamforming for power-splitting SWIPT cognitive radio networks,” *IEEE Access*, vol. 5, pp. 14141–14153, 2017.
 - [24] A. A. Al-Habob, A. M. Salhab, S. A. Zummo, and M.-S. Alouini, “Multi-destination cognitive radio relay network with SWIPT and multiple primary receivers,” in *Proceedings of the 2017 IEEE Wireless Communications and Networking Conference, WCNC 2017*, USA, March 2017.
 - [25] D. W. K. Ng, E. S. Lo, and R. Schober, “Multiobjective Resource Allocation for Secure Communication in Cognitive Radio Networks With Wireless Information and Power Transfer,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 5, pp. 3166–3184, 2016.
 - [26] Z. Yang, Z. Ding, P. Fan, and N. Al-Dhahir, “The Impact of Power Allocation on Cooperative Non-orthogonal Multiple Access Networks with SWIPT,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4332–4343, 2017.
 - [27] L. Mohjazi, I. Ahmed, S. Muhaidat, M. Dianati, and M. Al-Qutayri, “Downlink Beamforming for SWIPT Multi-User MISO Underlay Cognitive Radio Networks,” *IEEE Communications Letters*, vol. 21, no. 2, pp. 434–437, 2017.
 - [28] F. Zhou, N. C. Beaulieu, J. Cheng, Z. Chu, and Y. Wang, “Robust max-min fairness resource allocation in sensing-based wideband cognitive radio with SWIPT: Imperfect channel sensing,” *IEEE Systems Journal*, 2017.
 - [29] X. Chen, D. W. Ng, W. H. Gerstacker, and H. Chen, “A Survey on Multiple-Antenna Techniques for Physical Layer Security,” *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1027–1053, 2017.
 - [30] H. Ma, J. Cheng, X. Wang, and P. Ma, “Robust MISO Beamforming With Cooperative Jamming for Secure Transmission From Perspectives of QoS and Secrecy Rate,” *IEEE Transactions on Communications*, vol. 66, no. 2, pp. 767–780, 2018.
 - [31] D. W. K. Ng, E. S. Lo, and R. Schober, “Robust beamforming for secure communication in systems with wireless information

- and power transfer,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4599–4615, 2014.
- [32] Z. Zhu, Z. Chu, F. Zhou, H. Niu, Z. Wang, and I. Lee, “Secure Beamforming Designs for Secrecy MIMO SWIPT Systems,” *IEEE Wireless Communications Letters*, vol. 7, no. 3, pp. 424–427, 2018.
- [33] Z. Zhu, Z. Chu, N. Wang, S. Huang, Z. Wang, and I. Lee, “Beamforming and Power Splitting Designs for AN-Aided Secure Multi-User MIMO SWIPT Systems,” *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 12, pp. 2861–2874, 2017.
- [34] Y. Yuan and Z. Ding, “Outage Constrained Secrecy Rate Maximization Design With SWIPT in MIMO-CR Systems,” *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 5475–5480, 2018.
- [35] F. Zhou, Z. Li, J. Cheng, Q. Li, and J. Si, “Robust AN-Aided Beamforming and Power Splitting Design for Secure MISO Cognitive Radio with SWIPT,” *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2450–2464, 2017.
- [36] S. Ghosh, M. R. Bhatnagar, A. Singh, and B. K. Panigrahi, “Secrecy Capacity in CRN With Malicious Energy Harvester Using Game Theoretic Techniques,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 3, no. 3, pp. 343–360, 2017.
- [37] A. Singh, M. R. Bhatnagar, and R. K. Mallik, “Secrecy Outage of a Simultaneous Wireless Information and Power Transfer Cognitive Radio System,” *IEEE Wireless Communications Letters*, vol. 5, no. 3, pp. 288–291, 2016.
- [38] B. Fang, Z. Qian, W. Zhong, and W. Shao, “AN-Aided Secrecy Precoding for SWIPT in Cognitive MIMO Broadcast Channels,” *IEEE Communications Letters*, vol. 19, no. 9, pp. 1632–1635, 2015.
- [39] L. Liberti and N. Maculan, *Global Optimization*, vol. 84, Kluwer Academic Publishers, Boston, Mass, USA, 2006.
- [40] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.

Research Article

A Context-Aware Location Differential Perturbation Scheme for Privacy-Aware Users in Mobile Environment

Xuejun Zhang ^{1,2}, Haiyan Huang,¹ Shan Huang,³ Qian Chen,¹
Tao Ju,¹ and Xiaogang Du ¹

¹School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

²The Key Laboratory of Opto-Technology and Intelligent Control Ministry of Education, Lanzhou 730070, China

³School of Civil Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China

Correspondence should be addressed to Xuejun Zhang; zxjlyl_new@163.com

Received 14 March 2018; Revised 24 June 2018; Accepted 16 July 2018; Published 6 August 2018

Academic Editor: Wolfgang H. Gerstaecker

Copyright © 2018 Xuejun Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The proliferation of location-based services, representative services for the mobile networks, has posed a serious threat to users' privacy. In the literature, several privacy mechanisms have been proposed to preserve location privacy. Location obfuscation enforced using cloaking region is a widely used technique to achieve location privacy. However, it requires a trusted third-party (TTP) and cannot sufficiently resist various inference attacks based on background information and thus is vulnerable to location privacy breach. In this paper, we propose a context-aware location privacy-preserving solution with differential perturbations, which can enhance the user's location privacy without requiring a TTP. Our scheme utilizes the modified Hilbert curve to project every 2-d location of the user in the considered map to 1-d space and randomly generates the reasonable perturbation by adding Laplace noise via differential privacy. In order to solve the resource limitation of mobile devices, we use a quad-tree based scheme to transform and store the user context information as bit stream which achieves the high compression ratio and supports efficient retrieval. Security analysis shows that our proposed scheme can effectively preserve the location privacy. Experimental evaluation shows that our scheme retrieval accuracy is increased by an average of 15.4% compared with the scheme using standard Hilbert curve. Our scheme can provide strong privacy guarantees with a bounded accuracy loss while improving retrieval accuracy.

1. Introduction

As the indispensable parts of the communications and networks field, the green mobile networks are seen as a potential enabler to realize green communications and networks by minimizing energy consumption while guaranteeing the quality of service [1]. Recently, the rapid development of green wireless communication technologies and personal mobile devices equipped with GPS chips enable location-based services (LBSs) become very popular in almost all social and business domains. Some potential applications of LBS include location-aware information retrieval (e.g., Around Me), GPS navigator (e.g., TomTom), mapping application (e.g., Google Maps), and location-aware social networks (e.g., Foursquare) [2]. With the help of these applications, users can easily issue LBS queries from their smartphones to the LBS providers (LSP) and obtain services

related to their current locations. For example, users can search for their friends, share information with each other, and provide check-in data by using the Foursquare.

Despite the enormous benefits of LBSs provided to individual and society, they also raise major privacy concerns when location information has to leave users' devices to untrusted LSP. Location data contained into the LBS queries can be easily linked to a variety of other information about an individual and reveal his sensitive private information such as his home and work address, sexual preferences, political views, religious inclinations, and health conditions.

To address the privacy issues for mobile users in LBSs, a variety of privacy-preserving mechanisms and metrics have been proposed to allow users to make use of the LBSs while mitigating privacy concerns over the past few years [3–15]. These LBS privacy protection mechanisms (LPPMs) provide different privacy-utility trade-off, which offer alternatives

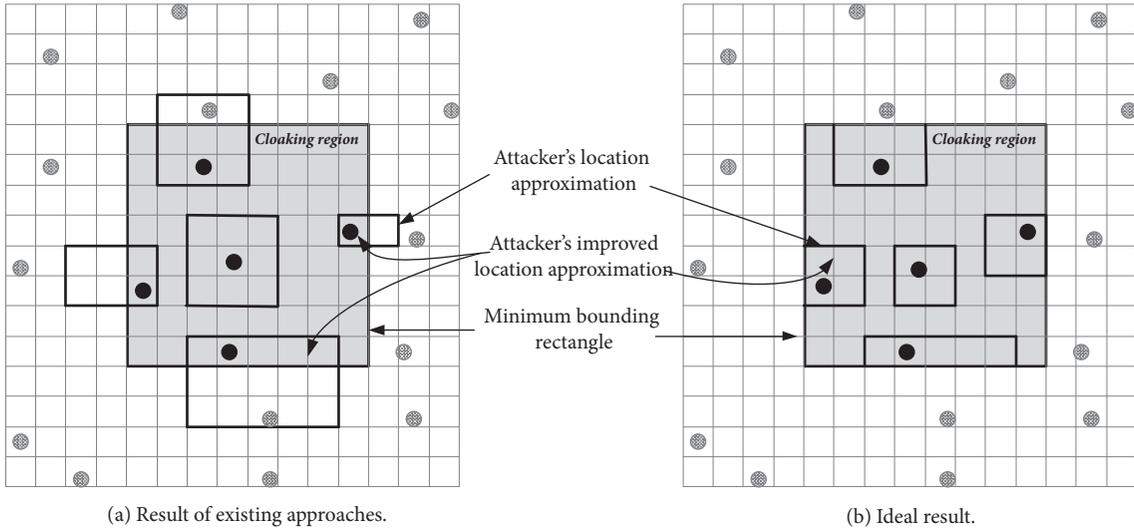


FIGURE 1: Location privacy as a result of using CR.

to better meet personal requirements of different mobile users. Roughly speaking, these LPPMs can be divided into two categories according to their architecture [16]: trusted anonymization server-based schemes [3, 5, 7, 8, 11, 13] and mobile devices-based schemes [4, 6, 9, 10, 14, 15]. In trusted anonymization server-based schemes, a trusted third-party server (e.g., anonymizer [3]) is employed to perturb, obfuscate, and cloak user's query location by using the notion of k -anonymity [3]. To achieve k -anonymity, a user issues his location to LSP via a trusted third-party server (TTP), which subsequently generates a cloaking regions (CR) that covers not only this user, but also $k-1$ other users geographically. Therefore, it is difficult for the untrusted LSP to distinguish a user among at least $k-1$ others. Although such schemes can indeed strengthen the location privacy of users, they heavily rely on the TTP, which would easily be a bottleneck due to handling query requests, frequent updates of user locations, and result postprocessing. Moreover, since the TTP knows the complete knowledge of the locations and queries of all users, it would suffer from a single point of failure. If the adversary seizes control of it, the privacy of all users will be compromised. Recent research [7] attempts to solve this problem by using dynamic grid system, while it requires changing the system mode of the client-side, TTP, and server-side. Furthermore, it incurs the high computation overhead at client-side. Mobile device-based approaches remove the requirement of a TTP by using k -anonymity [10, 15], location obfuscation and perturbation [4, 6, 9], and private information retrieval (PIR) [14]. However, PIR may incur high computation and communication costs unaffordable to mobile devices and LBS server. The k -anonymity [7, 10, 15] assumes that the adversary has no *side information* about the user [11, 12], such as approximate location, mobility profile, query frequency, and user profiles. In reality, since some adversary (e.g., the LSP) may possess such *side information*, these methods are inadequate to protect the user's location

privacy [8]. In Figure 1(a), for example, when approximate location knowledge (e.g., an area) is available to an adversary, he can exploit k -anonymous CR to enhance the precision of location knowledge of multiple users. The CR can therefore provide additional location knowledge to the adversary, thereby leading to a location privacy breach. As shown in Figure 1(b), the problem can be eliminated only if the cloaking regions are guaranteed to encompass the approximated regions corresponding to each of the k users. Unfortunately, it is difficult to judge the extent of knowledge that an adversary possesses. Furthermore, sometime it is difficult to find enough users in a reasonable CR. Thus, in order to achieve the desired level of privacy, CR may be unnecessary expansion. In the worst case, the services for users would be denied. Local obfuscation and differential perturbation approaches [4, 6, 9] may be used to protect user's privacy against an adversary with such *side information*, as they consider the adversary's knowledge and capability to better make a trade-off between location privacy and LBS utility. Further, the differential perturbation [9] abstracts from the *side information* of the adversary, which promises strong theoretical privacy guarantees with a bounded accuracy loss [17]. Nevertheless, these methods are unlikely suffice for LBS because they do not take the contextual information, such as map information, points of interest (POIs) density, the scale of location, and the user's privacy requirement into account. In real scenario, the LBS privacy protection level and accuracy, achieved by location obfuscation and differential perturbation approach, depend highly on the contextual information surround a user. For instance, intuition suggests that a LBS user should deviate from his query location in a rural area than in a downtown area in order to achieve the same privacy level and LBS utility. To the best of our knowledge, how to design a TTP-free and context-aware privacy-preserving LBS system suitable for mobile devices is still challenging.

In this paper, we propose a context-aware differentially private location perturbation solution for location privacy-preserving which operates solely on the devices and does not require any TTP. Different from existing approaches, our scheme considers the contextual information around the user's location and can prevent privacy breach against an adversary with some *side information*. We first use the modified Hilbert curve (MHC) to transform and store every 2-d geographical location in the considered map to 1-d space in terms of the contextual information of a user's location and then randomly perturb the user's location, by adding a controlled amount of noise from a carefully selected Laplace distribution, according to the desired level of privacy. The perturbed value is then submitted as the user's location to the LSP. To address the resource limitation of mobile devices, we use quad-tree based scheme to transform and to store users' context as bit stream. The generated bit stream can achieve a high compress ratio and support efficient retrieval. Our major contributions are as follows:

(1) We propose a context-aware differentially private location perturbation scheme that does not require a TTP and can protect a user's location privacy against an adversary with *side information*.

(2) We construct a MHC according to the density distribution of POIs in the considered local map and design a differential location perturbation algorithm based on it to protect user's location privacy in LBSs. This scheme provides strong privacy guarantees through the differential privacy. Due to the dimension reduction property of the modified Hilbert curve, the system overhead can also be reduced.

(3) We provide thorough security analysis and a comprehensive set of experiments to demonstrate the effectiveness of our approach to location privacy-preserving.

The remainder of this paper is organized as follows. In Section 2, we review the related works. Section 3 introduces some preliminaries of this paper. Section 4 presents the details of our proposed schemes. In what follows, we give the security and performance evaluation in Section 5. Finally, Section 6 concludes the paper.

2. Related Works

In the last few years, various privacy threats in terms of sharing location data have been identified in the literature. For instance, sharing location of a user not only diminishes his own privacy but also the privacy of others [18]. Even sharing the locations sporadically can still make adversary identify the user [19].

To cope with these threats, a variety of location privacy-preserving mechanisms and metrics have been proposed. In this section, we will review these related works.

2.1. Location Privacy Metrics. Since a location can be specified as single coordinate, to quantify the location privacy, we should find out how accurately an adversary might infer about this coordinate. Based on this principle, numerous privacy metrics have been proposed for quantifying the capability of the adversary. Location k-anonymity [3] and

its variation like l-diversity [20] and t-closeness [21] are proposed to measure the ability of the adversary to differentiate the real user from others within the anonymity set. To overcome the drawbacks of k-anonymity in quantifying location privacy, entropy-based metrics have been adopted in [5, 13, 22, 23] for quantifying the information an adversary can obtain from one (or a series) of location update(s). Nonetheless, Shokri et al. [24] show a lack of satisfactory correlation between these two metrics and the success of the adversary in inferring the users' actual position. Therefore, they proposed the expected distance error metric to quantify the degree of accuracy by which an adversary can estimate a user's real position. However, this metric is explicitly defined in terms of the adversary's *side information* [25]. Once the adversary has no such *side information*, the expected distance error is not sufficient for quantifying location privacy. As a result, differential privacy [6, 26] that abstract from the adversary's *side information* has been growing popularity in LBS privacy protection, which measures the ability of the adversary with arbitrary background knowledge to obtain the user's real location. However, as noted in [27], this metric can be problematic if prior is taken into account.

2.2. Location Privacy Protection. In the past few years, many approaches for protecting location privacy are proposed to allow users to enjoy the LBSs while limiting the amount of disclosed sensitive information [3–15, 22, 26–33]. Although, among them, policy-based approaches and cryptography-based approaches [14] have also been investigated, most existing works are based on location obfuscation. For location obfuscation mechanisms, most of them employ well-known location k-anonymity to protect user's privacy by blurring user's exact location into a sufficiently larger CR. Because of its simplicity, k-anonymity metric has been widely adopted in many different methods, including IntervalCloak [3], clique-based cloak [5], location differential perturbations [8], game-theoretic approach [12], dummy location selection [13], and hilbASR [28]. However, these methods suffer from the single point of failure due to the reliance on a TTP named anonymizer. If an adversary seizes control of the TTP, the privacy of all users will be breached. This TTP is also a performance bottleneck since all the submitted LBS queries have to go through it. Moreover, these methods are vulnerable to background knowledge attacks and homogeneity attacks [20].

To avoid the use of TTP, many mobile device-based schemes [4, 6, 9, 10, 14, 15, 29–33] are introduced into LBS privacy protection LBS system. However, k-anonymity based schemes [10, 13, 29–33] still need to generate CR via exchanged information from other encountered mobile users. Thus, they also cannot resist homogeneity attacks and background knowledge attacks. Expected distance error based schemes [4, 9] obfuscate user's location by taking the adversary's *side information* into account, which also suffer from background knowledge attacks. Differential privacy based schemes [6, 26] have gained popularity as they abstract from the adversary's side information and are capable of providing strong worst-case privacy guarantees. However,

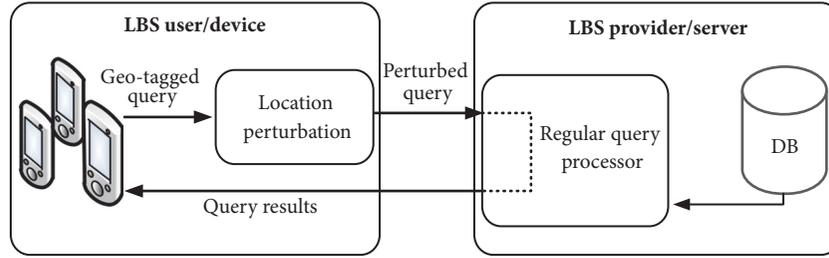


FIGURE 2: Architecture of our proposed framework.

these approaches do not take the contextual information of the user's location and are not sufficient to protect users from reidentification [34].

Different from existing works, our proposed method use MHC to store the context of the user's location, achieving robust privacy guarantee against the adversary with *some information*. It provides desired privacy level for mobile users without relying on any TTP.

Standard Hilbert curve (SHC) has been applied to some privacy protection schemes (e.g., [30, 32]), which is different from our MHC. Our MHC mapping is similar to the VHC-mapping [15], but there are several key differences. First, VHC-mapping is constructed from road density, but our MHC mapping is based on the density of POIs. Second, VHC-mapping is used to perturb a single location, but our MHC is used to select k POIs to preserve reciprocity [28].

3. Preliminaries

In this section, we first introduce the system model and some basic concepts used in this paper and then present the motivation and basic ideas of our scheme.

3.1. System Model and Basic Concepts. Our system model is composed of two parties: LBS user/device and the LBS provider/server, as shown in Figure 2. LBS user possesses a location-aware wireless device, capable of connecting to the network through a wireless protocol such as WiFi, GPRS, or 3G. LBS user uses location perturbation algorithm to perturb his location included in the LBS query, and he submits the perturbed query to LBS provider. The LBS provider is untrusted and considered as the adversary. He responds to the LBS user's requests and returns query results. He can also obtain all the side information by monitoring the queries issued from the LBS user. Additionally, he knows the location perturbation algorithm and noise distribution used in the system. Based on this information, he tries to perform inference attacks to deduce the user's location information.

In this paper, the side information is limited to the approximate location knowledge of users (an area instead of exact coordinates), which can be obtained by a variety of means, i.e., device communication logs such as cell towers used, public records such as parking violations, or social engineering methods such as during a casual conversation [8]. Unless regulated by legislations, the approximate location knowledge can more simply be inferred directly from the

information broadcast from cell towers and wireless access points.

3.2. Motivation and Basic Ideas. In Figure 2, location perturbation component perturbs the user's location contained in the geo-tagged query to generate the perturbed query. It also rearranges the query results returned from the regular query processor of the LSP, in order to provide better LBS utility. Location perturbation is a straightforward approach to achieve efficient location privacy-preserving. However, this method may lead to other challenges, e.g., how to achieve context-aware privacy protection without incurring the cost of storing and retrieving a full-scale map in a mobile device, and how to generate a reasonable perturbation to make a trade-off between privacy and LBS utility. Most of existing works generate the perturbation by adding a random noise (to the true location) drawn from a standard probability distribution. However, it is not a good way to protect user's privacy against the adversaries with *side information* (e.g., a set of likely positions including the true location). With the *side information* and noise distribution, the adversary can calculate the probability of generating the observed perturbation from each of the likely positions. If the probability is significantly high for the real location, the adversary will confidently infer the user real position. To enhance user privacy, these probabilities should be within a small constant factor of each other. Our main idea is to employ a MHC mapping and a carefully selected Laplace distribution to achieve effective privacy-preserving. Our approach can be presented in two parts: (1) we use a MHC based on POIs density in considered local map to achieve the contextual information of the user's location and store it as bit stream; (2) then, we employ a carefully selected Laplace noise distribution to generate a reasonable perturbation and transmit the perturbed value as the user's location to the LSP.

Specifically, we observe that in location-based applications such as nearby searches and check-in posts in geo-social networks (e.g., Foursquare and WeChat), users tend to query an LBS from places that are meaningful to them (e.g., offices and restaurants). In such places, users are most likely to perform an activity without too much movement. We call these places the points of interest of users and refer them to the real POIs in local map. In addition, we assume that users request LBSs from their POIs. Let \mathcal{R} be the (rectangle) boundary of the local map, Ψ be the set of all possible real POIs in \mathcal{R} , and $\Psi_u \in \Psi$ be the set of all POIs of user u . For

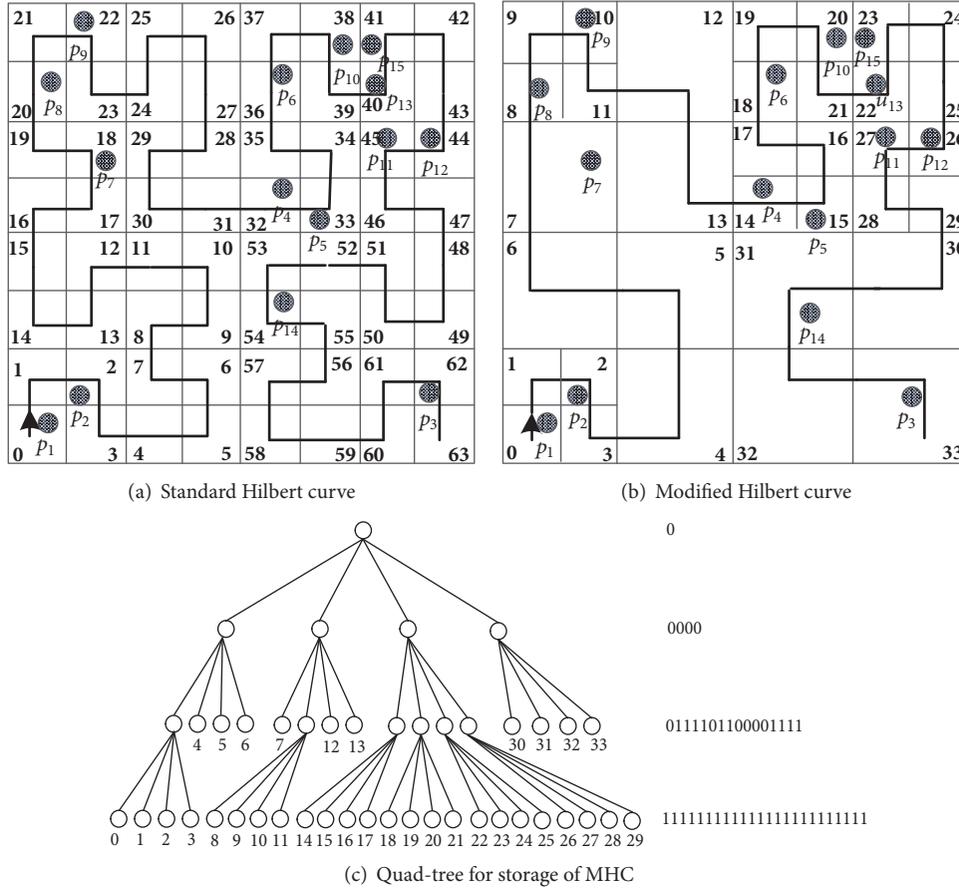


FIGURE 3: Our modified Hilbert curve and its quad-tree storage.

simplicity, each $\psi_i \in \Psi_u$ can be approximately represented as a (x_i, y_i, ζ_i) , where (x_i, y_i) is the location coordinate; ζ_i represents the semantic attribute of location coordinate (x_i, y_i) , i.e., its semantic location. In this way, user's exposed locations can be transformed into his exposed POIs where he queries LBSs [35]. Indeed, it has been demonstrated that inference of POIs leads to a sever privacy breach [36].

To protect privacy, user usually selects a perturbed POI around him as his real location to request LBSs. Intuitively, to achieve the same level of privacy and LBS accuracy, the perturbed POI should be far more from his real location in a rural area than in downtown. To capture the contextual information in this case, we modify the standard Hilbert curve according to the density of POIs in considering local map and use it to fill the local map (as shown in Figure 3). Figure 3(a) shows the standard Hilbert curve (SHC) that covers the local map. Correspondingly, Figure 3(b) shows the modified Hilbert curve (MHC), which projects every 2-dimension POI of users in local map to a 1-d space. Normally, higher density of POIs leads to finer gains such that every point in the 1-dimension space has homogenous context (e.g., equal density). As we will show in the section of location differential perturbation, this prevents a location at a high POIs density area from receiving too large a perturbation to remain utility for LBSs.

Because of the superior distance preserving properties of Hilbert curve that two adjacent points in 1-dimension space are likely to be close in the original space, and vice versa [15], given a particular point, we can easily discover the adjacent points around. With this property, we first project all POIs in considering local map to 1-dimension space by using MHC. Then we randomly perturb the user's POI where he queries LBS, based on a carefully selected Laplace distribution to guarantee that the probabilities to report the same perturbed POI from a set of likely noise POIs including the true POI are similar. However, the distribution of these noise POIs can affect the proximity of the perturbed POI to the true POI. For instance, if POIs are perturbed based on the locations of every known POI within a city, the scale parameter in the noise distribution will become considerably high, thus leading to heavy noise addition. To solve the problem, we compute the perturbation from a restricted set of k POIs by using the reciprocal framework algorithm [28]. In this way, the probabilities of any POI in these k POIs generating the same perturbed POI are within a small constant factor (up to e^ϵ) of each other.

Formally, let l_1, \dots, l_k represent a set of k noise POIs, one of which is the real POI $l_r = (x_r, y_r)$ of query user, and $p(\cdot)$ indicates the probability density function. For any two POIs l_i and l_j in these k noise, the perturbed POI $l_p = (x_p, y_p)$

Input: \mathcal{R} as the (rectangle) boundary of the map, Ψ as the set of all POIs in the map
Output: a quad-tree root node T

- (1) **if** $|\Psi|$ in \mathcal{R} is great than pre-determined threshold σ , then
- (2) partition \mathcal{R} equally into four sub-cells $\mathcal{R}_{nw}, \mathcal{R}_{ne}, \mathcal{R}_{se}, \mathcal{R}_{sw}$;
- (3) **for** $i=nw, ne, se, sw$ **do**
- (4) **return** 0 //recursively partition \mathcal{R}_i according to the condition $|\Psi_i|>\sigma$;
- (5) **end for**
- (6) **else**
- (7) **return** 1 //outputs quad-tree root node T;
- (8) **end if**

ALGORITHM 1: Modified Hilbert curve construction algorithm.

corresponding to the l_r is generated in a manner such that

$$p(l_p l_i) \leq e^\varepsilon p(l_p l_j), \quad (1)$$

where $\varepsilon > 0$ and $i, j \in [1, k]$. The privacy parameter ε corresponds to the strength of the privacy guarantee: smaller ε yield more privacy. It has been shown that adding noise to each coordinate independently (by applying Laplace noise to each coordinate) provides the stronger protection than adding noise to each point independently (by generating 2-dimensional noise vector) [37]. Therefore, we use a Laplace distribution with scale $b>0$ to perturb each coordinate of the $l_i=(x_i, y_i)$ independently such that

$$\begin{aligned} p(x_p | x_i) &= \frac{1}{2b} e^{-|x_i - x_p|/b}, \\ p(y_p | y_i) &= \frac{1}{2b} e^{-|y_i - y_p|/b}. \end{aligned} \quad (2)$$

The amount of noise to be added to each coordinate is given as $-b * \text{sign}(\text{rnd}) * \ln(1 - 2|\text{rnd}|)$, where rnd is a uniform random value in $(-1/2, 1/2)$. Based on the following observation, x_p is generated by setting b as $(\max_n x_n - \min_n x_n) / \varepsilon$, and y_p is generated by setting b as $(\max_n y_n - \min_n y_n) / \varepsilon$. l_p is obtained as (x_p, y_p) .

Observation 1. Using the triangle inequality, we have $|l_j - l_p| \leq |l_j - l_i| + |l_i - l_p|$. After rearrangement, dividing by b , raising as a power of e and multiplying by $1/2b$, we get

$$\begin{aligned} \frac{1}{2b} e^{-|l_i - l_p|/b} &\leq \frac{1}{2b} e^{-|l_j - l_p|/b} e^{|l_i - l_j|/b} \\ \text{or } p(l_p | l_i) &\leq p(l_p | l_j) e^{|l_j - l_i|/b}. \end{aligned} \quad (3)$$

Therefore, for each coordinate, we have

$$\begin{aligned} p(x_p | x_i) &\leq e^{|x_j - x_i|/b} p(x_p | x_j), \\ p(y_p | y_i) &\leq e^{|y_j - y_i|/b} p(y_p | y_j), \end{aligned} \quad (4)$$

and the power of the exponent is bounded as

$$p(x_p x_i) \leq e^{|\max_n x_n - \min_n x_n|/b} p(x_p x_j), \quad (5)$$

$$p(y_p y_i) \leq e^{|\max_n y_n - \min_n y_n|/b} p(y_p y_j). \quad (6)$$

Consequently, the probability of a POI generating a certain perturbed POI is always with a factor e^ε of the probability of some other POIs in the set of k noise generating the same perturbed POI.

4. Location Differential Perturbation

In this section, we introduce the modified Hilbert curve construction algorithm and location differential perturbation algorithm in detail.

4.1. Modified Hilbert Curve Filling. Without loss of generality, we consider a set of users $\mathcal{U} = \{u_1, u_2, \dots, u_n\}$ who subscribe certain LBSs and move in a local map. The (rectangle) boundary \mathcal{R} of the local map is taken as a large cell. We recursively partition a cell into four equal-size cells if and only if the number of POIs within the original cell is greater than a predetermined threshold σ . One can see that each cell contains roughly σ or fewer POIs. Figure 3(b) depicts an example of such a partition. From the figure, we can see that each cell is either partitioned into four equal-size square cells, or not partitioned (i.e., becoming a base cell). The partitioning scheme can be readily represented as a quad-tree. Figure 3(c) depicts an example of such a quad-tree for the MHC mapping in Figure 3(b). In particular, each node in the tree either is a leaf node (if corresponding to a base cell) or contains four children (if further partitioned). Thus, to efficiently store the tree, we construct a breath-first traversal of the tree, storing 1 bit for each node indicating whether it is a leaf node or not. Since a quad-tree with n leaf nodes has at most $4n/3$ total nodes, the space required by the serialized map file is at most $4n/3$ bits. Thus, the total storage overhead is $O(n)$. One can see that MHC covers the regions of high density of POIs with finer gains.

Algorithm 1 depicts the offline construction of MHC. In the algorithm, we partition original map based on predetermined threshold parameter σ (line (1)), recursively partition their children according to the given conditions and store the quad-tree into a bit stream (line (4)). The computational complexity of Algorithm 1 is $O(n)$.

After the partitioning process, we construct the mapped 1-d space as variation of the Hilbert space-filling curve [38] to connect all cells in the original 2-d space. To assign a corresponding range in the 1-d space for each base cell, we

```

Input: the T obtained from Algorithm 1, starting point  $S_0$  and curve orientation  $\theta$ 
Output: a updated quad-tree root node T
(1) initializes  $S(T) = S_0, \theta(T) = \theta, m = 0;$ 
(2) push T into the stack;
(3) while (stack is not empty) do
(4)   N = pop the top element from the stack
(5)   if (N has child node) then
(6)     for ( $i = sw, se, ne, nw$ ) do
(7)       set  $S(N_i)$ , and  $\theta(N_i)$ 
(8)       push  $N_i$  into the stack
(9)     end for
(10)  else
(11)     $h_N = m$ 
(12)    set the values of all corresponding POIs in the node N as  $m$ 
(13)     $m = m + 1$ 
(14)  end if
(15) end while
(16) outputs the updated quad-tree root node T.

```

ALGORITHM 2: Hilbert value generation algorithm for each base cell.

need to traverse every leaf node. To this end, we conduct a depth-first traversal of tree T, assigning the Hilbert value in the 1-d projected space for each leaf node according to its visiting orders. Let $S(N), \theta(N)$ be the orientation and starting point of the Hilbert curve of the node N. The Hilbert value corresponding to the node N is denoted by h_N . The formal description of our Hilbert value generation algorithm can be found in Algorithm 2.

In Algorithm 2, we construct a depth-first traversal over the quad-tree. In particular, we start from the root node T (lines (1)-(2)) and determine its every child node's curve orientation θ and starting point S in the manner of drill-down according to the fractal rules used in our recent work [39] (lines (7)-(8)). We repeat this process until reaching a leaf node and set the Hilbert value of this leaf node as m (line (11)). In such way, every leaf node is assigned to a unique Hilbert value. Correspondingly, the Hilbert value of all POIs in every leaf node is also obtained (line (12)). The computational complexity of Algorithm 2 is $O(n)$.

4.2. Location Perturbation Algorithm. In Section 3.2, we provide a method to generate a perturbed POI for query point from the carefully selected k POIs by using the Laplace distribution. As mentioned before, the k POIs should be chosen to preserve reciprocity. That is, the same anonymous set should be obtained irrespective of which of the k POIs is the query point. This is achieved by using the reciprocal framework algorithm [28], which partitions the POIs of user into k size buckets based on the Hilbert value of the POIs. The anonymous set is selected as the bucket to which the query point belongs. Each of the k POIs is used for perturbation and the one having the minimum average distance to all POIs in the anonymous set is chosen as the user's location to issue the query. The formal description of our location perturbation algorithm can be found in Algorithm 3.

In Algorithm 3, we firstly index the all-possible POIs by a quad-tree spatial index and assign the Hilbert value for each POI (Line (1)). This step has time complexity of $O(n)$. Then we find the mapped value based on the 1-d value range of the base cell which contains ψ_u (line (2)). One can see that the retrieval process requires access at most $\log n$ (the depth of tree) nodes, leading to computational complexity of $O(\log n)$. Based on the Hilbert indices of the POIs, we determine the k size bucket to which the ψ_u belongs by using reciprocal framework (lines (3)-(9)), which has time complexity of $O(\log n)$. The locality preserving properties of Hilbert curves guarantee the formation of buckets with POIs that are at close proximity to each other. Lines (10)-(14) compute a perturbed value corresponding to the k POIs in the bucket by using Laplace distribution. Thus, each coordinate c of a POI is perturbed to $c - b * \text{sign}(\text{rnd}) * \ln(1 - 2|\text{rnd}|)$, where rnd is a uniform random value in $(-1/2, 1/2)$, and b is set as $(\max_n c_n - \min_n c_n) / \epsilon$. This makes perturbation Laplace distributed around c . In the following experiment, the retrieval of the MHC mapping requires less than 0.1s in our system, and the perturbation requires less than 0.5s.

4.3. Security Analysis. In this section, we provide security analysis. In the context of location privacy, we consider two types of adversaries: active adversary and passive adversary. The purpose of the passive adversary is to obtain sensitive information about a particular user by eavesdropping on the wireless channel or compromising the LBS provider. Actually we can use some cryptography tools such as public key infrastructure (PKI) to cope with the eavesdropping attacks on the wireless channel between users and other entities. Thus, we mainly focus on how to avoid collusion attacks and inference attacks from active adversary, both of which can cause serious privacy problems.

Adversary may be collusion with some users or the LBS server to capture the other user's private information.

Input: Query user u with associated k , his POI ψ_u where he queries LBSs, and the pre-computed MHC filling file $mhcFile$

Output: A perturbed location l_p for u

- (1) load a quad-tree T of the partition from $mhcFile$ and use Algorithm 2 to assign the Hilbert value for each POI;
- (2) find the leaf node N containing ψ_u ;
- (3) **while** (there is non-empty node at the same level as N with $< k$ POIs) **do**
- (4) $N = \text{parent of } N$ //bottom-up traversal
- (5) **end while**
- (6) **while** (N is not a leaf and (each child of N is either empty or contains $\geq k$ POIs)) **do**
- (7) $N = \text{child of } N$ that contains ψ_u //top-down traversal
- (8) **end while**
- (9) obtain the $L = \{l_1, \dots, l_r, \dots, l_k\}$ by splitting the POIs inside sub-tree of N into buckets containing between k and $2k-1$ POIs using reciprocal algorithm
- (10) $L_p = \phi$
- (11) **for** ($l \in L$) **do**
- (12) $l_p = l + z_i$, where z_i is additive noise generated by Laplace distribution
- (13) $L_p = L_p \cup \{l_p\}$
- (14) **end for**
- (15) output $l_p \in L_p$ such that l_p has minimum average distance from L .

ALGORITHM 3: Location perturbation algorithm.

Theorem 1. *Our scheme is collusion attack resistant.*

Proof. We contemplate that the collusion attack occurs between a set of users. On the one hand, each user is independent with others in our scheme. He only needs to use his position and the stored Hilbert index file to generate the perturbation instead of interacting with the other users. On the other hand, Algorithm 3 in our scheme guarantees that all the processes are executing locally, not dependent on other users at all. That is, it is useless for the adversary to capture and collude with nearby users. The best case to this kind of adversary is that he can obtain the global information by capturing the LBS server and all the users, but in this case he becomes an active adversary to perform inference attack.

In our scheme, we directly contemplate the untrusted LBS server as the active adversary to perform the inference attack. He can get *side information* by monitoring all the users in the system, including their interests, approximately location (e.g., a set of likely positions including the true location), LBS queries, and observed perturbation. His aim is to use this *side information* to confidently infer real position of the query user. \square

Theorem 2. *Our scheme is inference attack resistant under ϵ -differential privacy.*

Proof. In our scheme, users need to issue the queries to the adversary in order to enjoy the LBSs. Ideally, due to the perturbation, the adversary cannot construct any direct linkage from the perturbed locations to a user. However, the adversary knows the POIs density of the whole map, approximately locations for a user and noise distribution. Based on this information, the adversary can perform inference attacks to gain the real location of the query user. More formally, the adversary knows the set of all POIs, Ψ , a set of positions, $l_1, \dots, l_r, \dots, l_k$ (including query user's real location), location

perturbation mechanism, and the noise distribution $p(l_i)$. As certain position in the adversary's approximate knowledge is highly unlikely to generate the observed perturbation under the used noise distribution, the adversary can use newly learned distribution to improve its probability of successfully guessing the real location from these equally likely positions. In our algorithm, the inference attack is avoided by using reciprocity framework and ϵ -differential privacy. First, since the k POIs set $L = \{l_1, \dots, l_r, \dots, l_k\}$ generated by using MHC method satisfies reciprocity, the probability of identifying the query user's real POI does not exceed $1/k$ [28]. Second, as discussed above, due to the usage of differential privacy, the probability to report the same observed perturbed location l_p from the positions $l_1, \dots, l_r, \dots, l_k$ is within a small constant factor of each other. The Laplace noise added to a POI depends on the component-wise maximum distance between two positions. As long as the scale parameters use these maximum, the perturbed POI (x_p, y_p) will satisfy the probability ratio. In our scheme, we use differential privacy notion to abstract the *side information* and guarantee the security efficiently. For any two POIs l_i and l_j in the set L , the adversary's *side information* can be modelled by two prior distributions $p(l_i)$ and $p(l_j)$. After observing the perturbed POI l_p , the adversary could use the l_i and l_j as input to differential location perturbation algorithm and compute the conditional probabilities $p(l_p | l_i)$ and $p(l_p | l_j)$. For the purpose of modelling the adversary's observation, we use Bayes' rule to obtain the posterior distribution:

$$p(l_i = l_r | l_p) = \frac{p(l_p | l_i) p(l_i)}{\sum_{c=1}^k p(l_p | l_c) p(l_c)},$$

$$p(l_j = l_r | l_p) = \frac{p(l_p | l_j) p(l_j)}{\sum_{c=1}^k p(l_p | l_c) p(l_c)}. \quad (7)$$

We use multiplicative distance to metric the distance between two distributions as

$$d_p = \sup_{i \leq L} \left| \ln \frac{p(l_i | l_p)}{p(l_j | l_p)} \right| \leq \varepsilon \quad (8)$$

According to the definition of ε -differential privacy, the d_p should be at most ε . Substituting formula (7) into formula (8), we can get

$$p(l_p | l_i) \leq e^{R/b} p(l_p | l_j), \quad (9)$$

where R is the radius of maximum perturbation range which satisfies Laplace distribution, also the maximum distance between any two noise POIs, and b is the scale parameter of Laplace distribution. Thus, formula (9) can be extended to formulas (5) and (6).

As can be seen from formula (9), our scheme is independent of the prior distribution. This is to say, the probabilities that the adversary uses *side information* to report the same observed perturbed location l_p from the positions $l_1, \dots, l_r, \dots, l_k$ are within a small constant factor of each other. Thus, the adversary cannot use such *side information* to improve its probability of successful guessing the real location. \square

4.4. Query Accuracy Analysis. In this section, we provide LBS query accuracy analysis. Using perturbed locations do affect the accuracy of query results. However, difference in the results may or may not exist depending on the distance between the perturbed location and the true location. Therefore, one has to trade-off between location privacy and LBS utility. In order to formally analyze the query accuracy of our location perturbation scheme, we consider three metrics with respect to KNN query [8]: *Nearness*, *Resemblance*, and *Displacement*.

- (i) *Nearness*: it indicates the ratio of perturbation at close proximity to the true location.
- (ii) *Resemblance*: it depicts the accuracy rate of query results retrieved by a KNN query related to a perturbed location. Let $O = \{o_1, o_2, \dots, o_K\}$ be the objects retrieved by a KNN-query relative to the true location of user u , and $O' = \{o'_1, o'_2, \dots, o'_K\}$ be the objects retrieved relative to the perturbed location. The resemblance is the rate of common objects between O and O' , given as

$$Q_{AR} = \frac{|O \cap O'|}{|O|}, \quad (10)$$

where $|O|$ is the number of query objects in the real results set O , $|O \cap O'|$ is the number of common objects between O and O' .

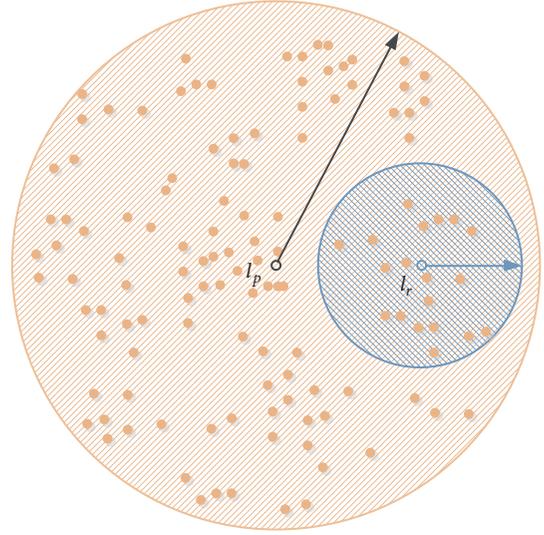


FIGURE 4: AOI and AOR with centre l_r and l_p .

- (iii) *Displacement*: it captures the average difference in distance between the actual results and real retrieval results of a KNN-query, given as

$$Q_P = \begin{cases} \frac{1}{K} \left[\sum_{i=1}^K \text{dist}(O'_i, q) - \sum_{i=1}^K \text{dist}(O_i, q) \right], & O \neq O' \\ 0, & O = O', \end{cases} \quad (11)$$

where q is the real query POIs of a user and the $\text{dist}(\cdot)$ is the Euclidean distance between an object's location and the true location of a user. It should be noted that the lowercase k is used to calculate anonymous set and the uppercase K is used to calculate KNN query.

These three metrics are used to measure the effectiveness of our scheme. The resemblance measures the query accuracy with respect to the perturbed location, while the displacement measures the expected distance error between the real query results and actual retrieval results. In this part of theoretical analysis, we adopt the *Resemblance* metric as the query accuracy measure. Nevertheless, in the following experimental evaluation, we also evaluate the *Nearness* and *Displacement* metrics.

As shown in Figure 4, we define the blue circle as the query area of interest (AOI) with regard to the real location l_r and the orange circle as the area of retrieval (AOR) with respect to the perturbed location l_p . In order to guarantee high *Resemblance* (see formula (10)), ideally, the AOR should always completely contain the AOI. Unfortunately, this condition cannot be guaranteed because of the nature of our location perturbation (note that the AOR is centred on a randomly generated location that can be arbitrarily distant from the real location). In order to measure the probability of such event, we introduce the notion of accuracy. Specifically, we use r_I and r_R to represent the radius of the AOI and

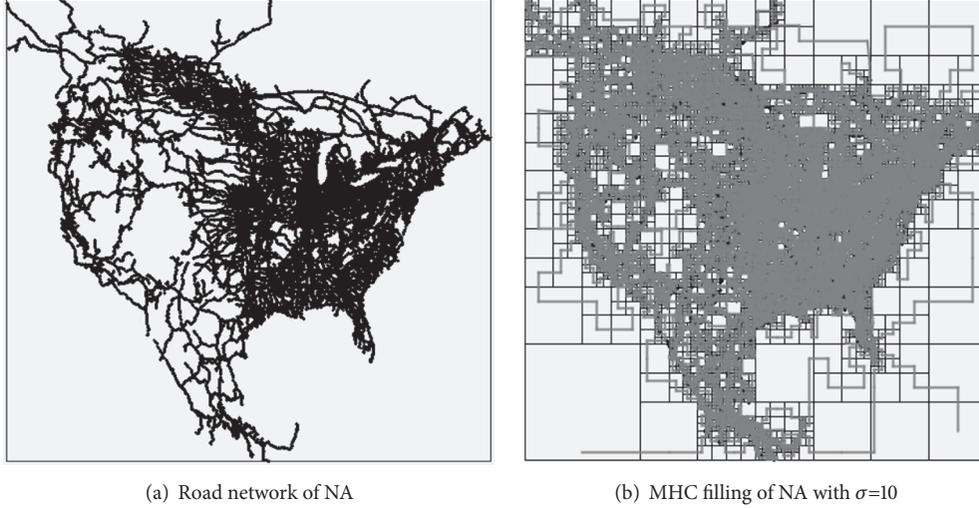


FIGURE 5: NA dataset.

the AOR, respectively, M to denote the location perturbation mechanism, and $C(x, r)$ to denote the circle with centre x and radius r .

Definition 3. An LBS perturbation is (c, r_I) -accurate iff for all locations x we have that $C(x, r_I)$ is fully contained in the $C(M(x), r_R)$ with probability at least c .

Give a privacy parameter ϵ and accurate parameters (c, r_I) , our goal is to obtain an LBS perturbation (M, r_R) satisfying both ϵ -differential privacy and (c, r_I) -accurate. As for a perturbation mechanism M , we use the Laplace perturbation M_ϵ discussed in Section 3.2, which satisfies ϵ -differential privacy. As for r_R , we attempt to find a minimum value validating the accurate condition. To achieve this goal, we use the notion of (α, δ) -usefulness, which was introduced in [40]. A location perturbation mechanism M is (α, δ) -usefulness if for every location x the perturbed location $z = M(x)$ satisfies $dist(x, z) \leq \alpha$ with probability at least δ . In our perturbation mechanism M_ϵ , we computer the perturbation from a restricted set of k POIs that preserve reciprocal. This guarantee that our M_ϵ can generate reasonable perturbation range. Therefore, the α and δ values which express M_ϵ usefulness are related by $-b \cdot \text{sign}(rnd) \cdot \ln(1-2|rnd|)$, the noise amount of our perturbation.

Observation 2. For any $\alpha > 0$, M_ϵ is (α, δ) -usefulness if $\alpha < \max_{i,j \in [1,k]} dist(l_i, l_j)$, where l_i and l_j are determined by $-b \cdot \text{sign}(rnd) \cdot \ln(1-2|rnd|)$.

In the following experimental evaluation (as shown in Table 2), we set various ϵ to compute the percentage of the perturbations which are within 1km, 0.5km, and 0.1km of the user's true position. As our running example, our perturbation mechanism M_ϵ (ϵ -differential privacy, with $\epsilon = 0.5$) generates a perturbed location l_p falling within 1km of the real position l_r with probability 0.9426.

According to the definition of usefulness, if M_ϵ is (α, δ) -usefulness, then the LBS perturbation (M_ϵ, r_R) is (δ, r_I) -accurate if $\alpha < \max_{i,j \in [1,k]} dist(l_i, l_j)$. The converse also holds if δ is maximal. By Observation 2, we have the following.

Proposition 4. The LBS perturbation (M_ϵ, r_R) is (δ, r_I) -accurate if $r_R \geq r_I + \delta \cdot \max_{i,j \in [1,k]} dist(l_i, l_j)$.

Therefore, it is sufficient to set $r_R = r_I + \delta \cdot \max_{i,j \in [1,k]} dist(l_i, l_j)$. Thus, our perturbation (M_ϵ, r_R) satisfies both ϵ -differential privacy and (δ, r_I) -accurate.

5. Experimental Evaluation

This section evaluates the proposed differential location perturbation algorithms. We implemented the algorithms using Java program. All experiments were executed on an Intel Core i7-4790 3.6GHz machine with 4G RAM and Windows OS. The perturbation scheme indexes the all-possible POIs of the considering local map, which are taken from the NA dataset (available at <http://www.cs.utah.edu/~lifeifei/SpatialDataset.htm>) containing 175813 real POIs of the North America road network (see Figure 5(a)). The parameter k is set from 10 to 1000. The results are obtained by taking the average of 100 times simulation of the corresponding algorithms.

Several parameters are employed in our evaluation. S_0 is the starting point of Hilbert curve, and its default value is $(0, 0)$. The θ represents the Hilbert curve direction, and its default value is D_1 (see Figure 6) [39]. Γ is the scale factor of Hilbert curve, and its default value is 1. k is related to k -anonymity. Figure 5(b) shows a real MHC filling for the North America road network with $\sigma=10$. From the figure, an intuitive observation is that the denser regions represent large cities, while the sparse regions represent the rural areas. Thus, MHC mapping captures the contextual information well. In [9], Shokri proposes an optimal location privacy preservation strategy by solving a linear program, which avoids TTP.

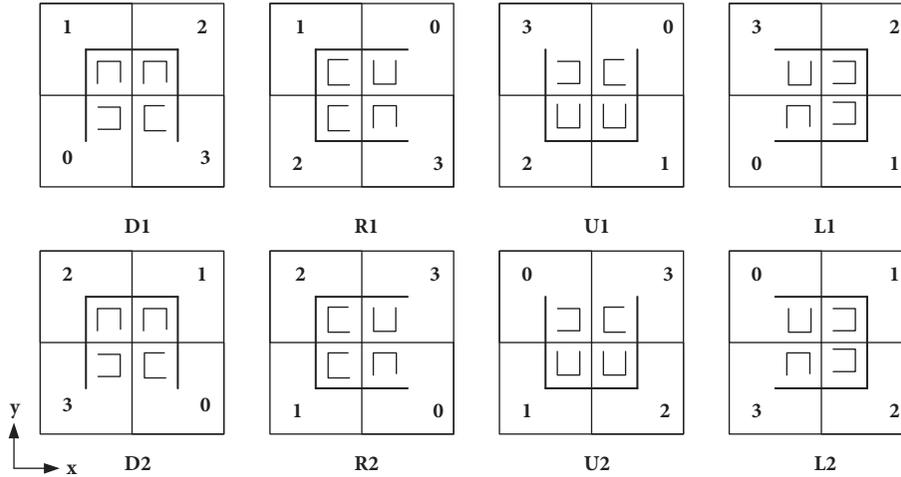
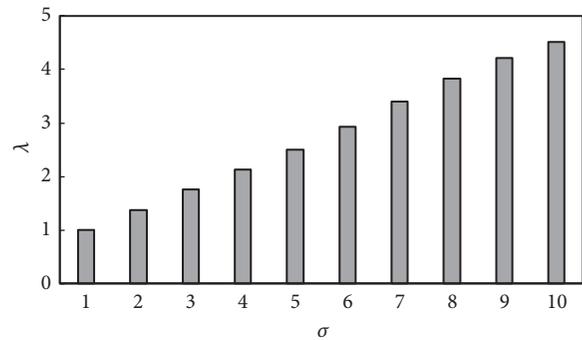


FIGURE 6: Fractal rules of Hilbert curve.

This optimal strategy computes location obfuscation probability distribution function to maximize the location privacy, subject to service quality constraints. However, this method depends on the modelling of adversary's side information and thus suffers from background knowledge attacks. As can be seen from previous analysis, our method abstracts adversary's side information. Furthermore, this optimal strategy has nothing to do with contextual information and Hilbert curve. Therefore, our method cannot be comparable to this optimal strategy. Location perturbation method in [8] is similar to our approach and uses SHC, whereas it still employs a TTP. For the purpose of comparing with [8], we implement the method in [8] under the same setting as our method. To generate the k -anonymous sets, the location perturbation scheme in [8] employs SHC to calculate the Hilbert index value of users' location online. Different from the method in [8], our scheme employs MHC to calculate the Hilbert indices of the POIs and stores them as a binary map file. We use the quad-tree recovered from the binary map file to find the node where the user's POIs are located and then generate the k -anonymous set. Therefore, in the following experiment, we evaluate the performance of the location perturbation by comparing MHC mapping with SHC mapping.

5.1. Parameters Selection for MHC and SHC. During the partitioning process, MHC and SHC employ different curve parameters. To carry out the following experiments under the same standard, we first examine the parameters selection for MHC and SHC. When the geographic space is filled by using MHC mapping, a unique index value is assigned to each atomic region according to the traversal order of the Hilbert curve. The index values of the POIs contained in the atomic region are also index value of the atomic region. Thus, we can obtain the Hilbert index of all POIs. If some POIs are in the same base cell, they are overlapped. We define the overlap factor λ to describe the overlap of the POIs for each base cell such that

$$\lambda = \frac{1}{M} \sum_{i=0}^H n_i \quad (12)$$

FIGURE 7: The relationship between σ and λ .

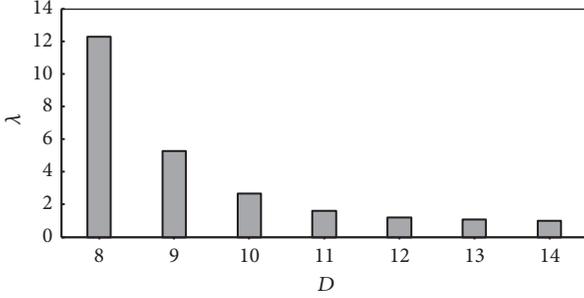
where M is the number of base cells that contain POIs, H is the upper bound of the Hilbert index value, and n_i represents the number of the POIs whose index value is equal to i .

Figure 7 illustrates the case that the MHC overlaps factor λ changes with the division threshold σ . We find that λ grows slowly as σ increases, and when σ is 1, the overlap factor is also 1. This is determined by the definition of the overlap factor and the division in Algorithm 2. The MHC can achieve finer grains via setting the threshold σ .

Figure 8 illustrates the relationship between the standard Hilbert curve degree D and the overlap factor λ when the map of the NA dataset is divided by the SHC. As can be seen from the figure, the larger D is, the smaller λ is and the finer grain that the partition leads to. Since the SHC employs the uniform standard to divide the space, the λ changes greatly with the changing of the D when D is small. In [8], the SHC mapping technique was employed to divide the entire map into a grid of $2^{14} * 2^{14}$ while calculating the Hilbert indices, which guarantees that there is not more than one user in each division. Objects in the same division have the same Hilbert index. This is because a larger curve degree D can lead to a finer granularity division of spatial maps. Nevertheless, the greater curve degree may lead to high computational overhead unaffordable to the server.

TABLE 1: Hilbert index generation time (ms).

Algorithm	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2.7$	$\lambda = 4.9$
	$\sigma=1,$ $D=13$	$\sigma=2,$ $D=11$	$\sigma=5,$ $D=10$	$\sigma=10,$ $D=9$
SHC	1237	1067	988	903
MHC	892	521	300	214

FIGURE 8: The relationship between λ and D .

The generation time of the index of POIs is an important measure when using the spatial filling curve to divide the considered map. We compare the index generation time of our scheme using MHC padding algorithm HVGA with the scheme using SHC padding algorithm EDHO in [8]. The HVGA represents the Hilbert value generation algorithm (see Algorithm 2). The results are shown in Table 1.

As seen from the table, in the case of the same λ , the efficiency of Hilbert index generation via using the MHC mapping technology in our scheme is significantly higher than that via using SHC mapping technology in [8], and with the increase of λ , the result of using MHC is more obvious. This is because the MHC partition considers the density distribution of POIs and uses different curve degree D for different density regions, which enable the partitioning of the lower density region not use high D , thus improving the efficiency of index generation.

When $\sigma \geq 10$, the MHC index generation time changes very slowly. Therefore, in all the following experiments we considered the $\lambda = 4.9$, $\sigma = 10$, $D = 9$.

5.2. Anonymous Evaluation. We compare the average anonymous set generation time of our scheme and the scheme in [9] for varying k (see Figure 9).

As can be seen from Figure 10, as the k increase, the anonymous set generation time for MHC perturbation (see Algorithm 3) and SHC perturbation (see [8]) does not vary significantly. This is because, in both perturbation algorithms, to select the k -anonymous set that satisfies reciprocity we only need to traverse the small subtree determined by node N in the quad-tree T , and the data structure of the intermediate nodes of the quad-tree T contains all the POIs in its subtree, so there is no need to traverse their subtrees to obtain this information. In the case of the same k , the anonymous sets generation time of MHC perturbation is much lower than that of SHC perturbation, with an average reduction about

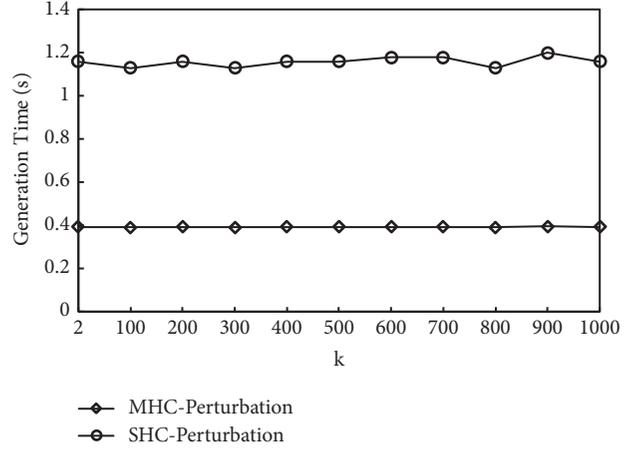
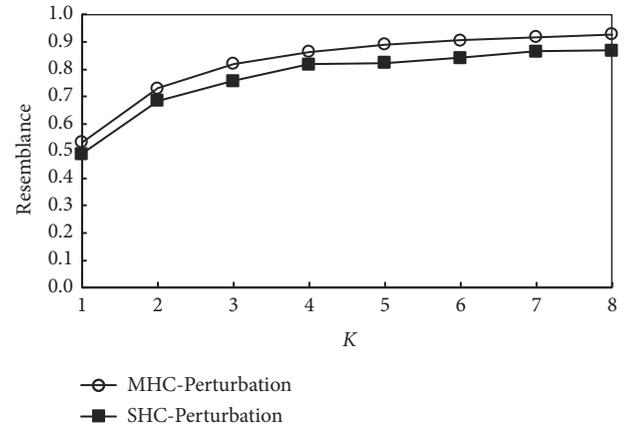
FIGURE 9: Anonymous set generation time for varying k .

FIGURE 10: Query accuracy rate of KNN retrieval.

66%. This is because when the location index is generated, the SHC perturbation is partitioned by the uniform granularity for all the regions; nevertheless, the MHC is divided according to the density distribution of the POIs. The MHC partition of the sparse regions uses lower curve orders and thus reduces the division time. That is to say, in the same case, the MHC partition traverses fewer subtrees than the SHC partition.

5.3. Differential Location Perturbation Evaluation. In this section, we evaluate the performance of our differential location perturbation algorithm by comparing MHC mapping with SHC mapping. From Section 3.2, we know that the probability ratio of generating the perturbed position (x_p ,

TABLE 2: Percentage of the generated perturbed positions that is at close proximity to true location.

ϵ	Nearness/%		
	$d \leq 1000\text{m}$	$d \leq 500\text{m}$	$d \leq 100\text{m}$
0.01	2.39	0.95	0.01
0.1	28.2	16.27	8.13
0.3	89.3	61.24	7.18
0.5	94.26	62.20	11.48
1.0	99.04	73.68	27.27
2.0	99.99	91.87	33.97

y_p) at any two POIs in the anonymous is bounded as e^ϵ . The amount of Laplace noise to be added to the position depends on the maximum distance of the corresponding coordinates of the two positions. As long as the Laplace scale parameter b uses these maximum distances, the probability ratio of the any two POIs in anonymous set generating a perturbed position (x_p, y_p) always satisfies the bounded e^ϵ . ϵ is privacy budget and smaller ϵ yields more privacy, but leading to less accuracy. In the following experiment, we evaluate the accuracy of our scheme for a scenario, where we issue a KNN query for nearest POIs. In particular, we use the *Nearness*, *Resemblance*, and *Displacement* metrics to measure LBS accuracy.

(1) **Nearness:** for the Nearness metric, we set the different privacy parameter ϵ to calculate the percentage of the perturbations that resulted in the perturbed point being generated within 1000 m, 500 m, and 100 m of the user's true position. The results are shown in Table 2.

As can be seen from the table, a value of $\epsilon = 0.01$ indicates that two users should have the same probability ($e^\epsilon = 1.01$) to generate perturbations. This is difficult to achieve most values of k . When the ϵ value reaches 0.5 ($e^{0.5} = 1.65$), more than 90 percent of the perturbed points are within 1000 meters of the real position. More than 60 percent of the perturbed points are within 500m of the true position. The number of perturbed points increases with increasing of the ϵ value. However, higher ϵ value reduces the practical significance of the approach. For example, the value of $\epsilon = 2.0$ means that a factor of 7 differences in the probability estimates ($e^{2.0} = 7.39$) must be accepted. Nonetheless, high nearness values with smaller value of ϵ are also possible as well.

(2) **Resemblance and Displacement:** as previously observed in Table 2, about 95% perturbed points fall within 1km of the real position when $\epsilon=0.5$. Therefore, for the resemblance metric Q_{AR} and displacement metric Q_p , we set $\epsilon = 0.5$ to generate the perturbations in the experiments. Figures 10 and 11 show the evaluation results of the *Resemble* and *Displacement* corresponding to different values of K (the number of the nearest neighbour objects retrieved by KNN).

From Figure 10, we can see that increasing the number of nearest neighbouring objects to search K enhances the similarities of the result set. As K increases, the query accuracy rate of KNN retrieval varies from around 60 percent to almost 90 percent. This is because that a greater number of retrieved results can be seen as enlarging the search radius, in which case, an object becomes more likely in the KNN

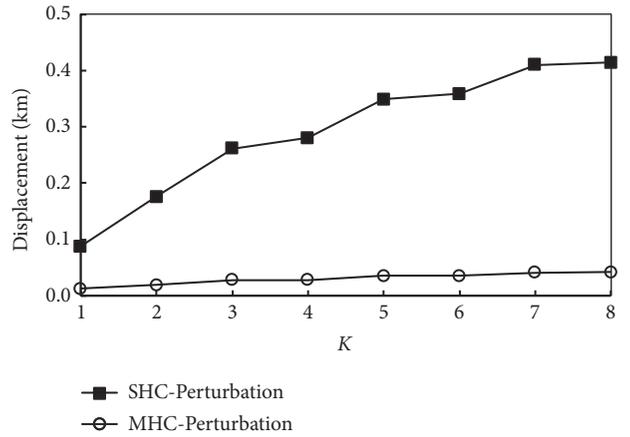


FIGURE 11: Query precision of the approximate KNN search.

set of more number of location-based queries. For example, searching for the nearest theatres from two different locations in a city, we can expect that these two different locations have higher overlap in their list of 10 nearest theatres. The extent of the overlap depends on how proximal the two locations are to each other. Therefore, the noise added to a location is important in this regard. Meanwhile, the retrieval accuracy of a KNN using MHC is increased by an average 15.4% compared with the approach using SHC. This is because the MHC partition considered the density distribution of the POIs, which needs only a small perturbation to achieve a high level of privacy-preserving in a densely populated area.

The query precision indicates the average difference in the distance between the actual results and real retrieval results of a KNN-query based on the real location and the perturbed position, which is more effective in measuring the quality of retrieved results. Figure 11 shows the results of the query precision.

From Figure 11, we can clearly see that the query precision of a KNN retrieval related to the perturbed location generated by using SHC perturbation varies from about 120m to 350m and that the query precision of a KNN retrieval related to the perturbed location generated by using MHC perturbation is within 50m. This shows that the query precision of a KNN retrieval related to MHC perturbation is smaller than that of a KNN retrieval related to SHC perturbation. The reason is that MHC considers the contextual information of the POIs, thereby resulting in smaller perturbation than

SHC. The results show that the quality of a KNN retrieval results related to MHC perturbation is higher than that of the results related to SHC, which also corresponds with the character that MHC conducts granularity partition of the defined spaces according to density distribution of the POIs.

6. Conclusion

Driven by the prosperity of smart mobile devices equipped with GPS, location-based services, as an import part of Green Mobile Communications and Networks (GMCNs), have become very popular recently in almost all business and society domain. Since these services access private position information, location privacy protection mechanisms are mandatory to ensure the user acceptance of such services. The location-based confounding mechanism based on the cloaking area is a wide range of research techniques to achieve location privacy protection, but most of these technologies rely on TTP and assume that the attacker does not have side information, thus easy to cause location privacy disclosure.

In this paper, we proposed a context-aware differential location perturbation technique to protect user privacy. Our scheme, the context information of the user's location is considered in the event of a perturbation, and the attack of the background information can be effectively prevented without depending on any TTP. We use MHC mapping technology to project each 2-d geographic location of the user on the map into 1-d space and combine the k anonymous with the differential privacy techniques to randomly disturb the user's location, and then to submit the perturbation as the user's real location to the location service provider. In order to solve the limited resources of mobile devices, we use a quad-tree based approach to transform and to store the users' context to support efficient retrieval and storage. Through the security analysis and experimental evaluation, we can find that our scheme can resist the inference attacks of approximate position knowledge. Using the perturbation position will not significantly improve the attacker's prior knowledge about the user's position, so it has strong privacy protection. However, the identification of some unreasonable perturbation is still a problem. In the future work, we will consider abandoning the anonymity sets to address this problem.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 61762058, in part by the Scientifically and Technological Project of Gansu

Province under Grant 1610RJZA056, in part by the Key Laboratory Opening Project of Opto-Technology and Intelligent Control Ministry of Education under Grant KFKT2016-7, and in part by the Youth Science Foundation of Lanzhou Jiaotong University under Grant 2014026.

References

- [1] M. Ismail, W. Zhuang, E. Serpedin, and K. Qaraqe, "A survey on green mobile networking: From the perspectives of network operators and mobile users," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1535–1556, 2015.
- [2] X. J. Zhang, X. L. Gui, and Z. D. Wu, "Privacy preservation for location-based services: a survey," *Journal of Software*, vol. 26, no. 9, pp. 2373–2395, 2015, (in Chinese with English abstract).
- [3] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proceedings of the 1st International Conference on Mobile Systems, Applications and Services, MobiSys 2003*, pp. 31–42, May 2003.
- [4] S. Oya, C. Troncoso, and F. Pérez-González, "Back to the drawing board: Revisiting the design of optimal location privacy-preserving mechanisms," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pp. 1959–1972, Dallas, Texas, USA, October 2017.
- [5] X. Pan, J. L. Xu, and X. F. Meng, "Protecting location privacy against location-dependent attacks in mobile services," *IEEE Transactions on Knowledge and Data Engineering*, vol. 8, no. 24, pp. 1506–1519, 2012.
- [6] N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Optimal geo-indistinguishable mechanisms for location privacy," in *Proceedings of the 21st ACM Conference on Computer and Communications Security, CCS 2014*, pp. 251–262, November 2014.
- [7] R. Schlegel, C.-Y. Chow, Q. Huang, and D. S. Wong, "User-defined privacy grid system for continuous location-based services," *IEEE Transactions on Mobile Computing*, vol. 14, no. 10, pp. 2158–2172, 2015.
- [8] R. Dewri, "Local differential perturbations: location privacy under approximate knowledge attackers," *IEEE Transactions on Mobile Computing*, vol. 12, no. 12, pp. 2360–2372, 2013.
- [9] R. Shokri, G. Theodorakopoulos, C. Troncoso, J.-P. Hubaux, and J.-Y. Le Boudec, "Protecting location privacy: Optimal strategy against localization attacks," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security, CCS 2012*, pp. 617–626, October 2012.
- [10] H. Lu, C. S. Jensen, and M. L. Yiu, "Pad: privacy-area aware, dummy-based location privacy in mobile services," in *Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access*, pp. 16–23, Vancouver, Canada, June 2008.
- [11] C. Y. Ma, D. K. Yau, N. K. Yip, and N. S. Rao, "Privacy vulnerability of published anonymous mobility trace," in *Proceedings of the 16th Annual International Conference on Mobile Computing and Networking*, pp. 186–196, Chicago, IL, USA, 2010.
- [12] X. Liu, K. Liu, L. Guo, X. Li, and Y. Fang, "A game-theoretic approach for achieving k -anonymity in location based services," in *Proceedings of the 32nd IEEE Conference on Computer Communications (INFOCOM '13)*, pp. 2985–2993, IEEE, Turin, Italy, April 2013.
- [13] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, "Achieving k -anonymity in privacy-aware location-based services," in *Proceedings of*

- the 33rd IEEE Conference on Computer Communications, IEEE INFOCOM 2014, pp. 754–762, Toronto, Canada, May 2014.
- [14] A. Khoshgozaran, C. Shahabi, and H. Shirani-Mehr, “Location privacy: Going beyond K-anonymity, cloaking and anonymizers,” *Knowledge and Information Systems*, vol. 3, no. 26, pp. 435–465, 2011.
- [15] A. Pingley, W. Yu, N. Zhang, X. Fu, and W. Zhao, “CAP: A context-aware privacy protection system for location-based services,” in *Proceedings of the 29th IEEE International Conference on Distributed Computing Systems Workshops (ICDCS '09)*, pp. 49–57, Montreal, Canada, June 2009.
- [16] K. G. Shin, X. Ju, Z. Chen, and X. Hu, “Privacy protection for users of location-based services,” *IEEE Wireless Communications Magazine*, vol. 19, no. 1, pp. 30–39, 2012.
- [17] C. Dwork, “Differential privacy,” in *Proceedings of the International Colloquium on Automata, languages and Programming*, vol. 4052, Springer, Berlin, Germany, 2006.
- [18] N. Vratonjic, K. Huguenin, V. Bindschaedler, and J.-P. Hubaux, “How others compromise your location privacy: The case of shared public IPs at hotspots,” in *Proceedings of the Privacy Enhancing Technologies Symposium – PETS*, vol. 7981, pp. 123–142, 2013.
- [19] J. Freudiger, R. Shokri, and J. Hubaux, “Evaluating the privacy risk of location-based services,” in *Proceedings of the Financial Cryptography and Data Security*, pp. 31–46, 2012.
- [20] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkatasubramanian, “L-diversity: privacy beyond k-anonymity,” in *Proceedings of the 22nd International Conference on Data Engineering (ICDE '06)*, pp. 1–12, Atlanta, Ga, USA, April 2006.
- [21] N. Li, T. Li, and S. Venkatasubramanian, “t-closeness: privacy beyond k-anonymity and l-diversity,” in *Proceedings of the 23rd International Conference on Data Engineering*, pp. 1–12, Istanbul, Turkey, April 2007.
- [22] J. Meyerowitz and R. Roy Choudhury, “Hiding stars with fireworks: location privacy through camouflage,” in *Proceedings of the 15th Annual ACM International Conference on Mobile Computing and Networking (MobiCom '09)*, pp. 345–356, Beijing, China, September 2009.
- [23] X. Zhang, X. Gui, and F. Tian, “A Framework for measuring query in Location-based Service,” *KSI Transactions on Internet and Information Systems*, vol. 5, no. 9, pp. 1717–1732, 2015.
- [24] R. Shokri, G. Theodorakopoulos, J. Boudec, and J. Hubaux, “Quantifying location privacy,” in *Proceedings of the IEEE Symposium on Security and Privacy*, pp. 247–262, Oakland, CA, USA, 2011.
- [25] X. Zhang, X. Gui, F. Tian, S. Yu, and J. An, “Privacy quantification model based on the Bayes conditional risk in Location-Based Services,” *Tsinghua Science and Technology*, vol. 5, no. 19, pp. 452–462, 2014.
- [26] Y. H. Xiao and X. Li, “Protecting location with dynamic differential privacy under temporal correlations,” in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, pp. 1298–1309, 2015.
- [27] K. Dong, T. Guo, H. Ye, X. Li, and Z. Ling, “On the limitations of existing notions of location privacy,” *Future Generation Computer Systems*, vol. 86, pp. 1513–1522, 2018.
- [28] G. Ghinita, K. Zhao, D. Papadias, and P. Kalnis, “A reciprocal framework for spatial K-anonymity,” *Information Systems*, vol. 3, no. 35, pp. 299–314, 2010.
- [29] C.-Y. Chow, M. F. Mokbel, and X. Liu, “A peer-to-peer spatial cloaking algorithm for anonymous location-based service,” in *Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems (ACM-GIS '06)*, pp. 171–178, New York, NY, USA, November 2006.
- [30] G. Ghinita, P. Kalnis, and S. Skiadopoulos, “MOBIHIDE: A mobile peer-to-peer system for anonymous location-based queries,” in *International Symposium on Spatial and Temporal Databases*, vol. 4605, pp. 221–238, Boston, MA, USA, 2007.
- [31] J. Manweiler, R. Scudellari, and L. P. Cox, “SMILE: Encounter-based trust for mobile social services,” in *Proceedings of the 16th ACM Conference on Computer and Communications Security, CCS'09*, pp. 246–255, November 2009.
- [32] G. Ghinita, P. Kalnis, and S. Skiadopoulos, “PRIVE: anonymous location-based queries in distributed mobile systems,” in *Proceedings of the 16th International World Wide Web Conference (WWW '07)*, pp. 371–380, May 2007.
- [33] B. Niu, Q. Li, X. Zhu, G. Cao, and H. Li, “Enhancing privacy through caching in location-based services,” in *Proceedings of the 34th IEEE Annual Conference on Computer Communications (IEEE INFOCOM '15)*, pp. 754–762, IEEE, May 2015.
- [34] V. Primault, S. B. Mokhtar, C. Lauradoux, and L. Brunie, “Differentially Private Location Privacy in Practice,” 2015, <https://arxiv.org/abs/1410.7744?context=cs>.
- [35] X. H. Chen, A. Mizera, and J. Pang, “Quantifying location privacy revisited: preliminary report,” 2014, <http://satoss.uni.lu/>.
- [36] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez, “Show me how you move and I will tell you who you are,” *Transactions on Data Privacy*, vol. 4, no. 2, pp. 103–126, 2011.
- [37] K. Jiang, D. Shao, S. Bressan, T. Kister, and K. Tan, “Publishing trajectories with differential privacy guarantees,” in *Proceedings of the 25th International Conference on Scientific and Statistical Database Management*, pp. 1–12, July 2013.
- [38] C. Faloutsos and S. Roseman, “Fractals for secondary key retrieval,” in *Proceedings of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pp. 247–252, New York, NY, USA, 1989.
- [39] F. Tian, X. Gui, J. An, P. Yang, and X. Zhang, “A density-based space filling curve for location privacy-preserving,” in *Proceedings of the 2014 IEEE International Conference on Services Computing (SCC)*, pp. 131–138, Atlanta, GA, USA, June 2014.
- [40] A. Blum, K. Ligett, and A. Roth, “A learning theory approach to non-interactive database privacy,” in *Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing (STOC '08)*, pp. 609–618, ACM, New York, NY, USA, 2008.

Research Article

Joint Range-Doppler-Angle Estimation for OFDM-Based RadCom System via Tensor Decomposition

Bo Kong ¹, Yuhao Wang ², Xiaohua Deng,³ and Dong Qin ²

¹*School of Mechanical and Electrical Engineering, Nanchang University, 330031, China*

²*School of Information Engineering, Nanchang University, 330031, China*

³*Institute of Space Science and Technology, Nanchang University, 330031, China*

Correspondence should be addressed to Yuhao Wang; wangyuhao@ncu.edu.cn

Received 17 March 2018; Accepted 4 July 2018; Published 1 August 2018

Academic Editor: Donatella Darsena

Copyright © 2018 Bo Kong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Radar and communication (RadCom) systems have received increasing attention due to their high energy efficiency and spectral efficiency. They have been identified as green communications. This paper is concerned with a joint estimation of range-Doppler-angle parameters for an orthogonal frequency division multiplexing (OFDM) based RadCom system. The key idea of the proposed method is to derive different factor matrices by the tensor decomposition method and then extract parameters of the targets from these factor matrices. Different from the classical tensor decomposition method via alternating least squares or higher-order singular value decomposition, we adopt a greedy based method with each step constituted by a rank-1 approximation subproblem. To avoid local extremum, the rank-1 approximation is solved by using a multiple random initialized tensor power method with a comparison procedure followed. A parameterized rectification method is also proposed to incorporate the inherent structures of the factor matrices. The proposed algorithm can estimate all the parameters simultaneously without parameter pairing requirement. The numerical experiments demonstrate superior performance of the proposed algorithm compared with the existing methods.

1. Introduction

The integrated radar and communication (RadCom) system has received much attention in recent years. By using a joint waveform, the occupied spectrum can be used efficiently and both radar and communication functions can be operated simultaneously. Due to the fact that the signal energy and frequency spectrum can be used efficiency and cognitively, the RadCom system is considered as a green communication system, which is a relatively new research discipline [1–4]. Such a RadCom system has been reported in many references [5–8]. Specifically, to embed the communication information into the radar waveforms efficiently, the performance of typical orthogonal frequency division multiplexing (OFDM) waveforms was analyzed [5]. Then, the single-input multiple-output (SIMO) and multiple-input multiple-output (MIMO) scenarios were extended in the RadCom system in order to estimate the directions of arrival (DOAs) of the targets [9, 10]. In general, the estimation methods can be classified into

sequential methods and simultaneous methods. The existing radar systems usually employ the sequential methods for computation cost reduction. However, estimation in separate dimensions encounters the pair-matching problem for different parameters, as well as signal-to-noise ratio (SNR) loss. With the rapid development of computing power, the simultaneous methods receive more attention. Since the simultaneous method can recover multidimensional parameters at the same time, how to avoid the pair-matching procedure and improve the performance is of great importance.

The existing joint estimation methods mainly focus on the problems in Doppler and angle domains. The algorithms can be mainly divided into two groups: subspace-based algorithms and sparse representation (SR) based algorithms. In general, the subspace-based methods need a reasonably large number of snapshots and high enough SNR to implement the eigenvalue decomposition (EVD) with a desirable performance. In recent years, the SR-based techniques exploit the sparsity of the radar target scenarios. However, most

of the SR-based techniques encounter the grid mismatch problem, which is caused by the solutions on discrete grids. The performance of such algorithms is directly affected by the grid resolution. On the other hand, these SR-based methods mainly focus on a one-dimensional problem and are usually extended to multidimensional cases by stacking operation [11, 12]. Nevertheless, the stacking operation ignores the inherent multidimensional structure of the received data.

Tensor based methods have been applied in radar applications [13]. With the benefits of multidimensional modeling and algorithms, the dimension of radar parameter estimation problem can be reduced and solved easily. The mainstream method is to convert the multidimensional problem into multiple one-dimensional problems with a low computational complexity. High-order singular value decomposition (HOSVD) algorithm and canonical polyadic (CP) decomposition algorithm, also known as CANDECOMP/PARAFAC decomposition, play an important role in processing multidimensional data. The alternating least squares (ALS) algorithm is still a workhorse for solving the CP decomposition problem [14, 15]. However, the ALS algorithm is faced with the troubles of local minimum and disappointing convergence properties. Another kind of algorithms is greedy algorithms, which is also known as rank-1 deflation. It is known that the greedy algorithms cannot generalize to tensor fields straightly [16]. In [17], the authors proposed a deflation method with a constraining procedure after each step. In [18], a similar method was proposed based on successive rank-1 approximations and an iterative process followed for eliminating the residue. The rank-1 approximation subproblem is usually computed by means of noniterative methods, including truncated high-order singular value decomposition (T-HOSVD) and sequential rank-one approximation and projection (SeROAP) [19]. However, they can only provide suboptimal solutions in spite of the low computational complexity.

In this paper, we introduce the tensor modeling for monostatic OFDM-SIMO based RadCom system. A data tensor is constructed from the demodulated OFDM symbols. Assuming a scenario with point scattering targets, the CP decomposition model is used to decompose the data tensor. Greedy CP decomposition (GCPD) algorithm combined with multiple random initialized tensor power method (TPM) is proposed for CP decomposition. Capitalizing on the inherent structure of the factor matrices, we present a parameterized rectification (PR) method to improve the target detection performance. The proposed algorithm deals with the received signals directly without multidimensional peak searching, covariance matrix estimation, or eigen-decomposition procedures which may bring error accumulation. Multidimensional parameter pairing is fulfilled automatically, avoiding the performance degradation caused by wrong pairing. The contributions of this paper can be summarized as follows:

- (i) A tensor model for OFDM-SIMO based RadCom system is proposed in order to jointly estimate target parameters in the range-Doppler-angle domain.
- (ii) A GCPD algorithm combined with multiple random initialized TPM is proposed for tensor decomposition. A globalization procedure is introduced to

avoid the locally optimal solutions. This algorithm can achieve much better performance compared with the traditional algorithms.

- (iii) A PR method is proposed to take advantage of the inherent structures of the factor matrices. The PR method can significantly improve the target detection performance, even when there are coherent targets.

The rest of the paper is organized as follows. In Section 2, the system model for OFDM-SIMO based RadCom system and the problem formulation are introduced. A novel GCPD algorithm with multiple random initialized TPM and the PR method is presented in Section 3. In Section 4, the results of simulation in a typical multitargets scene are given to verify the performance of the proposed method. Finally, in Section 5, a conclusion is drawn.

Notation: We denote the scalars and vectors with lowercase letters (a, b, \dots) and bold lowercase letters ($\mathbf{a}, \mathbf{b}, \dots$). The matrices are written as bold uppercase letters ($\mathbf{A}, \mathbf{B}, \dots$) and the symbol for tensors are calligraphic letters ($\mathcal{A}, \mathcal{B}, \dots$). The symbols $\circ, *$ denote the outer and Hadamard (element-wise) products. The transpose, conjugate, and conjugate-transpose are denoted by \cdot^T, \cdot^* and \cdot^H , respectively. $\|\cdot\|$ denotes the Euclidean (l_2) norm of a vector. $\|\cdot\|_F$ denotes the Frobenius norm of a tensor.

With respect to tensor $\mathcal{A} \in \mathbb{C}^{K \times N \times M}$ and vectors $\mathbf{u} \in \mathbb{C}^K$, $\mathbf{v} \in \mathbb{C}^N$ and $\mathbf{w} \in \mathbb{C}^M$, operator $A(\mathbf{u}, \mathbf{v}, \mathbf{w})$ is defined as

$$A(\mathbf{u}, \mathbf{v}, \mathbf{w}) := \sum_{i=1}^K \sum_{j=1}^N \sum_{l=1}^M u_i^* v_j^* w_l^* \mathcal{A}(i, j, l) \in \mathbb{C}. \quad (1)$$

In particular, when one of these vectors is absent, we have

$$A(I, \mathbf{v}, \mathbf{w}) := \sum_{j=1}^N \sum_{l=1}^M v_j^* w_l^* \mathcal{A}(:, j, l) \in \mathbb{C}^K. \quad (2)$$

It is similar for the definitions of $A(\mathbf{u}, I, \mathbf{w})$ and $A(\mathbf{u}, \mathbf{v}, I)$. Some preliminaries about tensor and its corresponding decomposition are given in the appendix.

2. System Model and Problem Formulation

Consider a monostatic OFDM-SIMO based RadCom system equipped with a single antenna for transmitter and a uniform linear array for receiver, as shown in Figure 1. The receiver array consists of N_R antennas uniformly spaced with the half wavelength separation, denoted by $\Delta_R = \lambda/2$, where λ is the wavelength of the transmitted signal.

The steering vector of the array is represented as

$$\mathbf{a}_R(\theta) = [a_R^0(\theta) \ a_R^1(\theta) \ \dots \ a_R^{N_R-1}(\theta)]^T, \quad (3)$$

where $a_R^m(\theta) = e^{j2\pi m \Delta_R \sin(\theta)/\lambda}$.

The transmitted waveform is modulated by Cyclic-Prefix OFDM (CP-OFDM) with the Quadrature Amplitude Modulation (QAM) or phase-shift keying (PSK) constellation mapping. The transmitted signal is given by

$$x(t) = \sum_{n=-\infty}^{\infty} \sum_{k \in K} d_k[n] e^{j\omega_k(t-nT)} p_T(t-nT), \quad (4)$$

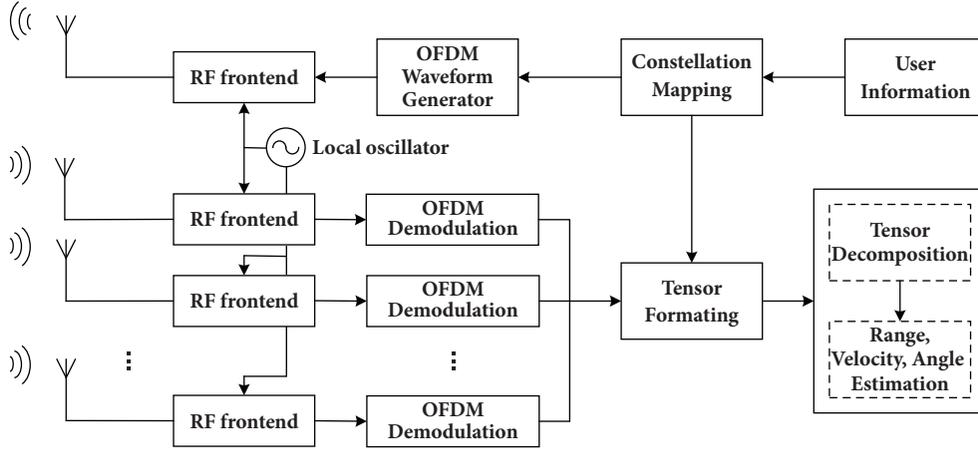


FIGURE 1: Block diagram of processing for RadCom system.

where $d_k[n]$ denotes the transmitted data over the k th subcarrier of the n th symbol with $k \in K$. ω_k is the angular frequency of the k th subcarrier. T is the symbol period. $p_T(t - nT)$ is the shaping filter of the transmitter, which is usually a time-domain rectangular filter for the general OFDM realization.

Suppose that there are H nonfluctuating (according to the Swerling-0 model) far-field point targets. The parameters of a target relative to the RadCom system are given by the quadruple $\{\alpha_h, r_h, v_h, \theta_h\}$, where α_h , r_h , v_h , and θ_h represent the complex amplitude, the range, the radial velocity, and the angle of arrival of the h th target, respectively. It is more convenient to equivalently consider the quadruple $\{\alpha_h, \tau_h, f_h, \sin(\theta_h)\}$ of complex amplitude, time delay, Doppler shift, and normalized angle parameters, where $\tau_h = 2r_h/c$ and $f_h = 2v_h/\lambda$. The delay, Doppler shift, and angle information can always be transformed back into the physical coordinates. In this paper, we consider the gridless scenario. That is, the true targets are likely to locate at any position in the delay-Doppler-angle domain.

The received continuous signals $y_m(t)$ at antennas $m = 0, \dots, N_R - 1$ consist of the superposition of the reflections from the targets of the transmitted probing signals, as well as the additive noise (includes the thermal noise, jamming and clutter), which are given by

$$y_m(t) = \sum_{h=1}^H \alpha_h e^{j2\pi m \Delta_R \sin(\theta_h)/\lambda} e^{j2\pi f_h t} x(t - \tau_h) + w_m(t), \quad (5)$$

where $w_m(t)$ is additive white Gaussian noise (AWGN) with zero-mean and variance σ^2 corresponding to the m th receiving antenna. α_h is the complex amplitude of the h th target affected by path loss, scattering, and processing gains.

Since the information is modulated in frequency domain with OFDM, the demodulated data corresponds to the m th receive antenna that is given by

$$\tilde{d}_k^{(m)}[n] = \sum_{h=1}^H d_k[n] v + \tilde{w}_m[n], \quad (6)$$

where $v = e^{j2\pi m \Delta_R \sin(\theta_h)/\lambda} e^{-j\omega_k \tau_h} e^{j2\pi f_h n T}$.

By implementing the CP decomposition model in Definition 4, the demodulated data in (6) can be formatted as a third-order tensor:

$$\widehat{\mathcal{D}} = \mathcal{D} * \mathcal{V} + \widetilde{\mathcal{W}}, \quad (7)$$

where $\widehat{\mathcal{D}} \in \mathbb{C}^{K \times N \times M}$. \mathcal{D} is a tensor from the transmitted data that is duplicated in the third mode. \mathcal{V} represents the response of the targets, expressed as

$$\mathcal{V} = \sum_{h=1}^H \mathbf{a}_h \circ \mathbf{b}_h \circ \mathbf{c}_h, \quad (8)$$

where

$$\begin{aligned} \mathbf{a}_h &= [e^{-j\omega_1 \tau_h} \quad \dots \quad e^{-j\omega_K \tau_h}]^T, \\ \mathbf{b}_h &= [e^{j2\pi f_h 0 T} \quad \dots \quad e^{j2\pi f_h (N-1)T}]^T, \\ \mathbf{c}_h &= [e^{j2\pi 0 \Delta_R \sin(\theta_h)/\lambda} \quad \dots \quad e^{j2\pi (M-1) \Delta_R \sin(\theta_h)/\lambda}]^T, \end{aligned} \quad (9)$$

and $\widetilde{\mathcal{W}}$ is the rearranged noise.

The objective is to estimate the target parameters $\{\alpha_h, \tau_h, f_h, \sin(\theta_h)\}$ from the demodulated data $\widehat{\mathcal{D}}$. Traditionally, the targets are firstly detected in the delay-Doppler plane with the matched filter, and then the angle estimation is performed. However, the sequential technique needs the pair-matching procedures to obtain the one-to-one relationship among the delay, Doppler shift, and angle.

Due to the sparse nature of the radar scene, the number of targets, H , is usually small relative to the dimensions of the tensor. Hence the response of the targets has an intrinsic low-rank structure. Different from the traditional sequential estimation methods, we hope to achieve a joint estimation procedure with the low-rank structure of the target response.

3. Parameter Estimation via Low-Rank Tensor Approximation

In this section, we present the joint parameter estimation algorithm using tensor decomposition. The estimation procedure includes two basic stages: (1) target separation, and (2) parameter estimation. The former is achieved by CP decomposition and the latter by a correlation-based estimation with the decomposed factor matrices.

In order to eliminate the influences of the transmitted user information in the RadCom system, we normalize the received data tensor with the transmitted data tensor by element-wise product as follows:

$$\check{\mathcal{D}} = \frac{\widehat{\mathcal{D}} * \mathcal{D}^*}{\mathcal{D} * \mathcal{D}^*} = \mathcal{V} + \check{\mathcal{W}}, \quad (10)$$

where $\check{\mathcal{W}} = (\check{\mathcal{W}} * \mathcal{D}^*) / (\mathcal{D} * \mathcal{D}^*)$ is an i.i.d. white Gaussian noise with the same distribution with $w_m(t)$.

Since the number of targets is usually small relative to the dimensions of the data tensor, $\check{\mathcal{D}}$ is low-rank but contaminated by additive noise. Therefore the target separation can be obtained by performing a CP decomposition of the normalized data tensor $\check{\mathcal{D}}$.

Here we assume that the number of targets H is known or estimated in advance. Then, the CP decomposition can be accomplished by solving

$$\min_{\hat{\alpha}, \hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}} \left\| \check{\mathcal{D}} - \sum_{h=1}^H \hat{\alpha}_h \hat{\mathbf{a}}_h \circ \hat{\mathbf{b}}_h \circ \hat{\mathbf{c}}_h \right\|_F^2, \quad (11)$$

$$\text{s.t. } \|\hat{\mathbf{a}}_h\| = \|\hat{\mathbf{b}}_h\| = \|\hat{\mathbf{c}}_h\| = 1, \quad \forall h = 1, \dots, H, \quad (12)$$

where $\hat{\alpha} = [\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_H]$, $\hat{\mathbf{A}} = [\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_H]$, $\hat{\mathbf{B}} = [\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \dots, \hat{\mathbf{b}}_H]$, $\hat{\mathbf{C}} = [\hat{\mathbf{c}}_1, \hat{\mathbf{c}}_2, \dots, \hat{\mathbf{c}}_H]$.

Remark 1. The uniqueness under mild conditions is a key feature of CP decomposition. The CP decomposition of the tensor $\check{\mathcal{D}}$ is essentially unique when the following sufficient condition is satisfied ([21], Theorem 10)

$$k(\mathbf{A}) + k(\mathbf{B}) + k(\mathbf{C}) \geq 2H + 2, \quad (13)$$

where $k(\mathbf{M})$ is the Kruskal rank of the matrix \mathbf{M} and is defined as the largest integer k such that any k columns of \mathbf{M} are linearly independent. When multiple noncoherent targets exist, i.e., different ranges, velocities, and angles, those three factor matrices are always full column rank so that (13) always satisfies.

Remark 2. It is clear that when there exists coherent targets, e.g., targets with the same range but different velocities and angles, (13) will not satisfy any more. In fact, the CP decomposition cannot ensure uniqueness and correctness in this case with third-order tensors even under a much milder condition ([21], Theorem 9). To improve the percentage of successful decomposition, we utilize inherent structures of the factor matrices. Refer to Section 3.3 for detail.

The most commonly used algorithm for solving the CP decomposition is ALS [22], which is quite simple and can be executed by updating each factor matrix alternately in each iteration. The ALS algorithm is extremely fast but not stable, and the global optimal solution is hard to obtain. It has been shown that tensors of order 3 or higher fail to have best rank- r approximation for $r \geq 2$ [23]. Fortunately, the rank-1 tensor approximation always exists and can be calculated by the high-order power method (HOPM), also known as tensor power method (TPM) [24].

Our proposed tensor decomposition method includes two main parts: (1) rank-1 tensor decomposition based on multiple random initialized TPM and (2) greedy iterations for removing the estimated components as well as residual error.

3.1. Tensor Rank-1 Approximation. There are several rank-1 tensor approximation algorithms, such as T-HOSVD and SeROAP. However, they are all suboptimal algorithms in spite of the low computational complexity. In this paper, we employ the TPM algorithm for the quasi-optimal rank-1 tensor approximation.

The problem of best rank-1 approximation of tensor $\check{\mathcal{R}}$ can be formulated as

$$\min_{\hat{\alpha}, \hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}} \left\| \check{\mathcal{R}} - \hat{\alpha} \hat{\mathbf{a}} \circ \hat{\mathbf{b}} \circ \hat{\mathbf{c}} \right\|_F^2, \quad (14)$$

$$\text{s.t. } \|\hat{\mathbf{a}}\| = \|\hat{\mathbf{b}}\| = \|\hat{\mathbf{c}}\| = 1. \quad (15)$$

This can be efficiently solved by TPM iterations in the complex domain, given as

$$\hat{\mathbf{a}}^{(t+1)} = \frac{R(I, \hat{\mathbf{b}}^{(t)}, \hat{\mathbf{c}}^{(t)})}{\|R(I, \hat{\mathbf{b}}^{(t)}, \hat{\mathbf{c}}^{(t)})\|}, \quad (16)$$

$$\hat{\mathbf{b}}^{(t+1)} = \frac{R(\hat{\mathbf{a}}^{(t+1)}, I, \hat{\mathbf{c}}^{(t)})}{\|R(\hat{\mathbf{a}}^{(t+1)}, I, \hat{\mathbf{c}}^{(t)})\|}, \quad (17)$$

$$\hat{\mathbf{c}}^{(t+1)} = \frac{R(\hat{\mathbf{a}}^{(t+1)}, \hat{\mathbf{b}}^{(t+1)}, I)}{\|R(\hat{\mathbf{a}}^{(t+1)}, \hat{\mathbf{b}}^{(t+1)}, I)\|}. \quad (18)$$

The initialization vectors $(\hat{\mathbf{a}}^{(0)}, \hat{\mathbf{b}}^{(0)}, \hat{\mathbf{c}}^{(0)})$ are usually randomly selected, where the values of $\hat{\mathbf{a}}^{(0)}$ and $\hat{\mathbf{b}}^{(0)}$ are uniformly chosen in the complex domain and vector $\hat{\mathbf{c}}^{(0)}$ is calculated by (18). Also, they can be given by any other algorithms, such as HOSVD and SeROAP. The iterations with (16)-(18) will repeat until t is larger than the maximum number of iterations t_{max} . In order to reduce the computational complexity, an additional option of the algorithm is used to stop the iterations when the following criteria satisfy

$$\|\hat{\mathbf{a}}^{(t)} - \hat{\mathbf{a}}^{(t-1)}\| \cdot \|\hat{\mathbf{b}}^{(t)} - \hat{\mathbf{b}}^{(t-1)}\| \cdot \|\hat{\mathbf{c}}^{(t)} - \hat{\mathbf{c}}^{(t-1)}\| \leq t_s, \quad (19)$$

where t_s is the stopping threshold.

On account that the rank-1 approximation is a nonconvex problem and many local optima exist, careful initialization is required for TPM iterations to ensure the convergence to the true rank-1 tensor components. Here we consider

Input: Tensor $\check{\mathcal{D}} \in \mathbb{C}^{K \times N \times M}$, $\mathcal{R} \in \mathbb{C}^{K \times N \times M}$, maximum number of iterations t_{max} , number of initializations L

- 1: **for** $\tau = 1$ **to** L **do**
- 2: Initialize unit vectors $\hat{\mathbf{a}}_\tau^{(0)} \in \mathbb{C}^K$, $\hat{\mathbf{b}}_\tau^{(0)} \in \mathbb{C}^N$, and $\hat{\mathbf{c}}_\tau^{(0)} \in \mathbb{C}^M$ as
 - * Option 1: random initialization.
 - * Option 2: Preset values or the last estimated results.
- 3: **for** $t = 1$ **to** t_{max} **do**
- 4:
$$\hat{\mathbf{a}}_\tau^{(t+1)} = \frac{R(I, \hat{\mathbf{b}}_\tau^{(t)}, \hat{\mathbf{c}}_\tau^{(t)})}{\|R(I, \hat{\mathbf{b}}_\tau^{(t)}, \hat{\mathbf{c}}_\tau^{(t)})\|},$$

$$\hat{\mathbf{b}}_\tau^{(t+1)} = \frac{R(\hat{\mathbf{a}}_\tau^{(t+1)}, I, \hat{\mathbf{c}}_\tau^{(t)})}{\|R(\hat{\mathbf{a}}_\tau^{(t+1)}, I, \hat{\mathbf{c}}_\tau^{(t)})\|},$$

$$\hat{\mathbf{c}}_\tau^{(t+1)} = \frac{R(\hat{\mathbf{a}}_\tau^{(t+1)}, \hat{\mathbf{b}}_\tau^{(t+1)}, I)}{\|R(\hat{\mathbf{a}}_\tau^{(t+1)}, \hat{\mathbf{b}}_\tau^{(t+1)}, I)\|}.$$
- 5: **end for**
- 6: **end for**
- 7: Choose $(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$ in $\{(\hat{\mathbf{a}}_\tau^{(t+1)}, \hat{\mathbf{b}}_\tau^{(t+1)}, \hat{\mathbf{c}}_\tau^{(t+1)}), \tau = 1, \dots, L\}$ that correspond to the largest $|\check{D}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})|$.
- 8: Amplitude estimation: $\hat{\alpha} = R(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$
- 9: **return** $(\hat{\alpha}, \hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$

ALGORITHM 1: Tensor rank-1 approximation via TPM.

an approximate globalization procedure. Multiple randomly generated initializations are used for the TPM iterations. In order to identify the best one among these initializations, we need a projection procedure to obtain the final estimates of the vectors. This procedure is performed with the estimated vectors that is projected to the original data tensor. The vectors corresponding to the largest absolute value of these projections are selected as the final results. It can be formulated as

$$(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}) = \arg \min_{\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}} |\check{D}(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})|, \quad \tau = 1, \dots, L, \quad (20)$$

where L is the number of initializations.

Because the vectors $(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}})$ are unit norm, the amplitude is estimated as

$$\hat{\alpha} = R(\hat{\mathbf{a}}, \hat{\mathbf{b}}, \hat{\mathbf{c}}). \quad (21)$$

Note that, in order to obtain the approximate global optimal solution, the operator in (20) is different from that in (21). Tensor rank-1 approximation via TPM algorithm is summarised in Algorithm 1.

3.2. Greedy CPD. In this section, we present the GCPD algorithm, which solves the problem of tensor decomposition in a greedy manner. The GCPD algorithm calculates the best rank-1 approximation and then removes the extracted component at each step. Since the best rank-1 approximation may not be the actual component of the tensor decomposition [16], additional iterations for refinement are employed. The idea of refinement is common for the greedy-like algorithms in the compressive sensing community.

The decomposition of tensor $\check{\mathcal{D}}$ can be formulated as

$$\check{\mathcal{D}} = \sum_{h=1}^H \alpha_h \mathbf{a}_h \circ \mathbf{b}_h \circ \mathbf{c}_h + \mathcal{E}, \quad (22)$$

where $\alpha_h \mathbf{a}_h \circ \mathbf{b}_h \circ \mathbf{c}_h$ corresponds to the h th rank-1 component with $1 \leq h \leq H$, and \mathcal{E} is the residual error tensor, e.g., the additive noise. In the first round iterations, we compute the H rank-1 components one by one and remove the extracted components after each computation. Let $\alpha'_h \mathbf{a}'_h \circ \mathbf{b}'_h \circ \mathbf{c}'_h$ be the extracted h th rank-1 component and the iterations can be formulated as

$$(\alpha'_h, \mathbf{a}'_h, \mathbf{b}'_h, \mathbf{c}'_h) = \arg \min_{\alpha, \mathbf{a}, \mathbf{b}, \mathbf{c}} \|\mathcal{E}_{h-1} - \alpha \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}\|_F^2, \quad (23)$$

$1 \leq h \leq H,$

where $\mathcal{E}_h = \check{\mathcal{D}} - \sum_{p=1}^h \alpha'_p \mathbf{a}'_p \circ \mathbf{b}'_p \circ \mathbf{c}'_p$.

On account that the extracted rank-1 components may not be the actual component of the tensor decomposition, the residual error \mathcal{E}_H usually contains the decomposition error and is not identical to the original residual error, i.e., \mathcal{E} in (22).

As a result, the refinement iterations are formulated as

$$(\alpha'_h, \mathbf{a}'_h, \mathbf{b}'_h, \mathbf{c}'_h) = \arg \min_{(\alpha, \mathbf{a}, \mathbf{b}, \mathbf{c})} \|\mathcal{E}_h - \alpha \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}\|_F^2, \quad (24)$$

$1 \leq h \leq H,$

where \mathcal{E}_h is the extracted rank-1 component in the last round iterations as well as the residual error given by $\mathcal{E}_h = \mathcal{E}_H + \alpha'_h \mathbf{a}'_h \circ \mathbf{b}'_h \circ \mathbf{c}'_h$. The refinement iterations will be implemented multiple rounds. Clearly, the refinement iterations play a role

TABLE 1: Computational complexity of different algorithms.

Algorithms	Complexity
FB-RootMUSIC	$\mathcal{O}\{KNM(K+H(K+N)(K+M))\}$ [20]
ALS	$\mathcal{O}\{3t_{max}HKNM\}$
GCPD (SeROAP init)	$\mathcal{O}\{2(k+1)HKNM^2M^2 + 3(t'_{max} + 1)t_{max}HKNM\}$ [19]
GCPD (random init)	$\mathcal{O}\{3(t'_{max} + 1)t_{max}HKNM\}$

Input: Tensor $\tilde{\mathcal{D}} \in \mathbb{C}^{K \times N \times M}$, number of targets H

- 1: $\mathcal{R} \leftarrow \tilde{\mathcal{D}}$
- 2: **for** $h = 1$ **to** H **do**
- 3: Calculate $(\hat{\alpha}_h, \hat{\mathbf{a}}_h, \hat{\mathbf{b}}_h, \hat{\mathbf{c}}_h)$ via Algorithm 1 (option 1).
- 4: $\mathcal{R} \leftarrow \mathcal{R} - \hat{\alpha}_h \hat{\mathbf{a}}_h \circ \hat{\mathbf{b}}_h \circ \hat{\mathbf{c}}_h$
- 5: **end for**
- 6: **repeat**
- 7: **for** $h = 1$ **to** H **do**
- 8: $\mathcal{R} \leftarrow \mathcal{R} + \hat{\alpha}_h \hat{\mathbf{a}}_h \circ \hat{\mathbf{b}}_h \circ \hat{\mathbf{c}}_h$
- 9: Update $(\hat{\alpha}_h, \hat{\mathbf{a}}_h, \hat{\mathbf{b}}_h, \hat{\mathbf{c}}_h)$ via Algorithm 1 (option 2).
- 10: $\mathcal{R} \leftarrow \mathcal{R} - \hat{\alpha}_h \hat{\mathbf{a}}_h \circ \hat{\mathbf{b}}_h \circ \hat{\mathbf{c}}_h$
- 11: **end for**
- 12: **until** a stopping criterion is met
- 13: **return** $(\hat{\alpha}_h, \hat{\mathbf{a}}_h, \hat{\mathbf{b}}_h, \hat{\mathbf{c}}_h)$, $h = 1, \dots, H$.

ALGORITHM 2: Greedy CPD.

on correction of the extracted components. Note that the estimated results in the last round are used as initialization for the next round iterations.

The refinement iterations will repeat until a stopping criterion is met. In this paper, we use the following criteria:

$$\|\mathcal{E}_H - \mathcal{E}'_H\|_F < t_G, \quad (25)$$

where \mathcal{E}_H and \mathcal{E}'_H are the residual errors corresponding to the current and the last round of refinement iterations, respectively. t_G is the stopping threshold.

The GCPD algorithm is summarised in Algorithm 2.

3.3. Target Parameter Estimation and Parameterized Rectification. We now discuss how to estimate the target parameters based on the estimated vectors from Algorithm 2. According to the definitions of these vectors in (8), each vector is characterized by the associated delay, Doppler shift, or angle of one target.

Hence, the delay of the h th target τ_h can be estimated via a correlation-based method given by

$$\hat{\tau}_h = \arg \max_{\tau_h} \frac{|\hat{\mathbf{a}}_h^H \mathbf{a}_h(\tau_h)|}{\|\hat{\mathbf{a}}_h\| \|\mathbf{a}_h(\tau_h)\|}. \quad (26)$$

With the additive white Gaussian noise, the correlation-based method is indeed a maximum likelihood (ML) estimator and provides the optimal solution.

The Doppler shift and angle of each target can be obtained similarly as

$$\hat{f}_h = \arg \max_{f_h} \frac{|\hat{\mathbf{b}}_h^H \mathbf{b}_h(f_h)|}{\|\hat{\mathbf{b}}_h\| \|\mathbf{b}_h(f_h)\|}, \quad (27)$$

$$\hat{\theta}_h = \arg \max_{\theta_h} \frac{|\hat{\mathbf{c}}_h^H \mathbf{c}_h(\theta_h)|}{\|\hat{\mathbf{c}}_h\| \|\mathbf{c}_h(\theta_h)\|}. \quad (28)$$

The maximization problems in (26)-(28) involve one-dimensional search and can be performed by zero-padded FFT efficiently combining the inherent structures of these vectors.

Vectors \mathbf{a}_h , \mathbf{b}_h and \mathbf{c}_h are inherently determined by a few parameters. However, Algorithms 1 and 2 do not take this into account. In order to make use of this structural information, we propose a method named parameterized rectification (PR). The PR method is based on parameter estimation and structure reconstruction. That is, an estimation process performed by (26)-(28) is inserted at the end of Algorithm 1. The returned vectors are regenerated with the desired structures and the estimated parameters. The PR method is summarised in Algorithm 3. The GCPD algorithm combined with PR is abbreviated as PR-GCPD.

Note that the PR method generally increases the rank, especially in the noisy case, and several initializations and iterations are necessary to obtain the global optimal solutions.

3.4. Computational Complexity Analysis. We use the number of complex multiplications (operations) as the complexity metric. Since GCPD is an iterative algorithm, the total complexity is unbounded. The complexity is mainly dominated by the rank-1 approximation, which is repeatedly computed several times. The major computing task of the rank-1 approximation is the TPM iteration. The computation of vectors $\hat{\mathbf{a}}^{(t+1)}$, $\hat{\mathbf{b}}^{(t+1)}$, and $\hat{\mathbf{c}}^{(t+1)}$ in (16)-(18) needs $3KNM + KN + KM + NM$ operations. The TPM iterations are repeated with the maximum t_{max} times. Thus, the complexity of the rank-1 approximation is given by $\mathcal{O}\{3t_{max}KNM\}$, where $\mathcal{O}\{\cdot\}$ is the big-O notation. Since the iterations with multiple initializations can be performed in parallel, the number of initializations is not considered in the computational complexity. Assume that the maximum number of iterations for refinement in the GCPD algorithm is t'_{max} , the total complexity of GCPD is given by $\mathcal{O}\{3(t'_{max} + 1)t_{max}HKNM\}$. Note that the stopping criteria in (19) and (25) are usually applied, the actual number of operations is much smaller.

In Table 1, we summarize the computational complexity of the algorithms presented in the next section, where k is a

Procedure PR($\hat{\mathbf{a}}_h, \hat{\mathbf{b}}_h, \hat{\mathbf{c}}_h$)
 Estimate parameters $\hat{\tau}_h, \hat{f}_h$ and $\hat{\theta}_h$ via Eq. (26)-(28).
 Regenerate $\hat{\mathbf{a}}_h, \hat{\mathbf{b}}_h$ and $\hat{\mathbf{c}}_h$ with the structures in Eq. (8) and estimated parameters.
return ($\hat{\mathbf{a}}_h, \hat{\mathbf{b}}_h, \hat{\mathbf{c}}_h$)

ALGORITHM 3: Parameterized rectification.

TABLE 2: OFDM waveform parameters.

Parameters	Values
Carrier frequency	5.9 GHz
Subcarrier spacing	90.909 kHz
OFDM symbol length	11 us
Cyclic prefix length	1.375 us
Bandwidth	93.1 MHz
Constellation Mapping	4 QAM

user-defined parameter. Since the PR method is performed on vectors, the number of operations per iteration needed by the PR method is negligible compared with the TPM iterations. However, on account that the PR method generally increases the rank, it usually slows down the convergence and may lead to more run time. The computational complexity is evaluated via simulation in the next section.

4. Numerical Results

In this section, some numerical results are used to illustrate the performance of the proposed method. Different from the related literature [18], the dimensions of the data are much larger and the simulated performance characteristics may be obviously distinct.

4.1. RadCom Parameters and Performance Metrics. The transmitted signal is modulated with CP-OFDM. The parameters for OFDM waveform generation are listed in Table 2. The receiver of the RadCom system is equipped with a uniform linear array with $M_R = 15$ receive antennas spaced at a distance $\Delta_R = 0.5\lambda_0$. The surveillance field is in the far field of the RadCom system. In each experiment, targets are placed randomly on the predefined unambiguous range-Doppler-angle parameter space. The target amplitudes are chosen with constant absolute value and random phase. White Gaussian noise is added to the data tensor with variance σ^2 corresponding to the specified output SNR. The SNR is changed from -40 dB to -20 dB in step of 1 dB.

In all simulations, the maximum number of iterations t_{max} for the TPM method is set to be 1000. Also, the stopping threshold t_S in (19) is assigned to be 10^{-10} . For the GCPD algorithm, the stopping threshold t_G in (25) is set to be 10^{-8} . In order to avoid the lengthy refinement iterations in the GCPD algorithm, we restrict the number of refinement rounds no greater than 10. In each experiment, 64 OFDM symbols are collected for signal processing. A total number of 1000 experiments are conducted.

Here we use the following performance metrics:

- (1) Root mean-square errors (RMSE) in the delay, Doppler, and angle estimation are given by

$$\text{RMSE}_\tau = \left(\frac{1}{HL} \sum_{l=1}^L \sum_{h=1}^H (\hat{\tau}_{h,l} - \tau_{h,l})^2 \right)^{1/2}, \quad (29)$$

$$\text{RMSE}_f = \left(\frac{1}{HL} \sum_{l=1}^L \sum_{h=1}^H (\hat{f}_{h,l} - f_{h,l})^2 \right)^{1/2}, \quad (30)$$

$$\text{RMSE}_\theta = \left(\frac{1}{HL} \sum_{l=1}^L \sum_{h=1}^H (\hat{\theta}_{h,l} - \theta_{h,l})^2 \right)^{1/2}, \quad (31)$$

where H is the number of targets and L is the total number of experiments.

- (2) Detection probability (P_d): the fraction of the total number of targets that are correctly detected in an ellipsoid area. A target is correctly detected when the estimated location of the target in the delay-Doppler-angle space falls within the ellipsoid with axes equivalent to ± 3 times the classical delay, Doppler, and angle resolution bins, Δ_τ , Δ_f , and Δ_θ , respectively.

$$P_d = \frac{1}{HL} \left| \left\{ \hat{\tau}_{h,l}, \hat{f}_{h,l}, \hat{\theta}_{h,l} \mid \Phi_{h,l} \leq \epsilon_{th} \right\} \right|, \quad (32)$$

where

$$\Phi_{h,l} = \left(\left(\frac{\hat{\tau}_{h,l} - \tau_{h,l}}{\Delta_\tau} \right)^2 + \left(\frac{\hat{f}_{h,l} - f_{h,l}}{\Delta_f} \right)^2 + \left(\frac{\hat{\theta}_{h,l} - \theta_{h,l}}{\Delta_\theta} \right)^2 \right)^{1/2}. \quad (33)$$

H and L are the same as those in (29)-(31). ϵ_{th} is the normalized tolerance factor, determined by application. We chose $\epsilon_{th} = 3$, which is three times the resolution bins.

When calculating (29)-(32) in the multiple target scenarios, $\hat{\tau}_{h,l}$, $\hat{f}_{h,l}$, and $\hat{\theta}_{h,l}$ are selected as the ones closest to $\tau_{h,l}$, $f_{h,l}$, and $\theta_{h,l}$, respectively, among all the estimated parameters of H targets, i.e., $\{\hat{\tau}_{h,l}, \hat{f}_{h,l}, \hat{\theta}_{h,l} \mid h = 1, 2, \dots, H\}$.

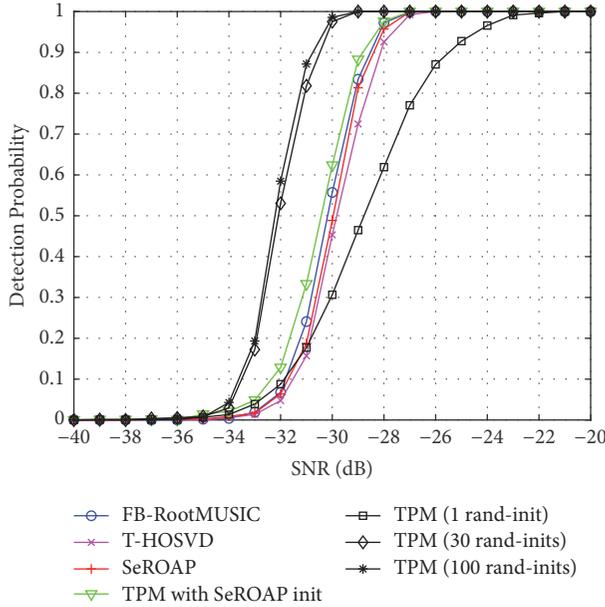


FIGURE 2: Performance comparison of different rank-1 approximation algorithms.

4.2. Performance with Single Target. In this case, the problem corresponds to the rank-1 situation. Figure 2 compares different rank-1 approximation methods as well as different number random initializations for TPM method. Because the dimensions of the tensor are high in this paper, we do not compare the results with the best rank-1 approximation described in [25].

In Figure 2, the number of random initializations for TPM method is selected in the collection $\mathcal{F} = \{1, 30, 100\}$. The methods for comparison include T-HOSVD and SeROAP. The former is a classical approach for tensor decomposition and the latter was recently proposed by Alex P. da Silva et al. in [19]. TPM initialized by the results of SeROAP is also considered. In addition, the subspace-based forward-backward root-MUSIC (FB-RootMUSIC) method [20] incorporating the inherent signal structures is selected for comparison.

The target detection probability versus SNRs for different methods is shown in Figure 2. This simulation shows that the TPM method with 100 random initializations performs much better than the others. TPM with 30 random initializations performs very close to the case with 100 random initializations. The TPM method with multiple random initializations is better than that with initialization generated by the SeROAP method.

In order to further analyze the impact of the number of random initializations, another experiment is performed. The number of random initializations is selected in the collection $\mathcal{F} = \{1, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100\}$. Figure 3 shows the detection probability versus the number of random initializations for SNR = -30, -31, and -32 dB, respectively. The detection probability tends to a fixed value with the increasing number of initializations. When the initialization number is sufficiently large, the output is close to the global optimal

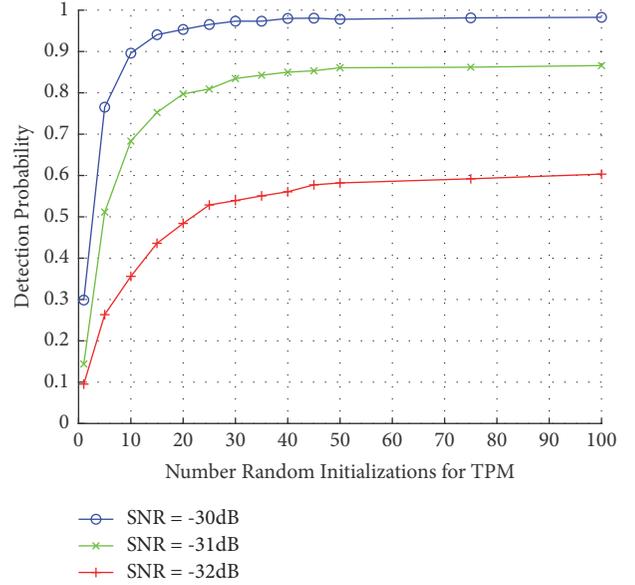


FIGURE 3: Impact of initial number of TPM.

solution. However, in view of the computational burden, we choose the number of initializations to be 30 as an acceptable suboptimal solution. It is worth noting that the TPM solving processes with multiple initializations can be easily realized by parallel computing, so that this is not a limiting factor for practical applications.

To simplify the analysis, the number of random initializations of TPM is fixed to be 30 in the subsequent analyses.

4.3. Performance with Multiple Noncoherent Targets. Consider five targets selected uniformly at random in the RadCom's unambiguous region. The delay, Doppler shift, and angle of each target do not overlap with others. The spacings between different targets in each dimension are larger than three times of the corresponding resolution bin. The amplitude of the targets are chosen such that they are with the same magnitude and random phases.

Figure 4 depicts the RMSE performance of the aforementioned methods in the range, Doppler and angle domains. When the SNR is low, the proposed GCPD and PR-GCPD methods with 30 random initializations have much better performance than the other methods. As the SNR increases, GCPD can achieve similar performance compared to PR-GCPD. The FB-RootMUSIC method cannot achieve similar performance compared to the GCPD and PR-GCPD methods, even at high SNRs. The ALS method performs the worst due to its instability. Figure 6 presents the detection probability of multiple targets. It can be seen that our proposed GCPD and PR-GCPD methods have better detection probability than the classical ALS and FB-RootMUSIC methods. Also, the PR-GCPD method initialized by SeROAP has an acceptable detection probability but is slightly lower than the multiple random initialized ones. The facts show that the globalization procedure with multiple random initializations, as well as the PR method, can improve performance significantly.

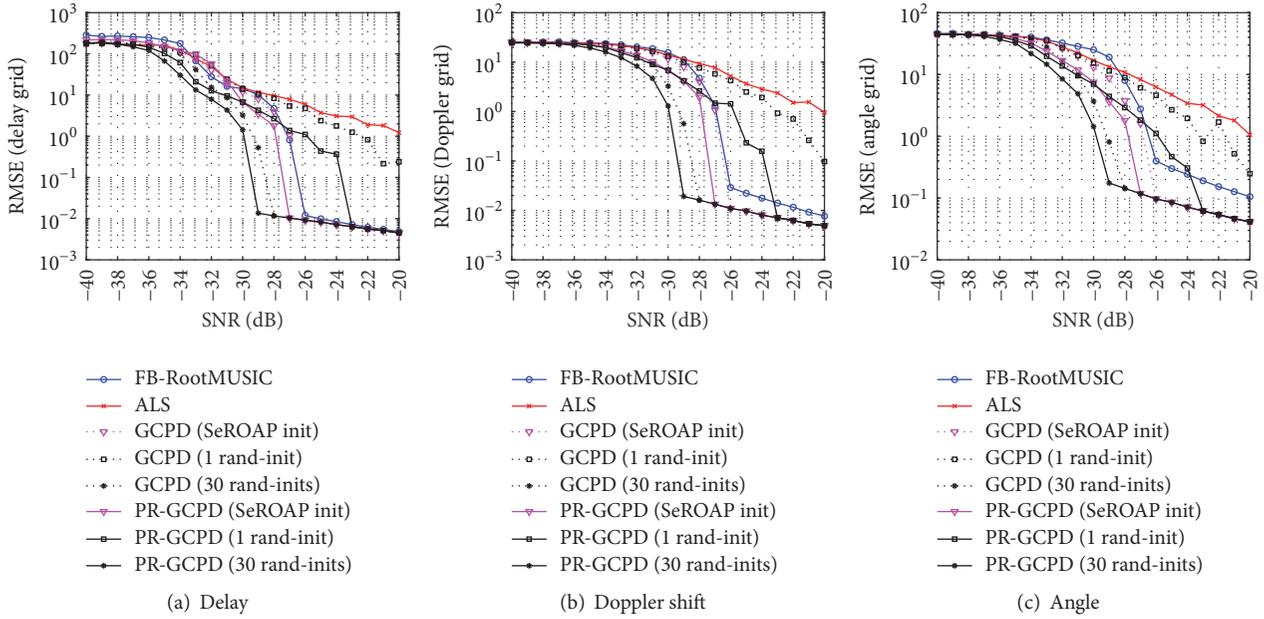


FIGURE 4: RMSE performance of different methods for noncoherent targets.

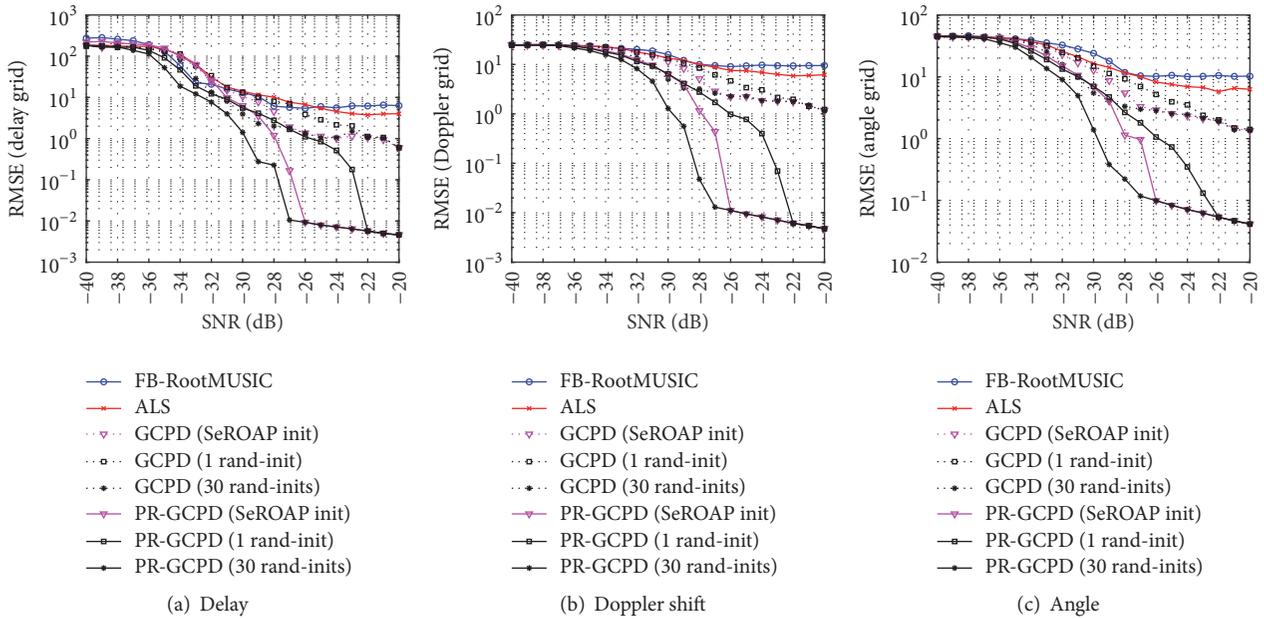


FIGURE 5: RMSE performance of different methods when coherent target exists.

4.4. *Influence of Coherent Targets.* Consider scenarios with five targets and two of them are coherent targets in the range dimension. The coherence indicates that the parameters of different targets in one dimension are the same, e.g., the same delay, Doppler shift, or angle. The coherence destroys the uniqueness conditions of CP decomposition and significantly influences the performance of parameter estimation and target detection.

From Figure 5, we can see that, in this condition, the proposed PR-GCPD method combined with multiple random

initialized TPM provides the best performance. Methods except for the PR-GCPD have much worse performance on account that they cannot resolve the coherent targets robustly. Figure 7 shows the detection probability of all the methods. It is observed that the classical ALS and FB-RootMUSIC methods all have miss detection even when the SNR is high. This is primarily because these methods cannot distinguish the coherent targets. The GCPD with 30 random initializations performs better, though it cannot achieve the best performance at high SNRs. The proposed PR-GCPD

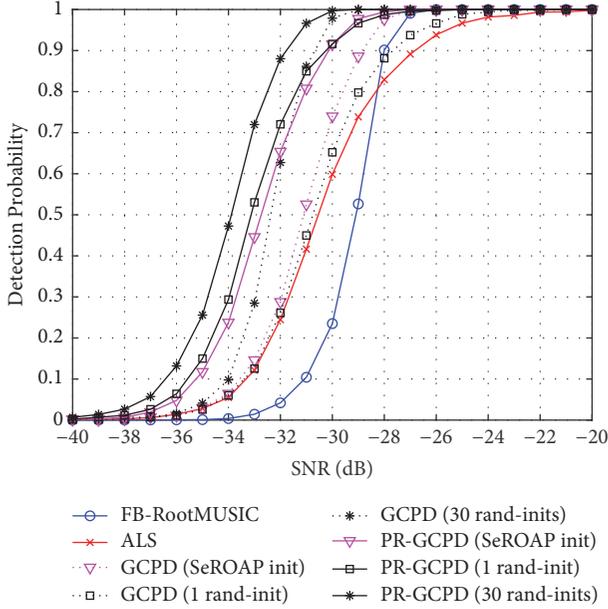


FIGURE 6: Detection probability performance of different methods for noncoherent targets.

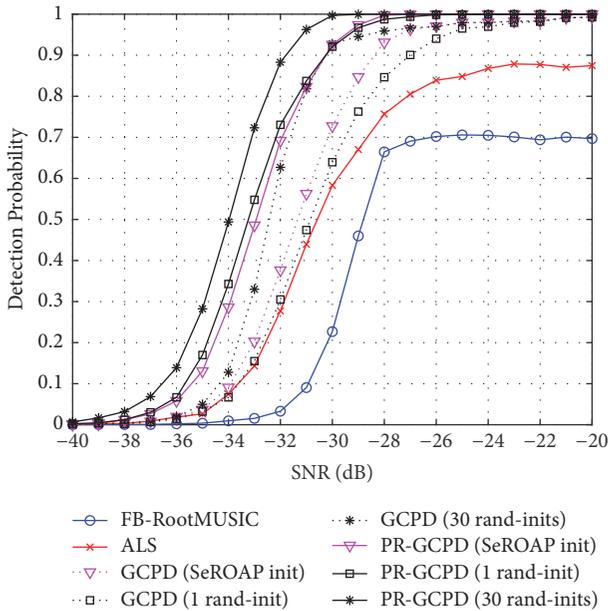


FIGURE 7: Detection probability performance of different methods when coherent target exists.

method performs much better than all the other methods, and the detection probability is 1 at high SNRs.

Figure 8 is given to evaluate the run times of different algorithms relative to the scenario existing five targets and two of them are coherent. The scenarios with noncoherent targets have similar results and will not be shown here. The run times are obtained by using a computer with Intel(R) Xeon(R) CPU E5-2682 v4 CPU, 16 GB RAM. As has been mentioned before, the proposed GCPD algorithm

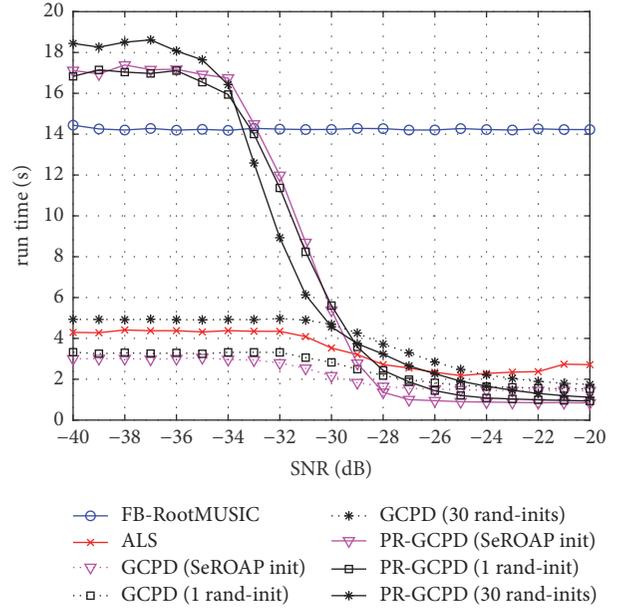


FIGURE 8: Comparison of the required run times of the evaluated algorithms.

with multiple random initializations has the ability to run in parallel. Here we calculate the equivalent run time of GCPD by selecting the largest run time among all these initializations.

From Figure 8 we can see that the required run time of FB-RootMUSIC is constant as it is a finite algorithm. The required run times of GCPD (SeROAP init) and GCPD (1 rand-init) are both smaller than that of the ALS algorithm, especially when the SNR is low. This indicates that the complexities of GCPD (SeROAP init) and GCPD (1 rand-init) are lower than that of ALS. The required run time of the GCPD (30 rand-inits) is slightly larger than that of the ALS algorithm. The reason is that the run times with different initializations are not all equal and the largest one determines the required time.

When the PR method is applied, the run times of all three PR-GCPD algorithms become larger in the low SNR region. This is mainly because that the PR method generally increases the rank and slows down the convergence. However, when the SNR is high, the PR-GCPD algorithm performs much faster. As has been mentioned, the PR method increases the detection and estimation performance of the GCPD algorithm both in low and in high SNR regions, although it may result in a higher computational cost.

5. Conclusion

In this article, we investigated joint range-Doppler-angle estimation in an OFDM-SIMO based RadCom system using CP decomposition. The signal model with tensor algebra was developed and a novel algorithm for CP decomposition was presented. Different from the classical ALS algorithm, the proposed one adopts a greedy strategy with each step solved by TPM with multiple random generated initializations

and a globalization procedure. This globalization procedure alleviates the local optimal problem to some extent. A PR method was proposed to make use of the inherent structures of the factor matrices. We demonstrated that our methods can estimate parameters for multiple targets, both in noncoherent and in coherent cases, and require no pair matching. The multiple random initialized TPM can be easily realized by parallel computing and it is beneficial for realistic applications.

Appendix

Definition 3 (rank-1 tensor). A third-order tensor \mathcal{F} has rank 1 if it can be expressed by the following form:

$$\mathcal{F} = \mathbf{a} \circ \mathbf{b} \circ \mathbf{c}, \quad (\text{A.1})$$

where \mathbf{a} , \mathbf{b} , and \mathbf{c} are three vectors, with $\mathbf{a} \in \mathbb{C}^{I \times 1}$, $\mathbf{b} \in \mathbb{C}^{J \times 1}$, $\mathbf{c} \in \mathbb{C}^{K \times 1}$, and $\mathcal{F} \in \mathbb{C}^{I \times J \times K}$.

Definition 4 (CP decomposition). The CP decomposition of the third-order tensor $\mathcal{F} \in \mathbb{C}^{I \times J \times K}$ is a decomposition of \mathcal{F} with a summation of minimal number of rank-1 tensors:

$$\mathcal{F} = \sum_{r=1}^R \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \quad (\text{A.2})$$

where \mathbf{a}_r , \mathbf{b}_r , \mathbf{c}_r are the r th columns of factor matrices $\mathbf{A} \in \mathbb{C}^{I \times R}$, $\mathbf{B} \in \mathbb{C}^{J \times R}$, and $\mathbf{C} \in \mathbb{C}^{K \times R}$.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest and the received funding did not lead to any conflicts of interest regarding the publication of this paper.

Acknowledgments

The research reported in this article was supported in part by the National Natural Science Foundation of China (61661028 and 61561034).

References

- [1] P. Gandotra, R. K. Jha, and S. Jain, "Green Communication in Next Generation Cellular Networks: A Survey," *IEEE Access*, vol. 5, pp. 11727–11758, 2017.
- [2] F. Zhou, Y. Wang, D. Qin, Y. Wang, and Y. Wu, "Secure EE maximisation in green CR: guaranteed SC," *IET Communications*, vol. 11, no. 16, pp. 2507–2513, 2017.
- [3] F. Zhou, N. C. Beaulieu, Z. Li, J. Si, and P. Qi, "Energy-efficient optimal power allocation for fading cognitive radio channels: ergodic capacity, outage capacity, and minimum-rate capacity," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2741–2755, 2016.
- [4] F. Zhou, Z. Chu, H. Sun, R. Q. Hu, and L. Hanzo, "Artificial Noise Aided Secure Cognitive Beamforming for Cooperative MISO-NOMA Using SWIPT," *IEEE Journal on Selected Areas in Communications*, 2018, to be published.
- [5] C. Sturm and W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1236–1259, 2011.
- [6] J. Moghaddasi and K. Wu, "Multifunctional transceiver for future radar sensing and radio communicating data-fusion platform," *IEEE Access*, vol. 4, pp. 818–838, 2016.
- [7] X. Tian, T. Zhang, Q. Zhang, and Z. Song, "High Accuracy Doppler Processing With Low Complexity in OFDM-Based RadCom Systems," *IEEE Communications Letters*, vol. 21, no. 12, pp. 2618–2621, 2017.
- [8] F. Liu, L. Zhou, C. Masouros, A. Li, W. Luo, and A. Petropulu, "Towards Dual-functional Radar-Communication Systems: Optimal Waveform Design," *IEEE Transactions on Signal Processing*, pp. 1-1.
- [9] W. Wiesbeck, L. Sit, M. Younis, T. Rommel, G. Krieger, and A. Moreira, "Radar 2020: the future of radar systems," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '15)*, pp. 188–191, Milan, Italy, July 2015.
- [10] J. Zhang, I. Podkurkov, M. Haardt, and A. Nadeev, "Efficient multidimensional parameter estimation for joint wideband radar and communication systems based on OFDM," in *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2017*, pp. 3096–3100, USA, March 2017.
- [11] X. Wang, M. Huang, X. Wu, and G. Bi, "Direction of Arrival Estimation for MIMO Radar via Unitary Nuclear Norm Minimization," *Sensors*, vol. 17, no. 4, p. 939, 2017.
- [12] R. Heckel, "Super-resolution MIMO radar," in *Proceedings of the 2016 IEEE International Symposium on Information Theory, ISIT 2016*, pp. 1416–1420, Spain, July 2016.
- [13] D. Nion and N. D. Sidiropoulos, "Tensor algebra and multidimensional harmonic retrieval in signal processing for MIMO radar," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5693–5705, 2010.
- [14] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Review*, vol. 51, no. 3, pp. 455–500, 2009.
- [15] A. Uschmajew, "Local convergence of the alternating least squares algorithm for canonical tensor approximation," *SIAM Journal on Matrix Analysis and Applications*, vol. 33, no. 2, pp. 639–652, 2012.
- [16] A. Stegeman and P. Comon, "Subtracting a best rank-1 approximation may increase tensor rank," *Linear Algebra and its Applications*, vol. 433, no. 7, pp. 1276–1300, 2010.
- [17] A.-H. Phan, P. Tichavský, and A. Cichocki, "Tensor deflation for CANDECOMP/PARAFAC-part I: Alternating subspace update algorithm," *IEEE Transactions on Signal Processing*, vol. 63, no. 22, pp. 5924–5938, 2015.
- [18] A. P. Da Silva, P. Comon, and A. L. F. De Almeida, "An iterative deflation algorithm for exact CP tensor decomposition," in *Proceedings of the 40th IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2015*, pp. 3961–3965, Australia, April 2014.
- [19] A. P. Da Silva, P. Comon, and A. L. F. De Almeida, "A Finite Algorithm to Compute Rank-1 Tensor Approximations," *IEEE Signal Processing Letters*, vol. 23, no. 7, pp. 959–963, 2016.

- [20] L. Huang, Y. Wu, H. C. So, Y. Zhang, and L. Huang, "Multidimensional sinusoidal frequency estimation using subspace and projection separation approaches," *IEEE Transactions on Signal Processing*, vol. 60, no. 10, pp. 5536–5543, 2012.
- [21] N. D. Sidiropoulos, L. De Lathauwer, X. Fu, K. Huang, E. E. Papalexakis, and C. Faloutsos, "Tensor decomposition for signal processing and machine learning," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, pp. 3551–3582, 2017.
- [22] P. Comon, X. Luciani, and A. L. F. de Almeida, "Tensor decompositions, alternating least squares and other tales," *Journal of Chemometrics*, vol. 23, no. 7-8, pp. 393–405, 2009.
- [23] V. de Silva and L.-H. Lim, "Tensor rank and the ill-posedness of the best low-rank approximation problem," *SIAM Journal on Matrix Analysis and Applications*, vol. 30, no. 3, pp. 1084–1127, 2008.
- [24] L. Wang and M. T. Chu, "On the global convergence of the alternating least squares method for rank-one approximation to generic tensors," *SIAM Journal on Matrix Analysis and Applications*, vol. 35, no. 3, pp. 1058–1072, 2014.
- [25] J. B. Lasserre, "Global optimization with polynomials and the problem of moments," *SIAM Journal on Optimization*, vol. 11, no. 3, pp. 796–817, 2000/01.

Research Article

Performance and Power Consumption Analysis of IEEE802.11ah for Smart Grid

Zhe Zheng , Wenpeng Cui, Lei Qiao, and Jinghong Guo

Department of Information and Communication, Global Energy Interconnection Research Institute, State Grid Key Laboratory of Information & Network Security, State Grid Laboratory of Electric Power Communication Network Technology, Beijing, China

Correspondence should be addressed to Zhe Zheng; zhengzhe@geiri.sgcc.com.cn

Received 19 April 2018; Revised 27 June 2018; Accepted 9 July 2018; Published 25 July 2018

Academic Editor: Fuhui Zhou

Copyright © 2018 Zhe Zheng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

IEEE802.11ah is a Wireless Local Area Network (WLAN) designed for the application of Internet of Things (IoT) and Machine to Machine (M2M), mainly used in sensor network, smart metering, car network, health care, and other emerging fields. IEEE802.11ah inherits the IEEE802.11n/ac technology. At present, smart grid has completed the installation of optical fiber communication as its backbone network; WLAN can be used to build new wireless sensor network for smart grid by improving the transmission distance, speed, and power efficiency. The critical features of 802.11ah make it a powerful candidate for WLAN in smart grid, such as intelligent substation sensor network and automatic metering system (AMS). This paper simulates the new added highly robust 1MHz bandwidth and Modulation Coding Scheme (MCS) 10 in the 802.11ah physical layer and analyzes the coverage range and energy-saving performance of 802.11ah based on the simulation results. The analysis shows that the 802.11ah at 1 MHz MCS 10 mode can obtain 2.5-3 dB gain. Combined with advantages of Sub-1GHz working frequency, 802.11ah could enlarge the coverage range by five times, compared to 2.4 GHz Wi-Fi. On the other side, 802.11ah module used in the smart grid can greatly reduce power consumption, especially in the AMS with a 1.1 kW.H power savings per year. 802.11ah not only provides the flexibility and low cost features of wireless communication, but also brings coverage and energy-saving performance improvements, which leads to good economic benefit.

1. Introduction

IEEE 802.11ah is a wireless networking protocol published in 2017 called Wi-Fi HaLow as an amendment of the IEEE 802.11-2007 wireless networking standard [1]. It operates at the Sub-1 GHz (SIG) industrial, scientific, medical (ISM) radio-band [2]. Due to the lower center frequencies, more robust PHY layer, and enhancement on MAC layer protocol, the transmission range is extended to 1km, the data rates are at least 100 kbps, and the number of nodes permitted is raised up to 8191 [1, 3]. Benefiting from the lower power consumption, it can support the application of the Internet of Things (IoT) by creating large groups of STAs and sensors that cooperate to share signals [4]. It aims at providing connectivity to thousands of devices under one access point (AP).

Due to the advantages of wide coverage, flexible access, low investment, and maintenance costs, as well as good scalability [5], wireless communication has been widely used for many years. However, traditional grid does not prefer

the wireless communication for its weakness of reliability. As the development of smart grid, demands for sensor network have been raised to support massive data collection for smart decisions. Therefore, combined with its advantages and improvement on reliability, wireless communication should have a priority for building new network for smart grid.

The networks in substation and automatic metering system (AMS) are most typical in smart grid. WLAN is suitable for the two scenarios by providing the connection between the sensor and the access point in the control room or between the meter and the concentrator. Because smart grid needs to realize better single hop coverage and faster reaction than the other IoT application, the WLAN should improve its transmission distance and speed in order to be admitted [6]. 802.11ah, which claims to extend the coverage to 1 km and provide at least 150 kbps data rate while maintaining a low power consumption, could be a competitive candidate for the new wireless communication of smart grid [7].

To discuss the feasibility of 802.11ah protocol used as a communication standard for smart grid, the data transmission features of the smart grid should be analyzed and compared to the specifications of 802.11ah. In the investigation, a simulation of the 802.11ah physical layer and a derivation of coverage range had been done to prove the feasibility. From the simulation results and analysis, 802.11ah can cover larger area than other 802.11 protocols while still maintaining a enough data rate. A further power consumption analysis was accomplished to prove economic benefit by adopting 802.11ah in smart grid.

2. Improvements on Physical Layer and Power Consumption

2.1. Physical Layer. In the physical layer, 802.11ah operates at the SIG. It not only supports the 2 MHz, 4 MHz, 8 MHz, and 16 MHz bandwidth transmission mode, but also adds 1MHz bandwidth with MCS0-rep2 to achieve >1km coverage range [1]. It also supports relaying to obtain further improvement on the coverage range [8]. Directional antenna is also supported to solve the hidden terminal problem.

The target spectrum of 802.11ah is as follows: 863-868.6 MHz (Europe), 950.8 MHz-957.6 MHz (Japan), 755-787 MHz (China), 917-923.5 MHz (Korea), and 902-928 MHz (USA) [1]. For example, the spectrum in China is shown in Figure 1. 802.11ah adopts OFDM to improve the transmission data rate. The subcarriers could be divided into 1MHz and 2 MHz and above bandwidth system. For 2 MHz and above bandwidth, the distribution of subcarriers follows the rules of 802.11ac with one-tenth of the frequency. For 1MHz bandwidth, 802.11ah utilizes 32 subcarriers and 32-point IFFT. It includes 1 DC component, 2 pilot components, 24 data carriers, and 5 zero components [3].

The subcarrier spacing Δf is 31.25 kHz, one-tenth of 802.11ac. The guard interval (GI) duration TGI is 8us and 4us for short GI, respectively. The OFDM symbol duration T_{SYML} with long GI is 40 us and 36 us for short GI, respectively.

The modulation and coding scheme (MCS) of 802.11ah are presented in Table 1. Differing from 802.11ac, a robust mode MCS10 has been added.

2.2. Power Optimization Technology. The IEEE802.11ah standard defines two power consumption management modes, namely, activation mode and low power consumption mode. In the activation mode, the STA keeps the RF module in an open state, and the data can be continuously transmitted when there are upstream and downstream data operations. On the other hand, in low power mode, STA switches between hibernation and activation [2]. When hibernating, the STA completely shuts down the RF module and stops receiving signal; in this case, even if the AP has downstream data transmission to the STA, the data can only be cached by AP. When the STA wakes up, switches to the active mode, and sends the downstreaming request, the AP can send the cached data to the STA.

In the IEEE802.11 system, AP periodically broadcasts Beacon frames, and STA periodically examines the TIM IE in

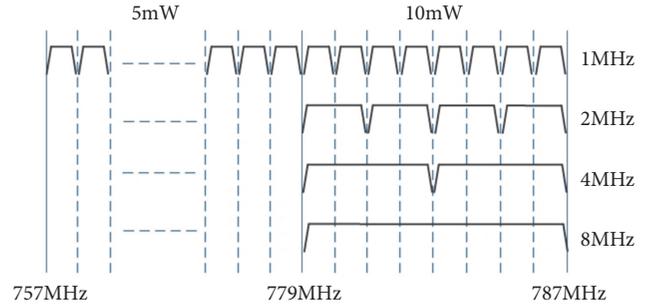


FIGURE 1: 802.11ah frequency band and transmit power limits.

TABLE 1: Modulation and coding schemes.

MCS	Mod	Rate	Data-Rate (kbps)
0	BPSK	1/2	300
1	QPSK	1/2	600
2	QPSK	3/4	900
3	16-QAM	1/2	1200
4	16-QAM	3/4	1800
5	64-QAM	2/3	2400
6	64-QAM	3/4	2700
7	64-QAM	5/6	3000
8	256-QAM	3/4	3600
9	256-QAM	5/6	4000
10	BPSK	1/4	150

Beacon to check if there is data to be received. If so, a PS-Poll must be sent to request the AP to start a packet transmission. After completing the transmission, STA goes to sleep mode.

However, under this mechanism, problems arise when there are a large number of STAs in the network. The TIM IE Partial Virtual Bitmap of Beacon frames will become very long. If overmuch packets are cached in AP, many low power STA may not be able to complete the data transmission in a Beacon interval and always maintain activation patterns, which causes huge power consumption.

In order to solve this problem, IEEE802.11ah standard puts forward a new traffic information indicator (TIM) diagram and the Page Segmentation strategy, the AP change the virtual bitmap from one page into multiple pages, and each Beacon carries their corresponding Page Segmentation STA. A new Segment Count IE is introduced, containing the splitting distribution and boundary information for each page segment.

DTIM Beacon is a new Beacon frame that contains the new TIM IE for grouping processing packets. When DTIM Beacon is broadcasted or multibroadcasted, all Group STA need to be woken up to see if there is cache data in the group. If the entire group does not have cached data, the whole group stays asleep throughout the DTIM cycle until the next TIM Beacon arrives. DTIM cycle is an integer multiple of TIM, giving STA more time to sleep, resulting in a significant reduction on power consumption.

In addition, 802.11ah introduces two other optimization strategies [2].

Maximum Sleep Time Expansion: use the maximum two bits of the maximum idle cycle field as a scaling factor .00,01,10,11, which represents the scale factor of 1,10,1000 and 10000. The maximum sleep time can be extended up to 2500 times. Second, allow the AP to set the maximum free period for different values at different STAs, and the STA can request the allocation of specific values in the associated request.

Target Wakeup Time (TWT): TWT mechanism allows AP to arrange STA wakeup time, which makes different stations wake up at different time, thereby reducing competition and collision, improving the system efficiency and also greatly reducing the power consumption of the STA. Each time the STA wakes up for several TWT Service Periods (SP), the station sends data through the traditional channel access mode during the TWT SP. STA can stay awake outside of TWT SP time slots, and may even receive the Beacon frame when sleeping. 802.11ah also proposes TWT grouping mechanism, AP builds TWT group, and notifies STA grouping information.

3. Demands of Smart Grid

As the development of smart grid, the communication network has been getting increasing attention. Compared with the wired communication technology, wireless communication technology has the advantages on mobility, simplicity, and flexibility of installation, cost reduction, and expandability.

The emerging M2M wireless communication standards, IEEE 802.11ah, are expected to be used in intelligent substation sensor network and AMS, which is the most common WLAN in smart grid.

3.1. Sensor Network in Substation. Online monitoring is one of the most important applications in power system. As the development of the online monitoring technology, it has been widely admitted and adopted in condition overhaul in the substation [5, 9].

The online monitoring system must include large number of diverse sensors, like temperature sensor, humidity sensor, CT sensor, PT sensor, vibration sensor, stress sensor, etc. 802.11ah, with the characteristics of long coverage range, low power consumption, large number of access nodes, and fairly high transmission data rate, is very suitable for the communication network of online monitoring in substation to form a connection among the sensor nodes. 1km communication distance guarantees the whole coverage for one substation. 8000+ number of stations guarantees that the capacity for new sensors is plenty. Low power consumption provides the possibility of ten years' duration with small battery. High data rate provides capability for some light video stream nodes. With some modified in the MAC layer, 802.11ah will meet the real-time requirements for industrial application.

3.2. Automatic Metering System. Traditional power meter reading is simple due to the small amount of transmission

data. The requirement for real-time transmission is not urgent [10].

The development of smart grid requires real-time mutual interactions between grid and users. Grid needs more data from users to regulate the generation and consumption. Therefore, a new communication system for the automatic metering system, with higher transmission speed, real-time performance, and reliability, should be deployed. For example, traditional AMS only needs to update the user data once in a day, but future AMS needs to update once in 5 minutes. Only in this way is the grid able to realize new function, such as demand response, relying on the real-time data collected.

Now in China, Broadband Power Line Communication (PLC) is being rapidly applied. Compared to the narrowband PLC, Broadband PLC has raised the transmission data rate from 5 kbps-50 kbps [11] to 100 kbps-2 Mbps, which is a great promotion. But similar to narrowband PLC, broadband PLC malfunctions when noise or impedance mismatch on power lines becomes too big, which always happens when power electronic devices, such as electronic vehicles, are connected to grid. Thus, wireless communication could act as a backup to avoid problems above. The present main wireless communication protocol in China is micropower wireless communication, which only supports a 1-20 kbps transmission data rate. 802.11ah supports 150 kbps at least when using the most robust mode by adopting OFDM and repetition technology. It also improves the real-time performance and reliability at the same time. These features of 802.11ah provides great potential for using it in the AMS communication of smart grid.

4. Simulation of PHY Layer

According to the analysis above, for smart grid application, the transmission data rate of 802.11ah 1MHz mode meets the demands for most of the sensor networks and AMS. And the robustness provided by the 1MHz MCS10 could also promote the reliability of grid communication system. Therefore, 1 MHz mode is simulated in this paper.

This paper use Simulink® tools to simulate the physical layer of 802.11ah 1MHz mode. The communication system toolbox in the Simulink has also been used.

4.1. Structure of PHY Layer Simulation. The structure of 802.11ah physical layer [1, 3] is as in Figure 2, which includes blocks as PHY service data unit (PSDU) generator, MCS setting, transmitter, receiver, error rate calculation (ERC), and signal to noise (SNR) calculation.

PSDU generator block uses a random integer generator to simulate the data from the MAC layer. MCS setting block controls the modulation scheme. According to the MCS chosen, transmitter block processes the data and passes it to channel model block which adds the channel affection to the signal; receiver block processes the data from the channel model block in a reverse flow. ERC block at the end of the data link provides the bit error rate (BER) by delaying the original data from the PSDU block and comparing it to the output data of the receiver block. SNR estimation block calculates

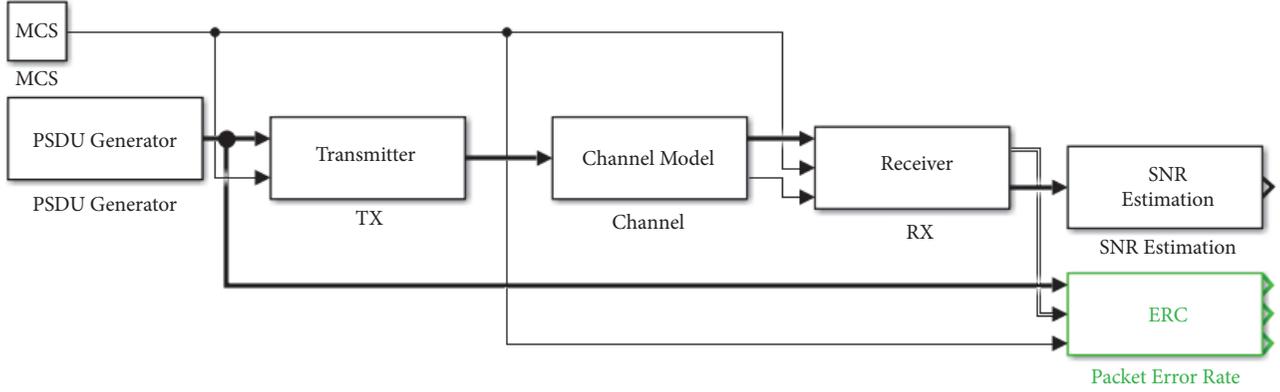


FIGURE 2: Structure of whole PHY layer simulation.

the SNR from the final signal. The transmitter and receiver are the critical blocks of the whole module.

The structure of the transmitter block is shown in Figure 3. It includes modulation, pilots inserting, zeros adding, IFFT, and cyclic prefix appending blocks. The data stream follows the flow of the 802.11ah standards.

The structure of the receiver block is shown in Figure 4. It includes time domain synchronizing, FFT, frequency domain synchronizing, and demodulation blocks.

4.2. Structure of Repetition Module in MCS 10. The IEEE802.11ah provides two transmitter models. One is the regular nonrepetition MCSs, which is consistent with the IEEE802.11ac transmitter [1].

The second type of transmitter model is MCS0-rep2 model, named MCS 10 [1, 10]. As the new scheme proposed, the purpose of this model is to improve the coverage of both

indoor and outdoor scenes by using repetition technology when guaranteeing a 150 kbps transmission speed.

The most important feature of MCS 10 is the repetition mechanism adopted in the 24 data subcarriers. The realization by the transmitter is to place the copy of the front 12 bits to the 12 bits behind. On the receiver side, a comparison between upper 12 bits and down 12 bits is made, in order to further improve the transmission reliability.

According to the standard definition, the transmitter of MCS 10 inserts the repetition module after the encoder module and before the interleaver module, as shown in Figure 5. Formula (1) shows the logic operation.

$$R_{24} = \{R_{12}, \text{xor}(R_{12}, [1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1])\}. \quad (1)$$

The receiver of MCS 10 receiver inserts the derepetition module after the deinterleaver module and before the Viterbi decoder module, as shown in the Figure 6. Formula (2) shows the logic operation.

$$R_{12} = \frac{R_{u12} + (|R_{d12}| \times \{\text{xor}[(\text{sign}(R_{d12}) + 1) / 2], [1, 0, 0, 0, 0, 1, 0, 1, 0, 1, 1, 1]\})}{2}. \quad (2)$$

4.3. Simulation Methods. In smart grid, the reliability of the communication has higher priority than the transmission data rate; therefore 16-QAM, QPSK, BPSK, and 1 MHz bandwidth will be enough in most of the applications. This paper simulates the performance of the physical layer in the case of 1 MHz bandwidth and MCS 0,1,2,3,4,10. The parameters of the simulation are listed in Table 2.

In the simulation, the Channel Model is AWGN with multipath, SNR is from -5 to 20 with a step of 0.5, code/decode adopts BCC/Turbo, synchronization and frequency offset correction are assumed to be ideal, data length is 1440 which is 20 symbols, and antenna mode chooses SISO.

A Matlab .m script runs the simulation, records all the BER results under different conditions, and plots the results in one figure for analysis.

5. Results and Analysis

5.1. BER Statistics. The BER statistics versus the SNR for different MCS and 1 MHz bandwidth are shown in Figure 7. It is clear that the tolerance to SNR goes lower when more complicated modulation scheme is chosen.

When MCS is 4, modulation is 16-QAM, code rate is 3/4, and the BER reaches 10⁻¹ when SNR equals 13.5 dB. In most applications of smart grid, the SNR cannot reach 15 dB; therefore it is concluded that MCS 4,5,6,7,8,9 are not suitable for the smart grid for lack of tolerance to SNR.

According to the trend of the data, when SNR is good enough, the BER for different MCSs is too small to tell a performance difference. But when the SNR becomes lower, the effects could be discovered. There is about 3 dB improvement when MCS moves one step in the robust direction.

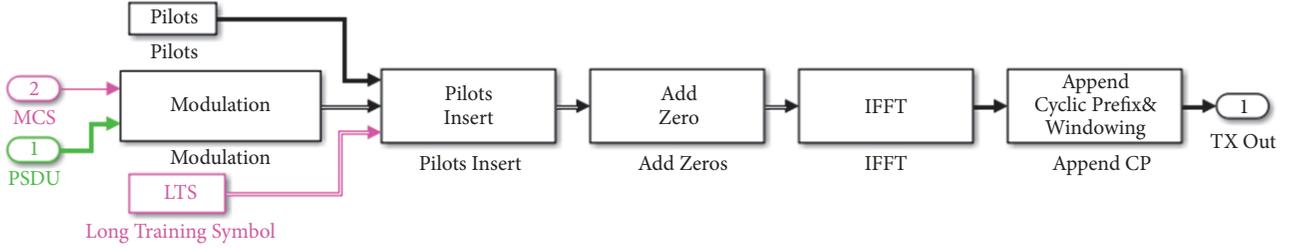


FIGURE 3: Structure of PHY layer transmitter.

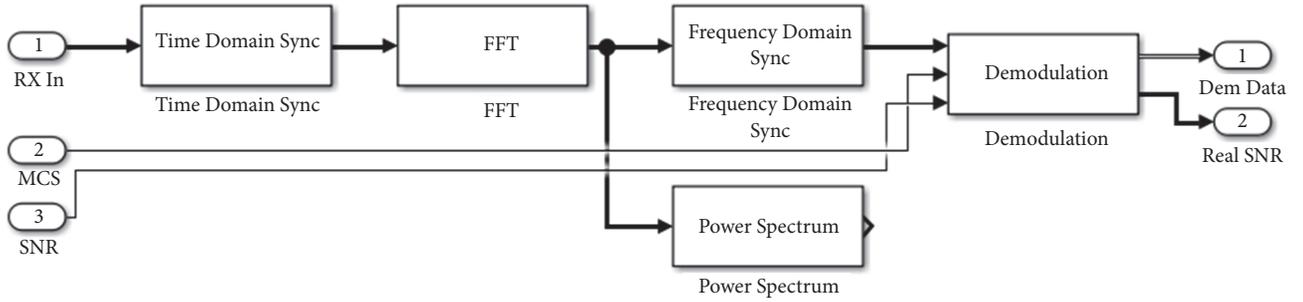


FIGURE 4: Structure of PHY layer receiver.

According to experience, BER equals 20% is set to the limit of malfunctions, and the SNR can reach a practical value of 10 dB in smart grid. When MCS equals 0,1,2, and 10, it satisfied the requirements. According to Table 1, 802.11ah could provide transmission capability of 150 kbps, 300 kbps, 600 kbps, and 900 kbps. This conclusion of transmission data rate is similar to the result of broadband PLC field test.

In the scheme of MCS 0, the most robust scheme without repetition, the BER will reach the limit when SNR reduces to 0. This paper also simulates the scheme of MCS 10, which applies the repetition. From result shown in Figure 4, a 2.5-3 dB improvement has been found compared to MCS 0 when the SNR is smaller than 0. When the SNR restores to 2-2.5 dB, there is no difference between MCS 10 and MCS 0.

5.2. Coverage Analysis. In the 802.11ah communication system, link budget can be expressed with formula (3) [4, 7], where P_{RX} is the receiving power, P_{TX} is the transmitting power, G_{RX} and G_{TX} are the antenna gain of receiver and transmitter, L_{RX} is the receiver system loss, L_{TX} is the transmitter system loss, and $PL(f, d)$ is the path loss.

$$P_{RX} = P_{TX} + G_{TX} - L_{TX} - PL(f, d) + G_{RX} - L_{RX}. \quad (3)$$

If we neglect system losses L_{RX} and L_{TX} and take fade margins, SNR, and thermal noise power density into account, formula (4) is obtained [4, 7].

$$R_{dB}(f, d) = P_{TX} + G_{TX} - PL(f, d) - FM(d) + G_{RX} - \left(\frac{S}{N}\right)_{dB} + N_o. \quad (4)$$

TABLE 2: Parameters of simulation.

Simulation	Parameter
MCS	0,1,2,3,4,10
Channel	AWGN+Multipath
SNR	(-5,0.5,20)
Code/Decode	BCC/Turbo
Bandwidth	1MHz
Synchronization	Ideal
Frequency Offset Correction	Ideal
Data Length	1440
Antenna	SISO

The SNR can be expressed as

$$\left(\frac{S}{N}\right)_{dB} = \left(\frac{E_b}{N_o}\right)_{dB} + 10 \log_{10} \left(\frac{R}{B}\right). \quad (5)$$

By substituting formula (5) into formula (4), we have

$$R_{dB}(f, d) = P_{TX} + G_{TX} - PL(f, d) - FM(d) + G_{RX} - \left(\frac{E_b}{N_o}\right)_{dB} - 10 \log_{10} \left(\frac{R}{B}\right) + N_o. \quad (6)$$

In (6), E_b is the required energy for each information bit transmission, N_o is the thermal noise power density, R is bit rate, and B is system bandwidth.

Since there is currently no 802.11ah commercial product for testing, it is practical to estimate its coverage performance by comparing it with 2.4 GHz Wi-Fi. For example, in China, the transmitting power is 10 mW, and the working frequencies

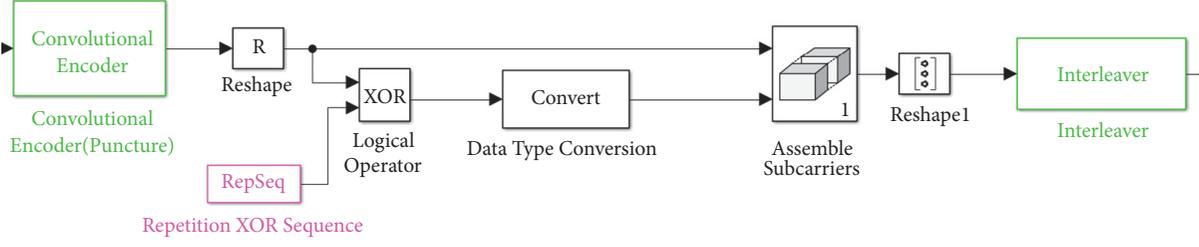


FIGURE 5: Repetition module in transmitter modulation.

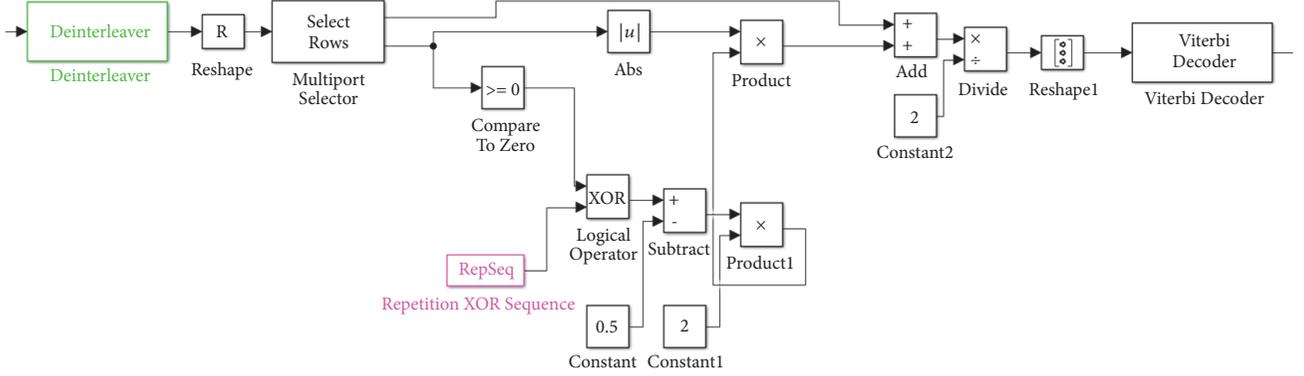


FIGURE 6: Derepetition module in receiver demodulation.

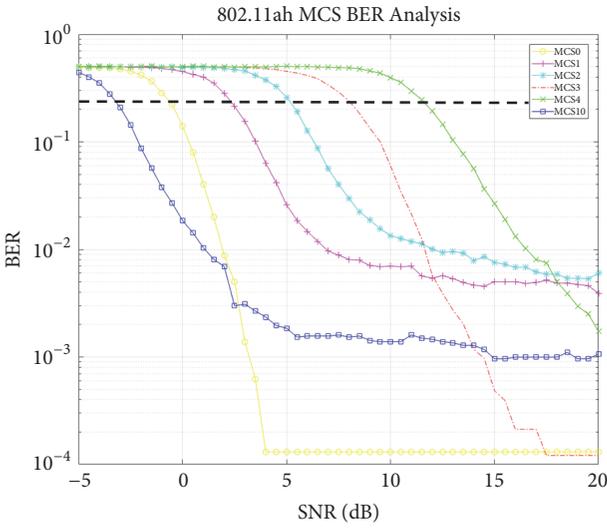


FIGURE 7: BER versus SNR with different MCS.

are 779-787 MHz. If using the most robust scheme, MCS 10, 1 MHz bandwidth and repetition mode are adopted.

$PL(f, d)$ can be expressed by formula (7), in which the f_{MHz} is the working frequency in MHz and the d_{km} is the transmission distance in kilometers.

$$PL(f, d) = 32.44 + 20 \log_{10} f_{MHz} + 20 \log_{10} d_{km}. \quad (7)$$

TABLE 3: Improvements of 11ah in dBs compared to 2.4 GHz Wi-Fi.

Parameter	Improvement of 780 MHz 11ah Over 2.4 GHz	Critical Impact Factor	
		11ah	WiFi
Transmit Power compared to 17 dBm	-7 dB	10 dBm	17 dBm
TX and RX antenna gain	0 dB	-	-
Free Space Path Loss	9.8 dB	780 MHz	2.4 GHz
Noise Bandwidth (1 MHz)	13 dB	1 MHz	20 MHz
Flat Fading	-4.5 dB	780 MHz	2.4 GHz
2x repetition coding	2.5 dB	MCS 10	-
Total	13.8 dB	-	-

Therefore, the corresponding difference value (dB) of each parameter compared with 2.4 GHz Wi-Fi can be expressed as

$$\Delta P_{TX} = (10mW)_{dBm} - 17dBm, \quad (8)$$

$$\Delta PL(f, d) = 20 \log_{10} f_{MHz,ah} - 20 \log_{10} f_{MHz,2.4G}, \quad (9)$$

$$\Delta \left(\frac{S}{N} \right)_{dB} = 10 \log_{10} \left(\frac{R}{B_{ah}} \right) - 10 \log_{10} \left(\frac{R}{B_{2.4G}} \right). \quad (10)$$

With the calculation of the above equations, the results are shown in the Table 3 [12]. The total improvement of 802.11ah compared with 2.4 GHz Wi-Fi is 13.8 dB. Formulas (11), (12), and (13) [4] approximately express the indoor and outdoor

$PL(d)$. It is concluded that the path loss of both indoor and outdoor situation has an exponential relation to the extreme transmission distance.

$$PL(d) = 30.29 + 20 \log_{10} d, \quad (11)$$

$$PL(d) = L_{FS} = 20 \log_{10} \left(\frac{(4\pi f d f)}{c} \right) \quad \text{for } d \leq d_{BP}, \quad (12)$$

$$PL(d) = L_{FS} + 3.5 \log_{10} \left(\frac{d}{d_{BP}} \right) \quad \text{for } d > d_{BP}, \quad (13)$$

13.8 dB improvement can be equivalent to about 5 times transmission distance. According to the field test experience of transformer substation, the transmission distance of 2.4 GHz Wi-Fi is about 30~50 m indoor and 100~200 m outdoor. Therefore, it can be inferred that the transmission distance of 802.11ah is about 150~250 m indoor, and about 500~1000 m outdoor, which is good enough for most of the scenes in smart grid.

5.3. Power Consumption Analysis. In the applications of AMS, in order to realize the demand response, the current metering frequency, one time per day, is required to upgrade to one time per 5 minutes. According to the electric power company estimation, each metering transmission for demands response has a quantity of 20k bits of data to report. If the transmission rate of 802.11ah is 100 kbps, assuming one concentrator manages up to 300 power meters, a 60-second duration, at most, is needed for data transmission, in which the communication module works at maximum power. Hence, a 802.11ah communication STA module will be in sleep mode for 80%, in receiving mode for 19.94%, and transmitting mode for 0.06% of the 5-minute period. As shown in the figure, the power consumption budget of the 802.11ah chip designed by the State Grid is shown in Table 4.

The average power consumption of 11ah STA module could be calculated by (14), which is 309.8 mW.s in 5 minutes. It equals 0.024 W.h per day.

$$P_{average} = P_{total} \times T_{tx} + (P_{rx} + P_{MCU}) \times T_{rx} + P_{sleep} \times T_{sleep}. \quad (14)$$

Compared to the broadband PLC module with the transmission rate of the same 100 kbps. The power of STA module is 3 watts in maximum load. Using equation (x), it consumes 0.528 W.h per day. Using 802.11ah module, each module will significantly save about 0.5 Wh per day. In China, about 600 million smart meters have been installed, which could save up to 300k kW.h per day, and 1.1 billion kW.h per year.

In the application of intelligent substation sensor network, referring to industrial automation work patterns, 11ah communication module is in sleep mode for 99.8% of the time. In the remaining 0.2%, receiving takes 95.7%, transmitting takes 0.1%, and idle states takes 4.2%. One 2200mAh lithium-ion battery can be used for five years [2]. Because the substation will be maintained two times a year, 802.11ah module fully meets the demand as a sensor network node in intelligent substation.

TABLE 4: Estimated power consumption of 11ah chip designed by state grid.

11ah Components		Power Consumption
TX	TX AFE	5 mW
	TX Power	10 mW
	TX Digital	0.8 mW
RX	RX AFE	5 mW
	RX Digital	1.2 mW
MCU(MAC,70 MHz)		15 mW
Total		37 mW
Sleep Mode		10 uW

6. Conclusion

According to the results and analysis above, it is shown that 802.11ah MCS 10 indeed improves the performance by raising the tolerance to SNR by 2.5-3 dB, which will benefit the distance and reliability of communication. Furthermore, except MCS 0,1,2,10, MCS 3,4,5,6,7,8,9 are not suitable for harsh communication condition.

When used in smart grid, 802.11ah is able to provide larger coverage range with a radius of up to 1km and significant energy-saving results with great economic benefits.

Therefore, 802.11ah standard could be adopted by smart grid communication, but it is suggested to choose the MCS 0,1,2,10 to achieve compromise between transmission data rate and reliability.

Data Availability

The authors declare that the data supporting the findings of this study are available within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work has been supported by the project of development on wireless communication chip of electric power system based on 802.11ah under Grant no. SGRIXTMMXS [2015]1128.

References

- [1] IEEE Standards for Information technology-Telecommunications and information exchange between systems Local and metropolitan area networks-Specific requirements Part 11: Wireless LAN Medium Access Control(MAC) and Physical Layer(PHY) Specifications. Amendment 2: Sub 1 GHz License Exempt Operation, IEEE Std 802.11ahTM-2016.
- [2] T. Adame, A. Bel, B. Bellalta, J. Barcelo, and M. Oliver, "IEEE 802.11AH: the WiFi approach for M2M communications," *IEEE Wireless Communications Magazine*, vol. 21, no. 6, pp. 144-152, 2014.

- [3] S. Aust, R. V. Prasad, and I. G. M. M. Niemegeers, "Outdoor Long-Range WLANs: A Lesson for IEEE 802.11ah," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1761–1775, 2015.
- [4] A. Hazmi, J. Rinne, and M. Valkama, "Feasibility study of IEEE 802.11ah radio technology for IoT and M2M use cases," in *Proceedings of the 2012 IEEE Globecom Workshops, GC Wkshps 2012*, pp. 1687–1692, USA, December 2012.
- [5] X. Bao, G. Wang, Z. Hou, M. Xu, L. Peng, and H. Han, "WDM switch technology application in smart substation communication network," in *Proceedings of the 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT)*, pp. 2373–2376, Changsha, China, November 2015.
- [6] Q. Zhang, X. Wang, J. Wang, C. Feng, and L. Liu, "Survey of demand response research in deregulated electricity markets," *Dianli Xitong Zidonghua/Automation of Electric Power Systems*, vol. 32, no. 3, pp. 97–106, 2008.
- [7] S. Aust, R. V. Prasad, and I. G. M. M. Niemegeers, "Performance evaluation of Sub 1 GHz wireless sensor networks for the smart grid," in *Proceedings of the 37th Annual IEEE Conference on Local Computer Networks, LCN 2012*, pp. 292–295, USA, October 2012.
- [8] B. Zhou, H. Hu, S.-Q. Huang, and H.-H. Chen, "Intracluster device-to-device relay algorithm with optimal resource utilization," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 2315–2326, 2013.
- [9] BIAN Baoyin, LI Xiucai, HUANG Xin. Research on wireless communication model and application analysis in substation [J]. *Electrotechnical Application*, 2015.
- [10] S. Aust, R. V. Prasad, and I. G. M. M. Niemegeers, "IEEE 802.11ah: Advantages in standards and further challenges for sub 1 GHz Wi-Fi," in *Proceedings of the 2012 IEEE International Conference on Communications, ICC 2012*, pp. 6885–6889, Canada, June 2012.
- [11] A. Haidine, B. Adebisi, A. Treytl, H. Pille, B. Honary, and A. Portnoy, "High-speed narrowband PLC in smart grid landscape—state-of-the-art," in *Proceedings of the IEEE International Symposium on Power Line Communications and Its Applications (ISPLC '11)*, pp. 468–473, Udine, Italy, April 2011.
- [12] Dr. Eldad Perahia, Enabling Wi-Fi Internet of Things with 802.11ah Technology. Internet of Things Group, July 2015.

Research Article

Energy Efficiency Maximization of Dynamic CoMP-JT Algorithm in Dense Small Cell Networks

Xuefei Peng, Jiandong Li , and Yifei Xu

State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China

Correspondence should be addressed to Jiandong Li; jdli@xidian.edu.cn

Received 2 April 2018; Revised 13 June 2018; Accepted 21 June 2018; Published 12 July 2018

Academic Editor: Shunqing Zhang

Copyright © 2018 Xuefei Peng et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We firstly formulate the energy efficiency (EE) maximization problem of joint user association and power allocation considering minimum data rate requirement of small cell users (SUEs) and maximum transmit power constraint of small cell base stations (SBSs), which is NP-hard. Then, we propose a dynamic coordinated multipoint joint transmission (CoMP-JT) algorithm to improve EE. In the first phase, SUEs are associated with the SBSs close to them to reduce the loss of power by the proposed user association algorithm, where the associated SBSs of each small cell user (SUE) form a dynamic CoMP-JT set. In the second phase, through the methods of fractional programming and successive convex approximation, we transform the EE maximization subproblem of power allocation for SBSs into a convex problem that can be solved by proposed power allocation optimization algorithm. Moreover, we show that the proposed solution has a much lower computational complexity than that of the optimal solution obtained by exhaustive search. Simulation results demonstrate that the proposed solution has a better performance.

1. Introduction

In future 5G wireless network, EE improvement is arisen as a challenging issue [1]. Therefore, how to promote EE is of great significance. With the explosive increasing data traffic demands of users in 5G wireless cellular network, one of the prospective solutions for satisfying the data rate requirement of users and improving EE is the deployment of low power, low cost, and small coverage range SBSs. Small cell tier is an integral part of 5G heterogeneous cellular network architecture that can provide more opportunities for users to connect to the networks close to them, which will decrease power consumption [2, 3].

However, due to the dense deployment and spectrum sharing of small cells, the interference among small cells becomes a key factor that influences EE of the network [4].

On the one hand, several recent works [5–7] have adopted scheduling schemes to mitigate interference in small cell networks. In [5], the authors studied the performance of different scheduling methods under distinct channel models with considering the intersite distance. In [6], the authors formulated a fast convergent speed algorithm to solve the energy-efficient

multijob scheduling function. In [7], dynamic clustering framework of multicell scheduling based on graph was proposed to mitigate intercell interference in dense small cell networks, and channel-aware resource allocation was incorporated in the dynamic clustering framework to provide different levels of tunable quality of service. However, EE optimization for dynamic coordination among SBSs is not considered [5–7].

On the other hand, some studies have discussed multicell CoMP transmission (CoMP) [8–15] of cellular network. CoMP-JT is a promising technique that guarantees data availability at multiple coordinated SBSs transmitted to a user simultaneously to improve the received signal quality of the user and reduce intercell interference, which is proposed in Third Generation Partnership Project (3GPP) LTE-Advanced systems [8]. In [9–11], the authors studied CoMP-JT from the aspects of throughput and coverage probability. However, EE optimization is not considered. In [12], the authors proposed two algorithms to tackle the problem of minimizing backhaul user data transfer by establishing CoMP joint processing beamforming matrix of multicell. In [13, 14], the authors formulated EE maximization problem with global precoding

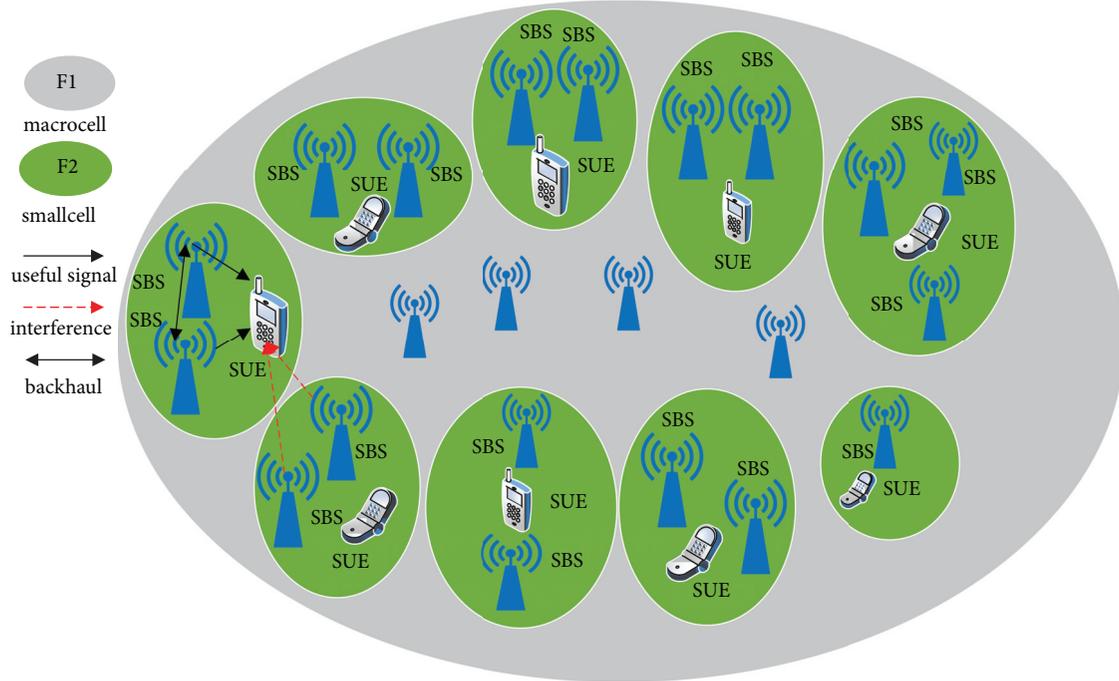


FIGURE 1: Network model.

matrix design. However, the beamforming and precoding matrix of global optimization are complex and difficult to realize in practical systems [12–14]. In [15], the authors proposed a distributed algorithm to solve the weighted EE maximin fairness problem for CoMP systems. However, dynamic coordinated transmission among SBSs is not considered.

Unlike existing literature, a dynamic CoMP-JT algorithm considering user association, power allocation, and minimum data rate requirement of SUEs is proposed in this paper to mitigate interference and maximize EE.

The main contributions of this paper are summarized as follows:

(i) We utilize dynamic CoMP-JT as an interference management technique to mitigate interference and improve EE, where minimum data rate requirements of SUEs and maximum transmit power constraint of SBSs are considered.

(ii) We propose to solve the NP-hard problem of EE maximization through the proposed dynamic CoMP-JT algorithm. In the first phase, each SUE is associated with SBSs close to them to reduce the loss of power by the proposed user association algorithm. In the second phase, we transform the formulated EE optimization problem into a convex problem by methods of fractional programming and successive convex approximation. Finally, we solve the formulated problem by the proposed power allocation optimization algorithm.

(iii) We analyze the computational complexities of the optimal solution and the proposed solution and compare the performance of near-optimal solution, no-CoMP, and random-CoMP with the proposed solution.

The rest of this paper organized as follows. The system model and problem formulation are given in Section 2. In Section 3, the dynamic CoMP-JT algorithm is proposed. In

Section 4, numerical results are given. Finally, the paper is concluded in Section 5.

2. System Model and Problem Formulation

Our system model is depicted in Figure 1. We consider a downlink orthogonal frequency division multiple access (OFDMA) network, where macrocell tier and small cell tier are allocated with orthogonal spectrums. Therefore, there is no cross-tier interference and we focus on the EE analysis of small cell tier, where only the interference between SBSs and SUEs is considered. Assume that SUEs are randomly and uniformly distributed in the coverage region of SBSs. Let $S = \{1, 2, \dots, S\}$ denote the set of all SBSs with low transmit power and $U = \{1, 2, \dots, U\}$ denote the set of SUEs. We will next derive the optimization problem of EE, where the EE is defined as the transmitted bits per unit energy consumption and equals the ratio of the sum data rate to the total power consumption. We define a $SU \times 1$ vector $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_s, \dots, \mathbf{p}_S)^T$ to represent the power of all SBSs allocate for all SUEs, where $\mathbf{p}_s = (p_{s1}, p_{s2}, \dots, p_{su})$ denotes the power of SBSs allocated for SUE 1 to SUE U . Let \mathbf{x} denote a binary vector to represent the association relationship between SBSs and SUEs. Its element is an association indicator decision variable $x_{su} \in \{0, 1\}$ denoting whether SBS s is associated with SUE u , $x_{su} = 1$, if SBS s is associated with SUE u ; otherwise $x_{su} = 0$.

From [16], we can derive the received signal strength of SUE u associated with one or more SBSs as follows:

$$p_u^r = \sum_{s \in S} (g_{su} \sqrt{x_{su} p_{su}})^2 = \sum_{s \in S} x_{su} p_{su} h_{su} \quad (1)$$

where p_{su} is SBSs transmit power for SUE u and g_{su} is the channel gain from SBS s to SUE u , which includes path loss, Rayleigh, and shadowing fading. Moreover, we assume that $h_{su} = g_{su}^2$.

Then, we can obtain the signal to interference and noise ratio (SINR) of SUE u as follows:

$$\Gamma_u(\mathbf{x}, \mathbf{p}) = \frac{P_u^r}{\sum_{l \in S, l \neq u} (x_{li} P_{lu} h_{lu}) + \delta^2}, \quad (2)$$

where the first part of the denominator denotes the interference of SUE u from the noncooperating SBSs. The second part of denominator δ^2 is the power of additive white Gaussian noise. Furthermore, we can get the data rate of SUE u as follows:

$$R_u(\mathbf{x}, \mathbf{p}) = \Delta f \log_2(1 + \Gamma_u(\mathbf{x}, \mathbf{p})), \quad (3)$$

where Δf is the bandwidth of the network.

The power consumption of SUE u can be also derived as

$$P_u(\mathbf{x}, \mathbf{p}) = x_{su} p_{bck} + \sum_{s \in S} x_{su} (\Delta_p p_{su} + p_{su}^c), \quad (4)$$

where p_{bck} is the power consumption of backhaul. Δ_p is a constant concerning the power amplifier efficiency and p_{su}^c denotes the circuit power consumption.

Then, we can formulate the EE maximization problem of joint user association and power allocation as follows:

$$\begin{aligned} \max_{\mathbf{x}, \mathbf{p}} \quad & \eta_{EE} = \frac{R_{tot}(\mathbf{x}, \mathbf{p})}{P_{tot}(\mathbf{x}, \mathbf{p})} = \frac{\sum_{u=1}^U R_u(\mathbf{x}, \mathbf{p})}{\sum_{u=1}^U P_u(\mathbf{x}, \mathbf{p})} \\ \text{s.t.} \quad & (5a) : R_u \geq R_u^{\min}, \quad \forall u \in U \\ & (5b) : \sum_{u=1}^U p_{su} \leq p_s^{\max}, \quad \forall s \in S \\ & (5c) : \sum_{u=1}^U x_{su} \leq 1, \quad \forall s \in S \\ & (5d) : x_{su} = \{0, 1\}, \quad \forall x_{su} \in \mathbf{x}, \end{aligned} \quad (5)$$

where (5a) specifies the minimum system data rate requirement R_u^{\min} of SUE u . Equation (5b) is the individual maximum transmit power constraint p_s^{\max} of SBS s . Equations (5c) and (5d) are CoMP-JT constraint, which guarantees that one SBS can only be associated with one SUE and one SUE can be served by one or more SBSs.

Owing to the NP-hardness of joint user association and power allocation, obtaining its optimal solution by exhaustive scheme will incur unaffordable computational complexity [17, 18]. Therefore, we propose a dynamic CoMP-JT algorithm to solve problem (5).

3. Dynamic CoMP-JT Algorithm

The proposed dynamic CoMP-JT algorithm consists of two phases, which are user association forming the CoMP-JT set and power allocation optimization.

3.1. Proposed User Association Algorithm. In this phase, we will determine the association relationship between SBSs and SUEs by the proposed user association algorithm.

According to the path loss from SBSs to SUEs, we firstly determine the size and the set of CoMP-JT SBSs for each SUE. Let M_u denote the set of associated SBSs of SUE u and S' denote the set of SBSs that has not been chosen by other SUEs. In each round, we add the SBS with the minimum path loss for SUE u from the set S' to the set of associated SBSs of SUE u according to the principle $M_u = M_u \cup \arg \min_{s \in S'} \{PL_{su}\}$, where the path loss from SBS s to SUE u is PL_{su} in our model. We will repeat the above procedure until $S' = \emptyset$. Since SBSs are densely deployed in our considered scenario, therefore, each SUE can be associated with at least one SBS in our proposed user association algorithm. Moreover, SUEs can be associated with the SBSs close to them because the path losses between SBSs and SUEs are small in dense small cell networks. The proposed user association algorithm ensures that SUEs associated with the SBSs near them are within the communication range through the condition $PL_{su} < PL_0$; i.e., SBSs and SUEs should be matched with each other through the bidirectional selection method that SUEs choose their serving SBSs; meanwhile SBSs choose the associated SUEs in their coverage range. To further reduce path loss and save power consumption, we traverse and find all the possible CoMP-JT user association cases through polling all the orders that SUEs choose their associated SBSs. The detail description of the proposed user algorithm is shown in Algorithm 1.

3.2. Power Allocation Optimization. In this phase, we will optimize power allocation of SBSs to improve EE. According to the established association relationship between SBSs and SUEs of the former phase, the EE maximization problem of the network with CoMP-JT can be transformed as follows:

$$\begin{aligned} \max_{\mathbf{p}} \quad & \eta_{EE} = \frac{R_{tot}(\mathbf{p})}{P_{tot}(\mathbf{p})} = \frac{\sum_{u=1}^U R_u(\mathbf{p})}{\sum_{u=1}^U P_u(\mathbf{p})} \\ \text{s.t.} \quad & (6a) : R_u \geq R_u^{\min}, \quad \forall u \in U \\ & (6b) : \sum_{u=1}^U p_{su} \leq p_s^{\max}, \quad \forall s \in S. \end{aligned} \quad (6)$$

Compared (6) with (5), we can observe that the EE maximization problem in (6) is determined by the power allocation vector of SBSs since user association vector \mathbf{x} has been determined in the proposed user association algorithm.

As the problem in [13, 14, 19], the optimization problem (6) with an objective function in fractional form, we use the method of fractional programming that equivalently convert the objective function to subtractive form as follows:

$$\begin{aligned} \max_{\mathbf{p}} \quad & R_{tot}(\mathbf{p}) - \eta_{EE}^n P_{tot}(\mathbf{p}) \\ \text{s.t.} \quad & (6a), (6b). \end{aligned} \quad (7)$$

The objective function in (7) can be transformed as

$$R_{tot}(\mathbf{p}) - \eta_{EE}^n P_{tot}(\mathbf{p}) = g(\mathbf{p}) - r(\mathbf{p}), \quad (8)$$

```

1: Initialization Set  $M_u = \phi$  and  $S' = S$ .
2: for ( $i = 1, i \leq U, i++$ )
3:   while  $S' \neq \phi$  do
4:     for ( $j = i - 1, j \leq U + i - 2, j++$ )
5:        $u = \text{mod}(j, U) + 1$ .
6:       If  $PL_{su} < PL_0$  then
7:          $M_u = M_u \cup \arg \min_{s \in S'} \{PL_{su}\}$ .
8:         Update  $S = S' \setminus M_u$ .
9:       end if
10:    end for
11:  end while
12: Update  $S' = S$ .
13: end for
14: Output All the CoMP-JT user association cases  $M_u (\forall u \in U)$ .

```

ALGORITHM 1: Proposed user association algorithm.

where

$$g(\mathbf{p}) = \Delta f \sum_{u=1}^U \log_2 \left(\sum_{s=1}^S p_{su} h_{su} + \delta^2 \right) - \eta_{EE}^n \times \sum_{u=1}^U P_u(\mathbf{p}), \quad (9)$$

and

$$r(\mathbf{p}) = \Delta f \sum_{u=1}^U \log_2 \left(\sum_{l=1, l \notin M_u}^S p_{lu} h_{lu} + \delta^2 \right). \quad (10)$$

Since (6a) is nonconvex constraint, we rearrange (6a) in (7) and an equivalent convex linear form (6a') can be derived as follows:

$$(6a') : \sum_{s \in M_u} p_{su} h_{su} + \left(1 - 2^{(R_u^{\min}/\Delta f)}\right) \times \left(\sum_{l=1, l \notin M_u}^S p_{lu} h_{lu} + \delta^2 \right) \geq 0. \quad (11)$$

Hence, optimization problem (7) is equivalent to

$$\max_{\mathbf{p}} g(\mathbf{p}) - r(\mathbf{p}) \quad (12)$$

$$s.t. \quad (6a'), (6b).$$

Therefore, the objective $g(\mathbf{p}) - r(\mathbf{p})$ in (12) is the difference between two concave functions program. The gradient of $r(\mathbf{p})$ at power \mathbf{p} is given by

$$\nabla r(\mathbf{p}) = \Delta f \sum_{u=1}^U \frac{1}{\sum_{l=1, l \notin M_u}^S p_{lu} h_{lu} + \delta^2} \mathbf{v}_u \quad (13)$$

where $\mathbf{v}_u(z) = 0$ and $z \in M_u$; $\mathbf{v}_u(l) = h_{lu}/\ln 2$, $l \notin M_u$. Then, we adopt successive convex approximation method to approximate $r(\mathbf{p})$ by its tangent function $r(\mathbf{p}) \approx r(\mathbf{p}^n) +$

$\nabla r^T(\mathbf{p}^n)(\mathbf{p} - \mathbf{p}^n)$ at each step; we can obtain the optimization problem as follows:

$$\max_{\mathbf{p}} g(\mathbf{p}) - \left(r(\mathbf{p}^n) + \nabla r^T(\mathbf{p}^n)(\mathbf{p} - \mathbf{p}^n) \right) \quad (14)$$

$$s.t. \quad (6a'), (6b).$$

Now, (14) is a standard convex optimization problem, and interior point method in [20] can be applied to solve it.

We transverse all the possible CoMP-JT user association cases obtained in Algorithm 1 and further get optimal EE solution for each user association case through power allocation optimization Algorithm 2. Finally, we can find the maximization value of EE in all the possible user association cases. The detail description of power allocation optimization process is shown in Algorithm 2.

3.3. Complexity Analysis. In this section, we analyze the computational complexities of the optimal and the proposed solutions. Finding optimal solution needs to exhaust all the association cases. Thus the computational complexity of optimal association is $O(S! / ((S - \sum_{u=1}^U w_u) \times U! \times w_u!))$, where w_u denotes the number of associated SBSs of SUE u . However, Algorithm 1 firstly chooses the SBSs in S' with the minimum path loss for SUEs, which requires a computational complexity $O(C_S^U)$. Then, Algorithm 1 needs to repeat the above procedure until $S' = \phi$, which requires a computational complexity not larger than $O(C_{S-(k-1)U}^U)$ in the k round. Moreover, Algorithm 1 needs to transverse all the user association cases in outer circulation with the computational complexity U . Therefore, Algorithm 1 requires a computational complexity no larger than $O(U * (C_S^U + \dots + C_{S-(k-1)U}^U))$, which is much lower than that of the optimal scheme. Suppose that the maximum outer power iterative of interior point scheme is N_{\max} and the computational complexity of applying the successive convex approximation scheme depends on the outer iteration and iterative solution of each step; thus the overall computational complexity of interior point scheme is $O(N_{\max} S U)$. So the overall computational complexity of

1: **Initialization:** Set $i = 0$, initial power vector $\mathbf{p}^0 \in (0, \mathbf{p}_s^{\max})$, minimum value of EE η_{EE}^{low} , maximum value of EE η_{EE}^{high} , and $\eta_{EE}^i = ((\eta_{EE}^{\text{low}} + \eta_{EE}^{\text{high}})/\theta)(\theta \in (N * \setminus 1))$, and iterative terminal condition of EE $\alpha > 0$.

2: **Repeat**

3: Solve (14) and obtain optimal value \mathbf{p}^* by interior point method.

4: Update $\mathbf{p}^{(i+1)} = \mathbf{p}^*$, $i = i + 1$.

5: **until** convergence.

6: **Obtain** $\mathbf{p}^{(i)}$ and the maximum value of EE η_{EE}^* .

7: **Until** $|R_{\text{tot}}(\mathbf{p}^i) - \eta_{EE}^i P_{\text{tot}}(\mathbf{p}^i)| \leq \alpha$ or $|\eta_{EE}^{\text{high}} - \eta_{EE}^{\text{low}}| \leq \alpha$.

8: **Output** $\mathbf{p}^* = \mathbf{p}^i$ and $\eta_{EE}^* = R_{\text{tot}}(\mathbf{p}^i)/P_{\text{tot}}(\mathbf{p}^i)$.

ALGORITHM 2: Power allocation optimization algorithm.

our proposed solution is $O(N_{\max} S U^2 (C_s^U + \dots + C_{s-(k-1)U}^U))$. However, we need to traverse all the power values for the optimal solution, which has a complexity of Q^U , where Q is the number of quantified power allocation values for each SUE of a SBS. So the overall computational complexity of the optimal solution is $O(Q^U S! / ((S - \sum_{u=1}^U w_u) \times U! \times w_u!))$. Consequently, our proposed solution has a notably reduced complexity compared with the optimal solution.

4. Simulation Results

In our simulation, all the SUEs and SBSs are randomly and uniformly distributed in the area of $40\text{m} \times 40\text{m}$. We consider that the channel model includes path loss, Rayleigh, and shadowing fading, where the path loss from SBS s to SUE u is given by $PL_{su} = 38.46 + 20 \times \log_{10}(d_{su})$, where d_{su} denotes the distance between SBS s and SUE u , with the unit m . The shadowing standard deviations are 8 dB for the link between SBS and the SUE [21]. Besides, $PL_0 = 58.46$ dB, which is calculated by assuming the coverage radius of each SBS is 10 m .

Figure 2 plots EE with respect to the iteration number under different maximum transmit powers of SBSs, and Figure 3 plots $g(\mathbf{p}) - r(\mathbf{p})$ with respect to the iteration number under different initial iterative transmit power values. We can observe that the proposed dynamic CoMP-JT Algorithm can quickly converge to the stable EE maximization solution, which confirms the fact that the proposed solution has a considerable convergence property.

Next, we compare the proposed solution with the following solutions:

(i) Near-optimal solution: Algorithm 1 is adopted as the user association scheme. Moreover, since the power allocation vector is continuous, therefore, we firstly equally quantify each power allocation variable into 20 values in $[0, p_s^{\max}]$ and then we need to exhaustively traverse the quantified 20^U values into the objective function to obtain near-optimal solution.

(ii) SecondNear-CoMP [16]: We assume that system bandwidth is shared by all the SUEs, and SUEs are associated with the nearest SBS and the SBS (except the serving SBS) nearest to the SUE from the set of SBSs which has not been chosen by other SUEs as their coordinated multipoint joint

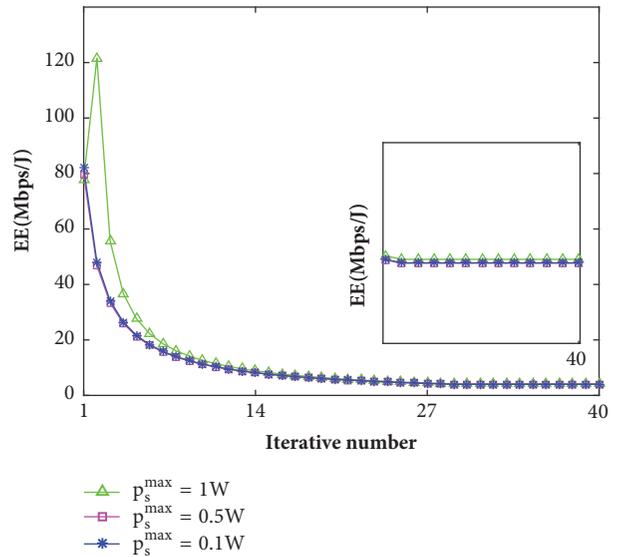


FIGURE 2: EE vs. iterative number.

transmission (CoMP-JT) SBSs. Moreover, Algorithm 2 is adopted as the power allocation optimization algorithm. We call this method SecondNear-CoMP solution.

(iii) No-CoMP [19]: Each SUE is associated with the nearest SBS that has not been associated with other SUEs. Moreover, Algorithm 2 is adopted as the power allocation optimization algorithm.

(iv) Random-CoMP: Each SUE is randomly associated with the SBSs close to it as Algorithm 1. Besides, Algorithm 2 is adopted as the power allocation optimization algorithm.

Figure 4 shows the EE with respect to different data rate requirement of SUEs, where $p_s^{\max} = 1\text{w}$ and the number of SUEs and SBSs is 16 and 22, respectively. Since obtaining the optimal solution of problem (5) by exhaustive search will incur prohibitive computational complexity in dense small cell network, thus, an efficient near-optimal solution of power allocation is used in (5), where 20 uniformly quantified power allocation values for each SUE of a SBS are traversed. We can observe the performance gap between the proposed and the optimal solution becomes a little larger when R_u^{\min}

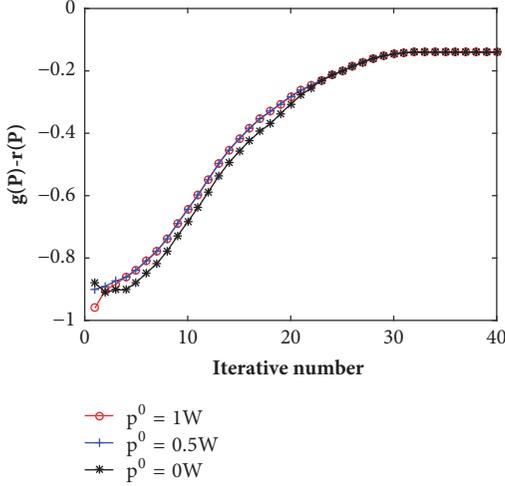
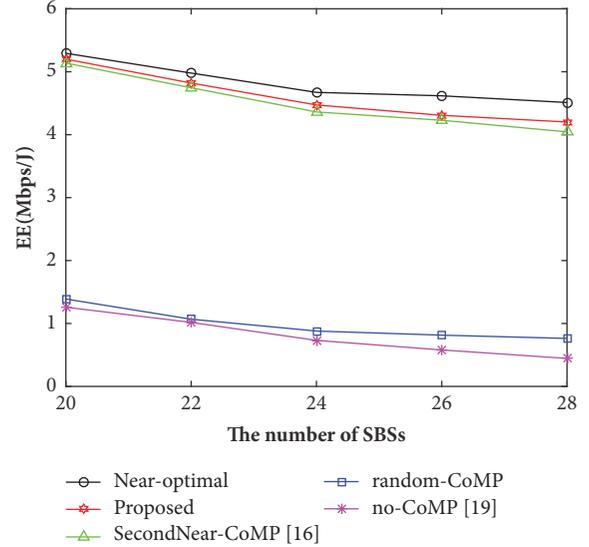
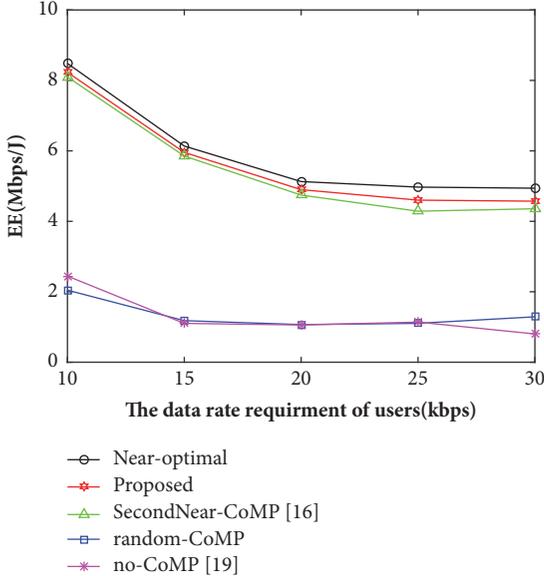
FIGURE 3: $g(\mathbf{p}) - r(\mathbf{p})$ vs. iterative number.

FIGURE 5: EE vs. The number of SBSs.

FIGURE 4: EE vs. R_u^{\min} .

increases. Nevertheless, the proposed solution has a significantly reduced computational complexity. Besides, we can intuitively observe that the proposed CoMP-JT scheme has a better performance than the other three schemes and random CoMP-JT scheme has a better performance than transmission without CoMP when R_u^{\min} becomes larger. The reason is that the proposed CoMP-JT ensures that SUEs are associated near SBSs with lower path losses. Moreover, CoMP-JT schemes can satisfy higher R_u^{\min} by exploiting interference through constructing CoMP-JT set.

Figure 5 depicts the EE with respect to different number of SBSs, where $p_s^{\max}=1\text{w}$, $R_u^{\min}=20\text{kbps}$, and the number of SUEs is 16. We can observe the performance gap between the proposed and the optimal solution becomes a little larger with the increasing number of SBSs. The reason is

that computational complexity increases with the number of deployed SBSs. Besides, we can observe that the performance of our proposed CoMP-JT scheme is better than the other three schemes. The reason is that our proposed scheme guarantees that SUEs are associated with SBSs closer to them, which will reduce the loss of power. Moreover, the EE decreases with increasing R_u^{\min} and the number of SUEs in Figures 4 and 5, respectively. The reason is that interference increases in dense small cell networks.

5. Conclusions

In this paper, we have firstly established EE maximization problem of joint user association and power allocation considering minimum data rate requirement of SUEs and maximum transmit power constraint of SBSs. Then, a dynamic CoMP-JT algorithm has been proposed to solve the formulated problem. In the first phase, SUEs have been associated with the SBSs close to them within the communication range to reduce the loss of power. In the second phase, through the methods of fractional programming and successive convex approximation, the EE maximization problem has been transformed into a convex problem that was solved by the given power allocation optimization algorithm. Moreover, we have analyzed the fact that the proposed dynamic CoMP-JT solution has a much lower computational complexity than the optimal solution. In the simulation, the performance of the proposed solution has been compared with near-optimal, SecondNear-CoMP, no-CoMP, and random-CoMP solutions, which demonstrate that the proposed algorithm has a better performance.

Data Availability

The data is based on our established network model.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This paper is supported in part by the National Natural Science Foundation of China [nos. 91638202, 61231008, and 61571351].

References

- [1] J. G. Andrews, S. Buzzi, and W. Choi, "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [2] A. Gupta and R. K. Jha, "A survey of 5G network: architecture and emerging technologies," *IEEE Access*, vol. 3, pp. 1206–1232, 2015.
- [3] C.-X. Wang, F. Haider, X. Gao et al., "Cellular architecture and key technologies for 5G wireless communication networks," *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, 2014.
- [4] N. Lee and R. W. Heath, "Advanced interference management technique: Potentials and limitations," *IEEE Wireless Communications Magazine*, vol. 23, no. 3, pp. 30–38, 2016.
- [5] A. H. Jafari, D. Lopez-Perez, M. Ding, and J. Zhang, "Study on Scheduling Techniques for Ultra Dense Small Cell Networks," in *Proceedings of the 2015 IEEE 82nd Vehicular Technology Conference (VTC Fall)*, pp. 1–6, Boston, MA, USA, September 2015.
- [6] X. Wang, Y. Wang, and H. Zhu, "Energy-efficient multi-job scheduling model for cloud computing and its genetic algorithm," *Mathematical Problems in Engineering*, vol. 2012, Article ID 589243, 16 pages, 2012.
- [7] E. Pateromichelakis, M. Shariat, A. Quddus, M. Dianati, and R. Tafazolli, "Dynamic clustering framework for multi-cell scheduling in dense small cell networks," *IEEE Communications Letters*, vol. 17, no. 9, pp. 1802–1805, 2013.
- [8] Q. Cui, H. Wang, P. Hu et al., "Evolution of limited-feedback CoMP systems from 4G to 5G: CoMP features and limited-feedback approaches," *IEEE Vehicular Technology Magazine*, vol. 9, no. 3, pp. 94–103, 2014.
- [9] M. Feng, X. She, L. Chen, and Y. Kishiyama, "Enhanced dynamic cell selection with muting scheme for DL CoMP in LTE-A," in *Proceedings of the 2010 IEEE 71st Vehicular Technology Conference, VTC 2010-Spring*, Taiwan, May 2010.
- [10] X. Li, Q. Cui, Y. Liu, and X. Tao, "An effective scheduling scheme for CoMP in heterogeneous scenario," in *Proceedings of the 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2012*, pp. 870–874, Australia, September 2012.
- [11] G. Nigam, P. Minero, and M. Haenggi, "Coordinated multipoint joint transmission in heterogeneous networks," *IEEE Transactions on Communications*, vol. 62, no. 11, pp. 4134–4146, 2014.
- [12] J. Zhao, T. Q. S. Quek, and Z. Lei, "Coordinated multipoint transmission with limited backhaul data transfer," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2762–2775, 2013.
- [13] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in multi-cell OFDMA systems with limited backhaul capacity," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3618–3631, 2012.
- [14] K. M. S. Huq, S. Mumtaz, J. Bachmatiuk, J. Rodriguez, X. Wang, and R. L. Aguiar, "Green HetNet CoMP: Energy Efficiency Analysis and Optimization," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4670–4683, 2015.
- [15] B. Du, C. Pan, W. Zhang, and M. Chen, "Distributed energy-efficient power optimization for CoMP systems with max-min fairness," *IEEE Communications Letters*, vol. 18, no. 6, pp. 999–1002, 2014.
- [16] S. Kim and C. Cho, "Power efficient CoMP-joint transmission with two stage QoSs," *Transactions on Emerging Telecommunications Technologies*, vol. 28, no. 4, p. e3014, 2017.
- [17] R. Sun, M. Hong, and Z. Luo, "Joint downlink base station association and power control for max-min fairness: computation and complexity," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1040–1054, 2015.
- [18] G. Ye, H. Zhang, H. Liu et al., "Energy efficient joint user association and power allocation in a two-tier heterogeneous network," in *IEEE Global Communications Conference*, pp. 1–5, 2016.
- [19] Y. Li, M. Sheng, C. Yang, and X. Wang, "Energy efficiency and spectral efficiency tradeoff in interference-limited wireless networks," *IEEE Communications Letters*, vol. 17, no. 10, pp. 1924–1927, 2013.
- [20] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, UK, 2004.
- [21] E. U. T. R. Access, "Further advancements for E-UTRA physical layer aspects," Tech. Rep. 3GPP TR 36, 2010.

Research Article

Resource Allocation for Green Cognitive Radios: Energy Efficiency Maximization

Zhou Yang, Wenqian Jiang , and Gang Li

Guangxi Power Grid Co., Power Grid Electric Power Research Institute, Nanning, China

Correspondence should be addressed to Wenqian Jiang; jwqqsky@126.com

Received 20 March 2018; Revised 26 April 2018; Accepted 10 May 2018; Published 5 July 2018

Academic Editor: Zheng Chu

Copyright © 2018 Zhou Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Green cognitive radios are promising in future wireless communications due to high energy efficiency. Energy efficiency maximization problems are formulated in delay-insensitive green cognitive radio and delay-sensitive green cognitive radio. The optimal resource allocation strategies for delay-insensitive green cognitive radio and delay-sensitive green cognitive radio are designed to maximize the energy efficiency of the secondary user. The peak interference power and the average/peak transmit power constraints are considered. Two algorithms based on the proposed resource allocation strategies are proposed to solve the formulated problems. Simulation results show that the maximum energy efficiency of the secondary user achieved under the average transmit power constraint is higher than that achieved under the peak transmit power constraint. It is shown that the design of green cognitive radio should take the tradeoff between its complexity and its achievable maximum energy efficiency into consideration.

1. Introduction

The unprecedented increase of mobile devices and escalating high data rate requirements have resulted in the rapid growth of energy consumption and greenhouse gas emission. It is reported in [1–3] that 2% to 10% of the global energy consumption and 2% of the greenhouse gas are generated by information and communication technologies. However, according to the Federal Communications Commission (FCC), there are 70% unused spectrum bands in the allocated spectrum bands in the USA [4–6]. Green cognitive radio (CR) is a promising and increasingly attractive technology that can improve spectrum efficiency (SE) and maximize energy efficiency (EE) simultaneously [7, 8]. In green CR, a secondary user (SU) is allowed to access a primary user (PU) band providing that the interference caused to the PU is acceptable, and the SU achieves high performance with the objective of EE maximization. In this paper, a green CR under spectrum sharing is our focus.

In CR, resource allocation is of great importance and has received wide attention [9–13]. An optimal resource allocation strategy for CR not only provides the SU a reasonably high transmission rate with limited power, but also

well protects the PU from harmful interference. Moreover, in green CR, an optimal resource allocation strategy can guarantee that the SU achieves maximum EE and reduces greenhouse emission. Although the designs of the resource allocation strategies have been well studied in non-CR and CR networks, the optimal resource allocation strategies designing for non-CR and CR networks are not optimal for green CR in terms of EE maximization. Thus, it is important to design resource allocation strategies for maximizing the EE of the SU in green CR.

1.1. Related Works and Motivation. Since the operation of a CR should protect the quality of service (QoS) of the PU, a metric that evaluates the performance of the protection of the PU should be imposed. Basically, there are three metrics for protecting the PU from intolerance interference caused by the SU, namely, a peak interference power constraint (PIP), an average interference power (AIP) constraint, and an outage probability (OP) constraint [9–13]. A PIP constraint requires that the instantaneous interference power caused by the SU is below a prescribed interference power threshold, whereas the AIP constraint regulates the long-term interference power over all fading states to be not more than the maximum

tolerably AIP of the PU [9, 10]. The OP constraint requires that the transmission OP of the PU due to its fading and the interference from the SU is not greater than its acceptable target [11–13]. When the AIP constraint or the OP constraint is applied, the designs of the optimal resource allocation strategies for CR are required to compute a parameter with respect to the AIP constraint or the OP constraint, such as a nonnegative dual variable [9–13]. However, the computation of the parameter is complex, especially when there are several parameters (such as several nonnegative dual variables) required for computing or when the probability distribution of the fading channel gain is complex [9, 10]. This high complexity computation consumes too much energy, which contradicts with the perspective of green CR. In contrast, it is not required to compute the parameter when the PIP constraint is used. In this paper, the PIP constraint is identified as the protection metric of the PU, which is more appropriate to green CR since it can facilitate the implementation of a green CR with low complexity.

The optimal resource allocation strategies have been well studied in CR with the spectrum sharing paradigm [9–18]. The designs of the optimal resource allocation strategies for CR networks combining different transmit power constraints and different interference power constraints were proposed in [9, 10], and those designs were extended to CR with imperfect CSI in [12]. In [11, 13], optimal resource allocation strategies were designed for CR with perfect CSI or CR with imperfect CSI, under an OP constraint. The ergodic capacity (EC), OC, and minimum-rate capacity based on the proposed optimal power control strategies were analyzed in [14]. In [15], the OP of a SU in spectrum sharing CR was minimized based on an optimal transmit power allocation scheme. Recently, adaptive gradient-based methods for power allocation in OFDM-based spectrum sharing CR have been proposed in [16]. Under both peak/average transmit power (PTP/ATP) and PIP/AIP constraints, optimal power control strategies that maximize the rates of CR with arbitrary input distributions were studied in [17]. In [18], the resource allocation problem was studied based on the proposed worst-case selective robust model.

There are some investigations of the design of the optimal resource allocation strategies for green CR [19–25]. The optimal resource allocation strategies for OFDM-based cognitive radio networks were proposed to maximize the EE of the SU in [19, 20]. An EE joint relay selection and power allocation scheme was proposed in [21]. The optimal relay selection and power allocation policy is determined using a distributed approach. In [22], the optimal power control that maximizes the EE of green CR with the opportunistic spectrum access paradigm was determined. However, the optimal power allocation strategies proposed in [19–22] are not adapted to green CR with fading channels since the EE of green CR with fading channels should be maximized based on all fading CSI instead of instantaneous CSI. Recently, although EE maximization problems have been analyzed in fading CR channels and the optimal power allocation strategies were proposed in [23–25], the EE maximization problem has not fully taken the impact of the fading CSI between the PU's transmitter and the SU's receiver into

consideration in [23]. In [24], the authors have studied secure EE maximization in green CR. Refs. [23–25] only considered EE maximization problem in delay-insensitive CR. Recently, in [26], the authors have studied the EE maximization problem in a device-to-device network. However, it did not consider the delay requirement. To our best knowledge, the EE maximization problem has not been analyzed in delay-sensitive CR.

Recently, the EE maximization problems were studied in delay-insensitive CR, delay-sensitive CR, and simultaneously delay-sensitive and delay-insensitive CR in [27]. In [27], the AIP constraint was regarded as the metric of the protection of the PU. Similar to the optimal power allocation strategies for conventional CR proposed in [9–18], which use the AIP constraint as the metric of the protection of the PU, the proposed optimal power allocation strategies for green CR also need to compute a complex nonnegative dual variable related to the AIP constraint. The computation of the nonnegative dual variable is extensively complex when there are several nonnegative dual variables required to simultaneously compute or when the probability distributions of the fading channels involved are complex. This complex computation limits its application when green CR requires fast and efficient determination of the optimal power allocation strategies. In this paper, different from the works in [25, 27], a PIP constraint is identified as the metric protecting the PU from harmful interference caused by the SU. EE maximization problems subject to the PIP constraint are analyzed in delay-insensitive green CR and delay-sensitive green CR. Optimal power allocation strategies that maximize the EE of the SU under the PIP constraint are found. The proposed optimal power allocation strategies have advantages in low complexity and fast computation.

1.2. Contributions and Organization. Different from the works in [25, 27], EE maximization problems subject to the PIP constraint are proposed in delay-insensitive green CR and delay-sensitive green CR, where the SU coexists with the PU and all the channels involved are fading. The EE maximization problem is first analyzed in delay-insensitive green CR. Then, the EE maximization problem is extended to delay-sensitive green CR. The main contributions of this work are summarized as follows.

- (1) The optimal resource allocation strategies for delay-insensitive green CR and delay-sensitive green CR that maximize the EE of the SU are found. Different from the work in [25, 27], the PIP constraint, along with the ATP/PTP constraint, is considered. It is shown that the optimal power allocation strategies for delay-insensitive green CR have similar form to the well-known “water-filling” power allocation strategies for delay-sensitive green CR have similar form to the well-known truncated channel inversion power allocation strategies.
- (2) Two algorithms based on the proposed optimal resource allocation strategies are presented. One is proposed to solve the EE maximization problem

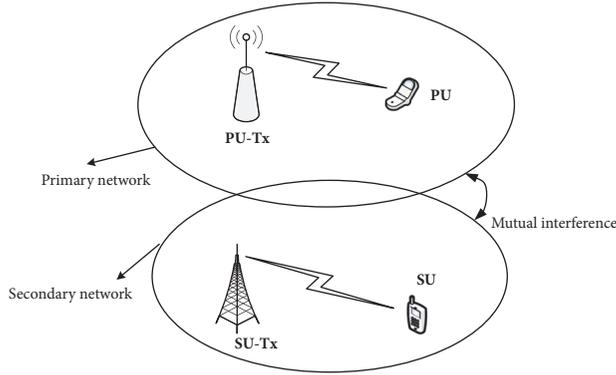


FIGURE 1: The system model.

when the ATP constraint and the PIP constraint are applied. The other one is given for solving the EE maximization when the PTP constraint and the PIP constraint are imposed. It is shown that the complexity of the proposed algorithm for the EE maximization problem subject to the ATP constraint is higher than that of the proposed algorithm for the EE maximization problem subject to the PTP constraint.

- (3) Simulation results show that the maximum EE of the SU achieved under the ATP constraint is larger than that achieved under the PTP constraint. The design of a green CR system should take the tradeoff between the achievable maximum EE and the implementation complexity into consideration.

The rest of this paper is organized as follows. Section 2 presents the system model. The EE maximization problem subject to constraints on the PIP and the ATP is presented in Section 3. Section 4 analyzes the EE maximization problem subject to the PIP and the PTP constraint. Section 5 presents simulation results. The paper concludes with Section 6.

2. System Model

As shown in Figure 1, a cognitive radio consisting of one PU and one SU is considered. The SU coexists with the PU under the spectrum sharing paradigm. There is one PU transmitter (PU-Tx) and one PU in the primary network while the secondary network has one SU transmitter (SU-Tx) and one SU. It is assumed that all the terminals have one antenna. The channel between the PU-Tx and the PU, the channel between the PU-Tx and the SU, the channel between the SU-Tx and the SU, and the channel between the SU-Tx and the PU are assumed to be block fading. The

corresponding channel power gains at fading state ν are denoted by $g_{ss}(\nu)$, $g_{sp}(\nu)$, $h_{pp}(\nu)$, and $h_{ps}(\nu)$, respectively. The fading index for all related channels is denoted by ν . All the channel power gains are assumed to be independent identically distributed (i.i.d.), ergodic, and stationary and have continuous probability density functions. It is assumed that perfect channel state information (CSI) related to $g_{ss}(\nu)$, $g_{sp}(\nu)$, and $h_{pp}(\nu)$ is available to the SU-Tx. The additive white Gaussian noise (AWGN) at the SU-Rx is circularly symmetric complex AWGN with mean zero and variance σ_w^2 . It is assumed that the PU does not have knowledge of the CSI of the PU link, and thus it transmits with a constant power, denoted by P_p .

3. EE Maximization under Average Transmit Power Constraint

In this section, EE maximization problems are formulated in delay-insensitive green CR and delay-sensitive green CR. Different from the works in [25, 27], the PIP constraint is considered. The optimal resource allocation strategies that maximize the EE of the SU, subject to the PIP constraint and the ATP constraint, are determined. An algorithm based on the proposed optimal power allocation strategy is presented to solve EE maximization problems.

3.1. EE Maximization in Delay-Insensitive Green CR. In CR, from the perspective of the PU, the interference caused by the SU should not be beyond the tolerable interference threshold of the PU. The PIP constraint is chosen as the protection metric of the PU since CR with PIP constraint has low complexity and facilitated implementation. In addition, the long-term power budget of the SU should be considered and can be evaluated by the ATP. From the perspective of the SU, the ATP should be below a threshold. Thus, the constraints on the PIP and ATP can be given as

$$g_{sp}(\nu) P_s(\nu) \leq P_{In} \quad (1)$$

$$\mathbb{E}\{P_s(\nu)\} \leq \overline{P_{th}} \quad (2a)$$

$$P_s(\nu) \geq 0 \quad (2b)$$

where $P_s(\nu)$ denotes the transmit power of the SU. P_{In} and $\overline{P_{th}}$ are the tolerable maximum PIP of the PU and the maximum ATP of the SU, respectively. $\mathbb{E}(\cdot)$ is the expectation operator.

In delay-insensitive CR, the EC is appropriately used to evaluate the performance of the SU [10]. The EC determines the maximum achievable long-term rates of the SU, which is averaged over all fading states. The EE maximization problem for delay-insensitive green CR under the PIP constraint and the ATP constraint can be formulated as problem \mathbf{P}_1 , given as

$$\mathbf{P}_1: \max_{P_s(\nu)} \eta_{EE}(P_s(\nu)) = \frac{\mathbb{E}\{\log_2(1 + g_{ss}(\nu)P_s(\nu) / (h_{ps}(\nu)P_p + \sigma_w^2))\}}{\mathbb{E}\{\zeta P_s(\nu) + P_C\}} \quad (3a)$$

$$\text{s.t.} \quad (1) \text{ and } (2a), (2b) \text{ are satisfied} \quad (3b)$$

where $\eta_{EE}(P_s(\nu))$ is the energy efficient function of $P_s(\nu)$. ζ and P_C represent the amplifier coefficient ($1/\zeta$ is also known as the power efficiency for the power amplifier) and the constant circuit power consumption of SU-Tx, respectively. Let S_1 denote the set $S_1 = \{P_s(\nu) \mid P_s(\nu) \in (1), P_s(\nu) \in (2a), (2b)\}$. Note that S_1 is a convex set. According to the fractional programming theory given in [28, 29], Theorem 1 can be stated as follows.

Theorem 1. *Problem \mathbf{P}_1 is a nonlinear strictly quasiconcave and strictly pseudoconcave fractional programming problem since the numerator of $\eta_{EE}(P_s(\nu))$ is strictly concave and differentiable, the denominator of $\eta_{EE}(P_s(\nu))$ is affine and differentiable, and the denominator $\mathbb{E}\{\zeta P_s(\nu) + P_C\} > 0$. Any local maximum of problem \mathbf{P}_1 is a global maximum of problem \mathbf{P}_1 , and \mathbf{P}_1 has at most one maximum since \mathbf{P}_1 is strictly quasiconcave.*

Proof. See Appendix A. \square

Based on Theorem 1 and Dinkelbach's method [30], problem \mathbf{P}_1 is equivalent to the parameter optimization problem, denoted by \mathbf{P}_2 , given as

$$\begin{aligned} \mathbf{P}_2: \max_{P_s(\nu) \in S_1} f(\eta) \\ = \mathbb{E} \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) \right\} \\ - \eta \mathbb{E} \{ \zeta P_s(\nu) + P_C \} \end{aligned} \quad (4)$$

where η is a nonnegative parameter. The following theorem can be obtained to solve problem \mathbf{P}_1 and \mathbf{P}_2 .

Theorem 2. *The global optimization solution of problem \mathbf{P}_1 is achieved if and only if*

$$\begin{aligned} \max_{P_s(\nu) \in S_1} f(\eta_{opt}) \\ = \mathbb{E} \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s^{opt}(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) \right\} \\ - \eta_{opt} \mathbb{E} \{ \zeta P_s^{opt}(\nu) + P_C \} = 0. \end{aligned} \quad (5)$$

η_{opt} is the optimal solution of \mathbf{P}_2 and also is the maximum EE of \mathbf{P}_1 . $P_s^{opt}(\nu)$ is the optimal transmit power of problem \mathbf{P}_1 and is the corresponding solution of η_{opt} .

Proof. See Appendix B. \square

Thus, on the one hand, problem \mathbf{P}_1 can be solved by solving problem \mathbf{P}_2 . On the other hand, for a given η_{opt} , the optimal transmit power can be obtained by solving (5).

For a given η , problem \mathbf{P}_2 can be solved by using the Lagrange duality method [31]. The Lagrangian with respect to the transmit power $P_s(\nu)$ of \mathbf{P}_2 is given as

$$\begin{aligned} L(P_s(\nu), \tau) = \mathbb{E} \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) \right\} \\ - \eta \mathbb{E} \{ \zeta P_s(\nu) + P_C \} \\ - \tau \left\{ \mathbb{E} \{ P_s(\nu) \} - \overline{P_{th}} \right\} \end{aligned} \quad (6)$$

where τ is the nonnegative dual variable with respect to (2a). Let S_2 denote the set $S_2 = \{P_s(\nu) \mid 0 \leq P_s(\nu) \leq P_{In}/g_{sp}(\nu)\}$. Then, the Lagrange dual function of \mathbf{P}_2 can be presented as

$$g(\tau) = \max_{P_s(\nu) \in S_2, \forall \nu} L(P_s(\nu), \tau). \quad (7)$$

Similar to [9, 11], the problem given by (7) can be decoupled into parallel subproblems by using the Lagrange dual-decomposition method [31]. Thus, the corresponding subproblem for a fading state can be given as

$$\max_{P_s \in S_2} y(P_s) = \log_2 \left(1 + \frac{g_{ss} P_s}{h_{ps} P_p + \sigma_w^2} \right) - \eta \zeta P_s - \tau P_s. \quad (8)$$

In order to solve \mathbf{P}_2 , (8) is required to iteratively solve for all fading states with respect to fixed τ and updating τ . τ can be updated by using the subgradient method [31], given as

$$\tau^{k+1} = [\tau^k - t_1 (\overline{P_{th}} - \mathbb{E} \{ P_{s,k}^{opt} \})]^+ \quad (9)$$

where $[a]^+ = \max(a, 0)$ and k is the iteration index. $P_{s,k}^{opt}$ denotes the optimal power level of the k th iteration of the subgradient method, and the parameter t_1 denotes the iteration step size. Since $y(P_s)$ is a concave function related to P_s , the optimal power allocation strategy of \mathbf{P}_2 , denoted by P_s^{opt} , can be found as given in Theorem 3. Let $\min(a, b)$ denote the minimum between a and b .

Theorem 3. *The optimal resource allocation strategy of \mathbf{P}_2 is given by*

$$\widehat{P}_s = \left[\frac{1}{(\eta \zeta + \tau) \ln 2} - \frac{(h_{ps} P_p + \sigma_w^2)}{g_{ss}} \right]^+ \quad (10a)$$

$$P_s^{opt} = \min \left(\widehat{P}_s, \frac{P_{In}}{g_{sp}} \right) \quad (10b)$$

Remark 4. In delay-insensitive green CR, the optimal power allocation strategy for EE maximization should take the achievable EE of the SU and the power amplifier coefficient of the SU-Tx into consideration, which is different from the optimal power allocation strategy for EC maximization in the conventional delay-insensitive CR proposed in [10, 12, 14]. When $\eta = 0$, problem \mathbf{P}_2 degenerates into the conventional EC maximization problem. This can be seen from (4). In this case, the optimal power allocation strategy for EE maximization subject to the PIP constraint and the ATP constraint, given by (10a) and (10b), has a similar form to that given in [10, eq. (13)]. The difference lies in the fact that the influence of the fading of the channel between the PU-Tx and the SU is fully considered, whereas that influence is only partially considered in [10].

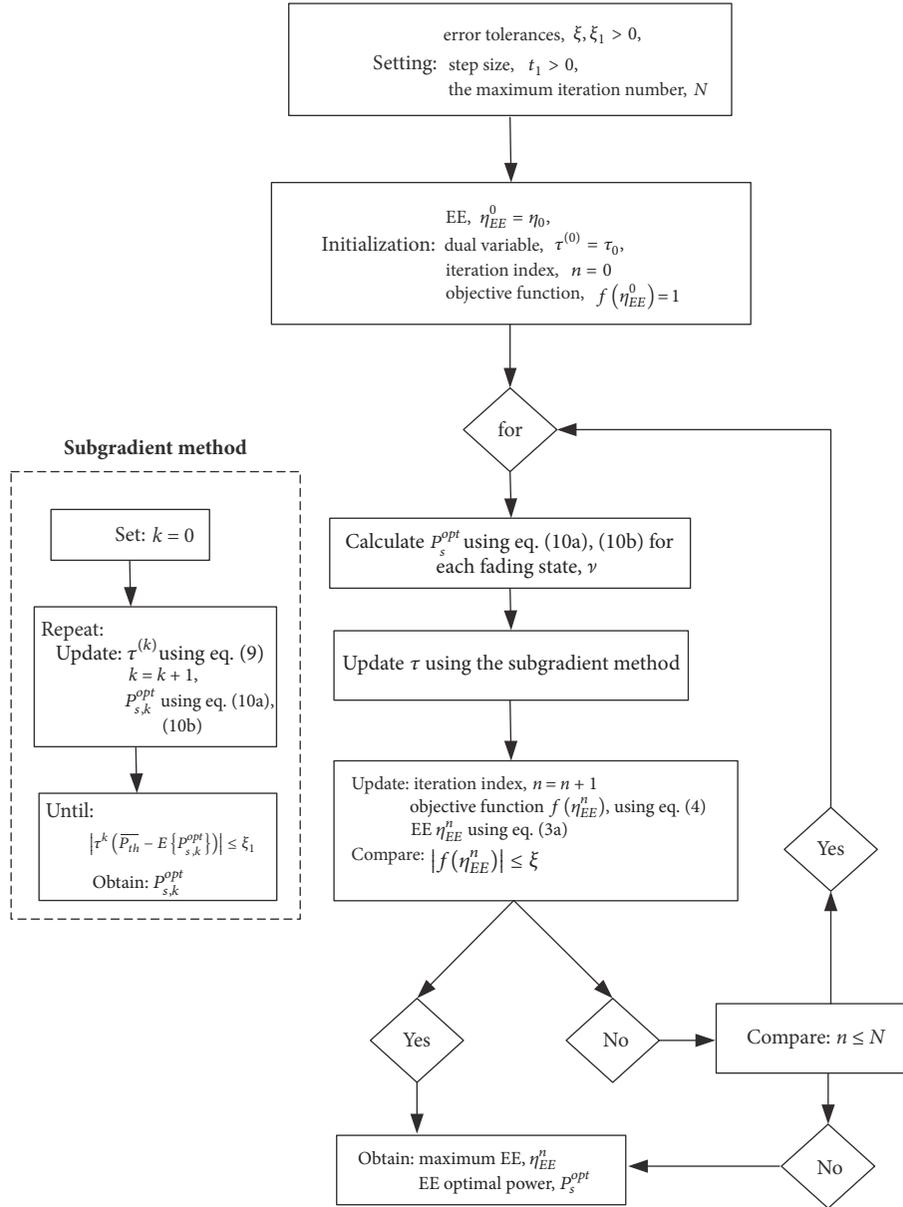


FIGURE 2: Flowchart of Algorithm 1 for EE maximization subject to the ATP constraint and the PIP constraint.

For a given η , \mathbf{P}_2 can be efficiently solved by using the proposed optimal power allocation strategy given by (10a) and (10b). In order to solve problem \mathbf{P}_1 and obtain the maximum EE of the SU, Dinkelbach's method is applied. An algorithm based on the subgradient method and Dinkelbach's method is proposed to solve problem \mathbf{P}_1 , denoted by Algorithm 1. The principle of this algorithm is as follows. The optimal transmit power is required to be iteratively updated by using Theorem 3. The subgradient method is used to compute the nonnegative dual variable until the optimal power level can satisfy the ATP constraint, namely, $|\tau^k (\overline{P_{th}} - E\{P_{s,k}^{opt}\})| \leq \xi_1$. k is the iterative index of the subgradient method and ξ_1 is the tolerable error for the ATP constraint. When the proposed optimal power allocation strategy can support $f(\eta_{EE}^n) = 0$,

the maximum EE and corresponding optimal power level of the SU are obtained. Otherwise, an ξ -optimal solution with an error tolerance ξ is adopted. In this case, the maximum EE and corresponding optimal power level are obtained when $|f(\eta_{EE}^n)| \leq \xi$. Note that if equations $f(\eta_{EE}^n) = 0$ and $|f(\eta_{EE}^n)| \leq \xi$ can not be satisfied, Algorithm 1 stops when the iteration number becomes N . The flowchart of Algorithm 1 for EE maximization subject to the PIP constraint and the ATP constraint is shown in Figure 2.

3.2. EE Maximization in Delay-Sensitive Green CR. In this subsection, the EE maximization problem is analyzed in delay-sensitive green CR, subject to constraints on the PIP and the ATP. In delay-sensitive green CR, the SU is sensitive

to the delay, such as voice and video applications. In delay-sensitive CR, the OC that evaluates the achievable constant rate for all fading states is a more appropriate metric. Thus, the EE definition in this green CR should be related to the OC of the SU.

According to the work in [9–13], the traditional OP minimization problem subject to the PIP constraint and the ATP constraint can be formulated as the problem, \mathbf{P}_3 , given as

$$\mathbf{P}_3: \min_{P_s(\nu) \geq 0, \forall \nu} \Pr \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) < r_s \right\} \quad (11a)$$

$$\text{s.t.} \quad (1) \text{ and } (2a), (2b) \text{ are satisfied} \quad (11b)$$

where r_s is the prescribed OC of the SU. It is straightforward to see that the problem \mathbf{P}_3 is equivalent to the following problem, \mathbf{P}_4 , given as

$$\mathbf{P}_4: \max_{P_s(\nu) \geq 0, \forall \nu} 1 - \Pr \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) < r_s \right\} \quad (12a)$$

$$\text{s.t.} \quad (1) \text{ and } (2a), (2b) \text{ are satisfied.} \quad (12b)$$

According to [27], the EE in delay-sensitive green CR is given as

$$\eta_{EE}(P_s(\nu)) = \frac{r_s \{1 - \Pr \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) < r_s \right\}}{\mathbb{E} \{ \zeta P_s(\nu) + P_C \}} \quad (13)$$

where $\eta_{EE}(P_s(\nu))$ denotes the EE function with respect to $P_s(\nu)$ in delay-sensitive green CR. In order to formulate the EE maximization problem, the OP given by (11a) is represented as

$$\Pr \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) < r_s \right\} = E \{ \chi_s(\nu) \} \quad (14a)$$

$$\chi_s(\nu) = \begin{cases} 1, & \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) < r_s \\ 0, & \text{otherwise} \end{cases} \quad (14b)$$

where $\chi_s(\nu)$ is an indicator function for the outage event of the SU at fading state ν . Thus, the EE maximization problem subject to constraints on the PIP and the ATP can be formulated as problem, \mathbf{P}_5 , given as

$$\mathbf{P}_5: \max_{P_s(\nu)} \eta_{EE}(P_s(\nu)) = \frac{r_s \mathbb{E} \{ 1 - \chi_s(\nu) \}}{\mathbb{E} \{ \zeta P_s(\nu) + P_C \}} \quad (15a)$$

$$\text{s.t.} \quad (1) \text{ and } (2a), (2b) \text{ are satisfied.} \quad (15b)$$

Since $\chi_s(\nu)$ is not a concave function, problem \mathbf{P}_5 is a general nonlinear fractional programming problem instead of a nonlinear concave fractional programming problem.

Although problem \mathbf{P}_5 is not a nonlinear concave fractional programming problem, Dinkelbach's method can nonetheless be applied to solve problem \mathbf{P}_5 [30]. The reason is that the numerator and the denominator of $\eta_{EE}(P_s(\nu))$ satisfy the condition of Dinkelbach's method; the numerator and the denominator of $\eta_{EE}(P_s(\nu))$ are continuous and satisfy $\mathbb{E} \{ \zeta P_s(\nu) + P_C \} > 0$ and $\mathbb{E} \{ 1 - \chi_s(\nu) \} \geq 0$ for all $P_s(\nu) \in S_1$. Thus, based on Dinkelbach's method, problem \mathbf{P}_5 can be equivalent to the parameter optimization problem, \mathbf{P}_6 , given as

$$\mathbf{P}_6: \max_{P_s(\nu) \in S_1} f(\eta) = r_s \mathbb{E} \{ 1 - \chi_s(\nu) \} - \eta \mathbb{E} \{ \zeta P_s(\nu) + P_C \} \quad (16)$$

where η is a nonnegative parameter. Using a similar method to that used for problem \mathbf{P}_2 , problem \mathbf{P}_6 can be solved by solving the following problem, given as

$$\max_{P_s \in S_2} y(P_s) = -r_s \chi_s(P_s) - \eta \zeta P_s - \tau P_s \quad (17)$$

where τ is the nonnegative dual variable with respect to the ATP constraint given by (2a) and $\chi_s(P_s)$ is an explicit function of P_s . Note that the fading state indicator, ν , is dropped. Similar to problem \mathbf{P}_2 , problem \mathbf{P}_6 can be solved by iteratively solving (17) for fixed τ , and then updating τ by using the subgradient method given in (9). $\chi_s(P_s)$ is in general a step function with respect to P_s . Let y denote the turning point of $\chi_s(P_s)$, given as

$$y = \frac{(2^{r_s} - 1) (h_{ps} P_p + \sigma_w^2)}{g_{ss}} \quad (18)$$

where $y \geq 0$. It is seen that y is the minimum power required for the SU to guarantee the OC of the SU, r_s . Note that $\chi_s(P_s) = 1$ for $P_s < y$ and $\chi_s(P_s) = 0$ for $P_s \geq y$. Let P_s^{opt} denote the optimal power of \mathbf{P}_6 . The following results can be obtained by solving (17).

Case 1 ($y > P_{In}/g_{sp}$). In this case, when the SU transmits with the minimum power required to maintain the OC r_s of the SU, the interference power caused for the PU is larger than the PIP constraint. The SU is always in outage if the SU transmits with the feasible power in S_1 . Thus, $P_s^{opt} = 0$.

Case 2 ($y \leq P_{In}/g_{sp}$). In this case, the maximum of $y(P_s)$ may be $-r_s$ when $P_s = 0$ or may be $-(\eta\zeta + \tau)y$ when $P_s = y$, which is the maximum depending on their relationship. When $y > r_s/(\eta\zeta + \tau)$, the minimum power required to maintain the OC r_s of the SU is very large, and the SU stops transmitting in order to save power. Thus, $P_s^{opt} = 0$. Otherwise, the SU transmits with the optimal power, $P_s^{opt} = y$.

Theorem 5 can now be formally stated based on the results obtained above.

Theorem 5. *The optimal resource allocation strategy of \mathbf{P}_6 is given by*

$$P_s^{opt} = \begin{cases} 0, & y > \frac{P_{In}}{g_{sp}} \\ 0, & \frac{r_s}{\eta\zeta + \tau} < y \leq \frac{P_{In}}{g_{sp}} \\ y, & y \leq \frac{P_{In}}{g_{sp}}, y \leq \frac{r_s}{\eta\zeta + \tau}. \end{cases} \quad (19)$$

Remark 6. It is seen from (16) that EE maximization problem in delay-sensitive CR is equivalent to the conventional OP minimization problem proposed in [10] when $\eta = 0$. Thus, it is seen that the optimal power allocation strategy given by (19) is similar to the form given in [10, eq. (28)] when $\eta = 0$. When $\eta = 0$, the difference of the optimal power allocation strategy for EE maximization given by (19) from the form given in [10, eq. (28)] lies in the fact that the effect of the PU on the SU is fully considered in this paper while that effect is only partially considered in [10]. It is also seen that the optimal power allocation strategy for EE maximization given by (19) is influenced by the EE of the SU and the amplifier coefficient of the SU-Tx. Furthermore, when the PIP is sufficiently loose and the equation $y \leq r_s/(\eta\zeta + \tau)$ always holds, the optimal power level is equal to the minimum transmitted power level

required to guarantee the OC of the SU, and the maximum EE of the SU is achieved at the minimum transmitted power level required to guarantee the OC. Intuitively, in delay-sensitive green CR, it is reasonable that the SU does not need to increase transmitted power when the OC can be guaranteed and thus the maximum EE is achieved when the SU transmits with the minimum power level required to guarantee the OC. Finally, when the ATP constraint is sufficiently loose compared with the PIP constraint and $y \leq r_s/(\eta\zeta + \tau)$ always holds, it is seen that the optimal power allocation strategy is only dependent on the PIP constraint and is equivalent to the optimal power allocation strategy for OP minimization given in [10, eq. (28)].

It is seen that problem \mathbf{P}_6 can be efficiently solved by using the proposed optimal power allocation strategy given in Theorem 5 for a given η . In order to solve problem \mathbf{P}_5 and obtain the maximum EE of the SU and the corresponding optimal power allocation strategy, Algorithm 1 can be applied and it only requires several modifications. In this case, the optimal power, P_s^{opt} , is calculated by using Theorem 5. The objective function, $f(\eta_{EE}^n)$, and EE, η_{EE}^n , are calculated by using (16) and (13), respectively. For brevity, the details are not given here.

4. EE Maximization under Peak Transmit Power Constraint

In this section, EE maximization problems subject to constraints on the PIP and the PTP are studied in delay-insensitive green CR and delay-sensitive green CR. The PTP constraint is related to the nonlinearity of power amplifiers. Another algorithm based on the derived optimal power allocation strategies is proposed to solve EE maximization problems under the PIP constraint and the PTP constraint.

4.1. EE Maximization in Delay-Insensitive Green CR. In this subsection, the peak transmit power constraint is considered, given as

$$P_s(\nu) \leq P_{th} \quad (20a)$$

$$P_s(\nu) \geq 0 \quad (20b)$$

where P_{th} is the maximum instantaneous transmit power of the SU. In delay-insensitive green CR, the EE maximization problem subject to the PIP constraint and the PTP constraint, denoted by \mathbf{P}_7 , is given as

$$\mathbf{P}_7: \max_{P_s(\nu)} \eta_{EE}(P_s(\nu)) = \frac{\mathbb{E} \left\{ \log_2 \left(1 + g_{ss}(\nu) P_s(\nu) / (h_{ps}(\nu) P_p + \sigma_w^2) \right) \right\}}{\mathbb{E} \{ \zeta P_s(\nu) + P_C \}} \quad (21a)$$

$$\text{s.t.} \quad (1) \text{ and } (20a), (20b) \text{ are satisfied.} \quad (21b)$$

Let S_3 be $S_3 = \{P_s(\nu) \mid P_s(\nu) \in (1), P_s(\nu) \in (20a), (20b)\}$. It is straightforward to show that S_3 is a convex set. And

thus, problem \mathbf{P}_7 is also a nonlinear strictly quasiconcave and strictly pseudoconcave fractional programming problem.

Based on Dinkelbach's method, problem \mathbf{P}_7 is equivalent to the parameter optimization problem, denoted by \mathbf{P}_8 , given as

$$\begin{aligned} \mathbf{P}_8: \max_{P_s(\nu) \in \mathcal{S}_3} f(\eta) \\ = \mathbb{E} \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) \right\} \\ - \eta \mathbb{E} \{ \zeta P_s(\nu) + P_C \} \end{aligned} \quad (22)$$

where η is a nonnegative parameter. Using a similar method as used in Section 3, problem \mathbf{P}_8 can be solved by solving the following problem, given as

$$\max_{P_s \in \mathcal{S}_3} y(P_s) = \log_2 \left(1 + \frac{g_{ss} P_s}{h_{ps} P_p + \sigma_w^2} \right) - \eta \zeta P_s. \quad (23)$$

For a given η , problem \mathbf{P}_8 can be solved by solving (23) for all fading states. Let P_s^{opt} denote the optimal power of \mathbf{P}_8 . Since $y(P_s)$ is concave with respect to P_s , the optimal power allocation strategy for a given η , P_s^{opt} , can be obtained by solving (23) and is formally stated in Theorem 7. Let $\min(a, b, c)$ denote the minimum of a , b , and c .

Theorem 7. The optimal resource allocation strategy of \mathbf{P}_8 is given by

$$\widehat{P}_s = \left[\frac{1}{\eta \zeta \ln 2} - \frac{(h_{ps} P_p + \sigma_w^2)}{g_{ss}} \right]^+ \quad (24a)$$

$$P_s^{opt} = \min \left(\widehat{P}_s, \frac{P_{In}}{g_{sp}}, P_{th} \right) \quad (24b)$$

Remark 8. When $\eta = 0$, the optimal power allocation strategy for delay-insensitive green CR subject to the PIP constraint and the PTP constraint, given in Theorem 7, is equivalent to the optimal power allocation strategy for EC maximization in conventional CR given in [10, eq. (11)]. In this case, the EE maximization problem is equivalent to the EC maximization problem in conventional CR. This also can be seen from (22). From (24a) and (24b), it is seen that the optimal power allocation strategy for maximizing the achievable EE of the SU is required to take the EE of the SU and the amplifier coefficient into consideration. Moreover, when \widehat{P}_s is larger than the maximum of P_{th} and P_{In}/g_{sp} , it is seen from (24a) and (24b) that the optimal power strategy given in Theorem 7 is also equivalent to the optimal power allocation strategy for EC maximization in conventional CR given in [10, eq. (11)].

For a given η , it is seen that problem \mathbf{P}_8 can be efficiently solved by using the proposed optimal power allocation strategy given in Theorem 7. In order to solve problem \mathbf{P}_7 and obtain the maximum EE of the SU, similar to problem \mathbf{P}_1 and problem \mathbf{P}_5 , Dinkelbach's method can be applied. An iterative optimal power algorithm based on Dinkelbach's method is proposed to solve problem \mathbf{P}_7 , denoted by Algorithm 2. The principle of this algorithm is as follows. The optimal power level is iteratively updated by using Theorem 7. The maximum

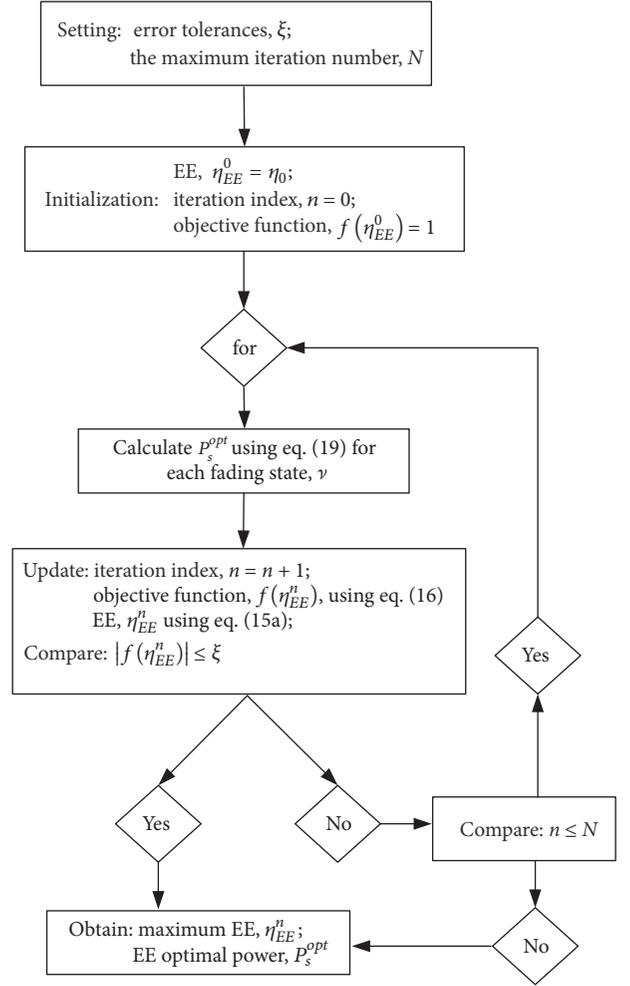


FIGURE 3: Flowchart of Algorithm 2 for EE maximization subject to the PTP constraint and the PIP constraint.

EE and the corresponding optimal power level of the SU are obtained as soon as the optimal power level can enable the fact that $f(\eta_{EE}^n) = 0$ holds. Otherwise, an ξ -optimal solution with an error tolerance ξ is adopted. If equations $f(\eta_{EE}^n) = 0$ and $|f(\eta_{EE}^n)| \leq \xi$ can not be satisfied, Algorithm 2 stops iterating when the maximum iteration number, N , is achieved. The flowchart of Algorithm 2 for EE maximization in delay-insensitive green CR, subject to the PIP constraint and the PTP constraint, is shown in Figure 3.

4.2. EE Maximization in Delay-Sensitive Green CR. In this subsection, the EE maximization problem in delay-sensitive green CR subject to constraint on the PIP and the PTP is studied. In this case, the EE maximization problem in delay-sensitive green CR, denoted by problem \mathbf{P}_9 , can be given as

$$\mathbf{P}_9: \max_{P_s(\nu)} \eta_{EE}(P_s(\nu)) = \frac{r_s \mathbb{E} \{ 1 - \chi_s(\nu) \}}{\mathbb{E} \{ \zeta P_s(\nu) + P_C \}} \quad (25a)$$

$$\text{s.t.} \quad (1) \text{ and } (20a), (20b) \text{ are satisfied} \quad (25b)$$

where $\chi_s(\nu)$ is an index associated with the outage event of the SU, given by (14b). Similar to problem \mathbf{P}_5 , problem \mathbf{P}_9 can be equivalent to a parameter optimization problem, denoted by problem \mathbf{P}_{10} , given as

$$\begin{aligned} \mathbf{P}_{10}: \max_{P_s(\nu) \in \mathcal{S}_3} f(\eta) \\ = r_s \mathbb{E} \{1 - \chi_s(\nu)\} - \eta \mathbb{E} \{\zeta P_s(\nu) + P_C\} \end{aligned} \quad (26)$$

where η is a nonnegative parameter. Similar to problem \mathbf{P}_6 , problem \mathbf{P}_{10} can be solved by solving the subproblem, given as

$$\max_{P_s \in \mathcal{S}_3} y(P_s) = -r_s \chi_s(P_s) - \eta \zeta P_s \quad (27)$$

where $\chi_s(P_s)$ is an explicit function of P_s . $\chi_s(P_s) = 1$ for $P_s < \gamma$ and $\chi_s(P_s) = 0$ for $P_s \geq \gamma$, and γ is given by (18). For a given η , problem \mathbf{P}_{10} can be solved by solving (27) for all fading states. Let P_s^{opt} denote the optimal power of \mathbf{P}_{10} . The following results can be obtained by solving (27). Let P_{\min} denote $P_{\min} = \min(r_s/(\eta\zeta), P_{In}/g_{sp}, P_{th})$.

Case 1 ($\gamma > P_{\min}$). In this case, when $\gamma > r_s/(\eta\zeta)$, the minimum power required to guarantee the OC of the SU is very large and results in very low EE; the SU stops transmitting in order to save energy. Thus, $P_s^{opt} = 0$. When $\gamma > P_{In}/g_{sp}$, the SU that transmits with the minimum power required to guarantee the OC of the SU results in unacceptable interference to the PU, and thus the SU stops transmitting in order to protect the PU. Hence, $P_s^{opt} = 0$. When $\gamma > P_{th}$ and the minimum power requiring the guarantee of the OC of the SU is larger than the PTP constraint, the SU stops transmitting in order to protect the SU transmitter. Thus, $P_s^{opt} = 0$.

Case 2 ($\gamma \leq P_{\min}$). In this case, $y(P_s)$ achieves the maximum, $-\eta\zeta\gamma$, at $P_s = \gamma$. Thus, $P_s^{opt} = \gamma$.

Theorem 9 can be formally stated based on the results obtained above.

Theorem 9. *The optimal resource allocation strategy of \mathbf{P}_{10} is given by*

$$P_s^{opt} = \begin{cases} \gamma, & \gamma \leq P_{\min} \\ 0, & \gamma > P_{\min} \end{cases} \quad (28a)$$

$$P_{\min} = \min\left(\frac{r_s}{\eta\zeta}, \frac{P_{In}}{g_{sp}}, P_{th}\right). \quad (28b)$$

Remark 10. In delay-sensitive green CR, under constraints on the PIP and the PTP, the optimal power allocation strategy for maximizing the EE given in Theorem 9 has a similar form to the well-known truncated channel inversion power allocation strategy [32]. The optimal power allocation strategy for green delay-insensitive CR subject to the PIP constraint and the PTP constraint, given in Theorem 9, is equivalent to the optimal power allocation strategy for OP minimization in conventional CR given in [10, eq. (24)] when $r_s/(\eta\zeta) \geq$

$\max(P_{In}/g_{sp}, P_{th})$. In this case, the maximum EEs achieved by using those two strategies are the same. It is seen that the optimal power allocation strategy for EE maximization in delay-sensitive green CR relates to the minimum power required to guarantee the OC of the SU. This outcome is explained by the fact that the SU does not need to increase transmit power when the OC can be guaranteed in delay-sensitive green CR.

For a given η , problem \mathbf{P}_{10} can be efficiently solved by the proposed optimal power allocation strategy given in Theorem 9. In order to solve problem \mathbf{P}_9 and obtain the optimal power strategy associated with the achievable maximum EE of the SU, the modified Algorithm 2 can be applied. In this case, the optimal power level, P_s^{opt} , is calculated by Theorem 9. The objective function, $f(\eta_{EE}^n)$, and EE, η_{EE}^n , are calculated by using (26) and (25a), respectively. For brevity, the details are not given here.

4.3. Complexity. As shown in Algorithms 1 and 2, a nonnegative dual variable related to the ATP is required to be updated by using the subgradient method when the ATP constraint is applied, whereas the for-loop is only activated when the PTP constraint is used. Let N_1 and N_2 denote the amount of for-loop activation and the number of the iterations required by using the subgradient method of Algorithm 1, respectively. Let N_3 denote the amount of for-loop activation of Algorithm 2. Let N_r denote the number of channel realizations. Let ℓ_1 denote the tolerance error for the subgradient method. Based on the complexity analysis for the subgradient method in [33, 34], the total complexity of Algorithm 1 is $O(N_1 N_2 N_r \times 1/\ell_1^2)$, whereas the total complexity of Algorithm 2 is $O(N_3 N_r)$. It is seen that the complexity of Algorithm 1 is higher than that of Algorithm 2 since it requires obtaining a nonnegative dual variable by using the subgradient method. N_2 is substantially increased when there are several constraints related to the average metrics, such as ATP. Thus, the decrease of the number of constraints on average metrics can enable the implementation of algorithms with low complexity.

5. Simulation Results

In this section, we give simulation results to evaluate the achievable maximum EE of the SU with the proposed optimal power allocation strategies in delay-insensitive green CR and delay-sensitive green CR. The achievable maximum EE with the proposed optimal power allocation strategies is compared with that achieved with the conventional power allocation strategies given in [10]. Simulation results are also presented to evaluate the performance of the two proposed algorithms. The constant circuit power of the SU transmitter and the amplifier coefficient, P_C and ζ , are set to be 0.05 W and 0.2. The variance of noise is set as 0.01. The transmit power of the PU, P_p , is set to be 60 mW. The iterative step size of τ is set as 0.1. The error tolerances, ξ and ξ_1 , are set as 10^{-4} . The EE and capacities are evaluated by using 10^6 channel realizations in ergodic fading channels. In all the simulations, the mean values of the channel power gains, g_{ss} , g_{sp} , and h_{ps} , are set as 2, 1.5, and 1.5. For the Nakagami- m channel, m is chosen as

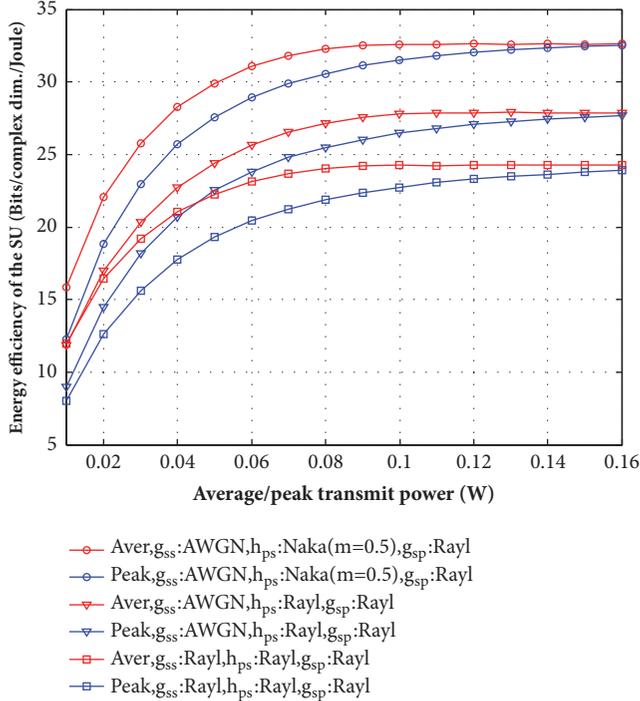


FIGURE 4: The EE of the SU versus the ATP/PTP constraint for different fading channel models with $P_{In} = 100 \text{ mW}$.

0.5. When the channel is modeled as an AWGN channel, the channel power gain is a constant. In delay-sensitive green CR, the OC of the SU is 1 bit/complex dimension (dim.).

Figure 4 shows the achievable maximum EE of the SU versus the ATP/PTP constraint for different fading channel models in delay-insensitive green CR. The PIP constraint is set as $P_{In} = 100 \text{ mW}$. It is seen that the SU can achieve EE gain under the ATP constraint compared with that achieved under the PTP constraint, irrespective of channel fading models. The reason is that the optimal power allocation strategy can benefit more from a flexibility on the ATP constraint than from that on the PTP constraint. Specifically, a large dynamic range of the transmit power can be obtained under the ATP constraint compared with that achieved under the PTP constraint. As shown in Figure 4, the maximum EE of the SU achieved under the PTP constraint converges to that achieved under the ATP constraint when the ATP/PTP constraint is sufficiently loose compared with the PIP constraint. This can be explained by the fact that the achievable maximum EE of the SU only depends on the PIP constraint when the ATP/PTP constraint is inactive, and the optimal power allocation strategies are the same in this case. This can be seen from Theorems 3 and 7. The same conclusion has been obtained in [27]. This further verifies our theoretical analysis.

Figure 5 shows the EE and the EC of the SU versus the PIP constraint for the EE maximization or for the conventional EC maximization under the PTP/ATP constraint. The PTP/ATP constraint is set as $P_{th} = \overline{P_{th}} = 100 \text{ mW}$. All the channels involved are Rayleigh fading. It is seen in Figure 5(a)

that the EE of the SU achieved for EE maximization by using the proposed optimal power allocation strategies is not less than that achieved for EC maximization by using the power allocation strategies proposed in [10]. This indicates that the proposed optimal power allocation strategies for EE maximization can guarantee that the SU obtains the maximum value of EE, whereas the optimal power allocation strategies proposed in [10] for EC maximization can not do that. Thus, those strategies proposed in [10] are not appropriate in delay-insensitive green CR in terms of EE maximization. It is interesting to note in Figure 5(a) that the EE of the SU achieved by using the conventional optimal power allocation strategy under the PTP constraint may be higher than that achieved by using the conventional optimal power allocation strategy under the ATP constraint. The reason is that the conventional optimal power allocation strategies for EC maximization are not optimal with respect to EE maximization in delay-insensitive green CR. However, it is seen in Figure 5(b) that the EC of the SU achieved by using the conventional optimal power allocation strategy under the PTP constraint is always smaller than that achieved by using the conventional optimal power allocation strategy under the ATP constraint. Similar results have been obtained for CR under spectrum sharing for EC maximization in [10, 14].

Figure 6 offers a deep insight into the maximum EE of the SU versus the PIP constraint and the ATP/PTP constraint in delay-insensitive green CR. All the channels involved are Rayleigh fading channels. It is seen that the maximum EE of the SU increases with the loose PIP constraint or with the loose ATP/PTP constraint. This is explained by the fact that the optimal power level increases with the looser constraints. It is also seen that the maximum EE of the SU achieved under the ATP constraint is larger than that achieved under the PTP constraint.

Figure 7 shows the EE of the SU versus the ATP/PTP constraint for different fading channel models in delay-sensitive green CR. The PIP constraint is set as $P_{In} = 100 \text{ mW}$. In delay-sensitive green CR, it is also seen that the EE of the SU achieved for EE maximization under the ATP constraint is larger than that achieved for EE maximization under the PTP constraint, irrespective of channel fading models. This can also be explained by the fact that the optimal power allocation strategy is more flexible under the ATP constraint than the optimal power allocation strategy under the PTP constraint.

Figure 8 presents a comparison of the EE and the OP of the SU achieved for the EE maximization with that achieved for the conventional OP minimization in delay-sensitive green CR. The PTP/ATP constraint is set as $P_{th} = \overline{P_{th}} = 50 \text{ mW}$. All the involved channels are Rayleigh fading. It is seen that the maximum EE of the SU for EE maximization by using the proposed optimal power allocation strategy under the ATP constraint is not less than that of the SU for OP minimization by using the optimal power allocation strategy given in [10, eq. (28)]. This indicates that the conventional optimal power allocation strategy for OP minimization given in [10, eq. (28)] is not always optimal in delay-sensitive green

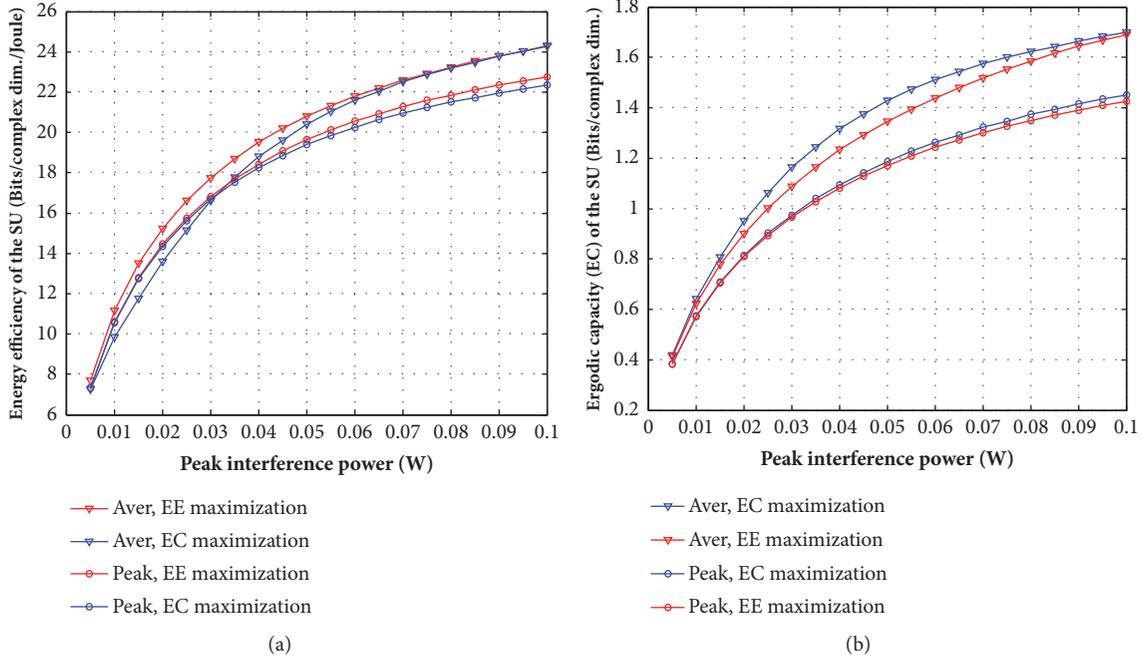


FIGURE 5: (a) The EE of the SU versus the PIP constraint for EE maximization or EC maximization under the PTP/ATP constraint, $P_{th} = \overline{P}_{th} = 100 \text{ mW}$. (b) The EC of the SU versus the PIP constraint for EE maximization or EC maximization under the PTP/ATP constraint, $P_{th} = \overline{P}_{th} = 100 \text{ mW}$.

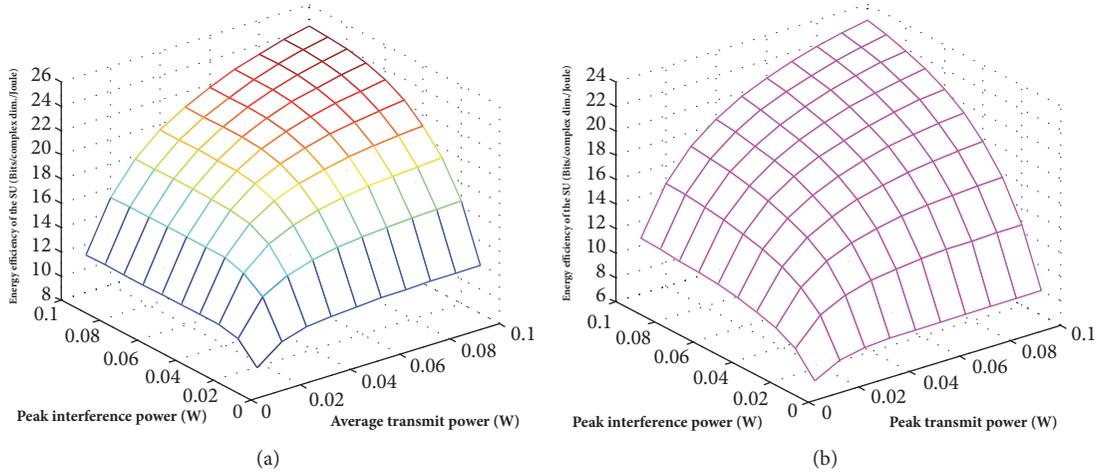


FIGURE 6: (a) The EE of the SU versus the PIP constraint and the ATP constraint in delay-insensitive green CR. (b) The EE of the SU versus the PIP constraint and the PTP constraint in delay-insensitive green CR.

CR in terms of EE maximization. As shown in Figure 8(a), when the PIP constraint is sufficiently tense compared with the ATP constraint, i.e., $P_{In} = 30 \text{ mW}$, the EE achieved by using the proposed optimal power allocation strategy is equal to that achieved by using the conventional optimal power allocation strategy. The reason is that only the PIP constraint is active and those two optimal strategies are equivalent in this case. This is consistent with our theoretical analysis given in Section 3.2. It is seen that the EE of the SU achieved by using the proposed optimal power allocation strategy for EE maximization given in Theorem 9 and that achieved by using

the optimal power allocation strategy for OP minimization given in [10, eq. (24)] are the same when the PIP constraint and the PTP constraint are very tense. This is explained by the fact that the proposed optimal power allocation strategy given in Theorem 9 and the optimal power allocation strategy given in [10, eq. (24)] are equivalent when the EE of the SU is small and the equation $r_s/(\eta\zeta) \geq \max(P_{In}/g_{sp}, P_{th})$ always holds. The results confirm our theoretical analysis given in Section 4.2.

Figure 9 is presented to show that the optimal power allocation strategy for OP minimization proposed in [10, eq.

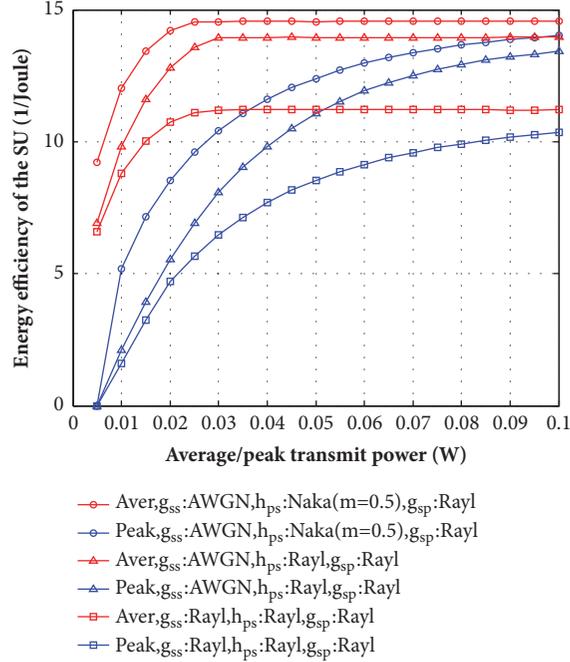


FIGURE 7: The EE of the SU versus the ATP/PTP constraint for different fading channel models with $P_{In} = 100 \text{ mW}$ in delay-sensitive green CR.

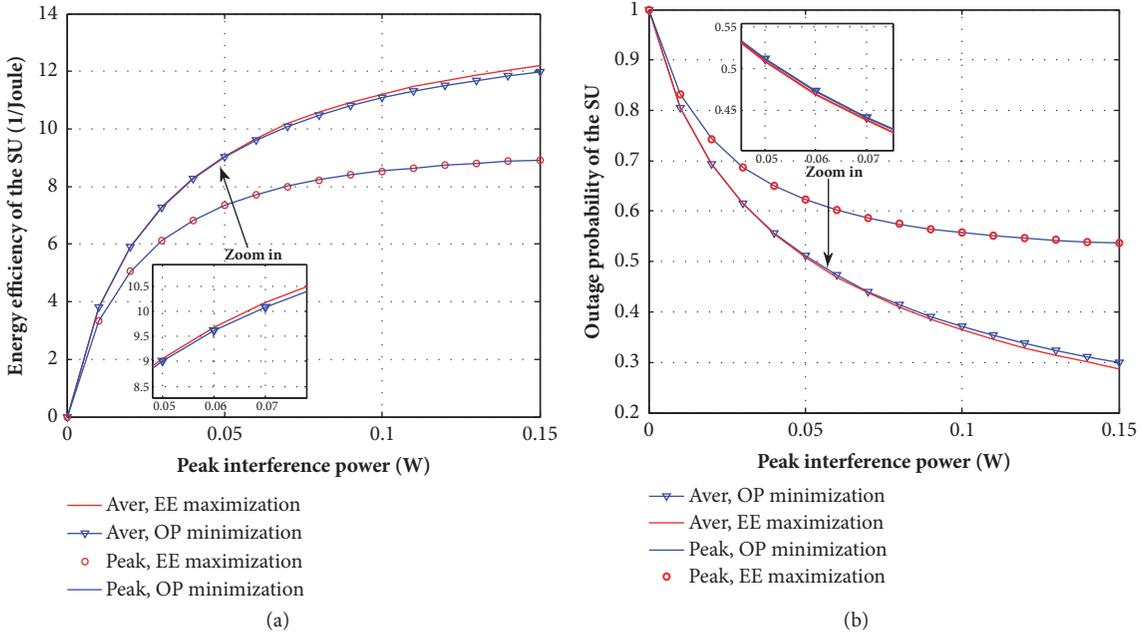


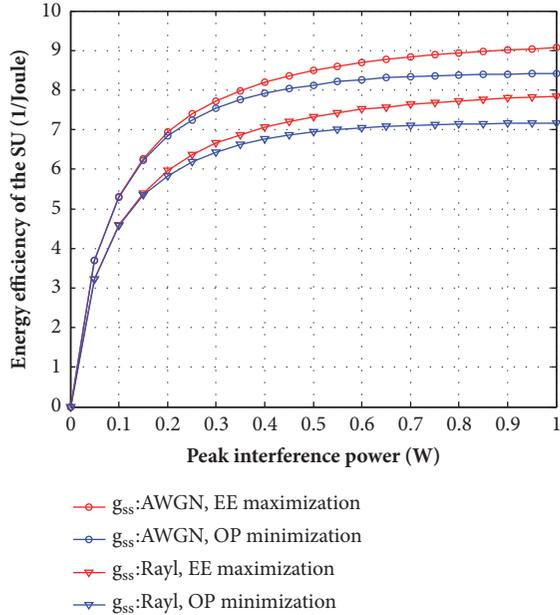
FIGURE 8: (a) The EE of the SU versus the PIP constraint for EE maximization or OP minimization under the PTP/ATP constraint in delay-sensitive green CR, $P_{th} = \bar{P}_{th} = 50 \text{ mW}$. (b) The OP of the SU versus the PIP constraint for EE maximization or OP minimization under the PTP/ATP constraint in delay-sensitive green CR, $P_{th} = \bar{P}_{th} = 50 \text{ mW}$.

(24)] is not always optimal in delay-sensitive green CR while the proposed power allocation strategy given in Theorem 9 is optimal with respect to EE maximization. Since those two strategies are equivalent when the PIP constraint and the PTP constraint are very tense, as shown in Figure 8, another green CR is considered. In this green CR, the transmit power

of the PU is set as $P_p = 500 \text{ mW}$. The PTP constraint is $P_{th} = 2 \text{ W}$. The channel gains, g_{sp} , and h_{ps} , model Rayleigh fading channels with mean 1.5. It is seen that the EE of the SU achieved for EE maximization by using the proposed power strategy given in Theorem 9 is larger than that achieved for OP minimization by using the power allocation strategy

TABLE I: The comparison of time (s) taken by Algorithm 1 with time taken by Algorithm 2.

Green CR		Transmit power							
		20 mW	30 mW	40 mW	50 mW	60 mW	70 mW	80 mW	90 mW
Delay-insensitive	Average	535.162	313.368	212.008	155.220	119.644	95.025	77.328	63.751
	Peak	0.198	0.203	0.199	0.199	0.208	0.200	0.199	0.198
Delay-sensitive	Average	856.223	733.604	242.051	198.239	130.344	81.732	61.740	50.934
	Peak	0.170	0.168	0.182	0.177	0.177	0.180	0.175	0.177


 FIGURE 9: The EE of the SU versus the PTP constraint for EE maximization or OP minimization under the PTP constraint with $P_{th} = 2$ W.

proposed in [10, eq. (24)] when the PIP is loose. It is also seen that the EE of the SU when g_{ss} models the AWGN channel is larger than that when g_{ss} models the Rayleigh channel. This also indicates that the fading of the channel between the SU-Tx and the SU-Rx is unfavorable to the SU in terms of EE maximization.

Table 1 shows comparisons of time taken by Algorithm 1 with the ATP constraint to time taken by Algorithm 2 with the PTP constraint in delay-insensitive green CR and delay-sensitive green CR. The unit of time is s. The PIP constraint is set as $P_{In} = 0.15$ W. All the channels involved are Rayleigh fading. Note that Algorithm 1 is proposed to solve the EE maximization problem under the ATP constraint, whereas Algorithm 2 is given for solving the EE maximization problem under the PTP constraint. It is seen that the time spent on solving the EE maximization problem under the ATP constraint is much more than the time spent on solving the EE maximization problem under the PTP constraint, irrespective of whether the case is delay-insensitive green CR or delay-sensitive green CR. The reason is that much time is spent on computing the nonnegative dual variable with respect to the ATP constraint. Thus, a green CR with the PTP constraint can be implemented with low complexity

compared with the green CR with the ATP constraint in terms of EE maximization. On the other hand, the SU can achieve EE gain under the ATP constraint in contrast with the PTP constraint. Thus, the design of a green CR should take the tradeoff between its complexity and its maximum EE into consideration.

6. Conclusion

Energy efficiency maximization problems were studied in delay-insensitive green CR and delay-sensitive green CR. Optimal power allocation strategies for delay-insensitive green CR and delay-sensitive green CR were designed to maximize the achievable EE of the SU. Two algorithms based on the proposed optimal resource allocation strategies were proposed. It is shown that CR with the instantaneous metric constraint can achieve implementation with low complexity in contrast with CR with the average metric constraint. Simulation results illustrated that the SU can achieve EE gain under the ATP constraint compared with that achieved under the PTP constraint in terms of EE maximization. The design of a green CR system should take the tradeoff between its complexity and its achievable maximum EE into consideration.

Appendix

A. Proof of Theorem 1

A strictly quasiconcave function is defined as follows. Let S be a nonempty convex set. f is a strictly quasiconcave function if, for each $x_1, x_2 \in S$ with $f(x_1) \neq f(x_2)$, one has

$$f[\lambda x_1 + (1 - \lambda) x_2] > \min \{f(x_1), f(x_2)\}, \quad (\text{A.1})$$

for $\lambda \in (0, 1)$.

Let $R_{EE}^{EC}(P_s)$ and $P_{EE}(P_s)$ denote the numerator and the denominator of $\eta_{EE}(P_s(\nu))$, respectively; namely, $R_{EE}^{EC}(P_s) = \mathbb{E}\{\log_2(1 + g_{ss}P_s/(h_{ps}P_p + \sigma_w^2))\}$ and $P_{EE}(P_s) = \mathbb{E}\{\zeta P_s + P_C\}$. $R_{EE}^{EC}(P_s)$ is a concave function of P_s since $\log_2(1 + g_{ss}P_s/(h_{ps}P_p + \sigma_w^2))$ is a concave function of P_s [31]. $P_{EE}(P_s)$ is a convex function of P_s and also is a concave function of P_s since $\zeta P_s + P_C$ is an affine function of P_s . Note that $P_{EE}(P_s) > 0$. For brevity, the index ν for the fading state is dropped. Let $P_s^1, P_s^2 \in S_1$ and $0 \leq \lambda \leq 1$. It is assumed that

$\eta_{EE}(P_s^1) < \eta_{EE}(P_s^2)$. Now, since $R_{EE}^{EC}(P_s)$ is a strictly concave function of P_s , one has

$$\begin{aligned} & R_{EE}^{EC}[\lambda P_s^1 + (1-\lambda)P_s^2] \\ & > \lambda R_{EE}^{EC}(P_s^1) + (1-\lambda)R_{EE}^{EC}(P_s^2) \\ & > \lambda R_{EE}^{EC}(P_s^1) + (1-\lambda)\frac{R_{EE}^{EC}(P_s^1)}{P_{EE}(P_s^1)}P_{EE}(P_s^2) \\ & = \frac{R_{EE}^{EC}(P_s^1)}{P_{EE}(P_s^1)}[\lambda P_{EE}(P_s^1) + (1-\lambda)P_{EE}(P_s^2)]. \end{aligned} \quad (\text{A.2})$$

Since $\lambda P_{EE}(P_s^1) + (1-\lambda)P_{EE}(P_s^2) = P_{EE}[\lambda P_s^1 + (1-\lambda)P_s^2]$, one has

$$\frac{R_{EE}^{EC}[\lambda P_s^1 + (1-\lambda)P_s^2]}{P_{EE}[\lambda P_s^1 + (1-\lambda)P_s^2]} > \frac{R_{EE}^{EC}(P_s^1)}{P_{EE}(P_s^1)}. \quad (\text{A.3})$$

Thus, $\eta_{EE}(P_s(\nu))$ is strictly quasiconcave on S_1 .

On the other hand, since $R_{EE}^{EC}(P_s)$ and $P_{EE}(P_s)$ are differentiable, the derivation of $\eta_{EE}(P_s)$ can be given as

$$\begin{aligned} & \frac{d\eta_{EE}(P_s)}{dP_s} \\ & = \frac{1}{P_{EE}^2(P_s)} \\ & \quad \times \left[\frac{dR_{EE}^{EC}(P_s)}{dP_s}P_{EE}(P_s) - \frac{dP_{EE}(P_s)}{dP_s}R_{EE}^{EC}(P_s) \right]. \end{aligned} \quad (\text{A.4})$$

Let P_v denote the value of P_s that makes the derivative of $\eta_{EE}(P_s)$ be equal to zero. One has

$$\frac{dR_{EE}^{EC}(P_v)}{dP_v} = \frac{dP_{EE}(P_v)}{dP_v}\eta_{EE}(P_v). \quad (\text{A.5})$$

Since $R_{EE}^{EC}(P_s)$ is a concave function of P_s , one has

$$\begin{aligned} R_{EE}^{EC}(P_s) & \leq R_{EE}^{EC}(P_v) + \frac{dR_{EE}^{EC}(P_v)}{dP_v}(P_s - P_v) \\ & = R_{EE}^{EC}(P_v) + \frac{dP_{EE}(P_v)}{dP_v}\eta_{EE}(P_v)(P_s - P_v). \end{aligned} \quad (\text{A.6})$$

Since $P_{EE}(P_s)$ is a convex function of P_s and also is a concave function of P_s , one has

$$\frac{dP_{EE}(P_v)}{dP_v}(P_s - P_v) = P_{EE}(P_s) - P_{EE}(P_v). \quad (\text{A.7})$$

According to (A.6) and (A.7), one has

$$\frac{R_{EE}^{EC}(P_s)}{P_s(P_s)} \leq \eta_{EE}(P_v). \quad (\text{A.8})$$

It is seen that P_v is the local maximum of $\eta_{EE}(P_s)$. Thus, $\eta_{EE}(P_s)$ is strictly pseudoconcave since $\eta_{EE}(P_s)$ is strictly quasiconcave and $R_{EE}^{EC}(P_s)$ is a concave function. Moreover, $\eta_{EE}(P_s) \leq \eta_{EE}(P_v)$ holds at any P_s if $d\eta_{EE}(P_v)/dP_v = 0$. Thus, P_v is the global maximum of $\eta_{EE}(P_s)$. The proof is complete.

B. Proof of Theorem 2

Let $P_s^{opt}(\nu)$ and $\eta_{opt}(P_s^{opt}(\nu))$ denote the optimal solution of \mathbf{P}_1 and the corresponding maximum value of EE. One has

$$\begin{aligned} \eta_{opt}(P_s^{opt}(\nu)) & = \max_{P_s(\nu) \in S_1} \\ & \quad \cdot \frac{\mathbb{E} \left\{ \log_2 \left(1 + g_{ss}(\nu) P_s(\nu) / (h_{ps}(\nu) P_p + \sigma_w^2) \right) \right\}}{\mathbb{E} \left\{ \zeta P_s(\nu) + P_C \right\}} \\ & = \frac{\mathbb{E} \left\{ \log_2 \left(1 + g_{ss}(\nu) P_s^{opt}(\nu) / (h_{ps}(\nu) P_p + \sigma_w^2) \right) \right\}}{\mathbb{E} \left\{ \zeta P_s^{opt}(\nu) + P_C \right\}}. \end{aligned} \quad (\text{B.1})$$

Thus, one has

$$\begin{aligned} & \frac{\mathbb{E} \left\{ \log_2 \left(1 + g_{ss}(\nu) P_s(\nu) / (h_{ps}(\nu) P_p + \sigma_w^2) \right) \right\}}{\mathbb{E} \left\{ \zeta P_s(\nu) + P_C \right\}} \\ & \leq \frac{\mathbb{E} \left\{ \log_2 \left(1 + g_{ss}(\nu) P_s^{opt}(\nu) / (h_{ps}(\nu) P_p + \sigma_w^2) \right) \right\}}{\mathbb{E} \left\{ \zeta P_s^{opt}(\nu) + P_C \right\}} \end{aligned} \quad (\text{B.2a})$$

$$\begin{aligned} & = \eta_{opt}(P_s^{opt}(\nu)) \\ & \mathbb{E} \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) \right\} \\ & \quad - \eta_{opt}(P_s^{opt}(\nu)) \mathbb{E} \left\{ \zeta P_s(\nu) + P_C \right\} \leq 0 \end{aligned} \quad (\text{B.2b})$$

$$\begin{aligned} & \mathbb{E} \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s^{opt}(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) \right\} \\ & \quad - \eta_{opt}(P_s^{opt}(\nu)) \mathbb{E} \left\{ \zeta P_s^{opt}(\nu) + P_C \right\} = 0. \end{aligned} \quad (\text{B.2c})$$

According to (B.2a), (B.2b), and (B.2c), the maximum of $f(\eta)$ is zero and is achieved when the optimal EE is obtained and the optimal power is adopted. It is proved that the optimized solution of \mathbf{P}_1 is the optimization of \mathbf{P}_2 . On the other hand, let P^{opt} denote the optimized solution of the problem, given as

$$\begin{aligned} & \max_{P_s(\nu) \in S_1} f(\eta_{opt}) \\ & = \mathbb{E} \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) \right\} \\ & \quad - \eta_{opt} \mathbb{E} \left\{ \zeta P_s(\nu) + P_C \right\} = 0. \end{aligned} \quad (\text{B.3})$$

One has

$$\begin{aligned} & \mathbb{E} \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P_s(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) \right\} \\ & \quad - \eta_{opt} \mathbb{E} \left\{ \zeta P_s(\nu) + P_C \right\} \\ & \leq \mathbb{E} \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P^{opt}(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) \right\} \\ & \quad - \eta_{opt} \mathbb{E} \left\{ \zeta P^{opt}(\nu) + P_C \right\} \end{aligned} \quad (\text{B.4a})$$

$$\mathbb{E} \left\{ \log_2 \left(1 + \frac{g_{ss}(\nu) P^{opt}(\nu)}{h_{ps}(\nu) P_p + \sigma_w^2} \right) \right\} \quad (\text{B.4b})$$

$$- \eta_{opt} \mathbb{E} \{ \zeta P^{opt}(\nu) + P_C \} = 0$$

$$\frac{\mathbb{E} \left\{ \log_2 \left(1 + g_{ss}(\nu) P_s(\nu) / (h_{ps}(\nu) P_p + \sigma_w^2) \right) \right\}}{\mathbb{E} \{ \zeta P_s(\nu) + P_C \}} \quad (\text{B.4c})$$

$$\leq \eta_{opt}.$$

Thus, it is seen that the optimized solution of \mathbf{P}_2 is also the optimized solution of \mathbf{P}_1 . The proof is complete.

Data Availability

The underlying data is not provided since it can be easily obtained by using the algorithms proposed in this article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] G. Y. Li, Z. Xu, C. Xiong et al., "Energy-efficient wireless communications: tutorial, survey, and open issues," *IEEE Wireless Communications Magazine*, vol. 18, no. 6, pp. 28–35, 2011.
- [2] F. Zhou, Y. Wu, R. Q. Hu, Y. Wang, and K. K. Wong, "Energy-efficient NOMA enabled heterogeneous cloud radio access networks," *IEEE Network*, vol. 32, no. 2, pp. 152–160, 2018.
- [3] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5G networks," *IEEE Wireless Communications Magazine*, vol. 24, no. 4, pp. 72–80, 2017.
- [4] F. Zhou, Z. Chu, H. Sun, R. Q. Hu, and L. Hanzo, "Artificial noise aided secure cognitive beamforming for cooperative MISO-NOMA using SWIPT," *IEEE Journal on Selected Areas in Communications*, 2018.
- [5] E. Z. Tragos, S. Zeadally, A. G. Fragkiadakis, and V. A. Siris, "Spectrum assignment in cognitive radio networks: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 3, pp. 1108–1135, 2013.
- [6] Z. Chu, F. Zhou, Z. Zhu, R. Q. Hu, and P. Xiao, "Wireless powered sensor networks for internet of things: maximum throughput and optimal power allocation," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 310–321, 2018.
- [7] F. Zhou, Y. Wu, Y. Liang, Z. Li, Y. Wang, and K. Wong, "State of the art, taxonomy, and open issues on NOMA in cognitive radio networks," *IEEE Wireless Communications Magazine*, In press.
- [8] S. Zhang, Q. Wu, S. Xu, and G. Y. Li, "Fundamental green tradeoffs: progresses, challenges, and impacts on 5G networks," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 33–56, 2017.
- [9] R. Zhang, "On peak versus average interference power constraints for protecting primary users in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 2112–2120, 2009.
- [10] X. Kang, Y. C. Liang, H. K. Garg, and R. Zhang, "Optimal power allocation for fading channels in cognitive radio networks: ergodic capacity and outage capacity," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 940–950, 2009.
- [11] X. Kang, R. Zhang, Y. Liang, and H. K. Garg, "Optimal power allocation strategies for fading cognitive radio channels with primary user outage constraint," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 374–383, 2011.
- [12] X. Gong, A. Ispas, G. Dartmann, and G. Ascheid, "Power allocation and performance analysis in spectrum sharing systems with statistical CSI," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1819–1831, 2013.
- [13] X. Gong, A. Ispas, G. Dartmann, and G. Ascheid, "Outage-constrained power allocation in spectrum sharing systems with partial CSI," *IEEE Transactions on Communications*, vol. 62, no. 2, pp. 452–466, 2014.
- [14] L. Musavian and S. Aissa, "Capacity and power allocation for spectrum-sharing communications in fading channels," *IEEE Transactions on Wireless Communications*, vol. 8, no. 1, pp. 148–156, 2009.
- [15] Y. He and S. Dey, "Power allocation for secondary outage minimization in spectrum sharing networks with limited feedback," *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2648–2663, 2013.
- [16] W.-C. Pao and Y.-F. Chen, "Adaptive gradient-based methods for adaptive power allocation in OFDM-based cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 2, pp. 836–848, 2014.
- [17] G. Ozcan and M. C. Gursoy, "Optimal power control for underlay cognitive radio systems with arbitrary input distributions," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, pp. 4219–4233, 2015.
- [18] S. Gong, P. Wang, and L. Duan, "Distributed power control with robust protection for PUs in cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3247–3258, 2015.
- [19] Y. Wang, W. Xu, K. Yang, and J. Lin, "Optimal energy-efficient power allocation for OFDM-based cognitive radio networks," *IEEE Communications Letters*, vol. 16, no. 9, pp. 1420–1423, 2012.
- [20] J. Mao, G. Xie, J. Gao, and Y. Liu, "Energy efficiency optimization for ofdm-based cognitive radio systems: a water-filling factor aided search method," *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2366–2375, 2013.
- [21] C. Luo, G. Min, F. R. Yu, M. Chen, L. T. Yang, and V. C. M. Leung, "Energy-efficient distributed relay and power control in cognitive radio cooperative communications," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 11, pp. 2442–2452, 2013.
- [22] A. Alabbasi, Z. Rezki, and B. Shihada, "Energy efficient resource allocation for cognitive radios: A generalized sensing analysis," *IEEE Transactions on Wireless Communications*, vol. 14, no. 5, pp. 2455–2469, 2015.
- [23] G. Ozcan and M. C. Gursoy, "Energy-efficient power adaptation for cognitive radio systems under imperfect channel sensing," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM '14)*, pp. 706–711, IEEE, Ontario, Canada, April-May 2014.
- [24] F. Zhou, Y. Wang, D. Qin, Y. Wang, and Y. Wu, "Secure EE maximisation in green CR: guaranteed SC," *IET Communications*, vol. 11, no. 16, pp. 2507–2513, 2017.
- [25] L. Wang, M. Sheng, X. Wang, Y. Zhang, and X. Ma, "Mean energy efficiency maximization in cognitive radio channels with PU outage constraint," *IEEE Communications Letters*, vol. 19, no. 2, pp. 287–290, 2015.

- [26] Q. Wu, G. Y. Li, W. Chen, and D. W. K. Ng, "Energy-efficient D2D overlaying communications with spectrum-power trading," *IEEE Transactions on Wireless Communications*, vol. 16, no. 7, pp. 4404–4419, 2017.
- [27] F. Zhou, N. C. Beaulieu, Z. Li, J. Si, and P. Qi, "Energy-efficient optimal power allocation for fading cognitive radio channels: ergodic capacity, outage capacity, and minimum-rate capacity," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2741–2755, 2016.
- [28] S. Schaible, "Fractional programming," *Zeitschrift für Operations Research*, vol. 27, no. 1, pp. 39–54, 1983.
- [29] I. Stancu-Minasian, *Fractional Programming: Theory, Methods and Applications*, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1997.
- [30] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.
- [31] S. P. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.
- [32] A. J. Goldsmith and P. P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Transactions on Information Theory*, vol. 43, no. 6, pp. 1986–1992, 1997.
- [33] Q. Wu, G. Y. Li, W. Chen, and D. W. K. Ng, "Energy-efficient small cell with spectrum-power trading," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 3394–3408, 2016.
- [34] Q. Wu, W. Chen, M. Tao, J. Li, H. Tang, and J. Wu, "Resource allocation for joint transmitter and receiver energy efficiency maximization in downlink OFDMA systems," *IEEE Transactions on Communications*, vol. 63, no. 2, pp. 416–430, 2015.

Research Article

Power-Splitting Scheme for Nonlinear Energy Harvesting AF Relaying with Direct Link

Xiaobo Bai ¹, Jingfeng Shao,¹ Jiengang Tian,² and Liqin Shi ¹

¹*School of Management, Xi'an Polytechnic University, Xi'an, China*

²*Army Academy of Border and Coastal Defence, Department of Information and Arms, Xi'an, China*

Correspondence should be addressed to Liqin Shi; connectliqin@126.com

Received 3 May 2018; Revised 6 June 2018; Accepted 10 June 2018; Published 2 July 2018

Academic Editor: Fuhui Zhou

Copyright © 2018 Xiaobo Bai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Simultaneous wireless information and power transfer (SWIPT) is a promising technique to prolong the lifetime of energy-constrained relay systems. Most previous works optimize power-splitting (PS) scheme based on a linear or a simple two-piecewise linear energy harvesting (EH) model, while the employed EH model may not characterize the properties of practical EH harvesters well. This leads to a mismatch between the existing PS scheme and the practical EH harvester available for relay systems. Motivated by this, this paper is devoted to the design of PS scheme in a nonlinear EH amplify-and-forward energy-constrained relay system in the presence of a direct link between the source and the destination. In particular, we formulate an optimization problem to maximize the system capacity according to the instantaneous channel state information, subject to a nonlinear EH model based on the logistic function. The objective function of the formulated problem is proven to be unimodal and there is no closed-form expression for the optimal PS ratio due to the complexity of logistic function. In order to reduce overhead cost of optimizing PS ratio, a simpler nonlinear EH model based on the inverse proportional function is employed to replace the nonlinear EH model based on the logistic function and we further derive the closed-form expression for the optimal PS ratio. Simulation results reveal that a higher system capacity can be achieved when the PS scheme is optimized based on nonlinear EH models instead of the linear EH model, and that there is only a marginal difference between the capacity under the two optimal PS schemes optimized for two different nonlinear EH models.

1. Introduction

The aim of Internet of Things (IoT) is to enable people and things to be connected anytime, anyplace, with anything and anyone, ideally using any path/network and any service [1]. It allows massive IoT devices such as low-power wireless sensors to access the wireless communication networks and communicate with each other. The limited lifetime of IoT devices is a fundamental problem for implementing IoT deployment. Motivated by this, simultaneous wireless information and power transfer (SWIPT) is proposed and devoted to the solution of this problem. On the other hand, relaying technology is highly beneficial in wireless communications in terms of the communication range, the energy efficiency, and the system capacity [2, 3]. However, when the relaying technology is employed in IoT networks, the IoT devices are unwilling to be a relay in order to avoid the extra energy

consumption since these devices usually have limited battery capacity. Thus, integrating SWIPT and relaying is a viable option to strike a tradeoff between information processing and power supply and gains much attention recently [4–6]. In this field, the design of energy harvesting (EH) scheme, which instructs the relay either to switch the received radio frequency (RF) signal in the time domain or to split the received RF signal in the power domain to provide signal processing and power transfer, is one of the most important issues.

Until now, there have been many reports [4–14] regarding how to design an appropriate EH scheme for SWIPT based relay systems. The works [4, 7] studied the effects of power-splitting (PS) ratio and time-switching (TS) ratio on the amplify-and-forward (AF) and decode-and-forward (DF) relay systems by deriving the expressions for outage probability and ergodic capacity. In [8], both the PS ratio and

TS ratio were optimized to maximize the system transmission rate in DF relay systems, where the relay has a certain amount of remaining energy. By combining both TS and PS, a hybrid scheme was proposed and further optimized in [9]. The authors of [10] designed two optimal PS schemes with full and partial channel state information (CSI) to minimize the system outage probability in an AF relay system. Assuming the availability of source-destination link, the optimal PS ratio was designed and the diversity gains for the relay and the destination were analyzed [11]. By means of the stochastic geometry, the authors proposed a dynamic PS scheme in a DF relay system with a random number of transmitter-receiver pairs and investigated its outage probability [12]. Recognizing the advantages of nonorthogonal multiple access (NOMA) in spectrum efficiency, a novel cooperative SWIPT-NOMA system was integrated [13], and an optimal PS scheme was further proposed [14].

These works [4–14] were based on a linear EH model, in which the RF-to-direct current (DC) power conversion efficiency is a fixed constant and independent of the input power of the energy harvester. As pointed out by [15–17], the practical energy harvester operates in a nonlinear mode and the linear EH model may not characterize the properties of practical EH harvesters well. Further, the optimal PS schemes based on the linear EH model may not be optimal for the practical scenario. As a result, the existing schemes based on a linear EH model may need to be redesigned to avoid the mismatch caused by the resource allocation under the linear EH model, and ever-increasing attention has been paid into the study of nonlinear EH model in wireless communications (see [18–30] and references therein). References [18–25] introduced the nonlinear EH model into the wireless powered communication networks, the SWIPT-NOMA system, and the cognitive radio networks with SWIPT, where the resource allocation scheme, including the transmit power of the transmitter, and the PS/TS ratio, is concentrated. The studies revealed that a higher system capacity could be achieved by designing the EH scheme based on the nonlinear EH model instead of the conventional linear one. Apart from the aforementioned networks, the researchers have also studied the design of EH scheme in nonlinear EH relay systems [26–30]. For example, the authors of [26, 27] focused on the design of PS scheme for nonlinear EH two-way relay systems. Since the low complexity of hardware is very vital to energy-constrained relay systems, the researches on one-way relay systems have attracted a lot of interests [28–30]. In particular, the works [28, 29] derived the outage probability of a PS enabled nonlinear EH relay system. Considering the perfect/imperfect CSI at the relay, an optimal PS scheme was developed to minimize the outage probability [30] in an AF nonlinear EH relay system. These aforementioned works have laid the foundation for the design of EH scheme in one-way relay systems. After careful analysis of the existing works [28–30], it can be found that a simple two-piecewise linear EH model was employed, and that the employed two-piecewise linear model cannot provide sufficient precision compared with the existing nonlinear EH models based on the logistic function and the inverse proportional function. Therefore, there still remains a large gap to be filled regarding

the design of EH scheme for nonlinear EH one-way relay systems.

Motivated by this observation, this paper is devoted to the design of PS schemes for an AF relay system with direct link in terms of system capacity, where the nonlinear EH models proposed in [16, 18] are used to characterize the properties of practical EH circuits. Our contributions are as follows.

- (i) We optimize the PS scheme to maximize the system capacity under the nonlinear EH model based on a logistic function. We prove that the objective function is unimodal and the optimal solution is obtained by the golden section search method.
- (ii) Employing the nonlinear EH model based on an inverse proportional function instead of the logistic function, a closed-form expression for the optimal PS ratio is derived to maximize the system capacity. Compared with the PS scheme optimized for the nonlinear EH model based on the logistic function, the PS scheme optimized for the nonlinear EH model based on the inverse proportional enjoys a lower computational complexity with the near-optimal performance.

It is worth pointing out that energy efficiency (EE) is also an important performance metric. Since the EE is defined as the ratio of system capacity to power consumption [31], the optimization of EE is equivalent to the optimization of system capacity for a fixed transmit power. Thus, our derived optimal solution of this paper is the same as the optimal solution to maximize EE. If the transmit power is adjustable and smaller than a maximum power transmit, we should optimize both PS ratio and transmit power simultaneously. In this case, the solution to maximize EE is different from the derived results of this paper, while how to obtain the optimal PS ratio and optimal transmit power is beyond the scope of this paper.

2. System Model and Working Flow

As shown in Figure 1, we consider a SWIPT based AF relay system, composed of a source node S , an energy-constrained relay node R , and a destination node D . To be general, we assume that there exists a direct link between the source and the destination. All nodes operate in a half-duplex mode and are equipped with single antenna. It is assumed that both the source and the destination are equipped with fixed power supply, and that both “harvest-then-forward” scheme and the PS scheme are employed to encourage the relay to be cooperative with the source’s transmission. Let h_{ij} , $\forall i, j \in \{S, R, D\}$ and $i \neq j$ denote the channel coefficients between i and j in a quasistatic fading model. Let P_s denote the transmit power of the source. All the channel state information (CSI) is available at the relay in order to investigate the system performance limits of the PS scheme. Moreover, we ignore the processing energy required by the transmit/receive circuitry at the relay [4–14].

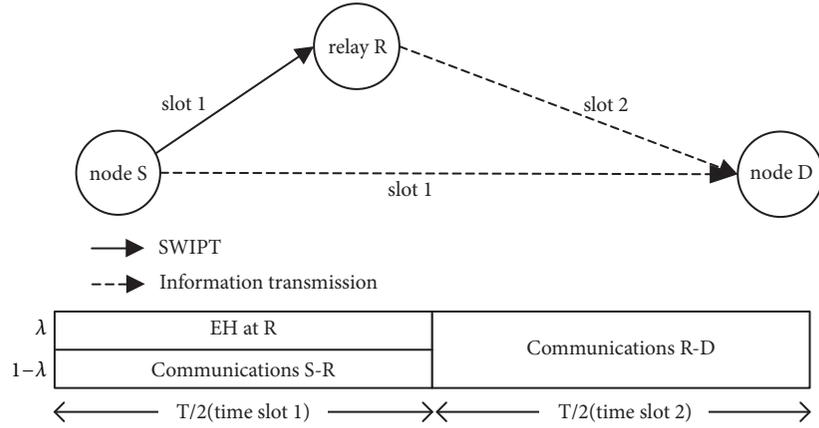


FIGURE 1: An illustration of PS scheme.

2.1. Energy Harvesting Model. Since the conventional linear EH model may not be accurate for modelling the practical energy harvester due to the nonlinearity of the diodes, inductors, and capacitors, by fitting over real measurement data based on the logistic function, the authors of [18] proposed an accurate nonlinear EH model, given by

$$P_H^{(1)} = \frac{E_{\max}/(1 + \exp(-a(P_{\text{RF}} - b))) - E_{\max}\Omega}{1 - \Omega} \quad (1)$$

where $P_H^{(1)}$ denotes the amount of harvested energy; P_{RF} is the input power of the energy harvester; a and b are the parameters determined by the resistance, capacitance, and so on; E_{\max} is the maximum harvestable power when the circuit becomes saturated and $\Omega = 1/(1 + \exp(ab))$.

Recently, using the inverse proportional function instead of the logistic function to fit over real measurement data, Chen et al. in [16] proposed a more mathematically tractable nonlinear EH model, given as

$$P_H^{(2)} = \frac{aP_{\text{RF}} + b}{P_{\text{RF}} + c} - \frac{b}{c} \quad (2)$$

where a, b , and c are the fixed parameters. The comparison between the above two nonlinear models will be presented in Section 4.

2.2. Working Flow. In the first time slot, S broadcasts information x_s to R and D with a transmit power P_s . Then the received signals at the relay and the destination node D are given by

$$y_{sr} = \sqrt{P_s} h_{\text{SR}} x_s + n_{sr} \quad (3)$$

$$y_{sd} = \sqrt{P_s} h_{\text{SD}} x_s + n_{sd} \quad (4)$$

where $\mathbb{E}\{|x_s|^2\} = 1$ and $n_{sr} \sim \mathcal{CN}(0, \sigma_{sr}^2)$ is the additive white Gaussian noise (AWGN) and $n_{sd} \sim \mathcal{CN}(0, \sigma_{sd}^2)$.

For the destination D , the received signal-to-noise ratio (SNR) can be calculated as

$$\gamma_{sd} = \frac{P_s |h_{\text{SD}}|^2}{\sigma_{sd}^2}. \quad (5)$$

For the relay R , after receiving the signal from the source S , the relay R splits it into two parts: $\sqrt{\lambda} y_{sr}$ for harvesting energy and $\sqrt{1 - \lambda} y_{sr}$ for information processing, where λ is the PS ratio at the relay. Thus, the received RF power used for energy harvesting, P_{RF} , and the signal used for amplification, x_r , are given by

$$P_{\text{RF}} = \lambda (P_s |h_{\text{SR}}|^2 + \sigma_{sr}^2) \quad (6)$$

$$x_r = \sqrt{1 - \lambda} \left(\sqrt{P_s} h_{\text{SR}} x_s + n_{sr} \right) + n_p \quad (7)$$

where $n_p \sim \mathcal{CN}(0, \sigma_p^2)$ is the additive white baseband Gaussian noise.

If the energy harvesting is completed, the total harvested energy is given by

$$E_{\text{total}} = \frac{T}{2} P_H \quad (8)$$

where P_H is the output power of the energy harvester and

$$P_H = \begin{cases} P_H^{(1)}, & \text{if the non-linear EH model based on the logistic function is employed} \\ P_H^{(2)}, & \text{if the non-linear EH model based on the inverse proportional function is employed.} \end{cases} \quad (9)$$

In the second time slot, R amplifies the signal x_r by using the harvested energy and forwards the amplified signal to D . Hence, the received signal at D is given by

$$y_{rd} = h_{RD}Gx_r + n_{rd} \quad (10)$$

where $G = \sqrt{P_H/((1-\lambda)(P_s|h_{SR}|^2 + \sigma_{sr}^2) + \sigma_p^2)}$ is the amplifier gain at R and $n_{rd} \sim \mathcal{CN}(0, \sigma_{rd}^2)$ is the AWGN at D .

Then the received SNR from R to D is given by

$$\gamma_{rd} = \frac{(1-\lambda)P_sP_H|h_{SR}|^2|h_{RD}|^2}{P_H|h_{RD}|^2((1-\lambda)\sigma_{sr}^2 + \sigma_p^2) + \sigma_{rd}^2((1-\lambda)K + \sigma_p^2)} \quad (11)$$

where $K = P_s|h_{SR}|^2 + \sigma_{sr}^2$.

By implementing the maximal ratio combining (MRC) scheme at the destination D , the overall system capacity is given by

$$C_{\text{total}} = \frac{T}{2} \log_2(1 + \gamma_{sd} + \gamma_{rd}). \quad (12)$$

3. Optimal Power-Splitting Scheme

3.1. Nonlinear EH Model Based on the Logistic Function. In this subsection, we design an optimal PS scheme to maximize the overall system capacity for an AF relay system with SWIPT based on the nonlinear EH model in [18].

Substituting (1) into (11), the received SNR from R to D under the nonlinear EH model, in (1), is

$$\gamma_{rd}^{(1)} = \frac{A(1-\lambda)P_H^{(1)}}{B[(1-\lambda)P_H^{(1)} + CP_H^{(1)} + D_1(1-\lambda) + E]} \quad (13)$$

where $A = P_s|h_{SR}|^2$, $B = \sigma_{sr}^2$, $C = \sigma_p^2/\sigma_{sr}^2$, $D_1 = K\sigma_{rd}^2/|h_{RD}|^2\sigma_{sr}^2$, and $E = \sigma_{rd}^2\sigma_p^2/|h_{RD}|^2\sigma_{sr}^2$.

Then the system capacity in this case can be computed as

$$C_{\text{total}}^{(1)} = \frac{T}{2} \log_2(1 + \gamma_{sd} + \gamma_{rd}^{(1)}). \quad (14)$$

Based on (14), the optimization problem can be formulated as

$$\begin{aligned} \mathcal{P}1 : \text{maximize} \quad & C_{\text{total}}^{(1)} \\ \text{subject to} \quad & 0 \leq \lambda \leq 1. \end{aligned} \quad (15)$$

It is obvious that the optimization problem of maximizing the overall system capacity is equivalent to maximizing the received SNR from R to D , which can be reformulated as

$$\begin{aligned} \mathcal{P}2 : \text{maximize} \quad & \gamma_{rd}^{(1)} \\ \text{subject to} \quad & 0 \leq \lambda \leq 1. \end{aligned} \quad (16)$$

By taking the first-order derivative of the objective function $\gamma_{rd}^{(1)}$ with respect to λ , we have

$$\frac{\partial \gamma_{rd}^{(1)}}{\partial \lambda} = \frac{-Af_1(\lambda)f_2(\lambda)}{B[(1-\lambda)P_H^{(1)} + CP_H^{(1)} + D_1(1-\lambda) + E]^2} \quad (17)$$

where

$$\begin{aligned} f_1(\lambda) &= \frac{aE_{\text{max}}(e^{ab} + 1) \exp(-a\lambda K)}{[1 + \exp(ab - a\lambda K)]^2} \\ f_2(\lambda) &= \frac{CE_{\text{max}}(1 - e^{-a\lambda K})^2 e^{a\lambda K}}{a(e^{ab} + 1)} - D_1(1-\lambda)^2 \\ &\quad + E \frac{(e^{ab} + e^{a\lambda K})(1 - e^{-a\lambda K})}{a(e^{ab} + 1)} - E(1-\lambda). \end{aligned} \quad (18)$$

Let $\partial \gamma_{rd}^{(1)}/\partial \lambda = 0$ and we have $f_1(\lambda)f_2(\lambda) = 0$. Since $f_1(\lambda) > 0$ always holds, $\partial \gamma_{rd}^{(1)}/\partial \lambda = 0$ is equivalent to $f_2(\lambda) = 0$. Further, taking the first-order derivative of $f_2(\lambda)$, we have

$$\begin{aligned} \frac{\partial f_2(\lambda)}{\partial \lambda} &= \frac{CE_{\text{max}}Ke^{a\lambda K}(1 - e^{-2a\lambda K})}{e^{ab} + 1} + 2D_1(1-\lambda) \\ &\quad + E + \frac{EK(e^{ab-a\lambda K} + e^{a\lambda K})}{e^{ab} + 1} > 0. \end{aligned} \quad (19)$$

Thus, $f_2(\lambda)$ increases with the increasing of λ . Since both $f_2(0) = -D_1 - E < 0$ and $f_2(1) = CE_{\text{max}}(1 - e^{-aK})^2 e^{aK}/a(e^{ab} + 1) + E(e^{ab} + e^{aK})(1 - e^{-aK})/a(e^{ab} + 1) > 0$ are satisfied, there exists a unique $\lambda \in [0, 1]$ for $f_2(\lambda) = 0$.

Let s_1 denote the solution to $f_2(\lambda) = 0$. Then s_1 is also the solution to $\partial \gamma_{rd}^{(1)}/\partial \lambda = 0$. Based on the expression of $\partial \gamma_{rd}^{(1)}/\partial \lambda$, $\gamma_{rd}^{(1)}$ increases with increasing λ when $\lambda \in [0, s_1]$ and decreases with increasing λ for $\lambda \in [s_1, 1]$. Therefore, $\gamma_{rd}^{(1)}$ is a unimodal function and s_1 is the optimal solution to $\mathcal{P}2$. Here, the value of s_1 can be determined by means of the golden section search method. The details of this method are summarized in Algorithm 1. According to [21], the computational complexity of this algorithm can be computed as $O(\log(1/\epsilon))$, where ϵ is the precision.

3.2. Nonlinear EH Model Based on the Inverse Proportional Function. In this subsection, we design an optimal PS scheme to maximize the overall system capacity for an AF relay system with SWIPT based on the nonlinear EH model in [16]. Based on (2), the received SNR from R to D under the nonlinear EH model, in (2), is given by

$$\gamma_{rd}^{(2)} = \frac{A(1-\lambda)P_H^{(2)}}{B[(1-\lambda)P_H^{(2)} + CP_H^{(2)} + D_1(1-\lambda) + E]}. \quad (20)$$

Then the optimization problem can be formulated as

$$\begin{aligned} \mathcal{P}3 : \text{maximize} \quad & C_{\text{total}}^{(2)} \\ \text{subject to} \quad & 0 \leq \lambda \leq 1, \end{aligned} \quad (21)$$

where $C_{\text{total}}^{(2)} = (T/2)\log_2(1 + \gamma_{sd} + \gamma_{rd}^{(2)})$. Similar to $\mathcal{P}1$, $\mathcal{P}3$ can be reformulated as

$$\begin{aligned} \mathcal{P}4 : \text{maximize} \quad & \gamma_{rd}^{(2)} \\ \text{subject to} \quad & 0 \leq \lambda \leq 1. \end{aligned} \quad (22)$$

Require: $a, b, E_{\max}, h_{SD}, h_{SR}, h_{RD}, P_s$
Ensure: Optimal PS ratio λ^*

- (1) Initialize the maximum tolerance ε ;
- (2) Set low = 0, high = 1;
- (3) **loop**
- (4) **If** |high – low| < ε **then**
- (5) Set $\lambda^* = \frac{\text{low} + \text{high}}{2}$
- (6) Break;
- (7) **end if**
- (8) Set $\lambda_{\min} = \text{low} + 0.382(\text{high} - \text{low})$ and $\lambda_{\max} = \text{low} + 0.618(\text{high} - \text{low})$;
- (9) Compute $\gamma_{rd}^{(1)}(\lambda_{\min})$ and $\gamma_{rd}^{(1)}(\lambda_{\max})$ based on (13);
- (10) **if** $\gamma_{rd}^{(1)}(\lambda_{\min}) > \gamma_{rd}^{(1)}(\lambda_{\max})$ **then**
- (11) Set high = λ_{\max} ;
- (12) **else**
- (13) Set low = λ_{\min} ;
- (14) **end if**
- (15) **end loop**
- (16) Compute the optimal $C_{\text{total}}^{(1)}(\lambda^*)$ based on (14).

ALGORITHM 1: Golden section search algorithm.

According to (20), the first-order derivative of $\gamma_{rd}^{(2)}$ with respect to λ is

$$\frac{\partial \gamma_{rd}^{(2)}}{\partial \lambda} = \frac{A f_3(\lambda) (M\lambda^2 - N\lambda + Q)}{B [(1 - \lambda) P_H^{(2)} + C P_H^{(2)} + D_1 (1 - \lambda) + E]^2} \quad (23)$$

where $f_3(\lambda) = c(a - b/c)/(x + c)^2$, $M = D_1 - CK(ac - b)/c^2 - EK/c$, $N = 2D_1 + 2E$, and $Q = D_1 + E$.

Based on (2), we have

$$\lim_{\lambda K \rightarrow +\infty} \left(\frac{a\lambda K + b}{\lambda K + c} - \frac{b}{c} \right) = a - \frac{b}{c} > 0. \quad (24)$$

Therefore, $f_3(\lambda) > 0$ always holds for any value of $\lambda \in [0, 1]$.

Let $\partial \gamma_{rd}^{(2)}/\partial \lambda = 0$ and we have $M\lambda^2 - N\lambda + Q = 0$. Thus, we need to obtain the solutions to $M\lambda^2 - N\lambda + Q = 0$ to achieve the optimal PS ratio.

According to the expression of M , there are two cases for the solutions to $M\lambda^2 - N\lambda + Q = 0$, which are **Case 1:** $M = 0$; **Case 2:** $M \neq 0$.

Case 1. Based on $M = 0$, we have $-N\lambda + Q = 0$. Then the solution to $M\lambda^2 - N\lambda + Q = 0$ is given by Q/N and $Q/N = 0.5$. Since $\gamma_{rd}^{(2)}$ increases with increasing λ when $\lambda \in [0, 0.5]$ and decreases with increasing λ for $\lambda \in [0.5, 1]$, the optimal solution to $\mathcal{P}4$ is 0.5 and $\lambda^* = 0.5$.

Case 2. Let $\Delta = N^2 - 4MQ$. When both $\Delta < 0$ and $M > 0$ are satisfied, $M\lambda^2 - N\lambda + Q > 0$ always holds and $\partial \gamma_{rd}^{(2)}/\partial \lambda > 0$ holds for any value of $\lambda \in [0, 1]$. Therefore, $\gamma_{rd}^{(2)}$ increases with the increasing of λ and the optimal PS ratio λ^* is 1.

When $\Delta \geq 0$ is satisfied, the solutions to $M\lambda^2 - N\lambda + Q = 0$ are given by

$$\begin{aligned} x_1 &= \frac{N - \sqrt{\Delta}}{2M}, \\ x_2 &= \frac{N + \sqrt{\Delta}}{2M}. \end{aligned} \quad (25)$$

Let $x_{\min} = \min(x_1, x_2)$ and $x_{\max} = \max(x_1, x_2)$.

For the case with $M > 0$, we have $x_{\min} > 0$ and $x_{\max} > 0$. (i) If $x_{\max} < 1$ is satisfied, $\gamma_{rd}^{(2)}$ increases with $\lambda \in [0, x_{\min}]$, decreases with $\lambda \in [x_{\min}, x_{\max}]$, and then increases with $\lambda \in [x_{\max}, 1]$. The maximum $\gamma_{rd}^{(2)}$ can be obtained by $\max(\gamma_{rd}^{(2)}(x_{\min}), \gamma_{rd}^{(2)}(1))$. Since $\gamma_{rd}^{(2)}(x_{\min}) > \gamma_{rd}^{(2)}(0) = 0 = \gamma_{rd}^{(2)}(1)$ holds, the optimal PS ratio is given by $\lambda^* = x_{\min}$. (ii) If $x_{\min} > 1$ is satisfied, $\gamma_{rd}^{(2)}$ is a monotonic increasing function and the optimal solution to $\mathcal{P}4$ is $\lambda^* = 1$. (iii) If $x_{\min} \leq 1 \leq x_{\max}$ is satisfied, $\gamma_{rd}^{(2)}$ increases with $\lambda \in [0, x_{\min}]$ and decreases with $\lambda \in [x_{\min}, 1]$. In this case, the optimal PS ratio is given by $\lambda^* = x_{\min}$.

For the case with $M < 0$, we have $x_{\min} < 0$ and $x_{\max} > 0$. (i) If $x_{\max} > 1$ is satisfied, $\partial \gamma_{rd}^{(2)}/\partial \lambda > 0$ always holds for $\lambda \in [0, 1]$. $\gamma_{rd}^{(2)}$ increases with the increasing of λ and the optimal PS ratio is given by $\lambda^* = 1$. (ii) If $x_{\max} \leq 1$ is satisfied, $\gamma_{rd}^{(2)}$ increases with $\lambda \in [0, x_{\max}]$ and decreases with $\lambda \in [x_{\max}, 1]$. In this case, the optimal PS ratio is given by $\lambda^* = x_{\max}$.

Thus, the optimal solution to $\mathcal{P}3$ is summarized as **Proposition 1**.

Proposition 1. The optimal PS ratio for $\mathcal{P}3$ is given by

$$\lambda^* = \begin{cases} 0.5, & \text{for } M = 0 \\ x_{\min}, & \text{for } M > 0, \Delta \geq 0 \text{ and } x_{\min} < 1 \\ x_{\max}, & \text{for } M < 0, \Delta \geq 0 \text{ and } x_{\max} < 1 \\ 1, & \text{otherwise.} \end{cases} \quad (26)$$

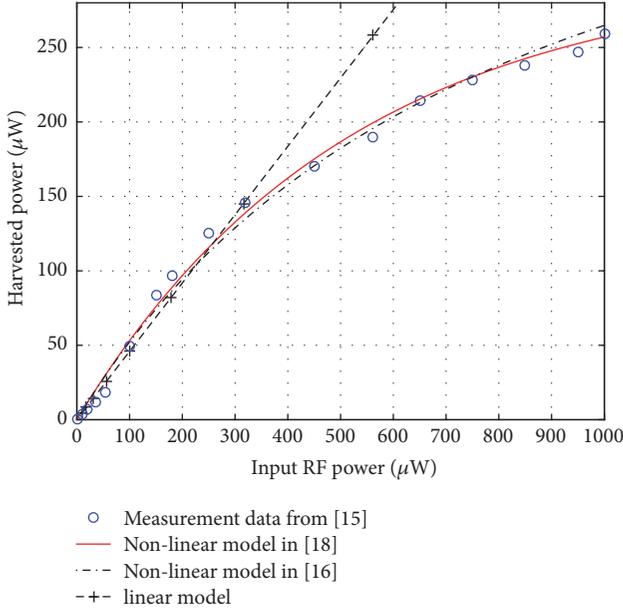


FIGURE 2: Comparison between the harvested powers among the nonlinear EH models in [16, 18], the linear EH model, and the experimental data from [15].

4. Numerical Results and Discussions

In this section, we evaluate the performance of the designed PS schemes based on two nonlinear EH models: the nonlinear EH model based on the logistic function and the nonlinear EH model based on the inverse proportional function, respectively. The basic parameters throughout the simulations, unless otherwise specified, are as follows. We consider the path model to be $|h_{ij}|^2 = |g_{ij}|^2 d_{ij}^{-\alpha_{ij}}$ ($ij = SR, SD$ or RD), where d_{ij} is the $i - j$ distance, $|g_{ij}|^2$ denotes the $i - j$ small-scale fast fading, and α_{ij} is the $i - j$ path loss exponent. Further, we assume that g_{SD} and g_{RD} follow the quasi-static Rayleigh fading with unit variance, and that g_{SR} follows the Rician fading with the Rician factor $3 + \sqrt{12}$. We set $d_{SD} = 20$ meters, $d_{SR} = 2$ meters, $d_{RD} = d_{SD} - d_{SR}$, $\alpha_{SR} = \alpha_{RD} = 3$, and $\alpha_{SD} = 4$. The channel noise power is set to be $\sigma_{sr}^2 = 0.0001$, $\sigma_p^2 = 0.001$, and $\sigma_{rd}^2 = \sigma_{sd}^2 = -40$ dBm, respectively.

Figure 2 compares the experimental data in [15] with the two nonlinear EH models and the conventional linear EH model. The parameters of the nonlinear model based on the logistic function are $a = 1.953 \times 10^{-3}$, $b = -3.571 \times 10^3$, and $E_{\max} = 299.5 \mu\text{W}$. The parameters for the nonlinear EH model based on the inverse proportional function are $a = 2.463$, $b = 1635$, and $c = 826$. For the linear EH model, the conversion efficiency is set to be 0.4591. As shown in this figure, compared with the linear EH model, the nonlinear models in [16, 18] are more accurate when modelling the practical RF-DC circuits. The root mean squared errors of the nonlinear model based on the logistic function, the nonlinear model based on the inverse proportional function, and the linear model are $9.5554 \mu\text{W}$, $11.2464 \mu\text{W}$, and $73.6854 \mu\text{W}$,

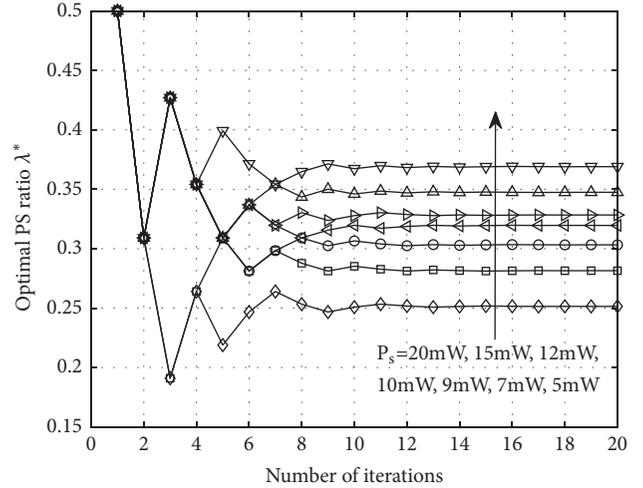


FIGURE 3: Convergence of Algorithm 1 under different sets of P_s .

respectively. The results show that the nonlinear model in [18] is the most accurate among them, and that the nonlinear model in [16] is also able to provide sufficient precision.

Figure 3 demonstrates the convergence of the Algorithm 1 for different sets of the transmit powers. The maximum tolerance ϵ is set to be 0.0001. From this figure, we can see that the optimal PS ratio λ^* always converges to the optimal value within a limited number of iterations. For example, for the case with $P_s = 10$ mW, the optimal PS ratio is obtained when the number of iterations is 11. Another observation is that the optimal PS ratio decreases with the increasing of P_s . The reason is as follows. With a given PS ratio, the input power of the energy harvester increases with the increasing of the transmit power. When the transmit power is large enough, the energy harvester is always saturated and more energy will be wasted. As a result, the optimal PS ratio decreases with the increasing P_s to avoid the waste of energy.

The PS schemes used for comparing are as follows: (1) the proposed PS scheme with the nonlinear EH model based on the logistic function where the optimal PS ratio is obtained by Algorithm 1; (2) the proposed PS scheme with the nonlinear EH model based on the inverse proportional function where the optimal PS ratio is determined by Proposition 1; (3) baseline scheme where the conventional linear EH model is employed. For convenience, we denote the proposed PS scheme with the nonlinear EH model based on the logistic function as LFPS and the proposed PS scheme with the nonlinear EH model based on the inverse proportional function as IPFPS. Since the nonlinear EH model based on the logistic function is closest to the practical EH circuit in [15], we use it to model the practical EH circuit. The optimal PS ratios obtained by LFPS, IPFPS, and baseline scheme will be applied to the practical EH circuit to investigate the performance of LFPS, IPFPS, and baseline schemes under the practical EH circuit. On this basis, the corresponding system capacities can be determined.

Figure 4 presents the overall system capacity achieved by the LFPS, IPFPS, and baseline schemes varying with the

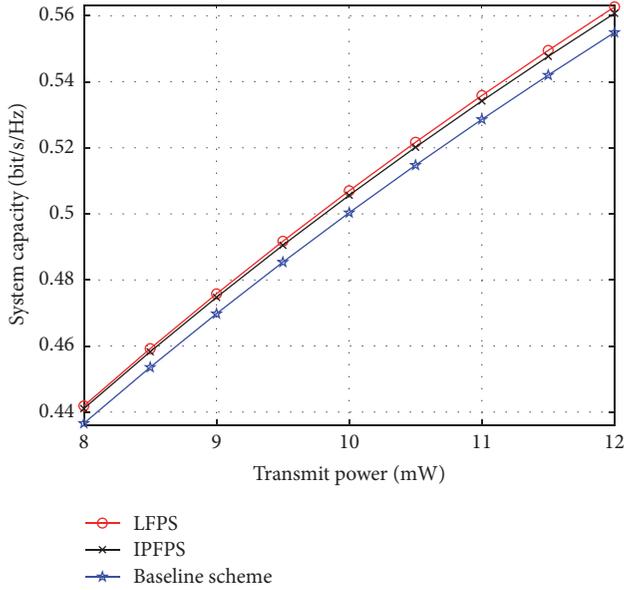


FIGURE 4: System capacity versus the transmit power P_s .

transmit power. It can be observed that, for all considered schemes, the system capacity increases monotonically as we increase the transmit power P_s at the source. Both LFPS and IPFPS schemes outperform the baseline scheme and LFPS can achieve a higher system capacity than IPFPS. This is because of the fact that the mismatches between the EH models and the practical EH circuits will lead to resource allocation mismatches which will bring a poor performance. The more accurate the energy harvesting model, the better the performance the PS scheme can achieve. These results also match the results in Figure 2 perfectly. Besides, we can also see that the performance gain between LFPS and IPFPS or baseline scheme increases with the increasing of P_s . This is due to the fact that, with the increasing of P_s , the gap between the nonlinear EH model based on the inverse proportional function or the linear EH model and the practical energy harvester becomes larger, leading to a worse performance.

5. Conclusion

In this paper, we have developed the optimal PS schemes for an AF relay system with direct link to maximize the system capacity, where both the nonlinear EH model based on the logistic function and the nonlinear EH model based on the inverse proportional function are employed to model the practical energy harvester. Through simulations, three insights were obtained as follows: (1) the PS schemes based on the nonlinear models can achieve superior performance than the one based on the linear model; (2) the PS scheme optimized for the former nonlinear EH model can achieve the best performance, while an extra computational complexity to obtain the optimal solution is incurred; (3) for the case with the latter nonlinear EH model, the PS scheme can achieve near-optimal performance with low computational complexity.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Research Program of Education Bureau of Shaanxi Province (17JK0699), Shaanxi Industrial Science and Technology Key Project (2017GY-039), and 2017 China Textile Industry Federation Science and Technology Guidance Project (2017067).

References

- [1] Z. Chu, F. Zhou, Z. Zhu, R. Q. Hu, and P. Xiao, "Wireless Powered Sensor Networks for Internet of Things: Maximum Throughput and Optimal Power Allocation," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 310–321, Feb. 2018.
- [2] L. Fan, X. Lei, N. Yang, T. Q. Duong, and G. K. Karagiannis, "Secure Multiple Amplify-and-Forward Relaying with Cochannel Interference," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 8, pp. 1494–1505, 2016.
- [3] L. Fan, X. Lei, N. Yang et al., "Secrecy cooperative networks with outdated relay selection over correlated fading channels," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7599–7603, 2017.
- [4] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Relaying protocols for wireless energy harvesting and information processing," *IEEE Transactions on Wireless Communications*, vol. 12, no. 7, pp. 3622–3636, 2013.
- [5] Y. Ye, Y. Li, D. Wang, F. Zhou, R. Q. Hu, and H. Zhang, "Optimal Transmission Schemes for DF Relaying Networks Using SWIPT," *IEEE Transactions on Vehicular Technology*, pp. 1–8, 2018.
- [6] Y. Ye, Y. Li, Z. Wang, X. Chu, and H. Zhang, "Dynamic Asymmetric Power Splitting Scheme for SWIPT-Based Two-Way Multiplicative AF Relaying," *IEEE Signal Processing Letters*, vol. 25, no. 7, pp. 1014–1018, 2018.
- [7] A. A. Nasir, X. Zhou, S. Durrani, and R. A. Kennedy, "Throughput and ergodic capacity of wireless energy harvesting based DF relaying network," in *Proceedings of the IEEE International Conference on Communications (ICC '14)*, pp. 4066–4071, Sydney, Australia, June 2014.
- [8] M. Ju, K.-M. Kang, K.-S. Hwang, and C. Jeong, "Maximum transmission rate of PSR/TSR protocols in wireless energy harvesting DF-based relay networks," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 12, pp. 2701–2717, 2015.
- [9] S. Atapattu and J. Evans, "Optimal energy harvesting protocols for wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 8, pp. 5789–5803, 2016.
- [10] L. Hu, C. Zhang, and Z. Ding, "Dynamic power splitting policies for AF relay networks with wireless energy harvesting," in *Proceedings of the 2015 ICC - 2015 IEEE International Conference on Communications Workshops (ICC)*, pp. 2035–2039, London, UK, June 2015.
- [11] H. Lee, C. Song, S. Choi, and I. Lee, "Outage probability analysis and power splitter designs for swipt relaying systems with direct

- link," *IEEE Communications Letters*, vol. 21, no. 3, pp. 648–651, 2017.
- [12] I. Krikidis, "Simultaneous information and energy transfer in large-scale networks with/without relaying," *IEEE Transactions on Communications*, vol. 62, no. 3, pp. 900–912, 2014.
- [13] Y. Liu, Z. Ding, M. ElKashlan, and H. V. Poor, "Cooperative Non-orthogonal Multiple Access with Simultaneous Wireless Information and Power Transfer," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 938–953, 2016.
- [14] Y. Ye, Y. Li, D. Wang, and G. Lu, "Power splitting protocol design for the cooperative NOMA with SWIPT," in *Proceedings of the IEEE International Conference on Communications*, pp. 1–5, May 2017.
- [15] T. Le, K. Mayaram, and T. Fiez, "Efficient far-field radio frequency energy harvesting for passively powered sensor networks," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 5, pp. 1287–1302, 2008.
- [16] Y. Chen, N. Zhao, and M. Alouini, "Wireless Energy Harvesting Using Signals From Multiple Fading Channels," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 5027–5039, 2017.
- [17] C. R. Valenta and G. D. Durgin, "Harvesting wireless power: survey of energy-harvester conversion efficiency in far-field, wireless power transfer systems," *IEEE Microwave Magazine*, vol. 15, no. 4, pp. 108–120, 2014.
- [18] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical non-linear energy harvesting model and resource allocation for SWIPT systems," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2082–2085, 2015.
- [19] D. Mishra, S. De, and D. Krishnaswamy, "Dilemma at RF Energy Harvesting Relay: Downlink Energy Relaying or Uplink Information Transfer?" *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 4939–4955, Aug, 2017.
- [20] F. Zhou, Z. Chu, H. Sun, R. Q. Hu, and L. Hanzo, "Artificial Noise Aided Secure Cognitive Beamforming for Cooperative MISO-NOMA Using SWIPT," *IEEE Journal on Selected Areas in Communications*, pp. 1-1, 2018.
- [21] L. Shi, L. Zhao, K. Liang, X. Chu, G. Wu, and H. Chen, "Profit maximization in wireless powered communications with improved non-linear energy conversion and storage efficiencies," in *Proceedings of the IEEE International Conference on Communications*, pp. 1–6, May, 2017.
- [22] H. Niu, D. Guo, Y. Huang, and B. Zhang, "Robust Energy Efficiency Optimization for Secure MIMO SWIPT Systems With Non-Linear EH Model," *IEEE Communications Letters*, vol. 21, no. 12, pp. 2610–2613, 2017.
- [23] Q. Yao, T. Q. S. Quek, A. Huang, and H. Shan, "Joint Downlink and Uplink Energy Minimization in WET-Enabled Networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6751–6765, Oct, 2017.
- [24] Y. Wang, Y. Wang, F. Zhou, Y. Wu, and H. Zhou, "Resource allocation in wireless powered cognitive radio networks based on a practical non-linear energy harvesting model," *IEEE Access*, vol. 5, pp. 17618–17626, 2017.
- [25] J.-M. Kang, I.-M. Kim, and D. I. Kim, "Mode Switching for SWIPT over Fading Channel with Nonlinear Energy Harvesting," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 678–681, Oct, 2017.
- [26] S. Wang, M. Xia, K. Huang, and Y. Wu, "Wirelessly Powered Two-Way Communication With Nonlinear Energy Harvesting Model: Rate Regions Under Fixed and Mobile Relay," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8190–8204, 2017.
- [27] T. Wang, G. Lu, Y. Ye, and Y. Ren, "Dynamic Power Splitting Strategy for SWIPT Based Two-Way Multiplicative AF Relay Networks with Nonlinear Energy Harvesting Model," *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 1802063, 9 pages, 2018.
- [28] A. Cvetkovic and V. Blagojevic, "Performance Analysis of Nonlinear Energy-Harvesting DF Relay System in Interference-Limited Nakagami-m Fading Environment," *ETRI Journal*, vol. 39, no. 6, pp. 803–812, Dec, 2017.
- [29] J. Zhang and G. Pan, "Outage Analysis of Wireless-Powered Relaying MIMO Systems with Non-Linear Energy Harvesters and Imperfect CSI," *IEEE Access*, vol. 4, pp. 7046–7053, 2016.
- [30] K. Wang, Y. Li, Y. Ye, and H. Zhang, "Dynamic Power Splitting Schemes for Non-Linear EH Relaying Networks: Perfect and Imperfect CSI," in *Proceedings of the IEEE 86th Vehicular Technology Conference*, pp. 1–5, Sept, 2017.
- [31] L. Wei, R. Q. Hu, Y. Qian, and G. Wu, "Energy efficiency and spectrum efficiency of multihop device-to-device communications underlying cellular networks," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 1, pp. 367–380, 2016.

Research Article

Security-Reliability Tradeoff Analysis in Multisource Multirelay Cooperative Networks with Multiple Cochannel Interferers

Weilong Hu ¹, Jiangbo Si ², and Hongyan Li¹

¹State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, Shaanxi 710071, China

²Collaborative Innovation Center of Information Sensing and Understanding, Xidian University, Xi'an, Shaanxi 710071, China

Correspondence should be addressed to Jiangbo Si; jbsi@xidian.edu.cn

Received 3 January 2018; Accepted 14 February 2018; Published 28 June 2018

Academic Editor: Zheng Chu

Copyright © 2018 Weilong Hu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Cooperative relaying communication is one of the green communication technologies since it shortens the communication distance and saves the transmit power. In this paper, the physical-layer security (PLS) of a multisource multirelay cooperative relaying communication network is investigated by considering the influence of cochannel interference from a security-reliability tradeoff (SRT) perspective. First, the SRT performance is characterized by the outage probability (OP) and the intercept probability (IP). In particular, the IP encountered at the eavesdropper is used to evaluate the security performance, while the reliability performance is analyzed in terms of the OP experienced at the destination. Then, under the impact of multiple cochannel interferers, the intercept probabilities and the outage probabilities of both the conventional direct transmission (DT) strategy and relay selection (RS) strategy are derived in closed-form expressions over Rayleigh fading channels, respectively. Simulation results are provided to validate the theoretical analysis. It is shown that when the OP (reliability) requirement is relaxed, the IP (security) performance improves and vice versa. It confirms that there is an SRT existing between the OP and the IP. Meanwhile, a better SRT performance can be achieved by increasing the number of sources, relays, and cochannel interferers. In addition, it is also shown that the RS strategy generally outperforms the conventional DT strategy in terms of the product of the IP and the OP.

1. Introduction

With an explosive growth of the number of wireless devices, such as smart phones, tablet computers, and wireless sensors, more and more energy has been consumed by wireless services. According to [1], the amount of the energy consumed by the information and communication technologies accounts for about 2% to 10% of the global energy consumption, which generate nonnegligible amount of the greenhouse gases. What is worst, this percentage will grow rapidly with the development of the information and communication technologies and the increase of number of wireless devices. It will result in more greenhouse gases emission and environment pollution [2]. One promising technique to alleviate such issue is to adopt green communication technology, which can improve both the spectrum efficiency and energy efficiency of wireless communication systems. Cooperative relaying communication is an energy efficient diversity technique, which has been recognized as a green communication

technology and attracts unprecedented research interest in both academic and industrial fields. A challenging issue in cooperative relaying communication is wireless security [3–5]. Because of the inherent broadcast nature of wireless channels, the destination may not successfully obtain source information, while the malicious eavesdropper may overhear and intercept the confidential information, which makes the wireless transmission insecure and vulnerable to eavesdropping attacks [6].

Motivated by the above fact, physical-layer security (PLS) was proposed and has attracted increasing research attention since it is an effective paradigm of achieving information-theoretic security for protecting wireless communications against the eavesdropping attacks by utilizing the physical characteristics of wireless channels [7]. The PLS was first investigated by Shannon [8] and further developed by Wyner, who examined a classical point-to-point discrete memoryless wiretap channel (WTC) scenario consisting of a source node and a destination node as well as an eavesdropper node [9]. It

was proven in [9] that perfectly secure and reliable transmission from the source node to the legitimate destination node can be achieved when the main channel from the source node to the destination node is an upgraded version of the wiretap channel from the source node to the eavesdropper node. Later on, in [10, 11], Wyner's conclusions were, respectively, extended from the discrete memoryless wiretap channel to the nondegraded wiretap channel and the Gaussian degraded wiretap channel, where the notion of secrecy capacity (SC) was introduced. It was derived as the difference between the channel capacity of the legitimate link and that of the wiretap link. Specifically, the SC can make the transmission from the source node to the legitimate destination node secure while achieving zero mutual information between the source node and malicious eavesdropper node. Based on this idea, extensive research efforts have been devoted to improving the SC from an information-theoretic perspective under different scenarios, for example, cooperative relaying [12–15] and beamforming techniques [16, 17], cooperative jamming (CJ) methods [18–20], and multiple-input multiple-output (MIMO) schemes [21, 22].

The previous works are mainly focused on enhancing wireless security without paying much attention to communication reliability. Hence, security-reliability tradeoff (SRT) was proposed to make best tradeoff between the outage probability (OP) and the intercept probability (IP). In particular, the IP encountered at the eavesdropper is used to evaluate the security performance, while the reliability performance is measured by the OP experienced at the destination. In [23], the authors studied the employment of various block cipher encryption algorithms from the perspective of both reliability and security and showed that there exists a tradeoff between communication reliability and security. Later on, the authors of [24] investigated the SRT for the downlink cloud radio access networks and the channel estimation errors were considered and the impact of the times of training on the security and reliability performance was also analyzed. The SRT of the cognitive amplify-and-forward (AF) relay network was investigated under imperfect channel estimation in [25]. As a further development, the authors of [26] characterized the SRT and quantified the benefits of opportunistic relay selection (ORS) for the purpose of improving the SRT. In [27], the authors proposed the single-relay and multirelay selection schemes for improving the SRT of general wireless networks. It was proved in [27] that in terms of the SRT the multirelay selection scheme outperformed the single-relay one.

It can be seen from the above works that the PLS of multisource multirelay cooperative networks under the impact of the cochannel interferers is not considered. Motivated by this fact, the main contributions of this paper are summarized as follows: firstly, the PLS of a multisource multirelay cooperative communication network is investigated from an SRT perspective and cochannel interferers are considered. Secondly, a signal-to-interference-plus-noise ratio- (SINR-) based method is proposed and the closed-form expressions of IP and OP are derived for the direct transmission (DT) and the relay selection (RS) schemes over Rayleigh fading, respectively. Finally, simulation results are provided to validate the theoretical analysis. It is shown

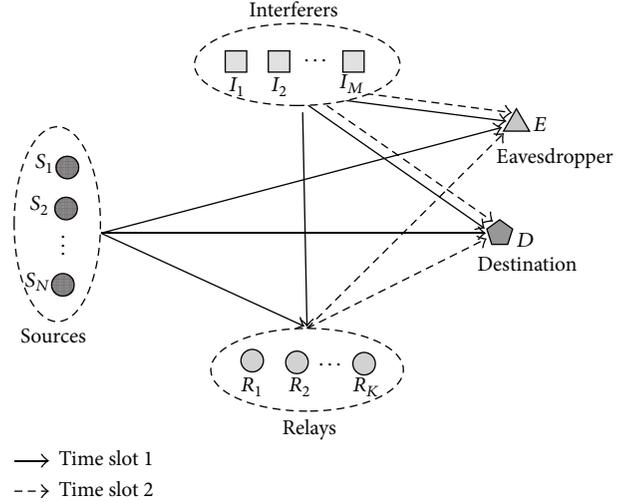


FIGURE 1: Multisource multirelay cooperative network system model.

that when the OP (reliability) requirement is relaxed, the IP (security) performance improves and vice versa. It confirms that there is an SRT existing between the OP and the IP. Meanwhile, a better SRT performance can be achieved by increasing the number of sources, relays, and cochannel interferers. In addition, it is also shown that the RS strategy generally outperforms the conventional DT strategy in terms of the product of the IP and the OP.

The remainder of this paper is organized as follows. In Section 2, the system models are described. The SRT performance analysis for both the conventional DT and RS schemes over Rayleigh fading channels is presented in Section 3. Section 4 presents simulation results to corroborate the proposed studies. Section 5 concludes the paper.

2. The System Model

2.1. System Model Description. Consider a multisource multirelay cooperative wireless network as shown in Figure 1, which consists of N sources S_n ($1 \leq n \leq N$), one eavesdropper E , one destination D , K relays R_k ($1 \leq k \leq K$), and M cochannel interferers I_m ($1 \leq m \leq M$). The sources communicate with the corresponding destination via the direct link or with the help of the intermediate relays. At a specific time, only the source having the highest direct-link channel quality is viewed as the best one and is selected to transmit with the aid of relays. Meanwhile, E will intercept the information from the selected source and relays. M interferers share the same bands with D and E and cause interferences to them. It can be observed that the system model is practical and can be applied to practical scenarios [28–30]. It is assumed that all nodes are equipped with single antenna and all channels are Rayleigh fading. Without loss of generality, we consider additive white Gaussian noise (AWGN) with zero mean and variance N_0 at each node in networks. We assume that the sources have the global channel state information (CSI) of both the main and wiretap

channels and in order to analyze the performance of the worst case, E is assumed to know all system parameters of the legitimate transmission from S to D, except for the signal. Typically, the linear minimum mean-square error (LMMSE) estimation method can be used to obtain the CSI by the destination and the eavesdropper [24, 25]. Note that this assumption has been widely used in [26, 27, 31].

2.2. Direct Transmission Strategy. In this subsection, the conventional DT strategy is considered for the purpose of performance comparison. A classical DT communication scenario consisting of N sources, one destination, and one eavesdropper with M cochannel interferers is considered. Assuming that all sources send messages at a power P_S while the interferers transmit at a power P_I . Let x_b ($E[|x_b|^2] = 1$) and x_m ($E[|x_m|^2] = 1$), respectively, denote the source signal from the selected best source S_b and the interfering signal transmitted by the m th interferer I_m . When S_b transmits x_b with the rate R_s at a particular time instant, I_m transmits x_m with the rate R_s . Hence, under the presence of M cochannel interferers, the signals received at D and E nodes can be, respectively, presented as

$$y_{S_b D}^{\text{DT}} = \sqrt{P_S} h_{S_b D} x_b + \sum_{m=1}^M \sqrt{P_I} h_{I_m D} x_m + n_D, \quad (1)$$

$$y_{S_b E}^{\text{DT}} = \sqrt{P_S} h_{S_b E} x_b + \sum_{m=1}^M \sqrt{P_I} h_{I_m E} x_m + n_E, \quad (2)$$

where $h_{S_b D}$, $h_{S_b E}$, $h_{I_m D}$, and $h_{I_m E}$, respectively, denote the fading gains of the channel from S_b to D, that from S_b to E, that from I_m to D, and that from I_m to E. n_D and n_E represent the AWGN encountered at D and E nodes, respectively. Using Shannon's capacity formula, the capacity of the channel spanning from S_b to D is given by

$$C_{S_b D}^{\text{DT}} = \log_2 \left(1 + \frac{\gamma_S |h_{S_b D}|^2}{\sum_{m=1}^M \gamma_I |h_{I_m D}|^2 + 1} \right), \quad (3)$$

where $\gamma_S = P_S/N_0$ and $\gamma_I = P_I/N_0$. Similarly, the channel capacity of S_b -E transmission is obtained from (2) as

$$C_{S_b E}^{\text{DT}} = \log_2 \left(1 + \frac{\gamma_S |h_{S_b E}|^2}{\sum_{m=1}^M \gamma_I |h_{I_m E}|^2 + 1} \right). \quad (4)$$

Since fading gains $h_{S_b D}$, $h_{S_b E}$, $h_{I_m D}$, and $h_{I_m E}$ are modeled as Rayleigh random variables, then $|h_{S_b D}|^2$, $|h_{S_b E}|^2$, $|h_{I_m D}|^2$, and $|h_{I_m E}|^2$ are exponentially distributed. Accordingly, $\sigma_{S_b D}^2$, $\sigma_{S_b E}^2$, $\sigma_{I_m D}^2$, and $\sigma_{I_m E}^2$ represent the means of $|h_{S_b D}|^2$, $|h_{S_b E}|^2$, $|h_{I_m D}|^2$, and $|h_{I_m E}|^2$, respectively.

2.3. Relay Selection Strategy. As shown in Figure 1, this subsection presents a multisource multirelay cooperative wireless network with multiple cochannel interferers existing at relays, D and E. Specifically, all sources share the relay nodes and the relays employ the decode-and-forward (DF)

relaying protocol. Without loss of generality, the total cooperative communication procedure is divided into two time slots. It can also be seen from Figure 1 that the solid and dash lines represent the transmission in the first time slot and that in the second time slot, respectively. In the first time slot, the selected best source node S_b transmits its signals x_b to D and all relays, and meanwhile E intercepts the transmission of the source. Under the presence of M cochannel interferers, the signals received at D, E, and R_k nodes can be, respectively, presented as

$$y_{S_b D}^{\text{RS}} = \sqrt{P_S} h_{S_b D} x_b + \sum_{m=1}^M \sqrt{P_I} h_{I_m D} x_m + n_D, \quad (5)$$

$$y_{S_b E}^{\text{RS}} = \sqrt{P_S} h_{S_b E} x_b + \sum_{m=1}^M \sqrt{P_I} h_{I_m E} x_m + n_E, \quad (6)$$

$$y_{S_b R_k}^{\text{RS}} = \sqrt{P_S} h_{S_b R_k} x_b + \sum_{m=1}^M \sqrt{P_I} h_{I_m R_k} x_m + n_{R_k}, \quad (7)$$

where $h_{S_b R_k}$ and $h_{I_m R_k}$, respectively, denote the fading gains of the channel from S_b to R_k and that from I_m to R_k . n_{R_k} represents the AWGN encountered at R_k node. Similar to [7], according to (5), (6), and (7), the capacities of the channel spanning from S_b to D, that spanning from S_b to R_k , and that spanning from S_b to E can, respectively, be obtained as

$$C_{S_b D}^{\text{RS}} = \frac{1}{2} \log_2 \left(1 + \frac{\gamma_S |h_{S_b D}|^2}{\sum_{m=1}^M \gamma_I |h_{I_m D}|^2 + 1} \right), \quad (8)$$

$$C_{S_b R_k}^{\text{RS}} = \frac{1}{2} \log_2 \left(1 + \frac{\gamma_S |h_{S_b R_k}|^2}{\sum_{m=1}^M \gamma_I |h_{I_m R_k}|^2 + 1} \right), \quad (9)$$

$$C_{S_b E}^{\text{RS}} = \frac{1}{2} \log_2 \left(1 + \frac{\gamma_S |h_{S_b E}|^2}{\sum_{m=1}^M \gamma_I |h_{I_m E}|^2 + 1} \right), \quad (10)$$

where $1/2$ arises from the fact that two orthogonal slots are needed for completing the overall transmission. Similarly, $|h_{S_b R_k}|^2$ and $|h_{I_m R_k}|^2$ are exponentially distributed and accordingly, and $\sigma_{S_b R_k}^2$ and $\sigma_{I_m R_k}^2$ represent the means of $|h_{S_b R_k}|^2$ and $|h_{I_m R_k}|^2$, respectively.

According to the DF protocol, only those relays that succeed in perfectly decoding the source signal x_b form a decoding set denoted by \mathcal{D} . Thus, in the second time slot, when \mathcal{D} is a nonempty set, a specific relay is chosen from \mathcal{D} for forwarding its received signal to D node with P_S denoting the transmit power. In particular, R_k is regarded as the selected relay node. Then, the signal received at D node can be presented as

$$y_{R_k D}^{\text{RS}} = \sqrt{P_S} h_{R_k D} x_b + \sum_{m=1}^M \sqrt{P_I} h_{I_m D} x_m + n_D, \quad (11)$$

where $h_{R_k D}$ represents the fading gain of the channel from R_k to D. Similarly, according to (11), the capacity of the channel spanning from R_k to D is given by

$$C_{R_k D}^{\text{RS}} = \frac{1}{2} \log_2 \left(1 + \frac{\gamma_S |h_{R_k D}|^2}{\sum_{m=1}^M \gamma_I |h_{I_m D}|^2 + 1} \right), \quad (12)$$

where $|h_{R_k D}|^2$ is exponentially distributed and $\sigma_{R_k D}^2$ represents the mean of $|h_{R_k D}|^2$. Based on the obtained capacity of the channel spanning from R_k to D, the selected relay node with the largest channel capacity is chosen from \mathcal{D} , that is,

$$R_{\text{best}} = \arg \max_{R_k \in \mathcal{D}} C_{R_k D}^{\text{RS}}, \quad (13)$$

where R_{best} represents the selected best relay node. It can be seen from (12) that the interferers and noise terms are same for the channel capacities of different relays. Then, (13) is simplified as $R_{\text{best}} = \arg \max_{R_k \in \mathcal{D}} |h_{R_k D}|^2$. Thus, the signal received at E node with the best relay node can be presented as

$$y_{R_{\text{best}} E}^{\text{RS}} = \sqrt{P_S} h_{R_{\text{best}} E} x_b + \sum_{m=1}^M \sqrt{P_{I_m}} h_{I_m E} x_m + n_E, \quad (14)$$

where $h_{R_{\text{best}} E}$ denotes the fading gain of the channel from R_{best} to E. Thus, according to (14), the capacity of the channel spanning from R_{best} to E is given by

$$C_{R_{\text{best}} E}^{\text{RS}} = \frac{1}{2} \log_2 \left(1 + \frac{\gamma_S |h_{R_{\text{best}} E}|^2}{\sum_{m=1}^M \gamma_I |h_{I_m E}|^2 + 1} \right), \quad (15)$$

where $|h_{R_{\text{best}} E}|^2$ is exponentially distributed and $\sigma_{R_{\text{best}} E}^2$ represents the mean of $|h_{R_{\text{best}} E}|^2$.

Across this paper, for ease of discussion, we assume that $\sigma_{S_b D}^2 = \sigma_{S_b R_k}^2 = \sigma_{R_k D}^2 = \sigma_S^2$, $\sigma_{S_b E}^2 = \sigma_{R_{\text{best}} E}^2 = \sigma_E^2$ and $\sigma_{I_m R_k}^2 = \sigma_{I_m D}^2 = \sigma_{I_m E}^2 = \sigma_I^2$. This assumption can be valid in a statistical sense when all relays are mobile and uniformly distributed around S and D nodes [7].

3. SRT Performance Analysis over Rayleigh Fading Channels

In this section, the SRT performance analysis of the conventional DT strategy as well as of the RS strategy with the presence of multiple cochannel interferers communicating over Rayleigh fading channels is presented. As discussed in [7], the tradeoff between the security and reliability, characterized by the intercept probability and by the outage probability, is analyzed. In particular, the outage probability represents the probability that the capacity of the main channel is lower than the data rate and the intercept probability represents the probability that the capacity of the wiretap channel is higher than the data rate. Then, the two performance metrics can be expressed as

$$P_{\text{out}} = \Pr(C_d < R_s), \quad (16)$$

$$P_{\text{int}} = \Pr(C_e > R_s), \quad (17)$$

where C_d and C_e represent, respectively, the capacity of the main channel achieved at the destination and that of the wiretap channel experienced by the eavesdropper. R_s is the data rate.

3.1. Direct Transmission Strategy. In what follows, the SRT performance of the conventional DT strategy is first analyzed as a benchmark. According to (16), using the law of total probability, the OP of the conventional DT strategy can be formulated as

$$P_{\text{out}}^{\text{DT}} = \sum_{b=1}^N \Pr \left(|h_{S_b D}|^2 > \max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2), C_{S_b D}^{\text{DT}} < R_s \right), \quad (18)$$

where $C_{S_b D}^{\text{DT}}$ is given by (3). Substituting $C_{S_b D}^{\text{DT}}$ into (18), the OP is given by

$$P_{\text{out}}^{\text{DT}} = \sum_{b=1}^N \Pr \left(\underbrace{|h_{S_b D}|^2 > \max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2)}_{P_1}, \frac{\gamma_S |h_{S_b D}|^2}{Y + 1} < \Delta \right), \quad (19)$$

where $Y = \sum_{m=1}^M \gamma_I |h_{I_m D}|^2$ and $\Delta = (2^{R_s} - 1)$. Note that, due to the common term $|h_{S_b D}|^2$, P_1 cannot be calculated as the conventional analysis directly. Hence, upon assuming $|h_{S_b D}|^2 = x$, P_1 can be expressed as

$$P_1 = \int_0^{\infty} \Pr \left(\max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2) < x \mid x \right) \times \Pr \left(\frac{\gamma_S x}{Y + 1} < \Delta \mid x \right) f_{|h_{S_b D}|^2}(x) dx, \quad (20)$$

in which the first term can be obtained as

$$\begin{aligned} & \Pr \left(\max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2) < x \mid x \right) \\ &= \prod_{1 \leq n \leq N, n \neq b} \Pr \left(|h_{S_n D}|^2 < x \mid x \right) \\ &= \sum_{n=0}^{N-1} C_{N-1}^n (-1)^n \exp \left(-\frac{nx}{\sigma_{S_n D}^2} \right). \end{aligned} \quad (21)$$

In (20), it can be found that, for $x < \Delta/\gamma_S$, the second-term $\Pr(\gamma_S x/(Y + 1) < \Delta \mid x) = 1$. Therefore, there are two cases for the term $\Pr(\gamma_S x/(Y + 1) < \Delta \mid x)$; that is,

$$\begin{aligned} & \Pr \left(\frac{\gamma_S x}{Y + 1} < \Delta \mid x \right) \\ &= \begin{cases} 1 & x < \frac{\Delta}{\gamma_S} \\ \exp \left(\frac{\Delta - \gamma_S x}{\gamma_I \sigma_{I_m D}^2 \Delta} \right) \sum_{m_1=0}^{M-1} \frac{((\gamma_S x - \Delta)/\gamma_I \sigma_{I_m D}^2 \Delta)^{m_1}}{m_1!} & x > \frac{\Delta}{\gamma_S} \end{cases} \quad (22) \end{aligned}$$

Then substituting (21) and (22) into (20) yields

$$\begin{aligned}
P_1 = & \left[\int_0^{\Delta/\gamma_S} \sum_{n=0}^{N-1} C_{N-1}^n (-1)^n \exp\left(-\frac{nx}{\sigma_{S_n D}^2}\right) \right. \\
& + \int_{\Delta/\gamma_S}^{\infty} \sum_{n=0}^{N-1} C_{N-1}^n (-1)^n \exp\left(-\frac{nx}{\sigma_{S_n D}^2}\right) \\
& \times \exp\left(\frac{1}{\gamma_I \sigma_{I_m D}^2} - \frac{\gamma_S x}{\gamma_I \sigma_{I_m D}^2 \Delta}\right) \\
& \left. \times \sum_{m_1=0}^{M-1} \frac{(\gamma_S x / \gamma_I \sigma_{I_m D}^2 \Delta - 1 / \gamma_I \sigma_{I_m D}^2)^{m_1}}{m_1!} \right] \\
& \cdot f_{|h_{S_b D}|^2}(x) dx.
\end{aligned} \tag{23}$$

Proceeding as in Appendix A, one has

$$\begin{aligned}
P_1 = & \sum_{n=0}^{N-1} C_{N-1}^n (-1)^n \left[\frac{1 - \exp\left(-\frac{(n+1)\Delta/\sigma_S^2 \gamma_S}{n+1}\right)}{n+1} \right. \\
& + \sum_{m_1=0}^{M-1} \exp\left(-\frac{(n+1)\Delta}{\gamma_S \sigma_S^2}\right) \\
& \left. \cdot \frac{\gamma_I \sigma_I^2 \Delta}{\sigma_S^2 \gamma_S \left((n+1) \gamma_I \sigma_I^2 \Delta / \sigma_S^2 \gamma_S + 1\right)^{m_1+1}} \right].
\end{aligned} \tag{24}$$

Similarly, according to (4) and (17), the IP of the conventional DT strategy is formulated as

$$\begin{aligned}
P_{\text{int}}^{\text{DT}} & \\
= & \sum_{b=1}^N \Pr\left(|h_{S_b D}|^2 > \max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2), C_{S_b E}^{\text{DT}} > R_s\right).
\end{aligned} \tag{25}$$

With the aid of $C_{S_b E}^{\text{DT}}$, the IP can be expressed as

$$\begin{aligned}
P_{\text{int}}^{\text{DT}} & \\
= & \sum_{b=1}^N \underbrace{\Pr\left(|h_{S_b D}|^2 > \max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2), \frac{\gamma_S |h_{S_b E}|^2}{T+1} > \Delta\right)}_{P_2},
\end{aligned} \tag{26}$$

where $T = \sum_{m=1}^M \gamma_I |h_{I_m E}|^2$. Similar to the analysis of P_1 , P_2 can be rewritten as

$$\begin{aligned}
P_2 = & \int_0^{\infty} \Pr\left(\max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2) < x \mid x\right) \\
& \times \Pr\left(\frac{\gamma_S |h_{S_b E}|^2}{T+1} > \Delta \mid x\right) f_{|h_{S_b D}|^2}(x) dx,
\end{aligned} \tag{27}$$

where the term $\Pr(\gamma_S |h_{S_b E}|^2 / (T+1) > \Delta \mid x)$ can be readily derived as

$$\begin{aligned}
& \Pr\left(\frac{\gamma_S |h_{S_b E}|^2}{T+1} > \Delta \mid x\right) \\
& = \exp\left(-\frac{\Delta}{\gamma_S \sigma_{S_b E}^2}\right) \frac{(\sigma_{S_b E}^2)^M}{(\sigma_{S_b E}^2 + \gamma_I \sigma_{I_m E}^2 \Delta / \gamma_S)^M}.
\end{aligned} \tag{28}$$

Then substituting (21) and (28) into (27), one has

$$\begin{aligned}
P_2 = & \int_0^{\infty} \sum_{n=0}^{N-1} C_{N-1}^n (-1)^n \exp\left(-\frac{nx}{\sigma_{S_n D}^2}\right) \\
& \times \exp\left(-\frac{\Delta}{\gamma_S \sigma_{S_b E}^2}\right) \frac{(\sigma_{S_b E}^2)^M}{(\sigma_{S_b E}^2 + \gamma_I \sigma_{I_m E}^2 \theta / \gamma_S)^M} \\
& \times f_{|h_{S_b D}|^2}(x) dx.
\end{aligned} \tag{29}$$

After some appropriate incorporations and necessary mathematical manipulations, P_2 can be obtained as

$$P_2 = \sum_{n=0}^{N-1} \frac{C_{N-1}^n (-1)^n (\sigma_E^2)^M}{(n+1) (\sigma_E^2 + \gamma_I \sigma_I^2 \Delta / \gamma_S)^M} \exp\left(-\frac{\Delta}{\gamma_S \sigma_E^2}\right). \tag{30}$$

3.2. Relay Selection Strategy. This subsection focuses on the SRT performance analysis of the RS strategy. According to (16) and using the theory of total probability, the OP of the RS strategy is formulated as

$$P_{\text{out}}^{\text{RS}} = \sum_{b=1}^N \sum_{k=0}^K \sum_{\mathcal{D} \subset \mathbb{R}} \Pr(|\mathcal{D}| = k, C_D^{\text{RS}} < R_s), \tag{31}$$

where $|\mathcal{D}|$ represents the number of elements in successful decoding set \mathcal{D} and $\mathbb{R} = \{R_1, R_2, \dots, R_K\}$ denotes the relay set. As can be observed, when \mathcal{D} is an empty set (i.e., $|\mathcal{D}| = 0$), it shows that no relay can be chosen for forwarding the received signal. In this case, only the direct link is available, that is, $C_D^{\text{RS}} = C_{S_b D}^{\text{RS}}$, where \mathcal{D} is a nonempty set and selection combining is considered to combine the received signal copies at D from the selected best source S_b and the selected best relay R_{best} during the two time slots. In this case, the capacity achieved by D is the higher one between $C_{S_b D}^{\text{RS}}$ and $C_{R_{\text{best}} D}^{\text{RS}}$, that is, $C_D^{\text{RS}} = \max(C_{S_b D}^{\text{RS}}, C_{R_{\text{best}} D}^{\text{RS}})$. Substituting these results into (31), one has

$$P_{\text{out}}^{\text{RS}} = \sum_{b=1}^N \left[\Pr(\mathcal{D} = \emptyset) P_3 + \sum_{k=1}^K \sum_{\mathcal{D} \subset \mathbb{R}} \Pr(|\mathcal{D}| = k) P_4 \right], \tag{32}$$

where the terms P_3 and P_4 are, respectively, given by

$$P_3 = \Pr\left(|h_{S_b D}|^2 > \max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2), C_{S_b D}^{\text{RS}} < R_s\right), \tag{33}$$

$$P_4 = \Pr\left(|h_{S_b D}|^2 > \max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2), C_D^{\text{RS}} < R_s\right). \tag{34}$$

In (32), the term $\Pr(|\mathcal{D}| = k)$ denotes the probability that there exist k relays decoding the source signal x_s successfully. Thus, considering $\sigma_{S_b R_i}^2 = \sigma_{S_b R_j}^2 = \sigma_S^2$ and $\sigma_{I_m R_i}^2 = \sigma_{I_m R_j}^2 = \sigma_I^2$, one has

$$\begin{aligned} \Pr(|\mathcal{D}| = k) &= \prod_{R_i \in \mathcal{D}} \Pr(C_{S_b R_i}^{\text{RS}} > R_s) \times \prod_{R_j \in \overline{\mathcal{D}}} \Pr(C_{S_b R_j}^{\text{RS}} < R_s) \\ &= \left[\exp\left(-\frac{\theta}{\gamma_S \sigma_S^2}\right) \frac{(\sigma_S^2)^M}{(\sigma_S^2 + \gamma_I \sigma_I^2 \theta / \gamma_S)^M} \right]^k \\ &\quad \times \left[1 - \exp\left(-\frac{\theta}{\gamma_S \sigma_S^2}\right) \frac{(\sigma_S^2)^M}{(\sigma_S^2 + \gamma_I \sigma_I^2 \theta / \gamma_S)^M} \right]^{K-k}, \end{aligned} \quad (35)$$

where $\theta = (2^{2R_s} - 1)$ and $\overline{\mathcal{D}}$ is the complementary set of the successful decoding set \mathcal{D} . By utilizing a similar way, the term $\Pr(\mathcal{D} = \emptyset)$ is calculated as

$$\begin{aligned} \Pr(\mathcal{D} = \emptyset) &= \prod_{R_k \in \mathbb{R}} \Pr(C_{S_b R_k}^{\text{RS}} < R_s) \\ &= \left[1 - \exp\left(-\frac{\theta}{\gamma_S \sigma_S^2}\right) \frac{(\sigma_S^2)^M}{(\sigma_S^2 + \gamma_I \sigma_I^2 \theta / \gamma_S)^M} \right]^K. \end{aligned} \quad (36)$$

As discussed before, by utilizing a similar way as P_1, P_3 can be obtained as

$$\begin{aligned} P_3 &= \sum_{n=0}^{N-1} C_{N-1}^n (-1)^n \left[\frac{1 - \exp(- (n+1) \theta / \sigma_S^2 \gamma_S)}{n+1} \right. \\ &\quad \left. + \sum_{m_1=0}^{M-1} \exp\left(-\frac{(n+1)\theta}{\gamma_S \sigma_S^2}\right) \right. \\ &\quad \left. \times \frac{\gamma_I \sigma_I^2 \theta}{\sigma_S^2 \gamma_S ((n+1) \gamma_I \sigma_I^2 \theta / \sigma_S^2 \gamma_S + 1)^{m_1+1}} \right]. \end{aligned} \quad (37)$$

Similar to (20), the term P_4 is expressed as

$$\begin{aligned} P_4 &= \int_0^\infty \Pr\left(\max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2) < x \mid x\right) \\ &\quad \times \Pr\left(\max\left(\frac{\gamma_S x}{Y+1}, \frac{\gamma_S |h_{R_{\text{best}} D}|^2}{Y+1}\right) < \theta \mid x\right) \\ &\quad \times f_{|h_{S_b D}|^2}(x) dx. \end{aligned} \quad (38)$$

Proceeding as in Appendix B, P_4 can be obtained as

$$\begin{aligned} P_4 &= \sum_{n=0}^{N-1} \sum_{k_1=0}^{|\mathcal{D}|} C_{N-1}^n (-1)^n C_{|\mathcal{D}|}^{k_1} (-1)^{k_1} \exp\left(-\frac{k_1 \theta}{\sigma_S^2 \gamma_S}\right) \\ &\quad \times \left[\frac{1 - \exp(- (n+1) \theta / \sigma_S^2 \gamma_S)}{(n+1) (1 + k_1 \theta \sigma_I^2 \gamma_I / \sigma_S^2 \gamma_S)^M} \right. \\ &\quad \left. + \sum_{m_1=0}^{M-1} \exp\left(-\frac{(n+1)\theta}{\sigma_S^2 \gamma_S}\right) \right. \\ &\quad \left. \times \frac{\gamma_I \sigma_I^2 \theta (1 + k_1 \theta \sigma_I^2 \gamma_I / \sigma_S^2 \gamma_S)^{m_1-M}}{\sigma_S^2 \gamma_S (1 + k_1 \gamma_I \sigma_I^2 \theta / \sigma_S^2 \gamma_S + (n+1) \gamma_I \sigma_I^2 \theta / \sigma_S^2 \gamma_S)^{m_1+1}} \right]. \end{aligned} \quad (39)$$

On the other hand, similarly, according to (17) and using the theory of total probability, the IP of the RS strategy is formulated as

$$P_{\text{int}}^{\text{RS}} = \sum_{b=1}^N \sum_{k=0}^K \sum_{\mathcal{D} \subset \mathbb{R}} \Pr(|\mathcal{D}| = k, C_E^{\text{RS}} > R_s). \quad (40)$$

During the two slots, note that E will intercept the message transmitted by both the selected best source S_b and the selected best relay R_{best} and perform detection using both received signal copies. Thus, (40) can be rewritten as

$$P_{\text{int}}^{\text{RS}} = \sum_{b=1}^N \left[\Pr(\mathcal{D} = \emptyset) P_5 + \sum_{k=1}^K \sum_{\mathcal{D} \subset \mathbb{R}} \Pr(|\mathcal{D}| = k) P_6 \right] \quad (41)$$

in which the terms P_5 and P_6 are, respectively, given by

$$P_5 = \Pr\left(|h_{S_b D}|^2 > \max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2), C_{S_b E}^{\text{RS}} > R_s\right), \quad (42)$$

$$P_6 = \Pr\left(|h_{S_b D}|^2 > \max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2), C_E^{\text{RS}} > R_s\right). \quad (43)$$

As discussed before, similar to P_2, P_5 can be obtained as

$$P_5 = \sum_{n=0}^{N-1} \frac{C_{N-1}^n (-1)^n (\sigma_E^2)^M}{(n+1) (\sigma_E^2 + \gamma_I \sigma_I^2 \theta / \gamma_S)^M} \exp\left(-\frac{\theta}{\gamma_S \sigma_E^2}\right). \quad (44)$$

Similar to (38), the term P_6 can be rewritten as

$$\begin{aligned} P_6 &= \int_0^\infty \Pr\left(\max_{1 \leq n \leq N, n \neq b} (|h_{S_n D}|^2) < x \mid x\right) \\ &\quad \times \Pr\left(\max\left(\frac{\gamma_S |h_{S_b E}|^2}{T+1}, \frac{\gamma_S |h_{R_{\text{best}} E}|^2}{T+1}\right) > \theta \mid x\right) \\ &\quad \times f_{|h_{S_b D}|^2}(x) dx. \end{aligned} \quad (45)$$

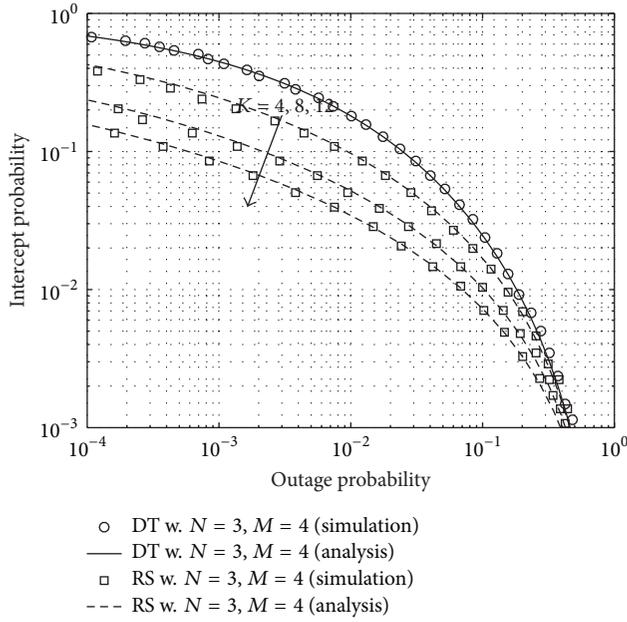


FIGURE 2: SRTs of the conventional DT strategy and RS strategy for different numbers of relays K with $\gamma_I = 15$ dB, $N = 3$, and $M = 4$.

After some appropriate substitutions and via utilizing a similar derivation for P_4 , P_6 can be obtained as

$$P_6 = \sum_{n=0}^{N-1} \frac{C_{N-1}^n (-1)^n}{(1+n)} \times \left[2 \exp\left(-\frac{\theta}{\gamma_S \sigma_E^2}\right) \frac{(\sigma_E^2)^M}{(\sigma_E^2 + \gamma_I \sigma_I^2 \theta / \gamma_S)^M} - \exp\left(-\frac{2\theta}{\gamma_S \sigma_E^2}\right) \frac{(\sigma_E^2 \sigma_E^2)^{2M}}{(\sigma_E^2 + \gamma_I \sigma_I^2 \theta / \gamma_S)^{2M}} \right]. \quad (46)$$

Therefore, after some incorporations and iterations, the closed-form outage probability and intercept probability expressions of both the DT and RS schemes with multiple cochannel interferers can be achieved.

4. Simulation Evaluations

In this section, the SRT performances of the DT and RS schemes are evaluated by simulations. The simulation parameters are set as follows: $R_s = 1$ bit/s/Hz, $\sigma_{S_b D}^2 = \sigma_{S_b R_k}^2 = \sigma_{R_k D}^2 = \sigma_S^2 = 1$, $\sigma_{I_m R_k}^2 = \sigma_{I_m D}^2 = \sigma_{I_m E}^2 = \sigma_I^2 = 0.1$, and $\sigma_{S_b E}^2 = \sigma_{R_{\text{best}} E}^2 = \sigma_E^2 = 0.2$.

Figures 2–4 show the curves of the theoretical SRT analysis. As can be seen from the figures the intercept probability is presented as a function of the outage probability. Obviously, it can be seen that the simulation results match well with the theoretical analysis. Figure 2 shows the intercept probabilities versus the outage probabilities of the conventional DT strategy as well as the RS strategy at

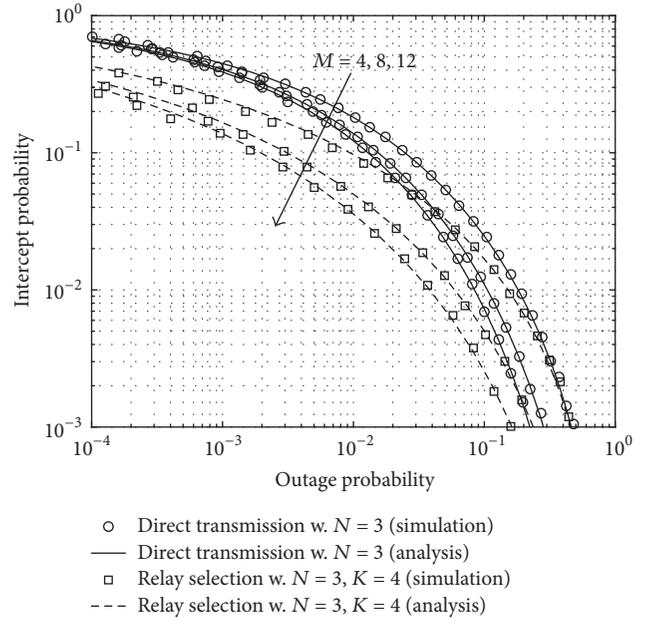


FIGURE 3: SRTs of the conventional DT strategy and RS strategy for different numbers of cochannel interferers M with $\gamma_I = 15$ dB, $N = 3$, and $K = 4$.

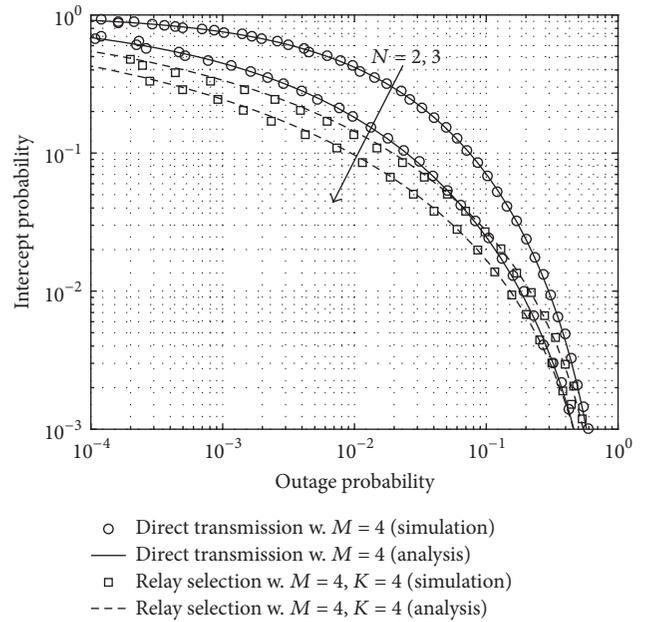


FIGURE 4: SRTs of the conventional DT strategy and RS strategy for different numbers of sources N with $\gamma_I = 15$ dB, $M = 4$, and $K = 4$.

different K ($K = \{4, 8, 12\}$). Figure 2 also shows that as the outage probabilities increase, the intercept probabilities of the conventional DT and the RS schemes decrease. This confirms that there exists a tradeoff between the intercept probability and the outage probability. Another phenomenon can be observed in Figure 2; that is, the SRT of the RS strategy always outperforms that of the conventional DT strategy. Moreover, the SRT of the RS strategy is also improved with increasing K

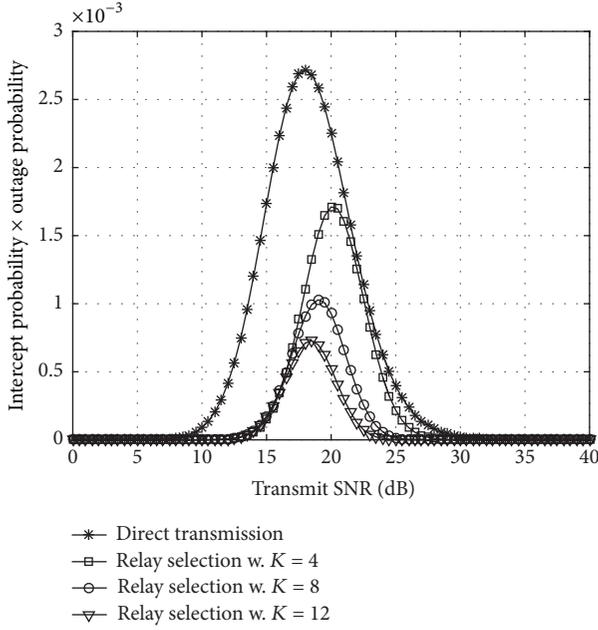


FIGURE 5: $IP \times OP$ versus the transmit SNR γ_S of the conventional DT strategy and RS strategy for different numbers of relays K with $\gamma_I = 15$ dB, $N = 3$, and $M = 4$.

(from 4 to 12). This is due to the reason that the diversity gain can be obtained with the increase of the number of relays.

Figure 3 depicts the intercept probabilities versus the outage probabilities of the conventional DT strategy as well as the RS strategy at different M ($M = \{4, 8, 12\}$). Figure 3 also shows that when the outage probabilities change from 10^{-4} to 10^0 , the intercept probabilities of the conventional DT and the RS schemes decrease correspondingly. Moreover, for a given M , the SRT of the RS strategy performs better than that of the conventional DT strategy. It is also seen that the SRTs of the conventional DT and the RS schemes are also improved with increasing M (from 4 to 12).

Figure 4 shows the intercept probabilities versus the outage probabilities of the conventional DT strategy as well as the RS strategy at different N ($N = \{2, 3\}$). Similar to Figure 3, the intercept probabilities of the conventional DT and the RS schemes decrease correspondingly, as the outage probabilities increase from 10^{-4} to 10^0 . Moreover, for a given N , the SRT of the RS strategy always outperforms that of the conventional DT strategy. The SRTs of the conventional DT and the RS schemes are also improved with increasing N (from 2 to 3). By jointly considering Figures 2–4, it is found that as the outage probabilities increase, the intercept probabilities of the conventional DT and the RS schemes decrease, implying that the SRT indeed exists between the intercept probability and the outage probability. Meanwhile, the improvement of the SRT is obtained with increasing the number of sources, relays, and cochannel interferers. Moreover, it is also shown that the SRT of the RS strategy consistently outperforms that of the conventional DT strategy.

In order to further evaluate the SRT, the products of the intercept probabilities and the outage probabilities of the conventional DT strategy as well as the RS strategy are plotted in Figures 5–7. Figure 5 shows the $(IP \times OP)$ products

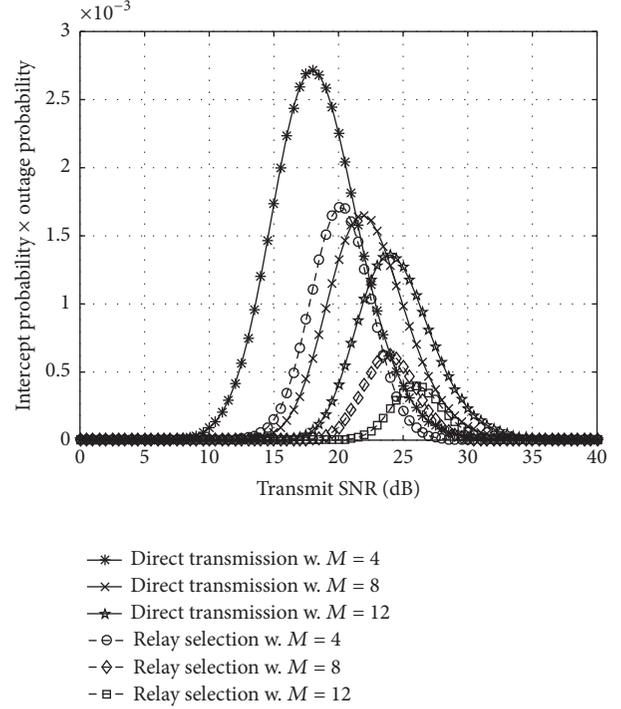


FIGURE 6: $IP \times OP$ versus the transmit SNR γ_S of the conventional DT strategy and RS strategy for different numbers of cochannel interferers M with $\gamma_I = 15$ dB, $N = 3$, and $K = 4$.

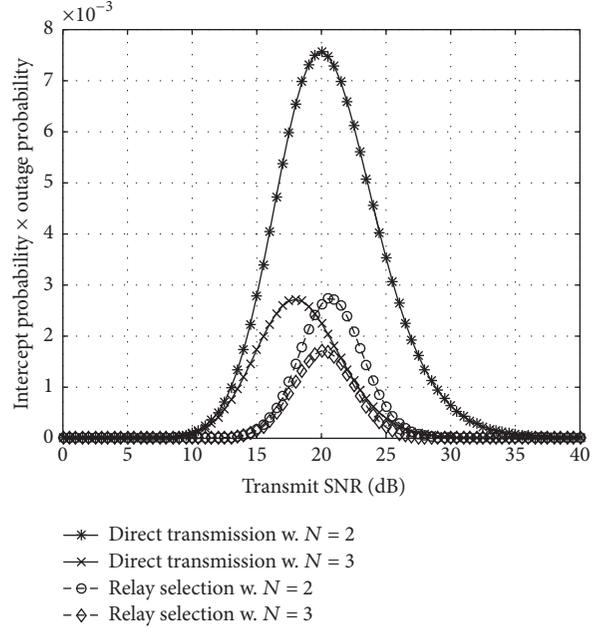


FIGURE 7: $IP \times OP$ versus the transmit SNR γ_S of the conventional DT strategy and RS strategy for different numbers of sources N with $\gamma_I = 15$ dB, $M = 4$, and $K = 4$.

of the conventional DT strategy as well as the RS strategy against the transmit SNR γ_S at different numbers of available relays K ($K = \{4, 8, 12\}$). Figure 5 also shows that there exists an $(IP \times OP)$ peak with the increase of the transmit SNR γ_S . This is due to the reason that the IP in the low SNR regime comes close to 0, and the OP in the high SNR

regime comes close to 0. Particularly, the presence of the (IP \times OP) peak is another perspective that the SRT indeed exists between the intercept probability and the outage probability. Clearly, it is also seen from Figure 5 that the maximum (IP \times OP) product of the RS strategy is smaller than that of the conventional DT strategy, which means the RS strategy is always better than the conventional DT strategy in terms of the (IP \times OP) product. Moreover, the maximum (IP \times OP) product decreases significantly with increasing K (from 4 to 12). It implies the SRT of the RS strategy is improved accordingly.

Figure 6 shows the (IP \times OP) products of the conventional DT strategy as well as the RS strategy against the transmit SNR γ_S at different M ($M = \{4, 8, 12\}$). Similar to Figure 5, there also exists an (IP \times OP) peak with the increase of the transmit SNR γ_S . Figure 6 also shows that, for a given M , the RS strategy outperforms the DT strategy in terms of the (IP \times OP) product. Moreover, the SRTs of the conventional DT and the RS schemes are also improved with increasing M (from 4 to 12).

Figure 7 shows the (IP \times OP) products of the conventional DT strategy as well as the RS strategy versus the transmit SNR γ_S at different N ($N = \{2, 3\}$). Similar to Figure 5, there also exists an (IP \times OP) peak with the increase of the transmit SNR γ_S . Meanwhile, for a given N , the RS strategy outperforms the conventional DT strategy in terms of the (IP \times OP) product, and when the number of sources increases from $N = 2$ to $N = 3$, the maximum (IP \times OP) product can be minimized so that the SRTs of the conventional DT and the RS schemes are improved. By jointly considering Figures 5–7, the improvement of the SRT performance in terms of the (IP \times OP) product can be achieved by increasing the number of relays, sources, and cochannel interferers. Moreover, it is also shown that the SRT performance of the RS strategy is better than that of the conventional DT strategy in terms of the product of the IP and the OP.

5. Conclusions

In this paper, we presented the PLS of a multisource multirelay cooperative communication network by considering the influence of cochannel interference from an SRT perspective. Under impact of cochannel interferers, we adopted an SINR-based method to analyze the SRT performance characterized by the OP and the IP. We derived the closed-form IP and OP expressions of both the conventional DT strategy and RS strategy over Rayleigh fading channels, respectively. We showed that when the OP (reliability) requirement is relaxed, the IP (security) performance improves and vice versa. It confirms that there is an SRT existing between the OP and the IP. We also showed that a better SRT performance can be achieved by increasing the number of sources, relays, and cochannel interferers. In addition, the RS strategy generally outperforms the conventional DT strategy in terms of the product of the OP and the IP.

Appendix

A. Derivation of (24)

According to (23), P_1 is rewritten as

$$\begin{aligned}
 P_1 &= \sum_{n=0}^{N-1} \frac{C_{N-1}^n (-1)^n}{\sigma_{S_b,D}^2} \int_0^{\Delta/\gamma_S} \exp\left(-\frac{nx}{\sigma_{S_n,D}^2} - \frac{x}{\sigma_{S_b,D}^2}\right) dx \\
 &+ \sum_{n=0}^{N-1} \sum_{m_1=0}^{M-1} \frac{C_{N-1}^n (-1)^n}{m_1! \sigma_{S_b,D}^2} \exp\left(\frac{1}{\gamma_I \sigma_{I_m,D}^2}\right) \\
 &\times \int_{\Delta/\gamma_S}^{\infty} \exp\left(-\left(\frac{n}{\sigma_{S_n,D}^2} + \frac{1}{\sigma_{S_b,D}^2} + \frac{\gamma_S}{\gamma_I \sigma_{I_m,D}^2 \Delta}\right)x\right) \\
 &\times \left(\frac{\gamma_S x}{\gamma_I \sigma_{I_m,D}^2 \Delta} - \frac{1}{\gamma_I \sigma_{I_m,D}^2}\right)^{m_1} dx.
 \end{aligned} \tag{A.1}$$

After some incorporations and necessary mathematical manipulations, P_1 can be calculated as

$$\begin{aligned}
 P_1 &= \sum_{n=0}^{N-1} \frac{C_{N-1}^n (-1)^n}{n \sigma_{S_b,D}^2 / \sigma_{S_n,D}^2 + 1} \\
 &\cdot \left[1 - \exp\left(-\frac{(n \sigma_{S_b,D}^2 / \sigma_{S_n,D}^2 + 1) \Delta}{\sigma_{S_b,D}^2 \gamma_S}\right)\right] \\
 &+ \sum_{n=0}^{N-1} \sum_{m_1=0}^{M-1} \frac{C_{N-1}^n (-1)^n}{m_1! \sigma_{S_b,D}^2} \exp\left(\frac{1}{\gamma_I \sigma_{I_m,D}^2}\right) \\
 &\times \exp\left(-\frac{(n/\sigma_{S_n,D}^2 + 1/\sigma_{S_b,D}^2 + \gamma_S/\gamma_I \sigma_{I_m,D}^2 \Delta)(1/\gamma_I \sigma_{I_m,D}^2)}{\gamma_S/\gamma_I \sigma_{I_m,D}^2 \Delta}\right) \\
 &\times \frac{(\gamma_S/\gamma_I \sigma_{I_m,D}^2 \Delta)^{m_1}}{(n/\sigma_{S_n,D}^2 + 1/\sigma_{S_b,D}^2 + \gamma_S/\gamma_I \sigma_{I_m,D}^2 \Delta)^{m_1+1}} \Gamma(m_1 + 1, 0).
 \end{aligned} \tag{A.2}$$

For notational convenience, considering $\sigma_{S_b,D}^2 = \sigma_{S_n,D}^2 = \sigma_S^2$ and $\sigma_{I_m,D}^2 = \sigma_I^2$ and using the equation $\Gamma(n, x) = (n-1)! \exp(-x) \sum_{m=0}^{n-1} (x^m/m!)$ [32], P_1 can be achieved as shown in (24).

This completes the derivation of (24).

B. Derivation of (39)

According to (38), the second term can be expressed as

$$\begin{aligned}
 &\Pr\left(\max\left(\frac{\gamma_S x}{Y+1}, \frac{\gamma_S |h_{R_{\text{best},D}}|^2}{Y+1}\right) < \theta \mid x\right) \\
 &= \begin{cases} \int_0^{\infty} \Pr\left(\frac{\gamma_S |h_{R_{\text{best},D}}|^2}{y+1} < \theta \mid x\right) f_Y(y) dy & x < \frac{\theta}{\gamma_S} \\ \int_{\gamma_S x/\theta-1}^{\infty} \Pr\left(\frac{\gamma_S |h_{R_{\text{best},D}}|^2}{y+1} < \theta \mid x\right) f_Y(y) dy & x > \frac{\theta}{\gamma_S} \end{cases} \tag{B.1}
 \end{aligned}$$

where the term $\Pr(\gamma_S | h_{R_{\text{best},D}}|^2 / (y+1) < \theta \mid x)$ can be calculated as

$$\begin{aligned} & \Pr\left(\frac{\gamma_S |h_{R_{\text{best},D}}|^2}{y+1} < \theta \mid x\right) \\ &= \Pr\left(\max_{R_k \in \mathcal{D}} (|h_{R_k,D}|^2) < \frac{\theta(y+1)}{\gamma_S} \mid x\right) \\ &= \prod_{R_k \in \mathcal{D}} \Pr\left(|h_{R_k,D}|^2 < \frac{\theta(y+1)}{\gamma_S} \mid x\right) \\ &= \sum_{k_1=0}^{|\mathcal{D}|} C_{|\mathcal{D}|}^{k_1} (-1)^{k_1} \exp\left(-\frac{k_1 \theta (y+1)}{\sigma_{R_{k_1},D}^2 \gamma_S}\right). \end{aligned} \quad (\text{B.2})$$

Substituting (B.2) into (B.1), one has

$$\begin{aligned} A_1 &= \sum_{k_1=0}^{|\mathcal{D}|} C_{|\mathcal{D}|}^{k_1} (-1)^{k_1} \exp\left(-\frac{k_1 \theta}{\sigma_{R_{k_1},D}^2 \gamma_S}\right) \\ &\quad \times \int_0^\infty \exp\left(-\frac{k_1 \theta y}{\sigma_{R_{k_1},D}^2 \gamma_S}\right) f_Y(y) dy. \end{aligned} \quad (\text{B.3})$$

Using the equation $\int_0^\infty x^m \exp(-\beta x^n) dx = \Gamma((m+1)/n) / n \beta^{(m+1)/n}$ [32], A_1 can be calculated as

$$\begin{aligned} A_1 &= \sum_{k_1=0}^{|\mathcal{D}|} C_{|\mathcal{D}|}^{k_1} (-1)^{k_1} \exp\left(-\frac{k_1 \theta}{\sigma_{R_{k_1},D}^2 \gamma_S}\right) \\ &\quad \cdot \frac{1}{\left(1 + k_1 \theta \sigma_{I_m,D}^2 \gamma_I / \sigma_{R_{k_1},D}^2 \gamma_S\right)^M}. \end{aligned} \quad (\text{B.4})$$

By utilizing a similar way, A_2 can be obtained as

$$\begin{aligned} A_2 &= \sum_{k_1=0}^{|\mathcal{D}|} \sum_{m_1=0}^{M-1} \frac{C_{|\mathcal{D}|}^{k_1} (-1)^{k_1}}{m_1!} \\ &\quad \cdot \exp\left(-\left(\frac{k_1}{\sigma_{R_{k_1},D}^2} + \frac{\gamma_S}{\gamma_I \sigma_{I_m,D}^2 \theta}\right) x\right) \end{aligned}$$

$$\begin{aligned} & \times \exp\left(\frac{1}{\gamma_I \sigma_{I_m,D}^2}\right) \\ & \cdot \frac{(\gamma_S x / \gamma_I \sigma_{I_m,D}^2 \theta - 1 / \gamma_I \sigma_{I_m,D}^2)^{m_1}}{\left(1 + k_1 \theta \sigma_{I_m,D}^2 \gamma_I / \sigma_{R_{k_1},D}^2 \gamma_S\right)^{M-m_1}}. \end{aligned} \quad (\text{B.5})$$

Substituting (21) and (B.1) into (38), P_4 can be expressed as

$$\begin{aligned} P_4 &= \underbrace{\int_0^{\theta/\gamma_S} \sum_{n=0}^{N-1} C_{N-1}^n (-1)^n \exp\left(-\frac{nx}{\sigma_{S_n,D}^2}\right) A_1 f_{|h_{S_b,D}|^2}(x) dx}_{B_1} \\ &+ \underbrace{\int_{\theta/\gamma_S}^\infty \sum_{n=0}^{N-1} C_{N-1}^n (-1)^n \exp\left(-\frac{nx}{\sigma_{S_n,D}^2}\right) A_2 f_{|h_{S_b,D}|^2}(x) dx}_{B_2}. \end{aligned} \quad (\text{B.6})$$

Then, incorporating (B.4) into (B.6) and employing some mathematical manipulations, B_1 can be calculated as

$$\begin{aligned} B_1 &= \sum_{n=0}^{N-1} \sum_{k_1=0}^{|\mathcal{D}|} C_{N-1}^n (-1)^n C_{|\mathcal{D}|}^{k_1} (-1)^{k_1} \\ &\quad \cdot \frac{\exp(-k_1 \theta / \sigma_{R_{k_1},D}^2 \gamma_S)}{\left(1 + k_1 \theta \sigma_{I_m,D}^2 \gamma_I / \sigma_{R_{k_1},D}^2 \gamma_S\right)^M} \times \frac{1}{\sigma_{S_b,D}^2} \\ &\quad \cdot \int_0^{\theta/\gamma_S} \exp\left(-\frac{nx}{\sigma_{S_n,D}^2} - \frac{x}{\sigma_{S_b,D}^2}\right) dx \\ &= \sum_{n=0}^{N-1} \sum_{k_1=0}^{|\mathcal{D}|} \frac{C_{N-1}^n (-1)^n C_{|\mathcal{D}|}^{k_1} (-1)^{k_1}}{n \sigma_{S_b,D}^2 / \sigma_{S_n,D}^2} \exp\left(-\frac{k_1 \theta}{\sigma_{R_{k_1},D}^2 \gamma_S}\right) \\ &\quad \times \frac{1 - \exp\left(-\left(n \sigma_{S_b,D}^2 / \sigma_{S_n,D}^2 + 1\right) \theta / \sigma_{S_b,D}^2 \gamma_S\right)}{\left(1 + k_1 \theta \sigma_{I_m,D}^2 \gamma_I / \sigma_{R_{k_1},D}^2 \gamma_S\right)^M}. \end{aligned} \quad (\text{B.7})$$

By utilizing a similar way, B_2 can be obtained as

$$\begin{aligned} B_2 &= \sum_{n=0}^{N-1} \sum_{k_1=0}^{|\mathcal{D}|} \sum_{m_1=0}^{M-1} \frac{C_{N-1}^n (-1)^n C_{|\mathcal{D}|}^{k_1} (-1)^{k_1} \gamma_I \sigma_{I_m,D}^2 \theta}{\sigma_{S_b,D}^2 \gamma_S} \times \exp\left(-\frac{k_1 \theta}{\sigma_{R_{k_1},D}^2 \gamma_S}\right) \exp\left(-\frac{(n \sigma_{S_b,D}^2 / \sigma_{S_n,D}^2 + 1) \theta}{\sigma_{S_b,D}^2 \gamma_S}\right) \\ &\quad \times \frac{\left(1 + k_1 \theta \sigma_{I_m,D}^2 \gamma_I / \sigma_{R_{k_1},D}^2 \gamma_S\right)^{m_1 - M}}{\left(1 + k_1 \gamma_I \sigma_{I_m,D}^2 \theta / \sigma_{R_{k_1},D}^2 \gamma_S + (n \sigma_{S_b,D}^2 / \sigma_{S_n,D}^2 + 1) \gamma_I \sigma_{I_m,D}^2 \theta / \sigma_{S_b,D}^2 \gamma_S\right)^{m_1 + 1}}. \end{aligned} \quad (\text{B.8})$$

Incorporating (B.7) and (B.8) into (B.6), and for simplicity, considering $\sigma_{S_b,D}^2 = \sigma_{S_n,D}^2 = \sigma_S^2$ and $\sigma_{I_m,D}^2 = \sigma_I^2$, P_4 can be obtained as shown in (39).

This completes the derivation of (39).

Data Availability

The underlying data comes from simulation results.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research reported in this article was supported by the National Natural Science Foundation of China (61501356 and 61501354), by the Key Project of National Natural Science Foundation of China (61631015), by the National Key Scientific and Technological Innovation Team Plan (2016KCT-01), and by the Fundamental Research Funds for the Central Universities (7215433803).

References

- [1] X. Huang, T. Han, and N. Ansari, "On green-energy-powered cognitive radio networks," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 2, pp. 827–842, 2015.
- [2] M. Zhang, K. Cumanan, and A. Burr, "Secure energy efficiency optimization for MISO cognitive radio network with energy harvesting," in *Proceedings of the 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP)*, pp. 1–6, Nanjing, October 2017.
- [3] A. G. Fragkiadakis, E. Z. Tragos, and I. G. Askoxyllakis, "A survey on security threats and detection techniques in cognitive radio networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 428–445, 2013.
- [4] F. Zhou, Z. Li, J. Cheng, Q. Li, and J. Si, "Robust AN-Aided Beamforming and Power Splitting Design for Secure MISO Cognitive Radio with SWIPT," *IEEE Transactions on Wireless Communications*, vol. 16, no. 4, pp. 2450–2464, 2017.
- [5] Y. Wu, R. Schober, D. W. Ng, C. Xiao, and G. Caire, "Secure massive MIMO transmission with an active eavesdropper," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 62, no. 7, pp. 3880–3900, 2016.
- [6] M. Bloch, J. Barros, M. R. Rodrigues, and S. W. McLaughlin, "Wireless information-theoretic security," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 54, no. 6, pp. 2515–2534, 2008.
- [7] Y. Zou, J. Zhu, X. Wang, and L. Hanzo, "A survey on wireless security: technical challenges, recent advances, and future trends," *Proceedings of the IEEE*, vol. 104, no. 9, pp. 1727–1765, 2016.
- [8] C. E. Shannon, "Communication theory of secrecy systems," *Bell Labs Technical Journal*, vol. 28, pp. 656–715, 1949.
- [9] A. D. Wyner, "The wire-tap channel," *Bell Labs Technical Journal*, vol. 54, no. 8, pp. 1355–1387, 1975.
- [10] I. Csisz and J. Köner, "Broadcast channels with confidential messages," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, 1978.
- [11] S. I. Y. Cheong and M. E. Hellman, "The Gaussian wire-tap channel," *Institute of Electrical and Electronics Engineers Transactions on Information Theory*, vol. 24, no. 4, pp. 451–456, 1978.
- [12] Y. Zou, Y. D. Yao, and B. Zheng, "Cooperative relay techniques for cognitive radio systems: spectrum sensing and secondary user transmissions," *IEEE Communications Magazine*, vol. 50, no. 4, pp. 98–103, 2012.
- [13] H. Sakran, M. Shokair, O. Nasr, S. El-Rabaie, and A. A. El-Azm, "Proposed relay selection scheme for physical layer security in cognitive radio networks," *IET Communications*, vol. 6, no. 16, pp. 2676–2687, 2012.
- [14] B. V. Nguyen and K. Kim, "Secrecy outage probability of optimal relay selection for secure AnF cooperative networks," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2086–2089, 2015.
- [15] L. J. Rodriguez, N. H. Tran, T. Q. Duong, T. Le-Ngoc, M. Elkashlan, and S. Shetty, "Physical layer security in wireless cooperative relay networks: State of the art and beyond," *IEEE Communications Magazine*, vol. 53, no. 12, pp. 32–39, 2015.
- [16] H. M. Wang, Q. Yin, and X. G. Xia, "Distributed beamforming for physical-layer security of two-way relay networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3532–3545, 2012.
- [17] H. M. Wang, F. Liu, and M. Yang, "Joint Cooperative Beamforming, Jamming, and Power Allocation to Secure AF Relay Systems," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4893–4898, 2015.
- [18] G. Zheng, L. C. Choo, and K. K. Wong, "Optimal cooperative jamming to enhance physical layer security using relays," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1317–1322, 2011.
- [19] F. Zhu, F. Gao, M. Yao, and H. Zou, "Joint information- and jamming-beamforming for physical layer security with full duplex base station," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6391–6401, 2014.
- [20] B. Han, J. Li, J. Su, M. Guo, and B. Zhao, "Secrecy capacity optimization via cooperative relaying and jamming for WANETs," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 4, pp. 1117–1128, 2015.
- [21] S. A. Fakoorian and A. L. Swindlehurst, "Solutions for the MIMO Gaussian wiretap channel with a cooperative jammer," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 5013–5022, 2011.
- [22] C. Jeong, I. M. Kim, and D. I. Kim, "Joint secure beamforming design at the source and the relay for an amplify-and-forward MIMO untrusted relay system," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 310–325, 2012.
- [23] R. Yin, S. Wei, J. Yuan, X. Shan, and X. Wang, "Tradeoff between reliability and security in block ciphering systems with physical channel errors," in *Proceedings of the IEEE Military Communications Conference (MILCOM '10)*, pp. 2156–2161, San Jose, Calif, USA, November 2010.
- [24] J. You, Z. Zhong, G. Wang, and B. Ai, "Security and Reliability Performance Analysis for Cloud Radio Access Networks with Channel Estimation Errors," *IEEE Access*, vol. 2, pp. 1348–1358, 2014.
- [25] Q. Gu, G. Wang, L. Gao, and M. Peng, "Security-reliability performance of cognitive AF relay-based wireless communication system with channel estimation error Security Challenges and Issues in Cognitive Radio Networks," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, article no. 28, 2014.

- [26] Y. Zou, X. Wang, W. Shen, and L. Hanzo, "Security versus reliability analysis of opportunistic relaying," *IEEE Transactions on Vehicular Technology*, vol. 63, no. 6, pp. 2653–2661, 2014.
- [27] J. Zhu, Y. Zou, B. Champagne, W.-P. Zhu, and L. Hanzo, "Security-reliability trade-off analysis of multi-relay aided decode-and-forward cooperation systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5825–5831, 2016.
- [28] L. Fan, X. Lei, N. Yang, T. Q. Duong, and G. K. Karagiannis, "Secure Multiple Amplify-and-Forward Relaying with Cochannel Interference," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 8, pp. 1494–1505, 2016.
- [29] M. Ahmadian-Attari, S. Vahidian, M. Mohammadi, and S. Hataminia, "Performance analysis of two-way decode-and-forward relaying in the presence of co-channel interferences," *IET Communications*, vol. 8, no. 18, pp. 3349–3356, 2014.
- [30] S. Vahidian, S. Aïssa, and S. Hatamnia, "Relay selection for Security-Constrained cooperative communication in the presence of eavesdropper's overhearing and interference," *IEEE Wireless Communications Letters*, vol. 4, no. 6, pp. 577–580, 2015.
- [31] W. Hu, Z. Li, H. Li, J. Si, and H. Huang, "SRT analysis of relay selection in the presence of multiple co-channel interferers," *IET Communications*, vol. 11, no. 6, pp. 809–816, 2017.
- [32] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*, Academic Press, Inc., San Diego, CA, USA, 7th edition, 1996.

Research Article

Dynamic Power Splitting Strategy for SWIPT Based Two-Way Multiplicative AF Relay Networks with Nonlinear Energy Harvesting Model

Tianci Wang ¹, Guangyue Lu,¹ Yinghui Ye ,^{1,2} and Yuan Ren¹

¹*Shaanxi Key Laboratory of Information Communication Network and Security, Xi'an University of Posts and Telecommunications, Xi'an, China*

²*Integrated Service Networks Lab of Xidian University, Xi'an, China*

Correspondence should be addressed to Yinghui Ye; connectyh@126.com

Received 8 March 2018; Accepted 7 May 2018; Published 11 June 2018

Academic Editor: Wolfgang H. Gerstaecker

Copyright © 2018 Tianci Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper investigates an energy-constrained two-way multiplicative amplify-and-forward (AF) relay network, where a practical nonlinear energy harvesting (NLEH) model is equipped at the relay to realize simultaneous wireless information and power transfer (SWIPT). We focus on the design of dynamic power splitting (DPS) strategy, in which the PS ratio is able to adjust itself according to the instantaneous channel state information (CSI). Specifically, we first formulate an optimization problem to maximize the outage throughput, subject to the NLEH. Since this formulated problem is nonconvex and difficult to solve, we further transfer it into an equivalent problem and develop a Dinkelbach iterative method to obtain the corresponding solution. Numerical results are given to verify the quick convergence of the proposed iterative method and show the superior outage throughput of the designed DPS strategy by comparing with two peer strategies designed for the linear energy harvesting (LEH) model.

1. Introduction

Internet of things (IoT) devices are usually powered by batteries with limited energy storage capacity, leading to a key constraint of the performance of energy-constrained wireless networks [1, 2]. To address this problem, simultaneous wireless information and power transfer (SWIPT) has been recently proposed as a promising solution to prolong the lifetime of energy-constrained wireless networks, where the wireless signal is either switched in the time domain or split in the power domain to provide signal transmission and power transfer using the same wireless carrier, i.e., time switching (TS) strategy and power splitting (PS) strategy. Accordingly, SWIPT is applicable in energy-constrained networks for striking a balance between information and energy [3–5].

Relaying techniques, including one-way relay networks (OWRN) and two-way relay networks (TWRN), are highly beneficial in wireless communications to overcome shadowing effects, to increase the communication range, to improve

the energy efficiency, and to increase the achievable throughput [6]. Of particular interest is the two-step (or three-step) TWRNs, in which one node shares its data with the other node via an intermediate relay. The system configuration may arise in many practical scenarios, e.g., data exchange between sensor nodes and the data through an immediate relay in IoT networks [7, 8]. However, in fact, the relay nodes may have limited battery capacity and thus rely on some external resources to charge in order to remain active. Further, due to the random positions of relay nodes, consistent power supply may be unavailable for energy-constrained relay nodes, leading to possible power outages. As a result, the aforementioned two promising techniques, SWIPT and two-step (or three-step) TWRNs, can be integrated to balance between information and energy [9].

Up to now, several works have been reported regarding this issue [10–15]. Authors of [10, 11] introduced decode-and-forward (DF) and amplify-and-forward (AF) into PS strategy based SWIPT with two-step TWRNs, respectively. Reference

[12] studied the optimal PS strategy to maximize the energy efficiency. Since the circuitry design of three-step is simpler than that of two-step, [13] studied the bounds performance for PS based SWIPT with three-step DF-TWRNs in terms of outage probability. Different from [13], the authors of [14] studied the PS based three-step multiplicative AF-TWRNs, due to the advantage of three-step multiplicative TWRNs in outage probability and investigated the corresponding outage performance with a static PS strategy, where the PS ratio is determined by statistic channel state information (CSI). This results in a room for improving by making full use of the instantaneous CSI. Due to this reason, the dynamic PS (DPS) strategy was further developed [15]. It was shown that the outage performance can be improved by employing the DPS strategy.

However, the above works discussed [10–15] were based on the assumption of a linear energy harvesting (LEH) model, which was shown to be inaccurate and not capable of capturing the nonlinear behaviour of RF energy harvesting (RF-EH) circuits [16]. As a result, those existing strategies based on the LEH model lead to significant performance loss in a real scenario owing to the mismatching between linear and nonlinear EH (NLEH) model. Even though several works [16–27] have been reported regarding the applications of the NLEH model for wireless communications, most of them (see [16–24]) focused on the wireless powered communication (WPC) networks and point-to-point/cognitive radio networks with SWIPT. Apart from the aforementioned networks, the applications of a NLEH model have also been studied to the ORNs [25–27]. In [25, 26], the authors investigated the outage performance of a NLEH relaying network with a PS strategy. Considering the perfect/imperfect CSI at the relay, the optimal PS strategy was developed in terms of outage performance [27]. However, there is no work in the existing literature studying the TWRNs with a NLEH harvester. This motivates our work.

In this paper, we study a DPS strategy for three-step multiplicative AF-TWRNs, where the relay is equipped with a NLEH harvester (this work extends the recent work [15] into the NLEH) to realize the SWIPT. To incentivize the relay to cooperate with the source, the harvest-then-forward scheme is adopted, i.e., the relay only uses the harvested energy from the source's signal to assist its transmissions. In order to investigate the upper bound outage throughput of the considered network, we assume that CSI is available. Our contributions are as follows.

We formulate an optimization problem to maximize the outage throughput by adjusting the PS ratio according to the instantaneous CSI. The optimization problem is equivalent to maximize worst end-to-end signal-to-noise ratio (SNR), which is nonconvex and difficult to solve. On this basis, we reformulate it as a fractional programming problem and employ the Dinkelbach method to derive a DPS strategy. The simulation results show, compared with the existing strategies, the proposed DPS strategy achieves a larger outage throughput.

The rest of this paper is organized as follows. In Section 2, we introduce the system model. In Section 3, we formulate an optimization problem to maximize outage throughput and

design an iterative method to obtain the optimal solution. Section 4 provides simulation results to verify our work. Finally, Section 5 concludes the paper.

2. System Model

2.1. Multiplicative AF-TWRNs. We consider a NLEH multiplicative AF-TWRNs, where an energy-constrained relay R coordinates the two-way communications for two terminals (i.e., node 1 and node 2) exchanging information by adopting the harvest-then-forward scheme, as shown in Figure 1. All nodes are equipped with one antenna due to the limited space (since the main focus of this work is on the novel dynamic power splitting (DPS) scheme design subject to the nonlinear model, for analytical tractability, we consider the single antenna AF relay networks). We ignore the direct transmission between two terminals due to the heavy fading [15]. For successful information exchange between the two nodes, we consider the following assumptions [14, 15]:

- (i) The total transmission time block $T = 1$ is divided into three consecutive equal time slots, as shown in Figure 2. First, node 1 sends the signal to R . And then, node 2 sends the signal to R . Finally, the relay R broadcasts the multiplied signal to nodes k ($k = 1$ or 2) using the harvested energy.
- (ii) The path-loss model is distance-dependent with a rate of $d_k^{-\alpha}$, where α is the path-loss exponent and d_k is the distance between node k and relay R . Let g_k denote the complex channel coefficients from the relay to the destination node k , respectively. Each link is independent with frequency nonselective Rayleigh block fading. Further, CSI is available in order to investigate the upper bound outage performance.
- (iii) The processing energy required by the transmit/receive circuitry at the relay is ignored since it is very small compared with the transmit energy.

2.2. Energy Harvesting Model. EH model is able to characterize the relationship between input power P_{in} and output power P_{out} by a function $f(\cdot)$, given by

$$P_{out} = f(P_{in}). \quad (1)$$

The conventional LEH model assumes a fixed constant to describe the relationship between input power P_{in} and output power P_{out} , given by

$$P_{out}^{linear} = f(P_{in}) = \eta P_{ER_k}, \quad 0 \leq \eta \leq 1, \quad (2)$$

where η is the fixed energy conversion efficiency of energy harvester equipped at relay.

However, the above LEH model cannot capture the properties of practical EH circuits, since the energy conversion efficiency improves as the input power rises, but for high input power there is a limitation on the maximum harvested

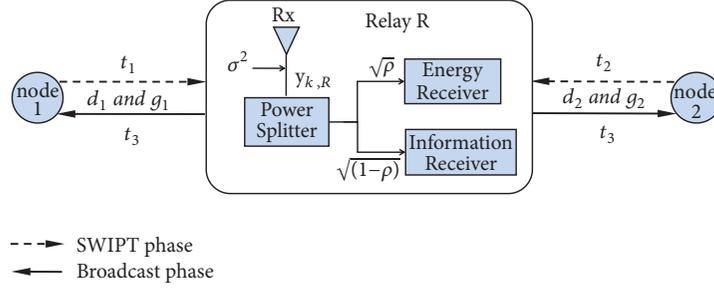


FIGURE 1: System model of three-step multiplicative AF-TWRNs with SWIPT.

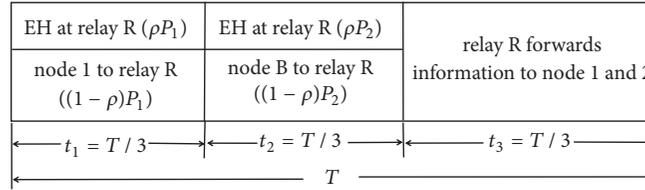


FIGURE 2: Transmission time-block structure for the DPS strategy.

energy [19]. To this end, a NLEH model was proposed by fitting the measurement data in [16], given by

$$P_{out}^{Non-linear} = f(P_{in}) = \frac{[M / (1 + \exp(-\bar{a}(P_{in} - \bar{b}))) - M\Omega]}{1 - \Omega}, \quad (3)$$

where M denotes the maximum harvested power at EH harvester when the EH circuit is saturated; \bar{a} and \bar{b} are constants related to the detailed EH circuit specifications such as the resistance, capacitance, and diode turn-on voltage; $\Omega = 1/(1 + \exp(\bar{a}\bar{b}))$.

More recently, the authors of [28] proposed another function to characterize the practical EH circuitry; that is,

$$P_{out}^{Non-linear} = f(P_{in}) = \frac{aP_{in} + b}{P_{in} + c} - \frac{b}{c}, \quad (4)$$

where parameters a , b , and c are constants determined by standard curve fitting tool.

Compared with the model in (3), the model in (4) is more mathematically tractable and is able to provide sufficient precision [28]. For the above reasons, this paper employs the model in (4) to develop the DPS strategy.

2.3. Working Flow. In the first or second slot, the signal from the node k received at relay R is given as

$$y_{k,R} = \sqrt{P_k} h_k x_k + z_k, \quad \forall k = 1, 2. \quad (5)$$

where P_k is the transmit power of node k , $h_k = g_k / \sqrt{d_k^\alpha}$, x_k denotes the normalized signal from the node k , and z_k is the additive white Gaussian noise (AWGN) at the receiving antenna.

The received RF signal is split into two streams, $\sqrt{(1-\rho)}y_{k,R}$ for information processing and $\sqrt{\rho}y_{k,R}$ for EH. Accordingly, based on model (4), the total harvested energy at R from the first and second slots can be calculated as

$$\begin{aligned} E_R &= \frac{T}{3} (P_{ER_1}^{Non-linear} + P_{ER_2}^{Non-linear}) \\ &= \frac{T}{3} \left(\frac{aP_{ER_1} + b}{P_{ER_1} + c} + \frac{aP_{ER_2} + b}{P_{ER_2} + c} - \frac{2b}{c} \right) \\ &= \frac{T}{3} \left(\frac{(ac - b) [2P_{ER_1} P_{ER_2} + c(P_{ER_1} + P_{ER_2})]}{cP_{ER_1} P_{ER_2} + c^2(P_{ER_1} + P_{ER_2}) + c^3} \right), \end{aligned} \quad (6)$$

where $P_{ER_k} = \rho P_k |h_k|^2$ is the received power from the node k ; $\rho \in [0, 1)$ is the PS ratio.

In the third slot, the received signals at the first slot and the second slot are multiplied and amplified, resulting in a hybrid signal at relay R , given by

$$\begin{aligned} y_R &= \sqrt{\frac{P_R}{\xi}} \left(h_1 \sqrt{(1-\rho)} P_1 x_1 + Z_1 \right) \\ &\quad \cdot \left(h_2 \sqrt{(1-\rho)} P_2 x_2 + Z_2 \right), \end{aligned} \quad (7)$$

where $\xi = \Pi_{k=1}^2 ((1-\rho)|h_k|^2 P_k + \sigma_k^2)$ is the power constraint factor; σ_k^2 is the power of the mixed AWGN noises Z_k , including the antenna noise and conversion noise; $P_R = 3E_R/T$ is the transmitted power at the relay.

According to the single channel separation signal technology [15], the desired signal from the target node can be obtained from the mixed signal. Therefore, the signal received at destination node 1 obtains the information from node 2; that is,

$$y_{1,2} = x_1^{-1} (h_1 y_R + \widehat{Z}_1), \quad (8)$$

where \widehat{Z}_1 denotes the mixed noise at node 1 with the variance $\widehat{\sigma}_1^2$, which includes the transmitter antenna noise, the receiver antenna noise, and the conversion noise; $x_1^{-1} = 1/x_1$.

$$y_{1,2} = \underbrace{\sqrt{\frac{P_R P_1 P_2}{\xi}} (1-\rho) |h_1|^2 h_2 x_2}_{\text{Signal Term}} + \underbrace{\sqrt{\frac{P_R}{\xi}} h_1 x_1^{-1} Z_1 Z_2 + \sqrt{\frac{P_R P_1 (1-\rho)}{\xi}} |h_1|^2 Z_2}_{\text{Noise Term I}} + \underbrace{\sqrt{\frac{P_R P_2 (1-\rho)}{\xi}} h_1 h_2 x_1^{-1} x_2 Z_1 + x_1^{-1} \widehat{Z}_1}_{\text{Noise Term II}}, \quad (9)$$

Therefore, the SNR from node 2 to node 1 can be described as

$$\gamma_1 = \frac{s}{n + o + i + e}, \quad (10)$$

where (in this paper, we assume that instantaneous CSI is only available at the relay in order to reduce the CSI feedback overhead. Thus, we assume that $P_1 = P_2 = P_S$, and the power allocation scheme for destination nodes has not been considered)

$$\begin{aligned} s &= \mu \left[2P_{ER_1} P_{ER_2} + c(P_{ER_1} + P_{ER_2}) \right] P_S^2 (1-\rho)^2 |h_1|^4 \\ &\quad \cdot |h_2|^2, \\ n &= \mu \left[2P_{ER_1} P_{ER_2} + c(P_{ER_1} + P_{ER_2}) \right] P_S (1-\rho) |h_1|^4 \sigma_2^2, \\ o &= \mu \left[2P_{ER_1} P_{ER_2} + c(P_{ER_1} + P_{ER_2}) \right] P_S (1-\rho) |h_1|^2 \\ &\quad \cdot |h_2|^2 \sigma_1^2, \\ i &= \mu \left[2P_{ER_1} P_{ER_2} + c(P_{ER_1} + P_{ER_2}) \right] |h_1|^2 \sigma_1^2 \sigma_2^2, \\ e &= (cP_{ER_1} P_{ER_2} + c^2(P_{ER_1} + P_{ER_2}) + c^3) \\ &\quad \cdot (P_S (1-\rho) |h_1|^2 + \sigma_1^2) (P_S (1-\rho) |h_2|^2 + \sigma_2^2) \widehat{\sigma}_1^2, \end{aligned} \quad (11)$$

$$\mu = ac - b,$$

$$P_1 = P_2 = P_S,$$

$$P_{ER_1} = \rho P_S |h_1|^2,$$

$$P_{ER_2} = \rho P_S |h_2|^2.$$

Likewise, the SNR from the node 1 to node 2 is γ_2 , which can be written as

$$\gamma_2 = \frac{s'}{n' + o' + i' + e'}, \quad (12)$$

where

$$\begin{aligned} s' &= \mu \left[2P_{ER_1} P_{ER_2} + c(P_{ER_1} + P_{ER_2}) \right] P_S^2 (1-\rho)^2 |h_2|^4 \\ &\quad \cdot |h_1|^2, \\ n' &= \mu \left[2P_{ER_1} P_{ER_2} + c(P_{ER_1} + P_{ER_2}) \right] P_S (1-\rho) |h_2|^4 \\ &\quad \cdot \sigma_1^2, \end{aligned}$$

Substituting (7) into (8), the received signal $y_{1,2}$ at the node 1 from the node 2 can be rewritten as

$$\begin{aligned} o' &= \mu \left[2P_{ER_1} P_{ER_2} + c(P_{ER_1} + P_{ER_2}) \right] P_S (1-\rho) |h_2|^2 \\ &\quad \cdot |h_1|^2 \sigma_2^2, \\ i' &= \mu \left[2P_{ER_1} P_{ER_2} + c(P_{ER_1} + P_{ER_2}) \right] |h_2|^2 \sigma_2^2 \sigma_1^2, \\ e' &= (cP_{ER_1} P_{ER_2} + c^2(P_{ER_1} + P_{ER_2}) + c^3) \\ &\quad \cdot (P_S (1-\rho) |h_2|^2 + \sigma_2^2) (P_S (1-\rho) |h_1|^2 + \sigma_1^2) \widehat{\sigma}_2^2. \end{aligned} \quad (13)$$

3. Dynamic Power Splitting

3.1. Optimization Problem Formulated. For SWIPT based TWRNs, the outage occurs if one destination node cannot decode the information from the other source node. Thus, the outage throughput τ can be described as

$$\tau = (1 - P_{out})U, \quad (14)$$

with

$$\begin{aligned} P_{out} &= P \{ (\gamma_1 < \gamma_{th}) \cup (\gamma_2 < \gamma_{th}) \} \\ &= 1 - P \{ \gamma_1 \geq \gamma_{th}, \gamma_2 \geq \gamma_{th} \}, \end{aligned} \quad (15)$$

where $\gamma_{th} = 2^U - 1$ is the minimum acceptable threshold value; U denotes the constant transmission rate of the source node.

It can be observed from (14) and (15) that the outage throughput is determined by the worse end-to-end SNR of the two links. Thus, the optimization problem to maximize the outage throughput can be written as

$$(P1): \quad \max_{\rho} \min \{ \gamma_1, \gamma_2 \} \quad (16)$$

$$\text{subject to: } \quad 0 \leq \rho < 1.$$

Lemma 1. P1 is proved to be equivalent to the following optimization problem P2, which is

$$(P2): \quad \max_{\rho} \gamma \quad (17)$$

$$\text{subject to: } \quad 0 \leq \rho < 1,$$

$$\text{where } \gamma = \begin{cases} \gamma_1, & \text{if } \gamma_1 \leq \gamma_2 \\ \gamma_2, & \text{if } \gamma_1 > \gamma_2 \end{cases}.$$

The proof is given in Appendix.

3.2. Iterative Method Design. In this subsection, we focus on the design of iterative method to solve (P2) based on the **Lemma 1**. Specifically, we assume that γ_1 is equal to γ and design an iterative method to obtain corresponding solution. Note that the case of $\gamma = \gamma_2$ can also be obtained by the same method.

We denote the maximum SNR of the node 2 to node 1 which equals q^* , which is

$$q^* = \max_{0 \leq \rho^{(l)} < 1} \frac{s}{n + o + i + e}, \quad (18)$$

It can be concluded that (18) is a typical fractional programming problem. Therefore, we can employ the Dinkelbach method to obtain the maximum q^* , as shown in Algorithm 1. Notice that q^* is available if and only if

$$\max_{0 \leq \rho^{(l)} < 1} \{s - q^*(n + i + s + e)\} = 0. \quad (19)$$

The proof can be found in [29].

Next, we solve the auxiliary problem in the step 4 of the Dinkelbach method, which is nonconvex. The equation in the auxiliary problem can be written as

$$F(q^{(l-1)}, \rho^{(l)}) = r_1 \rho^4 + r_2 \rho^3 + r_3 \rho^2 + r_4 \rho - r_5, \quad (20)$$

where

$$\begin{aligned} r_1 &= 2\mu m_1 m_3 - u_5 m_1^2, \\ r_2 &= \mu [-4m_1 m_3 + (c + 2u_4) m_2 m_3] \\ &\quad - u_5 (u_8 m_1 m_2 - 2m_1^2), \\ r_3 &= \mu [2(m_3 - m_8) m_1 + cm_2 m_7 - 2cm_3], \\ &\quad - k_5 [(u_9 + u_{10} m_2) m_1 + m_1^2 - u_3 m_2^2], \\ r_4 &= \mu (cm_3 - cm_8) \\ &\quad - u_6 [(m_2 - 2c) m_1 - u_3 m_2 + u_{11} m_2^2], \\ r_5 &= u_7 (m_1 + \sigma^2 m_2 + \sigma^4). \\ m_1 &= u_1 u_2, \\ m_2 &= u_1 + u_2, \\ m_3 &= u_1 u_2 |h_1|^2, \\ m_4 &= u_1 u_4 |h_1|^2, \\ m_5 &= u_2 u_4 |h_1|^2, \\ m_6 &= u_4 |h_1|^2 \sigma_1^2, \\ m_7 &= u_4 |h_1|^2 (u_1 + u_2), \\ m_8 &= k_4 |h_1|^2 (u_1 + u_2 + \sigma^2). \\ u_1 &= P_S |h_1|^2, \\ u_2 &= P_S |h_2|^2, \end{aligned}$$

$$u_3 = c\sigma^2,$$

$$u_4 = q^{(l-1)} \sigma^2,$$

$$u_5 = q^{(l-1)} \sigma^2 c,$$

$$u_6 = q^{(l-1)} \sigma^2 c^2,$$

$$u_7 = q^{(l-1)} \sigma^2 c^3,$$

$$u_8 = c - \sigma^2,$$

$$u_9 = \sigma^4 + c^2,$$

$$u_{10} = \sigma^2 - 2c,$$

$$u_{11} = \sigma^2 + \sigma^4. \quad (21)$$

The first-order derivative of the function $F(q^{(l-1)}, \rho^{(l)})$ can be written as

$$F'(q^{(l-1)}, \rho^{(l)}) = 4r_1 \rho^3 + 3r_2 \rho^2 + 2r_3 \rho + r_4 = 0. \quad (22)$$

Clearly, (22) is a typical cubic equation and the roots can be determined [30]. For the convenience of the following description, we denote that $A = 9r_2^2 - 24r_1 r_3$, $B = 6r_2 r_3 - 36r_1 r_4$, $C = 4r_3^2 - 9r_2 r_4$, and $\Delta = B^2 - 4AC$. The roots are as follows.

Case 1. If $A = B = 0$, there exists a triple root and the real root is $-3r_2/12r_1$.

Case 2. If $\Delta = B^2 - 4AC > 0$, there exist one real root and two complex roots. The real root is $(-3r_2 - (\sqrt[3]{Y_1} + \sqrt[3]{Y_2}))/12r_1$, where $Y_{1,2} = 3Ar_2 + 12r_1((-B \pm (\sqrt{B^2 - 4AC}))/2)$.

Case 3. If $\Delta = B^2 - 4AC = 0$, there exist a double real root, and a real root, i.e., two different real roots. Both two different roots are $-K/2$ and $-3r_2/4r_1 + K$, where $K = (B/A)(A \neq 0)$.

Case 4. If $\Delta = B^2 - 4AC < 0$, there exist three different real roots. Three different real roots are $(-3r_2 - 2\sqrt{A} \cos(\theta/3))/12r_1$, $(-3r_2 + \sqrt{A}(\cos(\theta/3) + \sqrt{3} \sin(\theta/3)))/12r_1$, and $(-3r_2 + \sqrt{A}(\cos(\theta/3) - \sqrt{3} \sin(\theta/3)))/12r_1$, where $\theta = \arccos T$ and $T = (6Ar_2 - 12r_1 B)/2\sqrt{A^3}(A > 0, -1 < T < 1)$.

Based on the above cases, we can obtain closed-form roots for the above cubic equation (22). Then the real root from 0 to 1 to maximize the function $F(q^{(l-1)}, \rho^{(l)})$ is selected as the optimal solution to the auxiliary problem by comparing values of the real roots.

4. Simulation Results

In this section, we present the numerical results to evaluate the outage throughput of the proposed DPS, two baseline schemes proposed in [14, 15] under a NLEH model. The fitted ones of the employed NLEH parameters are set as $a = 109.7$,

- (1): Initialize:
the maximum number of iterations L_{\max} , the iteration index $l = 1$, $q^{(0)} = 0$;
- (2): Set:
 $\eta, P_s, \delta, \sigma^2$ and input the instantaneous CSI, h_k ;
- (3): **repeat**
- (4): Solve the maximize optimization problem in (19) with $F(q^{(l-1)}, \rho^{(l)}) = \max_{0 \leq \rho^{(l)} < 1} \{s - q^{(l-1)}(n + o + i + e)\}$ and obtain $\bar{\rho}^{(l)}$;
- (5): Apply $\bar{\rho}^{(l)}$ into s, n, o, i, e to calculate $\bar{s}, \bar{n}, \bar{o}, \bar{i}, \bar{e}$;
- (6): Update $q^{(l)} = \bar{s}/(\bar{n} + \bar{o} + \bar{i} + \bar{e})$;
- (7): If $|F(q^{(l-1)}, \bar{\rho}^{(l)})| \geq \delta$, then $l = l + 1$;
- (8): **until** $|F(q^{(l-1)}, \bar{\rho}^{(l)})| < \delta$ or $l = L_{\max}$;
- (9): Obtain the maximum SNR $q^* = q^{(l)}$, and the optimal PS ratio $\rho^* = \bar{\rho}^{(l)}$.

ALGORITHM 1: The proposed iterative method.

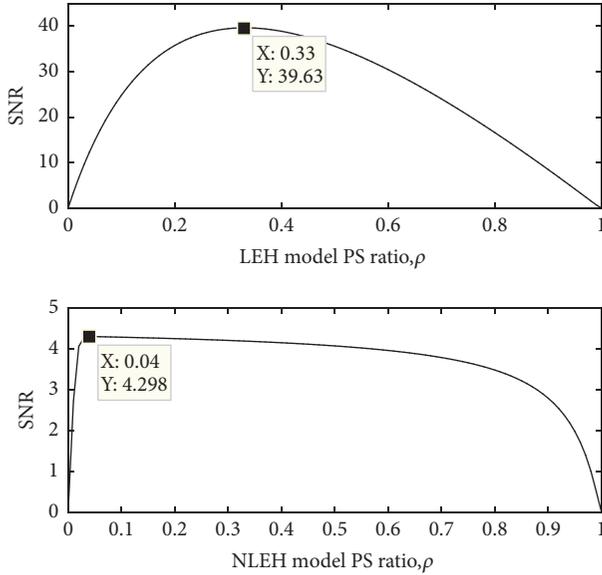


FIGURE 3: SNR versus PS ratio through exhaustive searching.

$b = 43.69$, and $c = 128.1$. For comparison, we set the parameters as follows: $\alpha = 2.7$, $d_1 = d_2 = 1\text{m}$, $\delta = 10^{-1}$, and $\sigma^2 = 0.01$.

Figure 3 shows the SNR versus PS ratio by exhaustive searching method with the channel gains $|h_1| = 0.9770$ and $|h_2| = 1.2594$. The transmit power and the transmission rate of source node are set to be 1.5 J/s and 3 bits/s/Hz , respectively. One can see that the value of optimal PS ratios for LEH and NLEH is unequal. This point illustrates that the optimal PS strategy designed for LEH is not an optimal one for NLEH, due to the mismatching between LEH and NLEH. Based on the same parameter as Figure 3, the PS ratio and SNR versus number of iterations are illustrated. It can be observed that only two iterations are required to obtain optimal solution by the proposed iterative method. In addition, the optimal solution obtained by the proposed iterative method is equal to the one obtained by exhaustive searching method in Figure 4. The above analyses verify our proposed iterative method.

Figure 5 depicts the outage throughput with different transmission rates U for the proposed DPS strategy, the static PS strategy [14], and the existing DPS strategy (here, we employ LEH-DPS scheme to calculate the optimal PS ratio under the LEH model and substitute it into NLEH model) [15]. Among those strategies, the static PS strategy and the existing DPS strategy are designed for LEH model and the existing DPS strategy is renamed as a baseline scheme for convenience. It can be observed that, in terms of outage throughput, the proposed DPS scheme outperforms the static PS and the baseline scheme at different transmission rates U , as expected. This is because our proposed strategy is designed for NLEH and can obtain the optimal solution optimized for a NLEH model.

Both Figures 6 and 7 illustrate the outage throughput versus SNR for two strategies under different scenarios. One can see from Figure 6 that the outage throughput decreases with the distance. The reason is that the larger distance leads to a smaller channel gain and a lower outage throughput. One also can see that the proposed strategy outperforms the baseline scheme, which shows that the baseline scheme is a suboptimal one for the NLEH. In addition, it can be observed from Figure 7 that a considerable performance gain can be achieved by selecting an appropriate transmission rate.

5. Conclusion

In this paper, we developed a DPS strategy to maximize the outage throughput for NLEH three-step multiplicative AF-TWRNs. In particular, we formulated a nonconvex optimization problem for maximizing the outage throughput and proposed a Dinkelbach method to obtain the solution with a few iterations. Although the auxiliary problem of the proposed Dinkelbach method is nonconvex, we derived the closed-form solutions and avoided the necessity of solving a sequence of nonconvex auxiliary problems. Simulations showed that the proposed DPS strategy achieves a higher outage throughput in NLEH three-step multiplicative AF-TWRNs compared with two recent schemes designed for the LEH model.

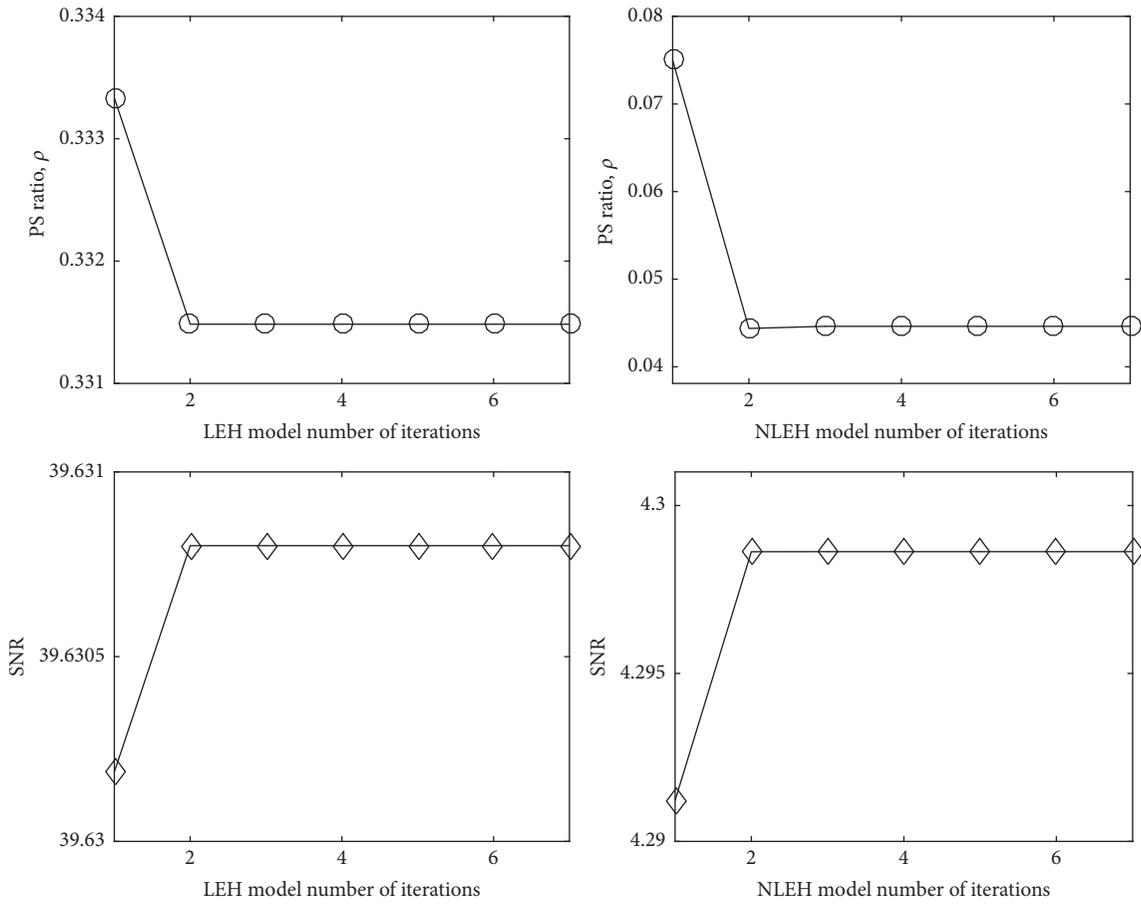


FIGURE 4: PS ratio and SNR versus number of iterations.

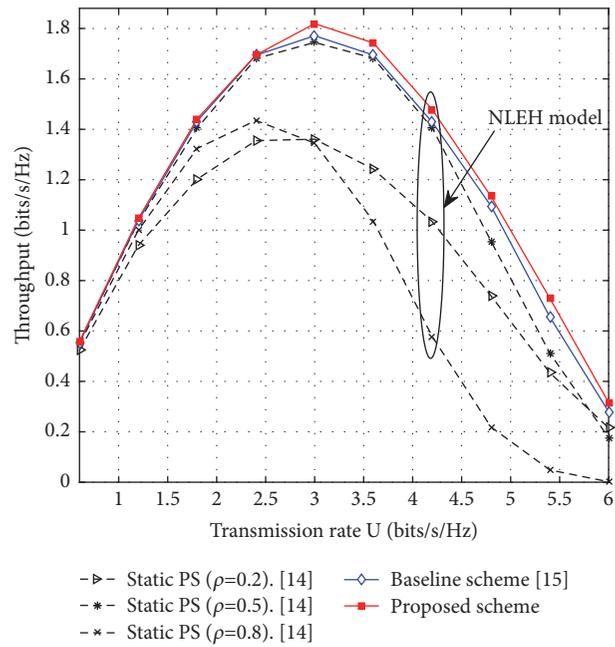


FIGURE 5: Outage throughput versus varying values of transmission rate.

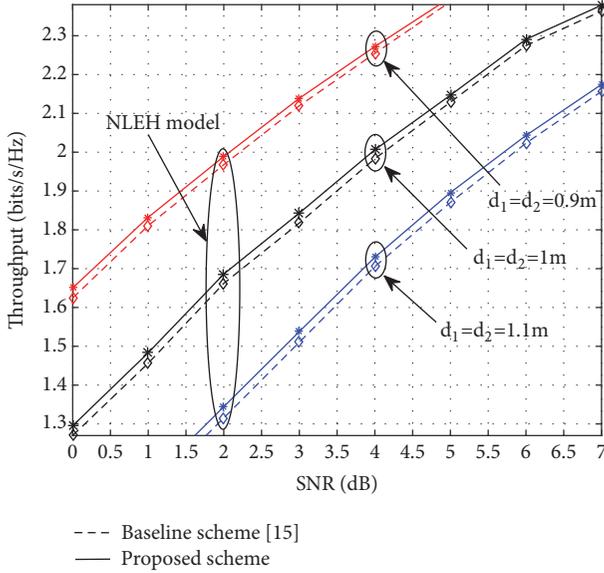


FIGURE 6: Outage throughput versus SNR with different distances.

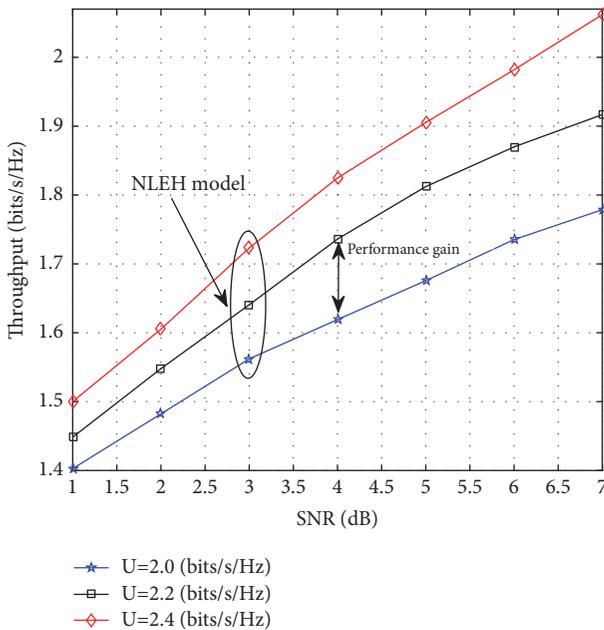


FIGURE 7: Outage throughput versus SNR with different transmission rates.

Appendix

Proof of the Lemma

For ease of analysis, we neglect the antenna noise and assume that $\sigma^2 = \sigma_1^2 = \sigma_2^2 = \bar{\sigma}_1^2 = \bar{\sigma}_2^2$ [14, 15].

Case i. If $|h_1|^2 = |h_2|^2$ or $\rho = 0$, it is obvious that $\gamma_1 = \gamma_2$. Thus, the optimization problem (P1) is equivalent to (P2).

Case ii. By means of reduction to absurdity, it can be proven that $\gamma_1 = \gamma_2$ holds when $|h_1|^2 \neq |h_2|^2$ and $\rho \neq 0$. We

assume that $\gamma_1 - \gamma_2 = 0$ holds. Through some convenient mathematical calculations, the equality $\gamma_1 - \gamma_2 = 0$ can be rewritten as

$$\begin{aligned} & |h_1|^2 \left((1 - \rho) P_S |h_1|^2 + \sigma^2 \right) \left((1 - \rho) P_S |h_2|^2 + \sigma^2 \right) \\ &= |h_2|^2 \left((1 - \rho) P_S |h_1|^2 + \sigma^2 \right) \\ &\cdot \left((1 - \rho) P_S |h_2|^2 + \sigma^2 \right) \end{aligned} \quad (\text{A.1})$$

Obviously, (A.1) holds if and only if $|h_1|^2 = |h_2|^2$ or $\rho = 0$, which indicates that the assumption $\gamma_1 - \gamma_2 = 0$ does not hold except for the case i. Based on the above cases, **Lemma 1** holds.

The proof is complete.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The research reported in this article was supported by the National Science and Technology Major Project (2016ZX03001016), the Science and Technology Innovation Team of Shaanxi Province for Broadband Wireless and Application under Grant 2017KCT-30-02, the Natural Science Foundation of China (61701399, 61501371), and the Research Program of Education Bureau of Shaanxi Province (17JK0699).

References

- [1] Z. Chu, F. Zhou, Z. Zhu, R. Q. Hu, and P. Xiao, "Wireless Powered Sensor Networks for Internet of Things: Maximum Throughput and Optimal Power Allocation," *IEEE Internet of Things Journal*, vol. 5, no. 1, pp. 310–321, Feb. 2018.
- [2] F. Zhou, Y. Wu, Y. Liang, Z. Li, Y. Wang, and K. K. Wong, "State of the art, taxonomy, and open issues on NOMA in cognitive radio networks," *IEEE Wireless Communications Magazine*, vol. 25, no. 2, pp. 100–108, 2018.
- [3] L. Shi, L. Zhao, K. Liang, and H. Chen, "Wireless Energy Transfer Enabled D2D in Underlying Cellular Networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1845–1849, Feb. 2018.
- [4] Y. Ye, Y. Li, D. Wang, and G. Lu, "Power splitting protocol design for the cooperative NOMA with SWIPT," in *Proceedings of the IEEE International Conference on Communications*, pp. 1–5, May 2017.
- [5] D. Wang, Y. Li, Y. Ye, H. Xia, and H. Zhang, "Joint Time Allocation and Power Splitting Schemes for DF Energy Harvesting Relaying Networks," in *Proceedings of the IEEE 86th Vehicular Technology Conference*, pp. 1–5, September 2017.
- [6] D. Gunduz, A. Yener, A. Goldsmith, and H. V. Poor, "The multiway relay channel," *IEEE Transactions on Information Theory*, vol. 59, no. 1, pp. 51–63, 2013.

- [7] F. Zhou, Y. Wu, R. Q. Hu, Z. Li, and K. K. Wong, "Energy-efficient NOMA heterogeneous cloud radio access networks," *IEEE Wireless Communications Magazine*, 2017, to be published.
- [8] B. Rankov and A. Wittneben, "Spectral efficient protocols for half-duplex fading relay channels," *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 379–389, 2007.
- [9] Z. Chu, M. Johnston, and S. Le Goff, "SWIPT for wireless cooperative networks," *IEEE Electronics Letters*, vol. 51, no. 6, pp. 536–538, 2015.
- [10] S. Ghosh, T. Acharya, and S. P. Maity, "Outage analysis in DF relay assisted two-way communication with RF energy harvesting," in *Proceedings of the WSA 2017; 21th International ITG Workshop on Smart Antennas*, pp. 1–8, March, 2017.
- [11] Z. Chen, B. Xia, and H. Liu, "Wireless information and power transfer in two-way amplify-and-forward relaying channels," in *Proceedings of the IEEE Global Conference on Signal and Information Processing*, pp. 168–172, December 2014.
- [12] C. Zhang, H. Du, and J. Ge, "Energy-efficient power allocation in energy harvesting two-way AF relay systems," *IEEE Access*, vol. 5, pp. 3640–3645, 2017.
- [13] N. T. P. Van, S. F. Hasan, X. Gui, S. Mukhopadhyay, and H. Tran, "Three-step two-way decode and forward relay with energy harvesting," *IEEE Communications Letters*, vol. 21, no. 4, pp. 857–860, 2017.
- [14] S. T. Shah, K. W. Choi, S. F. Hasan, and M. Y. Chung, "Energy harvesting and information processing in two-way multiplicative relay networks," *IEEE Electronics Letters*, vol. 52, no. 9, pp. 751–753, 2016.
- [15] Z. Wang, Y. Li, Y. Ye, and H. Zhang, "Dynamic Power Splitting for Three-Step Two-Way Multiplicative AF Relay Networks," in *Proceedings of the IEEE 86th Vehicular Technology Conference*, pp. 1–5, Sept, 2017.
- [16] E. Boshkovska, D. W. K. Ng, N. Zlatanov, and R. Schober, "Practical non-linear energy harvesting model and resource allocation for SWIPT systems," *IEEE Communications Letters*, vol. 19, no. 12, pp. 2082–2085, Dec, 2015.
- [17] D. Mishra, S. De, and D. Krishnaswamy, "Dilemma at RF Energy Harvesting Relay: Downlink Energy Relaying or Uplink Information Transfer?" *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 4939–4955, Aug, 2017.
- [18] E. Boshkovska, D. W. K. Ng, N. Zlatanov, A. Koelpin, and R. Schober, "Robust Resource Allocation for MIMO Wireless Powered Communication Networks Based on a Non-Linear EH Model," *IEEE Transactions on Communications*, vol. 65, no. 5, pp. 1984–1999, May, 2017.
- [19] L. Shi, L. Zhao, K. Liang, X. Chu, G. Wu, and H. Chen, "Profit maximization in wireless powered communications with improved non-linear energy conversion and storage efficiencies," in *Proceedings of the IEEE International Conference on Communications*, pp. 1–6, May, 2017.
- [20] K. Xiong, B. Wang, and K. J. R. Liu, "Rate-Energy Region of SWIPT for MIMO Broadcasting under Nonlinear Energy Harvesting Model," *IEEE Transactions on Wireless Communications*, vol. 16, no. 8, pp. 5147–5161, Aug, 2017.
- [21] Q. Yao, T. Q. S. Quek, A. Huang, and H. Shan, "Joint Downlink and Uplink Energy Minimization in WET-Enabled Networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6751–6765, Oct, 2017.
- [22] Y. Wang, F. Zhou, Y. Wu, and H. Zhou, "Resource allocation in wireless powered cognitive radio networks based on a practical non-linear energy harvesting model," *IEEE Access*, vol. 5, pp. 17618–17626, 2017.
- [23] J.-M. Kang, I.-M. Kim, and D. I. Kim, "Mode Switching for SWIPT over Fading Channel with Nonlinear Energy Harvesting," *IEEE Wireless Communications Letters*, vol. 6, no. 5, pp. 678–681, Oct, 2017.
- [24] F. Zhou, Z. Chu, H. Sun, R. Q. Hu, and L. Hanzo, "Artificial Noise Aided Secure Cognitive Beamforming for Cooperative MISO-NOMA Using SWIPT," *IEEE Journal on Selected Areas in Communications*, 2018, to be published.
- [25] A. Cvetkovic, V. Blagojevic et al., "Performance Analysis of Non-linear Energy-Harvesting DF Relay System in Interference-Limited Nakagami-m Fading Environment," *ETRI Journal*, vol. 39, no. 6, pp. 803–812, Dec, 2017.
- [26] J. Zhang and G. Pan, "Outage Analysis of Wireless-Powered Relaying MIMO Systems with Non-Linear Energy Harvesters and Imperfect CSI," *IEEE Access*, vol. 4, pp. 7046–7053, 2016.
- [27] K. Wang, Y. Li, Y. Ye, and H. Zhang, "Dynamic Power Splitting Schemes for Non-Linear EH Relaying Networks: Perfect and Imperfect CSI," in *Proceedings of the IEEE 86th Vehicular Technology Conference*, pp. 1–5, Sept, 2017.
- [28] Y. Chen, N. Zhao, and M. Alouini, "Wireless Energy Harvesting Using Signals From Multiple Fading Channels," *IEEE Transactions on Communications*, vol. 65, no. 11, pp. 5027–5039, 2017.
- [29] W. Dinkelbach, "On nonlinear fractional programming," *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.
- [30] S. Fan, "A new extracting formula and a new distinguishing means on the one variable cubic equation," *Natural Science Journal of Hainan Teachers College*, vol. 2, no. 2, pp. 91–98, 1989.

Research Article

Energy-Efficient Incentives Resource Allocation Scheme in Cooperative Communication System

Zi Yan Liu ¹, Pan Mao,^{1,2} Li Feng,³ and Shi Mei Liu¹

¹College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China

²Huaxin Consulting Co., Ltd., Hangzhou 310014, China

³State Grid Chongqing Electric Power Company, Chongqing 400014, China

Correspondence should be addressed to Zi Yan Liu; gzucomm@gmail.com

Received 21 January 2018; Revised 25 March 2018; Accepted 24 April 2018; Published 5 June 2018

Academic Editor: Zheng Chu

Copyright © 2018 Zi Yan Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Appropriate resource allocation has great significance to enhance the energy efficiency (EE) for cooperative communication system. The objective is to allocate the resource to maximize the energy efficiency in single-cell multiuser cooperative communication system. We formulate this problem as subcarrier-based resource allocation and solve it with path planning in graph theory. A two-level neural network model is designed, in which the users and subcarrier are defined as network nodes. And then we propose an improved intelligent water drops algorithm combined with Genetic Algorithm; boundary condition and initialization rules of path soil quantity are put forward. The simulation results demonstrate that the proposed resource allocation scheme can effectively improve the energy efficiency and enhance QoS performance.

1. Introduction

The rapid energy consumption due to the demands of mobile communication services has become a subject of global interests from environment perspective. On one hand, because of its slow development and limited capacity, battery technology becomes the bottleneck of limiting the development of the portable terminals [1]; on the other hand, enormous energy consumption of the communication industry indirectly leads to the greenhouse gas emission and increases the operators' operating costs. Statistics show that, in 2009, the power consumption of three service providers in China was 28.9 billion degrees, which equals 4.41 million tons of coal burning. By 2014, the energy consumption had been up to 6.71 million tons, with a 52% increase in 5 years [2, 3]. Compared to other industries, it is essential for the communication industry to reduce energy consumption. Meanwhile, as important support for social informatization, there is a very broad prospect for the communication industry to promote energy conservation in the society information industry. Therefore, designing a high energy-efficient communication system has become a consensus of the communication industry.

Cooperative communication [4, 5] is defined as follows: in a cell, the adjacent devices with single antenna create a virtual MIMO system by sharing their antennas with each other [6, 7] to achieve the goal of overcoming the multipath fading and gaining the benefit of multiantenna space diversity. In cooperative communication system, a reasonable resource allocation scheme has great significance in improving the spectral efficiency and reducing the energy consumption. The traditional design of cooperative communication system mainly focuses on the improvement of system capacity, outage probability, spectral efficiency, and other performances. With the scholars' attention to the energy consumption of communication industry and the concept of green communication, the energy efficiency of cooperative communication system is gaining widespread concern. Wong et al. [8] studied the network-level resource scheduling scheme in cooperative communication system and proposed a cooperative concept to obtain the higher energy efficiency; at the same time they also designed high energy-efficient network architecture. A resource allocation scheme on game theory is established; it is proven that the cooperation between users can effectively improve the energy efficiency of the system [9]. In [10],

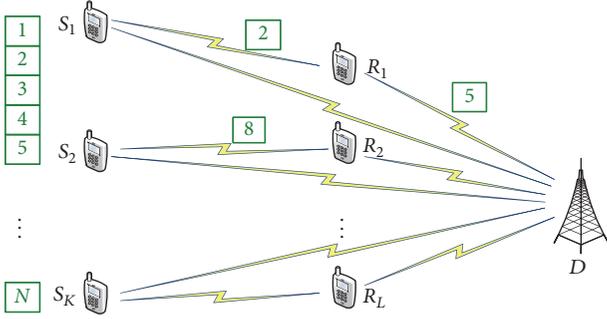


FIGURE 1: Cooperative communication model.

an algorithm is proposed to select the optimal relay in a single-relay cooperation system. Compared with the fixed cooperation method, the energy consumption is significantly reduced. The wireless sensor networks (WSNs) with energy harvesting and cooperative communication were studied in [11], and an energy-efficient scheduling strategy is proposed; the optimal scheduling problem is solved by using a Markov Decision Process (MDP). In [12], the process of collaboration was divided into two time slots. An algorithm of relay selection and power allocation is proposed for the minimum BER and maximum system capacity, respectively. However, these studies have achieved the purpose of improving the energy efficiency from the perspective of energy consumption and have not analyzed the energy efficiency in the form of quantitative indicators. QoS in the actual scenario is not considered, especially the demands of high transmission rate.

In this paper, we consider a single-cell multiuser cooperative communication system. To meet the demands of QoS performance, a two-level neural network model based on intelligent water drops (IWDs) algorithm is proposed to solve the problem of resource allocation with optimal energy efficiency. And then, the IWDs algorithm combined with Genetic Algorithm (GA) is improved to allocate resources flexibly and enhance performance.

This paper is organized as follows. Section 2 introduces the system model as well as the function of the proposed optimal problem. In Section 3, a novel algorithm is proposed. After that, the optimal energy-efficient resource allocation scheme is presented in Section 4. In Section 5, numerical results are depicted. Finally, Section 6 concludes the paper.

2. System Model and Problem Formulation

In this section, the proposed system model is presented, followed by the optimization problem.

2.1. System Model. We consider an uplink single-cell multiuser cooperative communication scenario as shown in Figure 1, and the radius is 600 meters. It consists of source node S , relay node R , and destination node D . In particular, the source node is K users with call requests, and L is the relay node that participates in the cooperative communication, and the destination node is the base station (BS). Orthogonal Frequency Division Multiple Access (OFDMA) is the multiple

access scheme, and the available bandwidth B is divided into N , in which the subcarriers are orthogonal to each other. It is assumed that the relay node has perfect channel state information (CSI). Further, the channels are considered as large-scale and small-scale fading. That is, large-scale fading is defined as path loss and small-scale fading is Rayleigh fading, respectively. Moreover, the forwarding mode is Decode-and-Forward (DF), and the relay network is operated in half-duplex mode.

2.2. Problem Formulation. Assume that the channel states between any two terminals (S_k, R_l, D) are independent of each other. h_{ij}^n ($i, j \in \{S_k, R_l, D\}$) indicates the channel fading between the device i and j . The channel fading between the nodes can be given by [13]

$$\begin{aligned} E\left(|h_{S_k D}^n|^2\right) &= d_{S_k D}^{-\alpha}, \\ E\left(|h_{S_k R_l}^n|^2\right) &= d_{S_k R_l}^{-\alpha}, \\ E\left(|h_{R_l D}^n|^2\right) &= d_{R_l D}^{-\alpha}, \end{aligned} \quad (1)$$

where $|h_{S_k D}^n|^2$, $|h_{S_k R_l}^n|^2$, $|h_{R_l D}^n|^2$ denote the CSI on the subcarrier n , from the user S_k to the destination node D , from the user S_k to the relay R_l , and from the relay R_l to the destination node D , respectively. d_{ij} ($i, j \in \{S_k, R_l, D\}$) represents the distance between the device i and j . $E(\cdot)$ is the average operator, and $\alpha \in [3, 5]$ is the channel fading factor.

The energy efficiency is defined as follows [14]:

$$\eta_{EE} = \frac{R}{P_{\text{tot}}}, \quad (2)$$

where R represents the total transfer rate and P_{tot} denotes the total energy consumption.

The transmission procedures of cooperative communication system are divided into two time slots [15]: the first is the broadcast slots, during which the source node S broadcasts the information to the relay node R and the destination node D ; the second is the forwarding slot, during which the relay node R processes the received signal and forwards it to the destination node D . In the case of dynamic allocation of transmission time slot, it is normalized. Particularly, t represents the transmission time in the broadcast slot, and $(1-t)$ denotes the time of the relay forwarding slot. Due to the different environments in which three terminals are located, the channel fadings are independent of each other in two different time slots. Two hops of information (broadcasting and forwarding) occupy different time slots, respectively. Therefore, in two hops, we can use the same subcarriers without considering the interference between them, but the system performance may be limited. So, the subcarriers should be allocated independently in two time slots, which involves the subcarrier pair matching and allocation issue.

The distribution coefficient of subcarrier pairs $C_{m,n}^k \in \{0, 1\}$ is introduced firstly. $C_{m,n}^k = 1$ represents the case where subcarrier m ($1 \leq m \leq N$) is paired with n ($1 \leq n \leq N$), which is noted as the subcarrier pair (m, n) . That is, user k

transfers the information on the subcarrier m at the first slot and on the subcarrier n at the second slot. On the contrary, $C_{m,n}^k = 0$ means that the subcarrier m is not paired with n .

$$R_{m,n}^k$$

$$= \min \left\{ tB \log_2 \left(1 + \frac{C_{m,n}^k \beta_{k,l} P_S}{N_{R_l}} |h_{S_k R_l}^m|^2 \right), tB \log_2 \left(1 + \frac{C_{m,n}^k P_S}{N_D} |h_{S_k D}^m|^2 \right) + (1-t) B \log_2 \left(1 + \frac{C_{m,n}^k \beta_{k,l} P_l}{N_{R_l}} |h_{R_l D}^n|^2 \right) \right\}, \quad (3)$$

where $|h_{S_k R_l}^m|^2$, $|h_{S_k D}^m|^2$, and $|h_{R_l D}^n|^2$, respectively, represent the channel coefficients on the subcarrier pair (m, n) between the devices i, j ($i, j \in \{S_k, R_l, D\}$). $\beta_{k,l}$ is the relay selection factor, and $\beta_{k,l} = 1$ indicates that the l th relay node participates in the k th user cooperative communication process and vice versa with no participation. The solution of $\beta_{k,l}$ is the problem of relay selection in cooperative communication system; we use the method “dichotomous map” proposed in [17]. P_S is the transmitting power of the user node, and P_l is the transmitting power of the relay node. Respectively, N_{R_l}, N_D are the noises at the relay node R and the destination node D .

According to the Shannon's equation, after the user S_k cooperates with the relay node R_l , the transfer rate on the subcarrier pair (m, n) can be written as [16]

$$P^k = (\zeta (tP_S + (1-t) \beta_{k,l} P_l) + P_{CT} + (t+1) P_{CR}), \quad (4)$$

$$k \in \{1, 2, \dots, K\},$$

where ζ is the reciprocal of the drain efficiency of the power amplifier at the transmitter, P_{CT} is the fixed circuit power at the transmitter, and P_{CR} is the fixed circuit power at the receiver.

From (2), the energy efficiency of user k can be defined as follows:

$$\eta_{EE}^k = \frac{R_{m,n}^k}{P^k}, \quad (5)$$

and the optimal energy-efficient resource allocation is given by

$$\eta_{EE} = \max_{C_{m,n}^k} \sum_{k=1}^K \sum_{m=1}^N \sum_{n=1}^N \left(\min \left\{ tB \log_2 \left(1 + \left(C_{m,n}^k \beta_{k,l} P_S / N_{R_l} \right) |h_{S_k R_l}^m|^2 \right), tB \log_2 \left(1 + \left(C_{m,n}^k P_S / N_D \right) |h_{S_k D}^m|^2 \right) \right. \right. \\ \left. \left. + (1-t) B \log_2 \left(1 + \left(C_{m,n}^k \beta_{k,l} P_l / N_{R_l} \right) |h_{R_l D}^n|^2 \right) \right\} \right. \\ \left. \cdot \left(\zeta (tP_S + (1-t) \beta_{k,l} P_l) + P_{CT} + (t+1) P_{CR} \right)^{-1} \right) \quad (6)$$

$$\text{s.t. C1: } C_{m,n}^k \in \{0, 1\};$$

$$\text{C2: } \sum_m \sum_n C_{m,n}^k = 1 \quad \forall k;$$

$$\text{C3: } \sum_k C_{m,n}^k \leq 1 \quad \forall m, n; \quad (7)$$

$$\text{C4: } 0 \leq P_S, P_l \leq P_{\max};$$

$$\text{C5: } R_{k,l} \geq R_{t_{\min}},$$

where C1 is the subcarrier matching and the allocation coefficient, indicating that the subcarriers have two states of cooperative communication and noncooperative communication; C2 means that a subcarrier is assigned to one user

only; C3 indicates that a user selects only one subcarrier to cooperate; C4 is defined as the power limitation between users and relay nodes; and C5 is the transmission rate, and $R_{t_{\min}}$ is the minimum transmission rate, which is QoS.

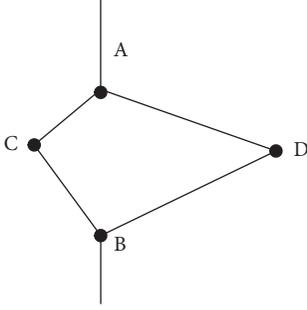


FIGURE 2: Intelligent water drops algorithm.

3. Proposed Optimal Resource Allocation Scheme

In this section, we proposed an improved algorithm to solve the optimal resource allocation formulated in (6).

3.1. Intelligent Water Drops Algorithm. Intelligent water drops (IWDs) algorithm is an intelligent algorithm introduced by Shah-Hosseini inspired by the flow of natural water drops, which construct a solution by cooperation with each other [18]. The IWDs are associated with two properties: the amount of soil in the path and the velocity of the IWDs. When the water drops pass through different paths, the change of the soil quantity on the different paths is different due to the difference of the path distance. When the subsequent water drops face different available paths, they are more likely to choose a path with less soil, in which the IWDs move faster. As shown in Figure 2, the initial amounts of soil at two paths are the same; when $t = 0$, different intelligent water drops will select the paths ACB and ADB with the same probability. After the intelligent water drops pass these two paths, due to a shorter path of ACB, the water drops selecting this path will run faster and carry a lot of soil, leading the soil on the path ACB to be less than that on ADB after the initial iteration. At $t = 1$, the intelligent water drops will select a shorter path, ACB, with a greater probability. After repeated feedback at many times, the intelligent water drops will find the shortest path between A and B. The intelligent water drops algorithm draws on the feedback mechanism of changing soil quantity on the path and completes iterative search.

3.2. Improved Intelligent Water Drops Algorithm. The main drawback of the IWDs algorithm is the low speed at the early stage of training. Because the total amount of soil is the same on all paths, the intelligent water drops will randomly select a path, even if that path is not the optimal one; this will change the amount of soil on that path, resulting in the phenomenon of path dependence in the subsequent iteration process. Therefore, other water drops are inclined more to select that path and many invalid searching paths appear.

Genetic Algorithm (GA) [19] is an adaptive heuristic search algorithm based on the biological evolution process, reserving well-adapted individuals in the process of crossover and mutation, and after several evolutions, the optimal

solution of the objective function is obtained. It starts the iterative process in individual population, which makes it easy to achieve expansion and algorithm fusion. GA has the characteristics of implicit parallelism and strong global search ability [20], which can quickly seek the solution in search space without trapping into the local optimal solution. The local search occurs in GA when the value of $gIter_{max}$ is too small, and GA will stop iterative search without finding the optimal solution; then IWDs algorithm begins, thus affecting the searching efficiency. In practice, GA encounters premature convergence problems.

As mentioned previously, IWDs algorithm is prone to many ineffective searches in the early stage, and the local search ability of the GA is limited in later period. We proposed a novel algorithm to improve the IWDs algorithm with GA. That is, in the early stage of the process, the global search ability of the GA is applied to achieve the rapid optimal solution, which is used as the initial solution of the IWDs algorithm. Finally, the global optimal solution is obtained by the characteristics of fast convergence of IWDs algorithm.

The flowchart of improved IWDs algorithm is shown in Figure 3.

Two main problems in improved IWDs algorithm should be mentioned as follows.

(1) *Value of Boundary Condition $gIter_{max}$.* In the calculation process, the convergence is different due to different scale of data. The boundary condition $gIter_{max}$ of GA and IWDs algorithm should be determined by the population size. The local search occurs in GA when the value of $gIter_{max}$ is too small, and GA will stop iterative search without finding the optimal solution; then the algorithm transfers to IWDs algorithm, thus affecting the searching efficiency. On the contrary, if the value of $gIter_{max}$ is too large, it leads to the slow convergence in GA due to redundant computing. Furthermore, the early maturing of GA causes the phenomenon of path dependence in IWDs algorithm.

According to the convergence analysis of GA based on the Markov chain model mentioned in [21], the value of $gIter_{max}$ is expressed as

$$gIter_{max} = \frac{N(f^* - f^*(X_1))}{2\sqrt{K}p_c p_m p_{s_{min}} f^*(X_1)}, \quad (8)$$

where N and K are the numbers of subcarriers and users, respectively. p_c and p_m are the probability of crossover and mutation, $p_{s_{min}}$ is the minimum selection probability of the nonoptimal individual, $f^*(X_1)$ is the fitness value of the best individual in initial population, and f^* is the best individual fitness value in current population.

(2) *Intelligent Water Drops Soil Quantity Initialization.* GA achieves an optimal solution and many relatively good solutions, which are received with different weights and are used for the soil initialization of intelligent water drops. That is,

$$soil(i, j) = Intsoil * \left(1 - L_*^\alpha - \sum_{\tau=1}^3 L_\tau^{\beta_\tau} \right), \quad (i, j) \in L. \quad (9)$$

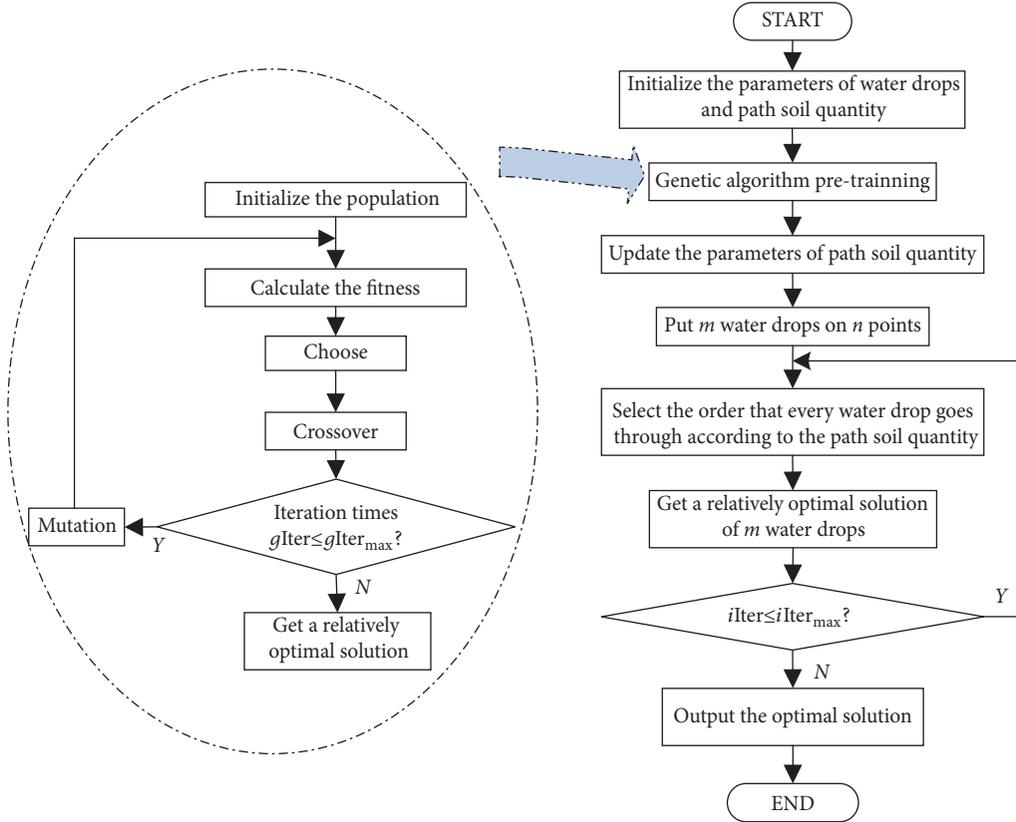


FIGURE 3: Flowchart of improved intelligent water drops algorithm.

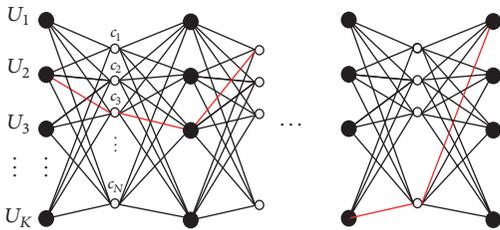


FIGURE 4: Two-level neural network model.

In particular, L_* is the optimal solution generated by GA. L_τ ($\tau = 1, 2, 3$) refers to the three optimal solutions generated by GA and sorted by the fitness value. α and β_τ , respectively, refer to the weight of solution at the initialization.

4. Optimal Resource Allocation Scheme

4.1. Optimal Resource Allocation Model. The model in (6) is the combinatorial optimization problem, which should meet a series of continuous or discrete conditions to obtain the optimal resource allocation. So, we present two-level neural network model to solve the problems of resource allocation in cooperative communication system and then seek the optimal path with improved IWDs algorithm.

The two-level neural network model is shown in Figure 4. It is composed of two types of network nodes and the connection edges between adjacent nodes. The basic unit is

rows. A row of K master nodes represents the K users and a row of N secondary nodes represents the N unallocated subcarriers in cooperative communication system. The user's primary nodes are a total of $K + 1$ rows. A new row inserted between two adjacent rows represents secondary nodes of subcarriers. A path between two roles of primary nodes presents the two selected subcarriers, and then this path is weighted according to (6). From the first row to the $K + 1$ column, a path passing $2K$ nodes represents a possible allocation scheme. In this way, the optimal energy-efficient resource allocation in cooperative communication system transfers into the path planning in this two-level neural network. The red path in this figure represents a possible allocation scheme.

4.2. Optimal Resource Allocation Scheme. The optimal resource allocation flow of improved IWDs algorithm is described as follows.

Step 1. Initialize global static parameters; the amount of IWDs $N_{IWD} = K$.

Step 2. Pretrain Genetic Algorithm.

Step 2.1. Initialize the population and the static parameters.

Step 2.2. Calculate the population fitness according to (6).

Step 2.3. Operate the heredity and mutation on the population according to the boundary conditions of (8); operate the mutation and iteration until the end of loop.

TABLE 1: Simulation parameters.

Parameters	Values
Total system bandwidth B	15 MHz
Number of subcarrier N	64
Subcarrier mean signal to noise ratio	38 dB
Path-loss factor	3.5
Maximum delay extension	4 μ m
Maximum Doppler shift	30 Hz
Channel status information update cycle	0.5 ms
Maximum transmit power P_{\max}	30 dBm
Circuit power P_c	27 dBm
Transmitter power amplifier efficiency $1/\xi$	38%
User minimum transfer rate R_{\min}	1.2 Mbps

Step 3. Place N_{IWD} water drops on primary nodes on the left as shown in Figure 4, and iterate each intelligent water drop according to Steps 4 and 5.

Step 4. Iterate intelligent water drops.

Step 4.1. Initialize the intelligent water drops' parameters and the soil quantity on the path according to (9).

Step 4.2. Empty the Tabu list, and list all subcarriers as assignable.

Step 4.3. With the Tabu list, calculate the probability of all selectable paths (subcarrier pair scheme), and select the most suitable subcarrier pair (m, n) as the user's allocation scheme.

Step 4.4. Put the subcarrier m in Tabu list 1 and subcarrier n in Tabu list 2, which indicates that this subcarrier pair has been occupied by the system in two time slots.

Step 4.5. Update the amount of path soil and soil carried by the intelligent water drops.

Step 4.6. Intelligent water drops pass through the path to the next primary node; set the primary node as the initial position of the next path selection, repeat the steps from Step 4.3 to Step 4.5 until the water drops reach the model on the right side as shown in Figure 4, and this loop ends.

Step 5. At the end of the current iteration, calculate the optimal solution of all paths of water drops, and update the total amount of the path soil according to the optimal solution.

Step 6. Determine whether the number of iterations satisfies $i\text{Iter} \leq i\text{Iter}_{\max}$; if it does, then repeat Steps 4 and 5; otherwise, it goes to the end of program and optimal solution is shown.

5. Numerical Analysis

In this section, we evaluate the performance of the proposed optimal resource allocation scheme via simulation on MATLAB 2014b. The parameters are shown in Table 1.

5.1. Performance Analysis of Proposed Optimal Scheme. In Figure 5, if the number of users in system is 10, the

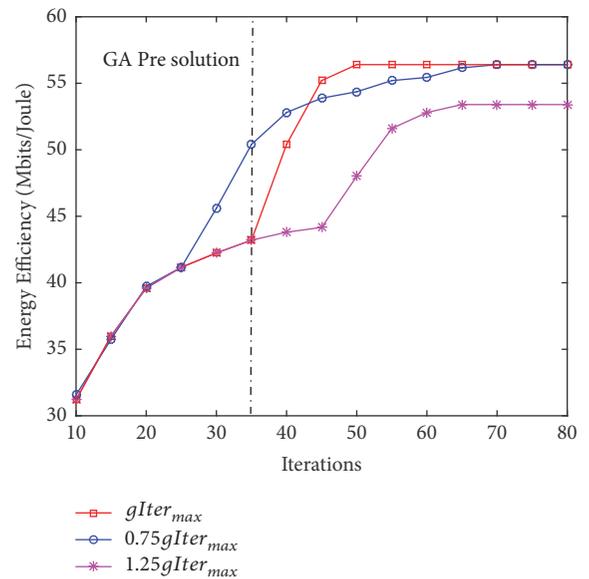


FIGURE 5: Energy efficiency comparison of different $g\text{Iter}_{\max}$.

relationship between energy efficiency and $g\text{Iter}_{\max}$ is depicted. The proposed algorithm is combined with GA and IWDs algorithm. That is, after pretraining by GA at early period of iterations, the performance is further improved by using IWDs at later period. It is obvious that when the value of boundary condition is $0.75 * g\text{Iter}_{\max}$, the premature convergence occurs. In this case, it is inefficient that IWDs are operated due to GA getting stuck at local optimal solution. When the value of boundary condition is $1.25 * g\text{Iter}_{\max}$, the performance declines, and the premature convergence of GA makes the phenomenon of path dependence in IWDs occur.

5.2. Performance Comparison of Variant Algorithms. Figure 6 illustrates the performance comparison of improved IWDs algorithm in terms of energy efficiency. Obviously, the energy efficiency obtained by employing the proposed algorithm is much higher than that of GA and Ant Colony Optimization (ACO) algorithm mentioned in [22]. That is, the better performance is achieved by using improved IWDs algorithm.

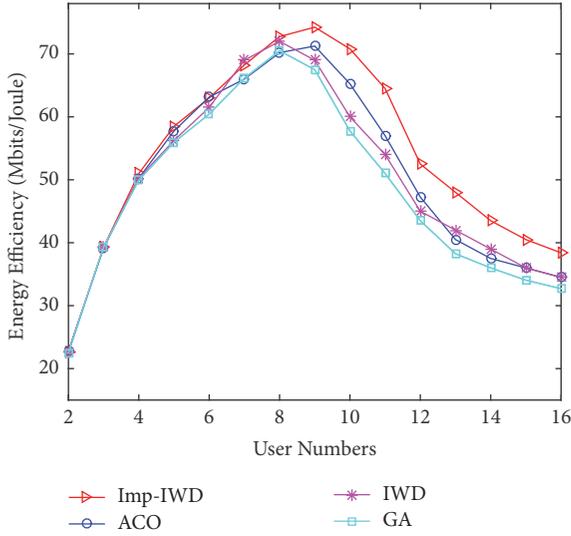


FIGURE 6: Energy efficiency comparison of different user numbers.

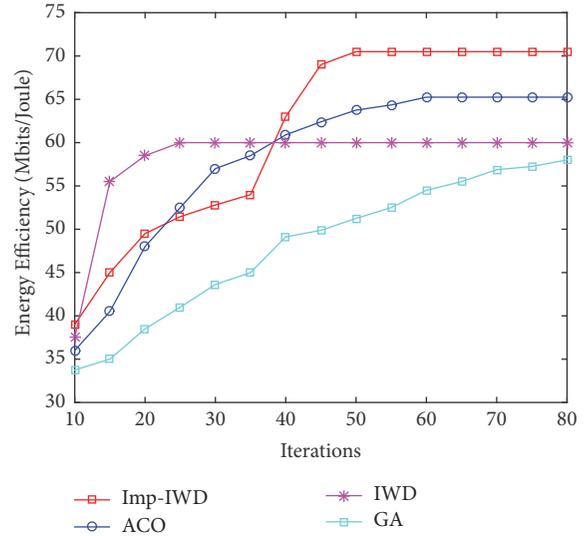


FIGURE 8: The convergence rate of different algorithms.

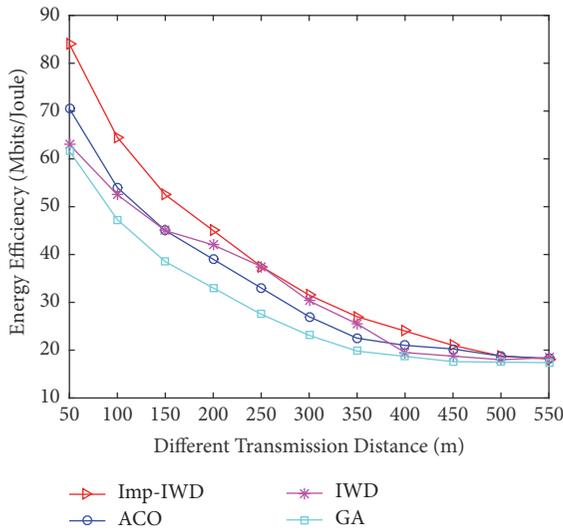


FIGURE 7: The system energy efficiency comparison of different transmission distance.

However, the algorithmic stability of IWDs is poor for its local optimum. Meanwhile, the performance increases at the beginning and then decreases, while the number of users increases. Due to limitation of system resources, when the number of users increases, it results in a lack of system resources and furthermore the performance degradation.

If the number of users is 10, the radius is 600 meters. Figure 7 shows the relationship between energy efficiency and the distance between the users and BS, comparison of different resource allocation schemes. It is proven that the proposed improved IWDs algorithm obtains better performance than that of algorithms in [22]. At the same time, with the increasing of transmission distance, the energy efficiency is gradually reduced. When the transmission distance is far, the channel condition becomes more severe, so the

transmitting power increases to overcome the path loss, resulting in reduced system energy efficiency.

Figure 8 shows the convergence rate of different algorithms if the number of users is 10. The improved IWDs algorithm has a faster convergence rate and obtains a better performance, while the original IWDs algorithm encounters premature convergence.

6. Conclusion

We have addressed the optimal resource allocation for uplink single-cell multiuser cooperative communication system. With the goal of optimizing energy efficiency, the improved IWDs algorithm has faster convergence rate, which achieves better resource allocation. The performance evaluation results demonstrate the effectiveness of the proposed solution.

It is important to notice that a subcarrier is assigned to a user only in this paper. In the practical communication scenario, users' data are abrupt and asymmetric. In the future, a real-time resource allocation strategy can be established to meet the demands of rapid development of mobile services.

Data Availability

The network parameter data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of Guizhou Province (Grant no. [2016]1054), Joint Natural Science Foundation of Guizhou Province (Grant

no. LH[2017]7226), Academic Talent Training and Innovation Exploration Project of Guizhou University (Grant no. [2017]5788), and Graduate Student Innovation Foundation of Guizhou University (Grant no. 2017015).

References

- [1] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, and M. Olsson, "How much energy is needed to run a wireless network," *Wireless Communications*, vol. 18, no. 5, pp. 40–49, 2011.
- [2] Y. Zhao, Y. Z. Wang, and H. Y. Yang, "Introduction to green communication technologies," *Information & Communications*, vol. 5, pp. 254–255, 2016.
- [3] J. C. Xu, "Current Situation and Countermeasures of Energy Conservation and Emission Reduction in Telecommunication Industry in China," *Modern Economic Information*, vol. 7, p. 304, 2014.
- [4] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity-part I: system description," *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.
- [5] J. N. Laneman, D. N. Tse, and G. Wornell, "Cooperative diversity in wireless networks: efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.
- [6] J. C. Xiao, *Development of Virtual Massive MIMO Channel Measurement System and Analysis of Large Scale Fading*, Beijing Jiaotong University, 2016.
- [7] L. Li and C. Chigan, "A Virtual MIMO based anti-jamming strategy for cognitive radio networks," in *Proceedings of the IEEE International Conference on Communications*, pp. 1–6, 2016.
- [8] C. Y. Wong, R. S. Cheng, K. B. Letaief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, 1999.
- [9] K. N. Pappi, P. D. Diamantoulakis, H. Otrok, and G. K. Karagiannidis, "Cloud compute-and-forward with relay cooperation," *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, pp. 3415–3428, 2015.
- [10] S. Yousaf, N. Javaid, U. Qasim, N. Alrajeh, Z. A. Khan, and M. Ahmed, "Towards reliable and energy-efficient incremental cooperative communication for wireless body area networks," *Sensors*, vol. 16, no. 3, 2016.
- [11] H. Li, N. Jaggi, and B. Sikdar, "Relay scheduling for cooperative communications in sensor networks with energy harvesting," *IEEE Transactions on Wireless Communications*, vol. 10, no. 9, pp. 2918–2928, 2011.
- [12] A. Doosti-Aref and A. Ebrahimzadeh, "Adaptive Relay Selection and Power Allocation for OFDM Cooperative Underwater Acoustic Systems," *IEEE Transactions on Mobile Computing*, vol. 17, no. 1, pp. 1–15, 2018.
- [13] X. Yin Long, *Research on Energy Efficiency Based Cooperative Communication System*, 2014.
- [14] V. Rodoplu and T. H. Meng, "Bits-per-joule capacity of energy-limited wireless networks," *IEEE Transactions on Wireless Communications*, vol. 6, no. 3, pp. 857–864, 2007.
- [15] W.-S. Lai, T.-H. Chang, and T.-S. Lee, "Distributed dynamic resource allocation for ofdma-based cognitive small cell networks using a regret-matching game approach," *Game Theory Framework Applied to Wireless Communication Networks*, 2016.
- [16] Y. Xu, Z. Bai, B. Wang et al., "Energy-efficient power allocation scheme for multi-relay cooperative communications," in *Proceedings of the International Conference on Advanced Communication Technology*, pp. 260–264, IEEE, 2014.
- [17] Z. Y. Liu, H. Tang, P. Mao, S. M. Liu, and L. Feng, "Relay selection in cooperative communication with bipartite graph," *Application Research of Computers*, vol. 4, 2018.
- [18] H. Shah-Hosseini, "Problem solving by intelligent water drops," in *Proceedings of the Evolutionary Computation, CEC*, IEEE Congress, pp. 3226–3231, 2007.
- [19] Y. J. Ma and W. X. Yun, "Research progress of genetic algorithm," *Application Research of Computers*, vol. 4, pp. 1201–1210, 2012.
- [20] M. Pei, *Ant Colony Optimization Algorithm in the Allocation of Cloud Computing Resources*, Shandong Normal University, 2015.
- [21] S. Q. Kuang, *Parameter Adaptive Controlling and Convergence Theory for Genetic Algorithms*, Central South University, 2009.
- [22] A. Zainaldin, H. Halabian, and I. Lambadaris, "Joint Resource Allocation and Relay Selection in LTE-Advanced Network Using Hybrid Co-Operative Relaying and Network Coding," *IEEE Transactions on Wireless Communications*, vol. 15, no. 6, pp. 4348–4361, 2016.

Research Article

Energy Efficiency Maximized Resource Allocation for Opportunistic Relay-Aided OFDMA Downlink with Subcarrier Pairing

Tao Wang , Chao Ma, Yanzan Sun, Shunqing Zhang, and Yating Wu

Shanghai Institute for Advanced Communication and Data Science, Key laboratory of Specialty Fiber Optics and Optical Access Networks, Joint International Research Laboratory of Specialty Fiber Optics and Advanced Communication, Shanghai University, Shanghai, China

Correspondence should be addressed to Tao Wang; twang@shu.edu.cn

Received 12 February 2018; Accepted 24 April 2018; Published 29 May 2018

Academic Editor: Paul Honeine

Copyright © 2018 Tao Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper studies the energy efficiency (EE) maximization for an orthogonal frequency division multiple access (OFDMA) downlink network aided by a relay station (RS) with subcarrier pairing. A highly flexible transmission protocol is considered, where each transmission is executed in two time slots. Every subcarrier in each slot can either be used in direct mode or be paired with a subcarrier in another slot to operate in relay mode. The resource allocation (RA) in such a network is highly complicated, because it has to determine the operation mode of subcarriers, the assignment of subcarriers to users, and the power allocation of the base station and RS. We first propose a mathematical description of the RA strategy. Then, a RA algorithm is derived to find the globally optimum RA to maximize the EE. Finally, we present extensive numerical results to show the impact of minimum required rate of the network, the user number, and the relay position on the maximum EE of the network.

1. Introduction

With rapid growth of multimedia services, requirements for high-speed wireless communications are growing fast. To meet these requirements, telecom operators arrange a large number of base stations, which lead to a high amount of energy consumption [1]. To address this issue, many scholars have proposed energy-saving methods to minimize total energy consumption for a variety of wireless communication systems, such as Device-to-Device (D2D) communications, wireless sensor networks (WSNs), and cellular networks [2, 3]. Recently, energy efficiency- (EE-) based optimization design, which aims to maximize the EE defined as the number of transmitted bits per Joule total energy consumption, has attracted much interest from academia and industry [4–10].

Orthogonal frequency division multiple access (OFDMA) has been widely recognized as one of the dominant wireless technologies for high-data-rate wireless multimedia services. One of the main reasons behind this fact is that performance of OFDMA systems can be significantly improved by proper

resource allocation (RA) when transmitter channel state information (CSI) is available [11–16]. Lately, relay-aided cooperation schemes have been widely used in combination with OFDMA networks to improve spectral efficiency. Under the constraint of guaranteeing users' communication rate, some works designed RA algorithms to minimize the total transmission power of networks [17–24]. In [17], to minimize the maximum value between transmission power of the BS and transmission power of all the RSs, Muller et al. designed a RA algorithm for decode-and-forward (DF) relay-aided OFDMA networks. In [21], knowing the relay selection, Huang et al. proposed an optimization algorithm of subcarriers and power allocation to minimize the total power of BS and all the RSs. Chen et al. designed a strategy of user assignment for subcarriers, RS's choice, and modulation scheme to minimize the total transmission power of networks for amplify-and-forward (AF) relay-aided downlink OFDMA networks [23]. The above works ignore the influence of circuit power of the BS and RSs, so these algorithms cannot ensure high EE of networks.

It is interesting to further study how to improve EE of relay-aided OFDMA networks. When users lie outside the BS's radio coverage, EE maximized RA problems for OFDMA networks using subcarrier-pair-based DF protocols have been addressed in [25–29]. In these works, every subcarrier in the first time slot is paired with a subcarrier in the second time slot for the relay-aided transmission. In most cases, BS can also transmit messages to users directly; designing EE maximized RA algorithm for flexible transmission protocols is more meaningful. In [30–34], the authors adopted more flexible transmission protocols and proposed EE maximized RA algorithms for downlink OFDMA networks when the total transmission power is constrained to be smaller than a prescribed value. In this case, when the network reaches the maximum EE, the total communication rate might be too small to meet the needs of users.

In this paper, we focus on the optimum energy-efficient RA for downlink OFDMA with a RS using subcarrier-pair-based DF relaying, when the sum rate is constrained above a prescribed value. An opportunistic relay-aided transmission protocol is considered. User message bits are transmitted during two consecutive equal-duration time slots. In the first slot, the BS broadcasts OFDM symbols to RSs and users. In the second slot, subcarriers in direct mode can transmit to users directly; other subcarriers can be paired with subcarriers in first slot to transmit messages with the help of the RS. To be more specific, our contributions are summarized as follows:

- (i) An EE maximized RA problem is formulated, and a polynomial-complexity algorithm is designed to find the optimum RA to maximize the EE based on the Dinkelbach method as well as the dual method to solve a subproblem.
- (ii) Extensive numerical results are shown to exhibit the impact of system parameters (including minimum communication rate required by the network, RS deployment, and user number) on the network EE.

The rest of this paper is organized as follows. In the next section, the transmission protocol of the network is described. After that, the EE maximized RA algorithm is developed in Section 3. Numerical experiments are shown in Section 4. Finally, Section 5 concludes the paper.

Notations. $\mathcal{E}(x) = \log_2(1 + x)$.

2. Network Model and Transmission Protocol

2.1. General Introduction of the Network and Protocol. We consider an OFDMA downlink network as shown in Figure 1. The network under consideration consists of a BS, a RS, and multiple users. Both the BS and the RS adopt OFDMA scheme using the same frequency band of bandwidth B Hz and with K subcarriers, which means that each OFDM symbol has a duration of K/B seconds.

For illustration purpose, the channel coefficient and noise-power normalized channel gain at any subcarrier k ($k \in \{1, \dots, K\}$) from BS to RS, from RS to any user u ,

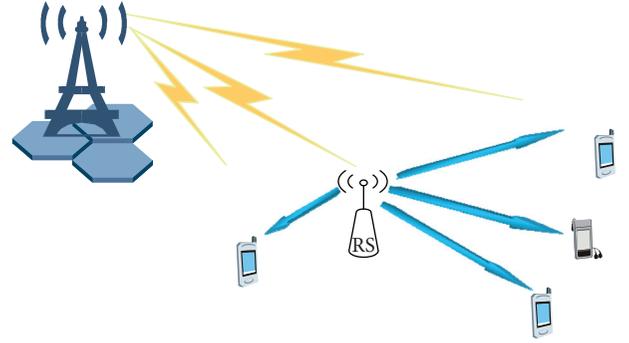


FIGURE 1: The relay-aided OFDMA downlink network under consideration.

TABLE 1: Channel coefficient and gain for subcarrier $k \in \{1, \dots, K\}$.

Channel for subcarrier k	Coefficient	Gain
BS to RS	h_{sr}^k	$G_{sr}^k = \frac{ h_{sr}^k ^2}{\sigma^2}$
BS to user u	h_{su}^k	$G_{su}^k = \frac{ h_{su}^k ^2}{\sigma^2}$
RS to user u	h_{ru}^k	$G_{ru}^k = \frac{ h_{ru}^k ^2}{\sigma^2}$

and from BS to any user u are defined in Table 1, where σ^2 is assumed as the power of additive white Gaussian noise at each subcarrier.

The transmission protocol under consideration is carried out as follows. Each transmission needs two time slots denoted by slot-1 and slot-2, respectively. Each subcarrier in every slot can either operate in direct mode (it is used by the BS for transmission to a user directly) or be paired with a subcarrier in another slot to operate in relay mode (the relay helps the transmission as explained in Section 2.2). The protocol is illustrated by Figure 2. In the following subsections, the transmission procedures for the direct and relay mode are explained in detail.

2.2. Transmission for a Subcarrier Pair in Relay Mode. Suppose that a subcarrier k in slot-1 is paired with a subcarrier l in slot-2 to operate in relay mode, and this relay-link is assigned to user u . Denote this subcarrier pair as (k, l, u) , which is shown in Figure 3. Over this link, the transmission procedure is carried out as follows.

In slot-1, the BS broadcasts a symbol x to both the relay and the user with power P_1 . The received signal at the relay is expressed as

$$y_r = \sqrt{P_1} h_{sr}^k x + n_r, \quad (1)$$

and that at the user is expressed as

$$y_{u,1} = \sqrt{P_1} h_{su}^k x + n_{u,1}. \quad (2)$$

At the end of slot-1, the relay decodes y_r and recovers x . To enable the relay to successfully decode the message bits, the

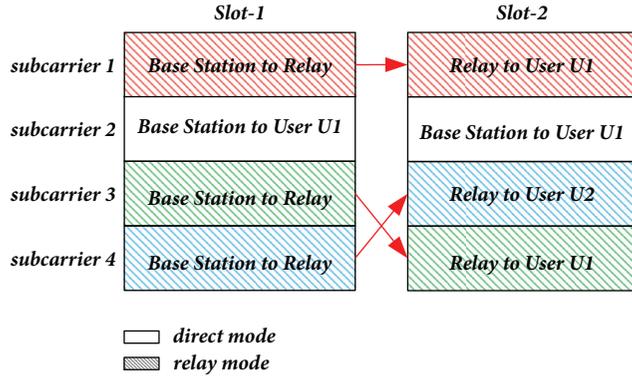


FIGURE 2: The transmission procedure over the subcarrier pair (k, l, u) .

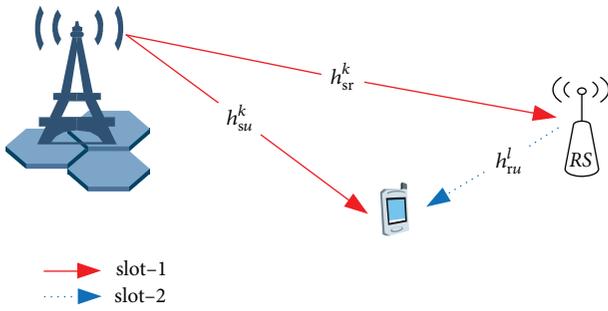


FIGURE 3: The transmission procedure over the subcarrier pair (k, l, u) .

maximum transmission rate is no greater than $\log_2(1 + P_1 G_{sr}^k)$ bits/symbol.

In slot-2, the relay simply emits x at subcarrier l to the user with power P_2 . The received signal at the user is expressed as

$$y_{u,2} = \sqrt{P_2} h_{ru}^l x + n_{u,2}. \quad (3)$$

At the end of slot-2, the source combines $y_{u,1}$ and $y_{u,2}$ with maximum-ratio combining (MRC) to maximize the received SNR. The final signal used for decoding at the user can be expressed as

$$z = \alpha y_{u,1} + \beta y_{u,2}. \quad (4)$$

It can be shown that when $\alpha = \sqrt{P_1} h_{su}^k$ and $\beta = \sqrt{P_2} h_{ru}^l$, the MRC is achieved [35]. The maximum received SNR is

$$\text{SNR}_{k,l,u} = G_{su}^k P_1 + G_{ru}^l P_2. \quad (5)$$

As a result, the maximum transmission rate over this relay link should be the minimum between the source-relay rate and the source-relay-user rate. It can be evaluated as

$$C_{klu} = \min \left\{ \mathcal{C}(\text{SNR}_{klu}), \mathcal{C}(G_{sr}^k P_1) \right\}, \quad (6)$$

in the unit of bits/symbol.

Suppose that P' is the sum power of P_1 and P_2 ; the optimums P_1 and P_2 for maximizing the rate are the optimum solution for

$$\begin{aligned} \max_{P_1, P_2} \quad & \min \{ P_1 G_{sr}^k, P_1 G_{su}^k + P_2 G_{ru}^l \}, \\ \text{s.t.} \quad & P_1 + P_2 = P', \quad P_1 \geq 0, \quad P_2 \geq 0. \end{aligned} \quad (7)$$

Using the same method as in [36], it can easily be shown that the optimums P_1 and P_2 are

$$\begin{aligned} P_1 &= \begin{cases} \frac{G_{ru}^l}{\Delta_{u,k} + G_{ru}^l} P' & \text{if } \min \{ G_{sr}^k, G_{ru}^l \} > G_{su}^k, \\ P' & \text{if } \min \{ G_{sr}^k, G_{ru}^l \} \leq G_{su}^k, \end{cases} \\ P_2 &= \begin{cases} \frac{G_{sr}^k - G_{su}^k}{\Delta_{u,k} + G_{ru}^l} P' & \text{if } \min \{ G_{sr}^k, G_{ru}^l \} > G_{su}^k, \\ 0 & \text{if } \min \{ G_{sr}^k, G_{ru}^l \} \leq G_{su}^k, \end{cases} \end{aligned} \quad (8)$$

where $\Delta_{u,k} = G_{sr}^k - G_{su}^k$.

The maximum rate associated with the above solution is equal to

$$\begin{aligned} C_{klu} &= \frac{B}{2K} \mathcal{C}(G_{klu} P') \quad (\text{bits/second}), \\ G_{klu} &= \begin{cases} \frac{G_{sr}^k G_{ru}^l}{\Delta_{u,k} + G_{ru}^l} & \text{if } \min \{ G_{sr}^k, G_{ru}^l \} > G_{su}^k, \\ \min \{ G_{sr}^k, G_{su}^k \} & \text{if } \min \{ G_{sr}^k, G_{ru}^l \} \leq G_{su}^k. \end{cases} \end{aligned} \quad (9)$$

2.3. Transmission for a Subcarrier in Direct Mode. Every subcarrier in either slot-1 or slot-2 can be assigned to operate in direct mode. In this mode, a direct link from the BS to a certain user is formed at this subcarrier. Suppose that a subcarrier k in either slot-1 or slot-2 is assigned to user u and operates in the direct mode, and BS uses power P for this subcarrier. Therefore, the average transmission rate over this direct link can be evaluated as

$$R(P) = \frac{B}{2K} \mathcal{C}(G_{su}^k P) \quad (\text{bits/second}). \quad (10)$$

3. Energy-Efficient RA Algorithm Design

Before data transmission, we assume that the BS controller knows all CSI, that is, $\{G_{sr}^k, G_{su}^k, G_{ru}^k \mid \forall k, \forall u\}$. Based on the a priori knowledge, the BS runs an algorithm to find the optimum RA strategy to maximize the network EE.

To be more specific, the RA algorithm for the transmission protocol needs to be optimized:

- (i) *Subcarrier operation mode*: how to decide whether each subcarrier should operate in direct or relay mode
- (ii) *Subcarrier pairing for relay mode*: how to pair subcarriers in relay mode
- (iii) *Subcarrier assignment to users*: how to allocate subcarriers to users

- (1) evaluate $G_{klu}, \forall k, l, u$;
- (2) $\theta_{\min} = 0$; θ_{\max} is set by a large value; $\theta = \theta_{\max}$; $\delta = 10^{-3}$;
- (3) **while** $|F(\theta)| \leq \delta$ **do**
- (4) solve (P2) for $\mathbf{S}(\theta)$ using Algorithm 2;
- (5) update θ by $\min\{\theta_{\max}, \max\{R(\mathbf{S}(\theta))/P(\mathbf{S}(\theta)), \theta_{\min}\}\}$;
- (6) **end while**
- (7) output θ as η^* and $\mathbf{S}(\theta)$ as the optimum RA strategy.

ALGORITHM 1: The algorithm to solve (P1).

- (iv) *Power allocation*: how to allocate BS's and RS's transmission power for each subcarrier

To design the RA algorithm, we proceed as follows. First, a mathematical description of the RA strategy and network EE is proposed in Section 3.1. Then, a Dinkelbach-method-based RA algorithm, namely, Algorithm 1, is designed in Section 3.2. Moreover, algorithms called by Algorithm 1 are designed in Sections 3.3 and 3.4.

3.1. Description of the RA Strategy and Network EE. We first define the following variables to describe the RA strategy:

- (i) $t_{klu} \in \{0, 1\}, \forall k, l, u$: indicating subcarrier k in slot-1 is paired with subcarrier l in slot-2 when $t_{klu} = 1$.
- (ii) $p_{klu} \geq 0, \forall k, l, u$: indicating the total transmission power (i.e., $P_1 + P_2$ as mentioned in Section 2.2) for the subcarrier pair (k, l, u) .
- (iii) $t_{klab} \in \{0, 1\}, \forall k, l, a, b$: indicating that subcarrier k in slot-1 and subcarrier l in slot-2 are, respectively, allocated to user a and user b when $t_{klab} = 1$.
- (iv) $p_{klab}^1 \geq 0$ and $p_{klab}^2 \geq 0$: respectively indicating the BS's transmission power for subcarrier k in slot-1 and l in slot-2 for the direct mode.

Let us define a RA strategy as $\mathbf{S} = \{\mathbf{I}, \mathbf{P}\}$, where \mathbf{I} collects all indicator variables and \mathbf{P} collects all power variables. A feasible \mathbf{S} must satisfy

$$\begin{aligned} t_{klu} &\in \{0, 1\}, \\ t_{klab} &\in \{0, 1\}, \\ &\forall k, l, u, a, b, \end{aligned} \quad (11)$$

$$\sum_l \left(\sum_u t_{klu} + \sum_{ab} t_{klab} \right) \leq 1, \quad \forall k, \quad (12)$$

$$\sum_k \left(\sum_u t_{klu} + \sum_{ab} t_{klab} \right) \leq 1, \quad \forall l, \quad (13)$$

$$\begin{aligned} p_{klu} &\geq 0, \\ p_{klab}^1 &\geq 0, \\ p_{klab}^2 &\geq 0, \\ &\forall k, l, u, a, b, \end{aligned} \quad (14)$$

where constraints (12) and (13) guarantee that each subcarrier in either slot-1 or slot-2 must operate in a single mode and be assigned to a single user.

For given \mathbf{S} , the network EE is formulated as

$$\eta(\mathbf{S}) = \frac{R(\mathbf{S})}{P(\mathbf{S})} \quad (\text{bits/Joule}), \quad (15)$$

where $R(\mathbf{S})$ represents the sum rate for the network:

$$\begin{aligned} R(\mathbf{S}) &= \frac{B}{2K} \left(\sum_{klu} t_{klu} \mathcal{C}(p_{klu} G_{klu}) \right. \\ &\quad \left. + \sum_{klab} t_{klab} \left(\mathcal{C}(p_{klab}^1 G_{su}^k) + \mathcal{C}(p_{klab}^2 G_{su}^l) \right) \right), \end{aligned} \quad (16)$$

and $P(\mathbf{S})$ is the sum power consumption for the network:

$$\begin{aligned} P(\mathbf{S}) &= \frac{1 + \alpha}{2} \left(\sum_{klu} t_{klu} p_{klu} + \sum_{klab} t_{klab} (p_{klab}^1 + p_{klab}^2) \right) \\ &\quad + P_{\text{cir}}, \end{aligned} \quad (17)$$

with α being the loss factor of the power amplifiers (PA) used by the BS and RS and P_{cir} representing the total power consumption by the BS's and RS's circuit devices.

3.2. Dinkelbach-Method-Based RA Algorithm Design. We consider the EE maximization problem when the sum rate must be greater than a prescribed value R_{req} . This problem can be expressed as (P1):

$$\begin{aligned} \max_{\mathbf{S}} \quad &\eta(\mathbf{S}) \\ \text{s.t.} \quad &(11), (12), (13), (14) \\ &R(\mathbf{S}) \geq R_{\text{req}}. \end{aligned} \quad (18)$$

It can be seen that solving problem (P1) is highly challenging due to the following reasons:

- (i) $\eta(\mathbf{S})$ has a fractional structure, which is highly non-linear.
- (ii) (P1) is a mixed-integer problem containing both binary and continuous variables.

To solve (P1) for the optimum RA (denoted by \mathbf{S}^*) and the maximum EE (denoted by η^*), we make use of the Dinkelbach method. To be more specific, we define a parameter θ and a function $F(\theta)$ as the optimum objective value for the following problem (P2):

$$\begin{aligned} F(\theta) &= \max_{\mathbf{S}} R(\mathbf{S}) - \theta P(\mathbf{S}) \\ \text{s.t.} \quad &(11), (12), (13), (14), (18), \end{aligned} \quad (19)$$

whose optimum solution is $\mathbf{S}(\theta)$.

Based on Dinkelbach method, η^* must satisfy $F(\eta^*) = 0$, and $\mathbf{S}(\eta^*)$ is the optimum solution for problem (P1). Moreover, η^* can be found by iterative procedures as elaborated in [6].

```

(1)  $\mu = 0$ ;
(2) solve (P3) for  $\mathbf{S}_\theta(0)$  using Algorithm 3;
(3) if  $\mathbf{S}_\theta(0)$  is feasible for (P2) then
(4)   output  $\mathbf{S}_\theta(0)$  as  $\mathbf{S}(\theta)$ 
(5) else
(6)    $\mu_{\min} = 0$ ;  $\mu_{\max}$  is set as a sufficiently large value;
(7)   while  $\mu_{\max} - \mu_{\min} > \delta$  do
(8)      $\mu = \frac{\mu_{\max} + \mu_{\min}}{2}$ ;
(9)     solve (P3) for  $\mathbf{S}_\theta(\mu)$  using Algorithm 3;
(10)    if  $\gamma_\theta(\mu) > 0$  then
(11)       $\mu_{\max} = \mu$ ;
(12)    else if  $\gamma_\theta(\mu) < 0$  then
(13)       $\mu_{\min} = \mu$ ;
(14)    else if  $\gamma_\theta(\mu) = 0$  then
(15)      break;
(16)    end if
(17)  end while
(18) end if
(19) Output  $\mathbf{S}_\theta(\mu)$  as  $\mathbf{S}(\theta)$ .

```

ALGORITHM 2: The algorithm to solve (P2).

```

(1) compute  $p_{klu}, p_{klab}^1, p_{klab}^2, \forall k, l, a, b, u$  with (25);
(2) compute  $A_{klu}, B_{klab}, \forall k, l, u, a, b$ ; and  $C_{kl}, \forall k, l$ ;
(3) solve (P5) for its optimum solution  $\{t_{kl}^* \mid \forall k, l\}$  with
    the Hungarian algorithm;
(4) construct the optimum  $\mathbf{I}$  for (P4) by assigning for every
    combination of  $k$  and  $l$ , all entries in  $\{t_{klu}, t_{klab} \mid \forall k, l\}$  to
    zero, except for the one with the metric equal to  $C_{kl}$  to  $t_{kl}^*$ .
(5)  $\mathbf{S}_1 = \{\mathbf{I}, \mathbf{P}_1\}$  is output as  $\mathbf{S}_\theta(\mu)$ .

```

ALGORITHM 3: The algorithm to find $\mathbf{S}_\theta(\mu)$.

Motivated by the above principle, the RA algorithm to solve (P1) for the EE maximized RA strategy is summarized in Algorithm 1. The iterative update of θ has a superlinear convergence rate [37]. We will elaborate on the design of Algorithm 2 to solve (P2) in Section 3.3.

3.3. Design of Algorithm 2 to Solve (P2). Using the same arguments as those in [38], we can show that (P2)'s duality gap is zero; therefore the dual method can be used to solve (P2). To this end, define

(i) μ as Lagrange multiplier for the constraint (18);

(ii) Lagrangian function

$$L_\theta(\mu, \mathbf{S}) = (R(\mathbf{S}) - \theta P(\mathbf{S})) + \mu (R(\mathbf{S}) - R_{\text{req}}); \quad (20)$$

(iii) Lagrange relaxation problem for (P2) as (P3):

$$\begin{aligned} \max_{\mathbf{S}} \quad & L_\theta(\mu, \mathbf{S}) \\ \text{s.t.} \quad & (11), (12), (13), (14) \end{aligned} \quad (21)$$

For (P3), we define its optimum solution as $\mathbf{S}_\theta(\mu)$ and its dual function as $d_\theta(\mu) = L_\theta(\mu, \mathbf{S}_\theta(\mu))$. It can be shown that the dual function is convex of $\mu \geq 0$, and

$$\gamma_\theta(\mu) = R(\mathbf{S}_\theta(\mu)) - R_{\text{req}} \quad (22)$$

is a subgradient of $d_\theta(\mu)$ satisfying

$$\forall \mu', d_\theta(\mu') \geq d_\theta(\mu) + (\mu' - \mu) \gamma_\theta(\mu). \quad (23)$$

The key for the dual method is to find the dual optimum:

$$\mu_\theta^* = \arg \min_{\mu \geq 0} d_\theta(\mu), \quad (24)$$

and then $\mathbf{S}_\theta(\mu_\theta^*)$ is the optimum solution for (P2).

According to the dual method, μ_θ^* is the μ satisfying the following two conditions: (1) $\mu \gamma_\theta(\mu) = 0$ (i.e., the complementary slackness condition) and (2) $\gamma_\theta(\mu) \geq 0$ (i.e., $\mathbf{S}_\theta(\mu)$ is feasible for (P2)). To find μ_θ^* , we use the following method:

(i) First, we compute $\mathbf{S}_\theta(0)$. If $\gamma_\theta(0) \geq 0$, $\mu = 0$ and $\mathbf{S}_\theta(0)$ satisfy the above conditions. Therefore, $\mathbf{S}_\theta(0)$ is the optimum for (P2).

- (ii) Otherwise, $\gamma_\theta(0) < 0$ and $\mu_\theta^* > 0$ must hold. Note that $\gamma_\theta(\mu)$ is increasing of μ . Once $\mu > 0$ satisfying $\gamma_\theta(\mu) = 0$ is found, μ and $\mathbf{S}_\theta(\mu)$ satisfy the above two conditions and hence can be taken as μ_θ^* and $\mathbf{S}(\theta)$, respectively. We will find $\mu > 0$ satisfying $\gamma_\theta(\mu) = 0$ with the bisection method.

To complete this subsection, the above procedures to solve (P2) are summarized in Algorithm 2 as follows. Algorithm 2 has a polynomial complexity with respect to K . We will elaborate on the design of Algorithm 3 to solve (P3) in Section 3.4.

3.4. Design of Algorithm 3 to Solve (P3). We show how to find $\mathbf{S}_\theta(\mu)$ as follows:

- (i) First, the optimum \mathbf{P} for (P3) with fixed \mathbf{I} is found and denoted by $\mathbf{P}_\mathbf{I}$.
- (ii) Second, define $\mathbf{S}_\mathbf{I} = \{\mathbf{I}, \mathbf{P}_\mathbf{I}\}$. Then we find the optimum \mathbf{I} to maximize $L_\theta(\mu, \mathbf{S}_\mathbf{I})$ subject to the constraints on \mathbf{I} in (P3).
- (iii) Finally, $\mathbf{S}_\mathbf{I}$ corresponding to this optimum \mathbf{I} can be taken as $\mathbf{S}_\theta(\mu)$.

As for the first step, the elements in the optimum $\mathbf{P}_\mathbf{I}$ can be computed according to KKT conditions as follows [39]:

$$\begin{aligned} p_{klu} &= \Lambda(\theta, \mu, G_{klu}), \quad \forall k, l, u \\ p_{klab}^1 &= \Lambda(\theta, \mu, G_{su}^k), \quad \forall k, l, a, b \\ p_{klab}^2 &= \Lambda(\theta, \mu, G_{su}^l), \quad \forall k, l, a, b, \end{aligned} \quad (25)$$

where

$$\Lambda(\theta, \mu, G) = \left[\frac{(\mu + 1)B \cdot \log_2(e)}{(1 + \alpha)\theta \cdot K} - \frac{1}{G} \right]^+ \quad (26)$$

As for the second step, it can readily be shown that

$$L_\theta(\mu, \mathbf{S}_\mathbf{I}) = -\mu R_{\text{req}} + \sum_{klu} t_{klu} A_{klu} + \sum_{klab} t_{klab} B_{klab}, \quad (27)$$

where

$$\begin{aligned} A_{klu} &= \frac{(\mu + 1)B}{2K} \mathcal{E}(G_{klu} \Lambda(\theta, \mu, G_{klu})) \\ &\quad - \frac{\theta(1 + \alpha)}{2} \Lambda(\theta, \mu, G_{klu}) \\ B_{klab} &= \frac{(\mu + 1)B}{2K} \mathcal{E}(G_{su}^k \Lambda(\theta, \mu, G_{su}^k)) \\ &\quad - \frac{\theta(1 + \alpha)}{2} \Lambda(\theta, \mu, G_{su}^k) \\ &\quad + \frac{(\mu + 1)B}{2K} \mathcal{E}(G_{su}^l \Lambda(\theta, \mu, G_{su}^l)) \\ &\quad - \frac{\theta(1 + \alpha)}{2} \Lambda(\theta, \mu, G_{su}^l) \end{aligned} \quad (28)$$

Finally, we find the optimum \mathbf{I} for maximizing $L_\theta(\mu, \mathbf{S}_\mathbf{I})$. This problem is equivalent to solving (P4):

$$\begin{aligned} \max_{\mathbf{I}, \{t_{kl} | \forall k, l\}} & \sum_{kl} \sum_{uab} (t_{klu} A_{klu} + t_{klab} B_{klab}) \\ \text{s.t.} & \quad t_{klu} \geq 0, \\ & \quad t_{klab} \geq 0, \end{aligned} \quad (29)$$

$\forall k, l, u, a, b,$

$$\sum_l t_{kl} = 1, \quad \forall k, \quad (30)$$

$$\sum_k t_{kl} = 1, \quad \forall l, \quad (31)$$

$$t_{kl} = \sum_u t_{klu} + \sum_{ab} t_{klab}, \quad \forall k, l \quad (32)$$

Note that

$$\sum_{uab} (t_{klu} A_{klu} + t_{klab} B_{klab}) \leq t_{kl} C_{kl} \quad (33)$$

holds, where $C_{kl} = \max\{\max_u A_{klu}, \max_{a,b} B_{klab}\}$. Call A_{klu} the metric for t_{klu} and B_{klab} the metric for t_{klab} ; the inequality is tightened when all entries of $\{t_{klu}, t_{klab} \mid \forall u, a, b\}$ are assigned to zero, except that the one with the metric equal to C_{kl} is assigned to t_{kl}^* .

Therefore, after problem (P5)

$$\begin{aligned} \max_{\{t_{kl} | \forall k, l\}} & \sum_{kl} t_{kl} C_{kl} \\ \text{s.t.} & \quad (30), (31), \quad t_{kl} \in \{0, 1\}, \quad \forall k, l \end{aligned} \quad (34)$$

is solved for its optimum solution $\{t_{kl}^* \mid \forall k, l\}$, an optimum (32) can be constructed by assigning for every combination of k and l all entries in $\{t_{klu}, t_{klab} \mid \forall k, l\}$ to zero, except for the one with the metric equal to C_{kl} to t_{kl}^* .

Most interestingly, (P5) is a standard assignment problem; hence every entry in $\{t_{kl}^* \mid \forall k, l\}$ is either 0 or 1 and $\{t_{kl}^* \mid \forall k, l\}$ can be found efficiently by the Hungarian algorithm [40]. After knowing Hungarian algorithm, the optimum \mathbf{I} can be constructed according to the way mentioned earlier.

Motivated by the above principle, the method to solve (P2) is summarized in Algorithm 3 as follows. The complexity of computing A_{klu} and B_{klab} is $O(K^2(U + U^2))$. We use the Hungarian algorithm to solve (P5); the complexity is $O(K^3)$. As a result, the complexity of Algorithm 3 is $O(K^2(U + U^2 + K))$.

4. Numerical Experiments and Discussions

We will first introduce system setup for numerical experiments, as well as two benchmark protocols for comparison purpose. Then, results and discussions are presented to show the impact of different parameters on the network EE.

TABLE 2: Network Parameters.

Meanings	Parameters	Values
Bandwidth of the network	B	20 MHz
Subcarriers' number	K	32
Loss factor of the PA	α	0.3
Circuit power of BS and RS	P_{cir}	500 w
Radius of UR	R	0.2 km
Distance between BS and the UR center	D	1.5 km

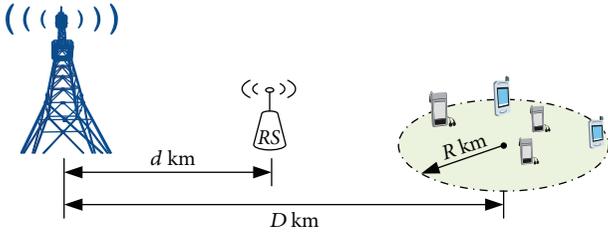


FIGURE 4: The downlink OFDMA system considered in numerical experiments.

4.1. System Setup and Benchmark Protocols. In numerical experiments, we consider the downlink OFDMA network with a RS exhibit in Figure 4. U represents the number of users in service and they are randomly distributed in a circular region of radius R km. The RS is located between the BS and the user-region (UR) center, and the BS-RS distance is d km. The distance between the BS and the user-region center is D km. The bandwidth of the system is B Hz, and the OFDMA uses K subcarriers. The network parameters are listed in Table 2.

The channels are independent of each other and are generated in the same way as in [7]. For every user u , the impulse response of the source-to- u channel is modeled as a tapped delay line with $L = 6$ taps, which are independently generated from circularly symmetric complex Gaussian distributions with zero mean and variance equal to $(1/L)(d_{su}/d_{\text{ref}})^{-4}$, where $d_{\text{ref}} = 1$ km and d_{su} represents the source-to- u distance. The source-to-relay and relay-to- u channels are generated in the same way, with each tap having variance as $(1/L)(d_{sr}/d_{\text{ref}})^{-4}$ and $(1/L)(d_{ru}/d_{\text{ref}})^{-4}$, respectively, where d_{ru} represents the relay-to- u distance. The CSI $\{h_{sr}^k | \forall k\}$, $\{h_{su}^k | \forall k, u\}$, and $\{h_{ru}^k | \forall k, u\}$ are computed by making K -point FFT over the impulse response of the associated channels.

In order to illustrate the benefit of optimized subcarrier pairing and opportunistic relaying, we also consider two other benchmark protocols, namely, BP-1 and BP-2. BP-1 is similar to the considered protocol, except that subcarrier k in slot-1 can only be paired with subcarrier k in slot-2 in relay mode. BP-2 is a simplified version of the considered protocol, and the simplification lies in the fact that each subcarrier in every slot should be allocated to users in direct mode. The RA algorithms for both BP-1 and BP-2 can be derived in the same way as that for the considered protocol, and therefore the derivation is omitted for the sake of clarity.

TABLE 3: Complexity comparison.

Algorithm	Complexity
Proposed algorithm	$O(K^2(U + U^2 + K))$
BP-1	$O(K(U + U^2))$
BP-2	$O(KU)$

The complexity of the three algorithms is shown in Table 3. It can be seen that the proposed algorithm has the highest complexity, while the BP-2 algorithm has the lowest complexity.

4.2. Impact of R_{req} on the Optimum EE and Corresponding Sum Rate. To show the influence of R_{req} on the EE, we choose $U = 10$ and $d = 0.6$ km and then evaluate the average optimum EE for every protocol over 1000 random channel realizations, when R_{req} increases from 0 to 40 Mbits/s. The results are shown in Figure 5.

Compared with BP-1 and BP-2, we can see that the proposed protocol and algorithm always correspond to a higher average EE as shown in Figure 5(a). Since BP-2 does not utilize opportunistic relaying as the proposed protocol and BP-1, it is reasonable that BP-2 corresponds to much lower average EE. The proposed protocol can achieve higher EE than BP-1, because a subcarrier in every slot can be paired with the other slot's subcarrier freely.

Figure 5(a) also shows that the average EE of these methods decreases with the increase of R_{req} . This is because the feasible set of the problem shrinks with the increase of R_{req} . From Figures 5(a), 5(b), and 5(c), when $R_{\text{req}} < 20$ Mbits/s and the average EE reaches the optimum value, the average communication rate is larger than R_{req} , the average total power remains stable, and the average EE of the network maintains high value. When $R_{\text{req}} \geq 20$ Mbits/s and the average EE reaches the optimum value, the average communication rate is equal to R_{req} , the average total power increases rapidly, and the average EE of the network decreases. The above phenomenon indicates that the restricted condition $R(\mathbf{S}) \geq R_{\text{req}}$ influences the choice of the optimum solution.

4.3. Impact of Relay Position on the Optimum EE and Corresponding Sum Rate. To show the impact of relay position on the EE, we choose $U = 10$ and $R_{\text{req}} = 20$ Mbits/s and then evaluate the average optimum EE for every protocol over 1000 random channel realizations, when d increases from 0.2 to 1.2 km. The results are shown in Figure 6.

It is shown that the proposed protocol leads to a higher average EE than the BP-1 and BP-2 for every relay position. Moreover, the average EE improves as the RS moves towards the middle region between the BS and the users. This can be interpreted as follows. In theory, the optimum EE enhances if $\forall k, l, u, G_{klu}$ is more likely to take a high value. Note that G_{klu} takes a high value only if both G_{sr}^k and G_{ru}^l are much higher than G_{su}^k . When RS lies in the middle between the BS and the users' region, it is more likely to have G_{sr}^k and G_{ru}^l , both much greater than G_{su}^k , and thus G_{klu} is more likely to take a high value. Moreover, BP-2 is a direct transmission protocol; the

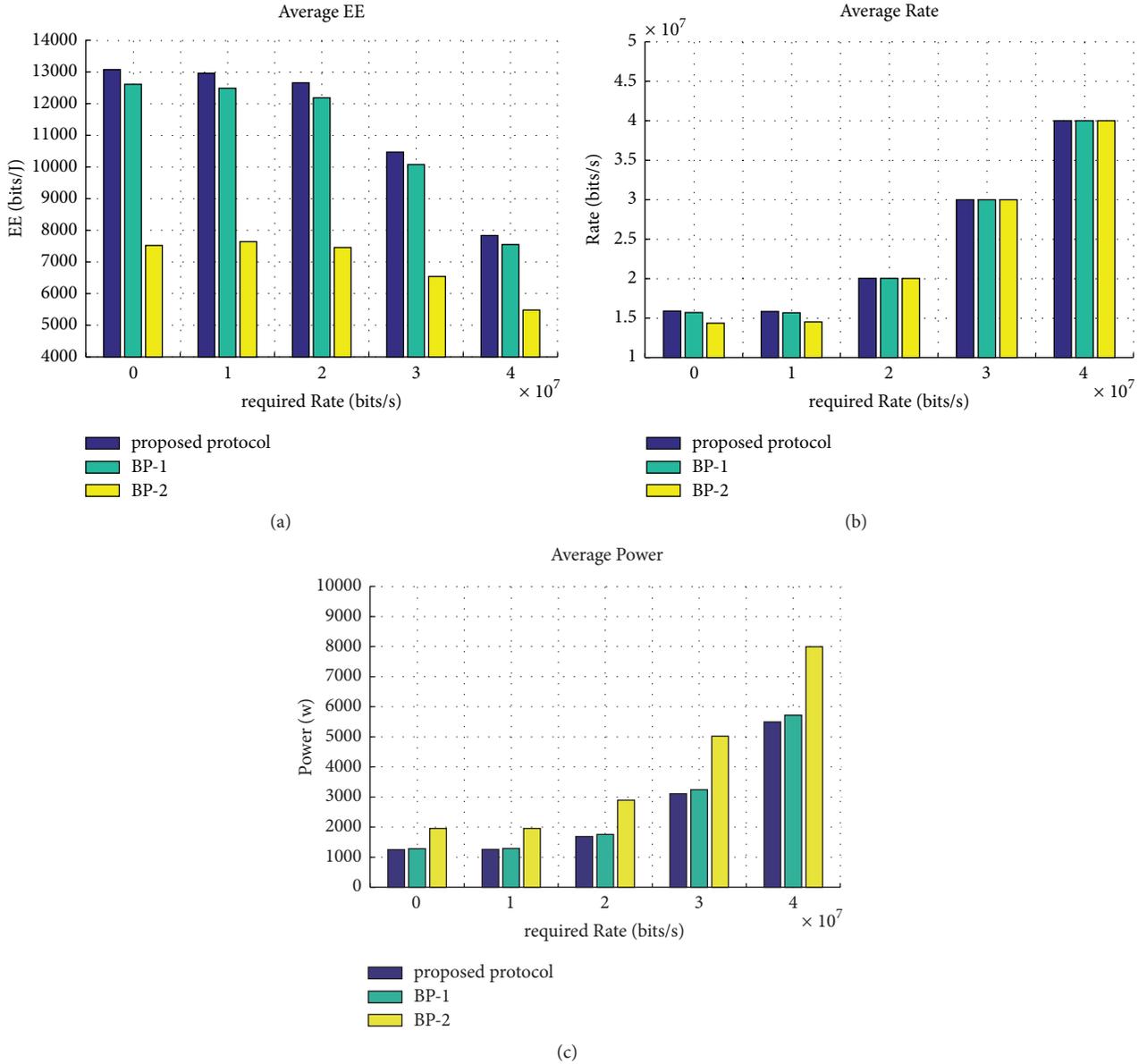


FIGURE 5: The average EE as the minimum rate changes.

RS does not help to transmit signals. It is reasonable that the average EE remains steady when the relay position changes.

4.4. Impact of User Number on the Optimum EE and Corresponding Sum Rate. To show the impact of user number on the EE, we choose $d = 0.6$ km and $R_{\text{req}} = 20$ Mbits/s and then evaluate the average optimum EE for every protocol over 1000 random channel realizations, when U increases from 5 to 30. The results are shown in Figure 7.

From Figure 7, we see that the average EEs of the three methods increase with the increase of user number. This is because when the number of users in the network increases, the subcarrier assignment has more flexibility. The numbers of A_{klu} and B_{klab} increase with the increase of user number,

which can improve the probability of C_{kl} taking a larger value. In this way, the average EE of the network will be improved.

5. Conclusions

We have addressed an EE maximized RA problem for cooperative OFDM transmission using the improved DF protocol with optimized subcarrier pairing when the network's communication rate is larger than a required value. The subcarrier-pair-based opportunistic DF relay-aided protocol has two operation modes: direct mode and relay mode. This scheme improves the flexibility of the communication network. Subcarriers can choose the mode that can improve the network's EE to send messages. Based on the above protocol,

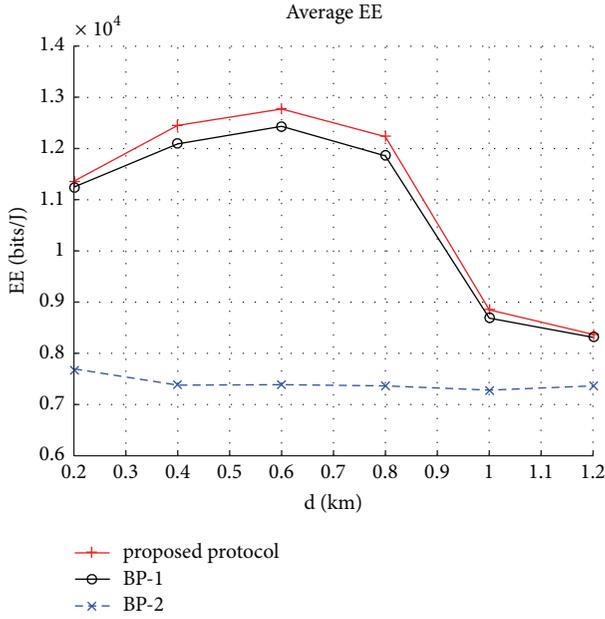


FIGURE 6: The average EE as the relay position changes.

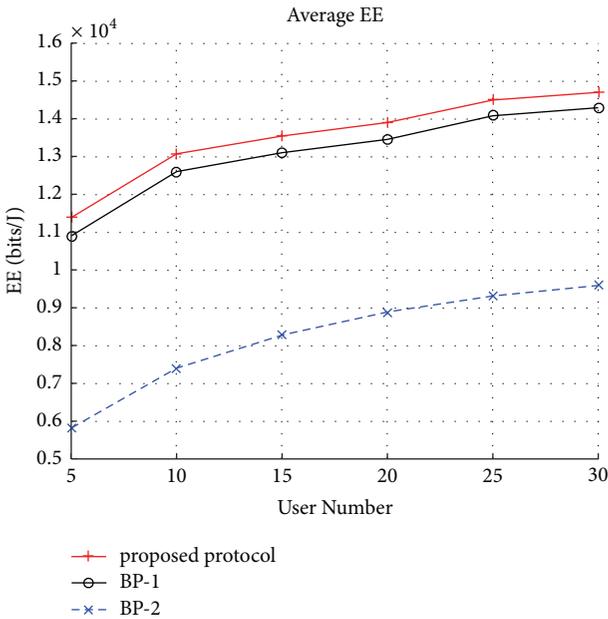


FIGURE 7: The average EE as the user number changes.

we formulate the optimization problem of maximizing the network's EE.

The problem is polynomial complexity, so we solve it with the following three steps. In the first step, we eliminate the fractional structure with the help of Dinkelbach method and transfer problem (P1) into problem (P2). In the second step, we get the Lagrangian function by using the dual method. In the third step, we use KKT conditions and Hungarian algorithm to solve the Lagrangian function. Then we can get the RA algorithm of maximizing the network's EE. Numerical experiments show that the proposed RA algorithm can

improve the EE of the downlink OFDMA networks. And the experiments also illustrated the impact of minimum required communication rate, relay position, and the user number. Theoretical analysis has been presented to interpret what is observed in numerical experiments.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants nos. 61671011, 61401266, 61501289, and 61701293), the National Science and Technology Major Project (Grant no. 2018ZX03001009), and research funds from Shanghai Institute for Advanced Communication and Data Science (SICS).

References

- [1] G. Fettweis and E. Zimmermann, "ICT Energy Consumption - Trends And Challenges," *Journal Abbreviation*, pp. 8–11, 2008.
- [2] J. Wu, S. Rangan, and H. Zhang, *Green Communications: Theoretical Fundamentals, Algorithms, and Applications*, CRC Press, 2012.
- [3] J. Wu, "Green wireless communications: From concept to reality [Industry perspectives]," *IEEE Wireless Communications Magazine*, vol. 19, no. 4, pp. 4–5, 2012.
- [4] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-efficiency of MIMO and cooperative MIMO techniques in sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1089–1098, 2004.
- [5] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy efficiency in the future internet: a survey of existing approaches and trends in energy-aware fixed network infrastructures," *IEEE Communications Surveys & Tutorials*, vol. 13, no. 2, pp. 223–244, 2011.
- [6] C. Isheden, Z. Chong, E. Jorswieck, and G. Fettweis, "Framework for link-level energy efficiency optimization with informed transmitter," *IEEE Transactions on Wireless Communications*, vol. 11, no. 8, pp. 2946–2957, 2012.
- [7] T. Wang, "Energy-efficiency maximization for OFDMA networks with total power constraint," in *Proceedings of the 2013 IEEE/CIC International Conference on Communications in China, ICCIC 2013*, pp. 357–361, China, August 2013.
- [8] Z. Wang and L. Vandendorpe, "Subcarrier Allocation and Precoder Design for Energy Efficient MIMO-OFDMA Downlink Systems," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 136–146, 2017.
- [9] S. Fu, H. Wen, J. Wu, and B. Wu, "Energy-Efficient Precoded Coordinated Multi-Point Transmission With Pricing Power Game Mechanism," *IEEE Systems Journal*, vol. 11, no. 2, pp. 578–587, 2017.
- [10] K. Yang, S. Martin, D. Quadri, J. Wu, and G. Feng, "Energy-Efficient Downlink Resource Allocation in Heterogeneous

- OFDMA Networks,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 6, pp. 5086–5098, 2017.
- [11] M. Mehrjoo, S. Moazeni, and X. Shen, “Resource allocation in OFDMA networks based on interior point methods,” *Wireless Communications and Mobile Computing*, vol. 10, no. 11, pp. 1493–1508, 2010.
- [12] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, “Energy- and spectral-efficiency tradeoff in downlink OFDMA networks,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3874–3886, 2011.
- [13] T. Wang and L. Vandendorpe, “On the SCALE algorithm for multiuser multicarrier power spectrum management,” *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4992–4998, 2012.
- [14] Q. Wu, W. Chen, M. Tao, J. Li, H. Tang, and J. Wu, “Resource allocation for joint transmitter and receiver energy efficiency maximization in downlink OFDMA systems,” *IEEE Transactions on Communications*, vol. 63, no. 2, pp. 416–430, 2015.
- [15] J. Tang, D. K. C. So, E. Alsusa, K. A. Hamdi, and A. Shojaefard, “On the energy efficiency-spectral efficiency tradeoff in MIMO-OFDMA broadcast channels,” *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5185–5199, 2016.
- [16] G. Zhang, J. Hu, W. Heng, and G. Wang, “Distributed power and resource allocation for weighted sum energy-efficiency maximization in OFDMA smallcell network,” in *Proceedings of the 27th IEEE Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC 2016*, pp. 1–5, September 2016.
- [17] C. Muller, A. Klein, and F. Wegner, “Dynamic subcarrier, bit and power allocation in OFDMA-based relay networks,” in *12th International OFDM Workshop*, pp. 1–5, 2007.
- [18] T. Girici, “Joint power, subcarrier and subframe allocation in Multihop relay networks,” *International Journal of Communication Systems*, vol. 22, no. 7, pp. 835–855, 2009.
- [19] T. Wang and L. Vandendorpe, “WSR maximized resource allocation in multiple DF relays aided OFDMA downlink transmission,” *IEEE Transactions on Signal Processing*, vol. 59, no. 8, pp. 3964–3976, 2011.
- [20] F. Lu, C. Liu, and H.-A. Zhao, “On the joint optimal power allocation for df relaying and beamforming communication systems,” *IEICE Transactions on Communications*, vol. E95-B, no. 3, pp. 889–897, 2012.
- [21] B. Huang, X. Fang, and Y. Chen, “Joint energy saving resource allocation and user scheduling in OFDMA relay networks,” in *Proceedings of the IEEE/CIC International Conference on Communications in China*, pp. 484–490, IEEE, Xi’an, China, August 2013.
- [22] B. Huang, X. Fang, Y. Zhao, Y. Chen, and R. He, “Dynamic energy saving subcarrier, bit and power allocation in OFDMA relay networks,” *China Communications*, vol. 10, no. 4, pp. 79–87, 2013.
- [23] Y. Chen, X. Fang, and B. Huang, “Energy-efficient relay selection and resource allocation in nonregenerative relay OFDMA systems,” *IEEE Transactions on Vehicular Technology*, vol. 63, no. 8, pp. 3689–3699, 2014.
- [24] X. Chen, F. He, L. Xiao, and S. Zhou, “Joint subcarrier and power allocation for DF-based multiuser two-way relay networks,” *China Communications*, vol. 14, no. 6, Article ID 7961373, pp. 179–188, 2017.
- [25] Y. Li, W. Wang, J. Kong, and M. Peng, “Subcarrier pairing for amplify-and-forward and decode-and-forward OFDM relay links,” *IEEE Communications Letters*, vol. 13, no. 4, pp. 209–211, 2009.
- [26] T. Wang, “Weighted sum power minimisation for multichannel decode-and-forward relaying,” *IEEE Electronics Letters*, vol. 48, no. 7, pp. 410–411, 2012.
- [27] E. S. Hassan, “Energy-efficient hybrid opportunistic cooperative protocol for single-carrier frequency division multiple access-based networks,” *IET Communications*, vol. 6, no. 16, pp. 2602–2612, 2012.
- [28] G. Zhou, T. Wang, Y. Wu, G. Zheng, and G. Yang, “Energy-Efficient Power Allocation for Decode-and-Forward OFDM Relay Links,” in *Mobile and Wireless Technologies 2016*, vol. 391 of *Lecture Notes in Electrical Engineering*, pp. 13–24, Springer Singapore, Singapore, 2016.
- [29] W. Yang and X. Zhao, “Resource allocation in two-way OFDM-based cognitive radio networks with QoE and power consumption guarantees,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2017, no. 1, article no. 216, 2017.
- [30] W. Yang, L.-H. Li, W.-L. Sun, and Y. Wang, “Energy-efficient relay selection and optimal relay location in cooperative cellular networks with asymmetric traffic,” *Journal of China Universities of Posts and Telecommunications*, vol. 17, no. 6, pp. 80–88, 2010.
- [31] C. Y. Ho and C.-Y. Huang, “Energy efficient subcarrier-power allocation and relay selection scheme for OFDMA-based cooperative relay networks,” in *Proceedings of the 2011 IEEE International Conference on Communications*, pp. 1–6, Japan, June 2011.
- [32] K. T. K. Cheung, S. Yang, and L. Hanzo, “Achieving maximum energy-efficiency in multi-relay OFDMA cellular networks: a fractional programming approach,” *IEEE Transactions on Communications*, vol. 61, no. 7, pp. 2746–2757, 2013.
- [33] Z. Song, Q. Ni, K. Navaie, S. Hou, S. Wu, and X. Sun, “On the spectral-energy efficiency and rate fairness tradeoff in relay-aided cooperative OFDMA systems,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6342–6355, 2016.
- [34] F. Heliot and R. Tafazolli, “Optimal Energy-Efficient Joint Resource Allocation for Multi-Hop MIMO-AF Systems,” *IEEE Transactions on Communications*, vol. 64, no. 9, pp. 3655–3668, 2016.
- [35] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge University Press, Cambridge, UK, 2005.
- [36] T. Wang, F. Glineur, J. Louveaux, and L. Vandendorpe, “Weighted sum rate maximization for downlink OFDMA with subcarrier-pair based opportunistic DF relaying,” *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2512–2524, 2013.
- [37] W. Dinkelbach, “On nonlinear fractional programming,” *Management Science*, vol. 13, no. 7, pp. 492–498, 1967.
- [38] Y. Wu and T. Wang, “Energy-efficient resource allocation for OFDM transmission with opportunistic DF relaying,” in *Proceedings of the 2014 IEEE/CIC International Conference on Communications in China, ICCIC 2014*, pp. 570–575, chn, October 2014.
- [39] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2013.
- [40] H. W. Kuhn, “The Hungarian method for the assignment problem,” *Naval Research Logistics Quarterly*, vol. 2, pp. 83–97, 1955.

Research Article

Average SEP of AF Relaying in Nakagami- m Fading Environments

Dong Qin ^{1,2}, Yuhao Wang ¹, and Tianqing Zhou ³

¹*School of Information Engineering, Nanchang University, Nanchang, China*

²*Postdoctoral Research Station of Environmental Science and Engineering, Nanchang University, Nanchang, China*

³*School of Information Engineering, East China Jiaotong University, Nanchang, China*

Correspondence should be addressed to Dong Qin; qindong@seu.edu.cn

Received 7 January 2018; Accepted 26 March 2018; Published 3 May 2018

Academic Editor: Wolfgang H. Gerstaecker

Copyright © 2018 Dong Qin et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper is devoted to an investigation of an exact average symbol error probability (SEP) for amplify and forward (AF) relaying in independent Nakagami- m fading environments with a nonnegative integer plus one-half m , which covers many actual scenarios, such as one-side Gaussian distribution ($m = 0.5$). Using moment generating function approach, the closed-form SEP is expressed in the form of Lauricella multivariate hypergeometric function. Four modulation modes are considered: rectangular quadrature amplitude modulation (QAM), M -ary phase shift keying (MPSK), M -ary differential phase shift keying (MDPSK), and $\pi/4$ differential quaternary phase shift keying (DQPSK). The result is very simple and general for a nonnegative integer plus one-half m , which covers the same range as integer m . The tightness of theoretical analysis is confirmed by computer simulation results.

1. Introduction

When the transceiver is far away, it is not a wise choice to increase transmit power in green communications. Energy efficiency is placed in a very important position in resource optimization configuration. The relay technology is used as one effective means to achieve diversity gain, reduce transmission power, and improve energy efficiency. Therefore, cooperative diversity technology has attracted considerable benefits for enhancing the performance of wireless networks, such as cell network and ad hoc network. SEP is considered to be an important performance measure. Integral-free bit error probability formulas were derived in [1, 2] for decode and forward (DF) cooperative systems. An asymptotic bit error probability formula was provided in [3–5] for all participate and selective AF cooperation. A new approximation to the symbol error rate was derived in [6]. Multiple input multiple output (MIMO) technology was introduced in [7, 8], which considered exact and asymptotic symbol error rate. Selection combining was added in [9] with the same configuration as that in [7]. Mobile to mobile communication scenario was shown in [10], where asymptotic symbol error rate bound was provided. Signal space cooperation was realized in [11] for full

rate transmission, where a tight bit error rate was obtained. Error probabilities of AF multihop variable gain relaying systems were analyzed in [12] by generalized hypergeometric functions. While the two-hop scenario was investigated in [13].

Although the error probability of the Nakagami- m relay channel has been studied extensively, they mostly addressed integer fading parameter m . The error probability formula for arbitrary m must resort to infinite series expansion or reasonable approximation in high signal to noise (SNR) region. Some articles provided error probability just by simulation. The actual situation often requires quickly grasping error probability of wireless communication, so it is impossible to carry out a large number of simulation experiments. The importance and generalization of Nakagami- m channel have not been fully exploited, which motivates our work from another perspective.

In this paper, we calculate the exact average SEP formulas in an AF relay system over Nakagami- m fading environment when m is a nonnegative integer plus one-half, while it includes as a special case the one-sided Gaussian distribution ($m = 0.5$). Such closed-form expressions are urgent because they allow fast and efficient evaluation of system reliability.

Note that this parameter m covers the same quantity range as the integer m in most literature, according to basic knowledge of number theory. Four prevailing modulations are investigated: MQAM, MPSK, MDPSK, and DQPSK. Using the properties of moment generating function, precise expressions are a combination of a series of special functions. Simulation results confirm the tightness compactness of the theoretical analysis. For ease of reading and searching, see Mathematical Notations.

2. System Model

Consider a cooperative system where a source node S communicates with a destination node D via a relay node R . There is not direct link between S and R due to obstacles. Assume that all links between transceivers are subject to independent but different Nakagami- m fading. When m is a nonnegative integer plus one-half, the probability density function of the end to end SNR γ at D is given by [14, eq. (7)]

$$f_\gamma(z) = \frac{\sqrt{\pi} e^{-(\sqrt{\beta_s} + \sqrt{\beta_r})^2 z}}{\Gamma(m_s) \Gamma(m_r)} \sum_{k_1=0}^{\lfloor m_s \rfloor} \sum_{k_2=0}^{\lfloor m_r \rfloor} \sum_{l=0}^{\lfloor (m_s - k_1 + m_r - k_2)/2 \rfloor} \frac{(\lfloor m_s \rfloor + k_1)! (\lfloor m_r \rfloor + k_2)! \beta_s^{m_s/2 - 1/4 - k_1/2} \beta_r^{m_r/2 - 1/4 - k_2/2}}{4^{k_1 + k_2} k_1! k_2! (\lfloor m_s \rfloor - k_1)! (\lfloor m_r \rfloor - k_2)! 2^{2l}} \times \frac{(m_s - k_1 + m_r - k_2)! z^{m_s - k_1 + m_r - k_2 - l - 3/2}}{l! (m_s - k_1 + m_r - k_2 - 2l)!} \times (-1)^l \left(\sqrt{\beta_s} + \sqrt{\beta_r} \right)^{m_s - k_1 + m_r - k_2 - 2l}, \quad (1)$$

where m and β are fading parameter and scale parameter, respectively. The subscripts s and r represent $S \rightarrow R$ link

and $R \rightarrow D$ link, respectively. Using [15, eq. (3.381.4)], the moment generating function of the SNR γ is given by

$$M_\gamma(s) = \int_0^\infty e^{-sz} f_\gamma(z) dz = \frac{\sqrt{\pi}}{\Gamma(m_s) \Gamma(m_r)} \sum_{k_1=0}^{\lfloor m_s \rfloor} \sum_{k_2=0}^{\lfloor m_r \rfloor} \sum_{l=0}^{\lfloor (m_s - k_1 + m_r - k_2)/2 \rfloor} \frac{(\lfloor m_s \rfloor + k_1)! (\lfloor m_r \rfloor + k_2)! \beta_s^{m_s/2 - 1/4 - k_1/2}}{4^{k_1 + k_2} k_1! k_2! (\lfloor m_s \rfloor - k_1)! (\lfloor m_r \rfloor - k_2)!} \times \frac{\beta_r^{m_r/2 - 1/4 - k_2/2} (m_s - k_1 + m_r - k_2)!}{2^{2l} l! (m_s - k_1 + m_r - k_2 - 2l)!} \times (-1)^l \left(\sqrt{\beta_s} + \sqrt{\beta_r} \right)^{m_s - k_1 + m_r - k_2 - 2l} \times \left(s + \beta_s + \beta_r + 2\sqrt{\beta_s \beta_r} \right)^{1/2 + k_1 + k_2 + l - m_s - m_r} \times \Gamma\left(m_s + m_r - k_1 - k_2 - l - \frac{1}{2}\right). \quad (2)$$

The above moment generating function is very useful in error probability analysis.

3. Average SEP Analysis

The average SEP is an important measure of communication reliability. Next, we prepare to study the SEP performances of four modulations.

3.1. Rectangular QAM. The average SEP of coherent rectangular $M_I \times M_Q$ QAM is given by

$$P_{e,\text{QAM}} = \int_0^\infty \left[2 \left(1 - \frac{1}{M_I} \right) Q(\sqrt{q_1 \gamma}) + 2 \left(1 - \frac{1}{M_Q} \right) \times Q(\sqrt{q_2 \gamma}) - 4 \left(1 - \frac{1}{M_I} \right) \times \left(1 - \frac{1}{M_Q} \right) Q(\sqrt{q_1 \gamma}) Q(\sqrt{q_2 \gamma}) \right] f_\gamma(\gamma) d\gamma, \quad (3)$$

where

$$q_1 = \frac{6}{M_I^2 - 1 + (M_Q^2 - 1)r^2}, \quad (4)$$

$$q_2 = \frac{6r^2}{M_I^2 - 1 + (M_Q^2 - 1)r^2}. \quad (5)$$

r is the decision distance ratio between constellations of in phase and quadrature components. Based on moment generating function approach, the error probability is written as

$$P_{e,\text{QAM}} = \frac{2}{\pi} \left(1 - \frac{1}{M_I} \right) \int_0^{\pi/2} M_\gamma \left(\frac{q_1}{2 \sin^2 \theta} \right) d\theta + \frac{2}{\pi} \left(1 - \frac{1}{M_Q} \right) \int_0^{\pi/2} M_\gamma \left(\frac{q_2}{2 \sin^2 \theta} \right) d\theta - \frac{2}{\pi} \left(1 - \frac{1}{M_I} \right) \left(1 - \frac{1}{M_Q} \right) \times \left[\int_0^{\pi/2 - \arctan \sqrt{b/a}} M_\gamma \left(\frac{q_1}{2 \sin^2 \theta} \right) d\theta + \int_0^{\arctan \sqrt{b/a}} M_\gamma \left(\frac{q_2}{2 \sin^2 \theta} \right) d\theta \right], \quad (6)$$

where the first and second integrals correspond to the error probability involving a single Q function and the third and

fourth integrals correspond to the error probability involving the product of two Q functions. We first begin by the integral containing one Gaussian Q function. Taking the first integral in (6) as an example, we encounter a kind of integral given by

$$J_1 = \int_0^{\pi/2} \left(\frac{q_1}{2 \sin^2 \theta} + \beta_s + \beta_r + 2\sqrt{\beta_s \beta_r} \right)^{1/2+k_1+k_2+l-m_s-m_r} d\theta. \quad (7)$$

By change of the variable $u = \cos^2 \theta$, after some manipulations, J_1 can be written as

$$J_1 = \int_0^1 \left(\frac{q_1}{2(1-u)} + \beta_s + \beta_r \right)$$

$$+ 2\sqrt{\beta_s \beta_r} \Big)^{1/2+k_1+k_2+l-m_s-m_r} \times \frac{1}{2\sqrt{u(1-u)}} du. \quad (8)$$

Applying [15, eq. (9.100)], (8) can be expressed in closed form in terms of Gauss hypergeometric function.

Next, we cope with the integral containing the product of two Gaussian Q functions. Making change of the variable $u = 1 - b^2/a^2 \tan^2 \theta$, we obtain one kind of integral given by

$$J_2 = \int_0^1 u^{a-1} (1-u)^{c-a-1} (1-vu)^{-b_1} (1-vu)^{-b_2} du. \quad (9)$$

Using [16, eq. (11)], J_2 can be expressed in closed form in terms of Appell hypergeometric function. Finally, combining the results in (8) and (9), the average SEP of rectangular QAM is given by

$$\begin{aligned} P_{e,\text{QAM}} = & \frac{1}{\Gamma(m_s)\Gamma(m_r)} \sum_{k_1=0}^{\lfloor m_s \rfloor} \sum_{k_2=0}^{\lfloor m_r \rfloor} \sum_{l=0}^{\lfloor (m_s-k_1+m_r-k_2)/2 \rfloor} \frac{(\lfloor m_s \rfloor + k_1)! (\lfloor m_r \rfloor + k_2)! \beta_s^{m_s/2-1/4-k_1/2}}{4^{k_1+k_2} k_1! k_2! (\lfloor m_s \rfloor - k_1)! (\lfloor m_r \rfloor - k_2)!} \\ & \times \frac{\beta_r^{m_r/2-1/4-k_2/2} (m_s - k_1 + m_r - k_2)!}{2^{2l} l! (m_s - k_1 + m_r - k_2 - 2l)!} \times \Gamma\left(m_s + m_r - k_1 - k_2 - l - \frac{1}{2}\right) \times (-1)^l \left(\sqrt{\beta_s} + \sqrt{\beta_r}\right)^{m_s-k_1+m_r-k_2-2l} \times \left\{ \left(1 - \frac{1}{M_I}\right) \frac{\Gamma(m_s - k_1 + m_r - k_2 - l)}{\Gamma(m_s - k_1 + m_r - k_2 - l + 1/2)} \times \left(\frac{q_1}{2} + \beta_s + \beta_r + 2\sqrt{\beta_s \beta_r}\right)^{1/2+k_1+k_2+l-m_s-m_r} \times {}_2F_1\left(\frac{1}{2}, m_s - k_1 + m_r - k_2 - l - \frac{1}{2}; m_s - k_1 + m_r - k_2 - l + \frac{1}{2}; \frac{2(\sqrt{\beta_s} + \sqrt{\beta_r})^2}{2(\sqrt{\beta_s} + \sqrt{\beta_r})^2 + q_1}\right) + \left(1 - \frac{1}{M_Q}\right) \frac{\Gamma(m_s - k_1 + m_r - k_2 - l)}{\Gamma(m_s - k_1 + m_r - k_2 - l + 1/2)} \times \left(\frac{q_2}{2} + \beta_s + \beta_r + 2\sqrt{\beta_s \beta_r}\right)^{1/2+k_1+k_2+l-m_s-m_r} \times {}_2F_1\left(\frac{1}{2}, m_s - k_1 + m_r - k_2 - l - \frac{1}{2}; m_s - k_1 + m_r - k_2 - l - \frac{1}{2}; \frac{2(\sqrt{\beta_s} + \sqrt{\beta_r})^2}{2(\sqrt{\beta_s} + \sqrt{\beta_r})^2 + q_2}\right) - \frac{\sqrt{q_1 q_2} (1 - 1/M_I)(1 - 1/M_Q)}{\sqrt{\pi} (q_1 + q_2) (m_s - k_1 + m_r - k_2 - l)} \left(\frac{q_1 + q_2}{2} + (\sqrt{\beta_s} + \sqrt{\beta_r})^2\right)^{1/2+k_1+k_2+l-m_s-m_r} \cdot \left[F_1\left(1, m_s - k_1 + m_r - k_2 - l - \frac{1}{2}, 1; m_s - k_1 + m_r - k_2 - l + 1; \frac{q_1 + 2(\sqrt{\beta_s} + \sqrt{\beta_r})^2}{q_1 + q_2 + 2(\sqrt{\beta_s} + \sqrt{\beta_r})^2}, \frac{q_1}{q_1 + q_2}\right) + F_1\left(1, m_s - k_1 + m_r - k_2 - l - \frac{1}{2}, 1; m_s - k_1 + m_r - k_2 - l + 1; \frac{q_2 + 2(\sqrt{\beta_s} + \sqrt{\beta_r})^2}{q_1 + q_2 + 2(\sqrt{\beta_s} + \sqrt{\beta_r})^2}, \frac{q_2}{q_1 + q_2}\right) \right] \right\}, \quad (10) \end{aligned}$$

where ${}_2F_1(\cdot, \cdot; \cdot; \cdot)$ and $F_1(\cdot, \cdot, \cdot; \cdot; \cdot, \cdot)$ stands for the Gauss hypergeometric function defined in [15, eq. (9.100)] and the Appell hypergeometric function [17].

3.2. MPSK. Following similar steps in QAM, the average SEP of coherent MPSK is given by

$$\begin{aligned} P_{e,\text{MPSK}} = & \frac{1}{\pi} \int_0^{\pi-\pi/M} M_\gamma \left[\frac{\sin^2(\pi/M)}{\sin^2 \theta} \right] d\theta = \frac{1}{\pi} \int_0^{\pi/2} M_\gamma \left[\frac{\sin^2(\pi/M)}{\sin^2 \theta} \right] d\theta + \frac{1}{\pi} \int_{\pi/2}^{\pi-\pi/M} M_\gamma \left[\frac{\sin^2(\pi/M)}{\sin^2 \theta} \right] d\theta \\ = & \frac{1}{\Gamma(m_s)\Gamma(m_r)} \sum_{k_1=0}^{\lfloor m_s \rfloor} \sum_{k_2=0}^{\lfloor m_r \rfloor} \sum_{l=0}^{\lfloor (m_s-k_1+m_r-k_2)/2 \rfloor} \frac{(\lfloor m_s \rfloor + k_1)! (\lfloor m_r \rfloor + k_2)! \beta_s^{m_s/2-1/4-k_1/2}}{4^{k_1+k_2} k_1! k_2! (\lfloor m_s \rfloor - k_1)! (\lfloor m_r \rfloor - k_2)!} \times \frac{\beta_r^{m_r/2-1/4-k_2/2} (m_s - k_1 + m_r - k_2)!}{2^{2l} l! (m_s - k_1 + m_r - k_2 - 2l)!} \end{aligned}$$

$$\begin{aligned}
& \times (-1)^l \left(\sqrt{\beta_s} + \sqrt{\beta_r} \right)^{m_s - k_1 + m_r - k_2 - 2l} \times \Gamma \left(m_s + m_r - k_1 - k_2 - l - \frac{1}{2} \right) \\
& \times \left(\left(\sqrt{\beta_s} + \sqrt{\beta_r} \right)^2 + \sin^2 \left(\frac{\pi}{M} \right)^{1/2 + k_1 + k_2 + l - m_s - m_r} \right) \times \left[\frac{\Gamma(m_s - k_1 + m_r - k_2 - l)}{2\Gamma(m_s - k_1 + m_r - k_2 - l + 1/2)} \right. \\
& \times {}_2F_1 \left(m_s - k_1 + m_r - k_2 - l - \frac{1}{2}, \frac{1}{2}; m_s - k_1 + m_r - k_2 - l + \frac{1}{2}; \frac{(\sqrt{\beta_s} + \sqrt{\beta_r})^2}{(\sqrt{\beta_s} + \sqrt{\beta_r})^2 + \sin^2(\pi/M)} \right) + \frac{\cos(\pi/M)}{\sqrt{\pi}} \\
& \left. \cdot F_1 \left(\frac{1}{2}, m_s - k_1 + m_r - k_2 - l - \frac{1}{2}, 1 + k_1 + k_2 + l - m_s - m_r; \frac{3}{2}; \frac{\cos^2(\pi/M) (\sqrt{\beta_s} + \sqrt{\beta_r})^2}{(\sqrt{\beta_s} + \sqrt{\beta_r})^2 + \sin^2(\pi/M)}, \cos^2 \left(\frac{\pi}{M} \right) \right) \right]. \tag{11}
\end{aligned}$$

3.3. *MDPSK*. With the aid of the common moment generating function, the average SEP for MDPSK is given by

$$\begin{aligned}
P_{e, \text{MDPSK}} &= \frac{1}{\pi} \int_0^{\pi - \pi/M} M_\gamma \left[\frac{\sin^2(\pi/M)}{1 + \cos(\pi/M) \cos \theta} \right] d\theta = \frac{2 \cos(\pi/2M)}{\sqrt{\pi} \Gamma(m_s) \Gamma(m_r)} \\
& \cdot \sum_{k_1=0}^{\lfloor m_s \rfloor} \sum_{k_2=0}^{\lfloor m_r \rfloor} \sum_{l=0}^{\lfloor (m_s - k_1 + m_r - k_2)/2 \rfloor} \frac{(\lfloor m_s \rfloor + k_1)! (\lfloor m_r \rfloor + k_2)! \beta_s^{m_s/2 - 1/4 - k_1/2}}{4^{k_1 + k_2} k_1! k_2! (\lfloor m_s \rfloor - k_1)! (\lfloor m_r \rfloor - k_2)!} \times \frac{\beta_r^{m_r/2 - 1/4 - k_2/2} (m_s - k_1 + m_r - k_2)!}{2^{2l} l! (m_s - k_1 + m_r - k_2 - 2l)!} \times \Gamma(m_s + m_r \\
& - k_1 - k_2 - l - \frac{1}{2}) \times (-1)^l \left(\sqrt{\beta_s} + \sqrt{\beta_r} \right)^{m_s - k_1 + m_r - k_2 - 2l} \times \left(\beta_s + \beta_r + 2\sqrt{\beta_s \beta_r} + 2 \sin^2 \left(\frac{\pi}{2M} \right) \right)^{1/2 + k_1 + k_2 + l - m_s - m_r} \\
& \times F_D^{(3)} \left[\frac{1}{2}, m_s + m_r - k_1 - k_2 - l - \frac{1}{2}, \frac{1}{2} + k_1 + k_2 + l - m_s \right. \\
& \left. - m_r, \frac{3}{2}; \frac{\cos(\pi/M) (\sqrt{\beta_s} + \sqrt{\beta_r})^2}{(\sqrt{\beta_s} + \sqrt{\beta_r})^2 + 2 \sin^2(\pi/2M)}, \cos \left(\frac{\pi}{M} \right), \cos^2 \left(\frac{\pi}{2M} \right) \right], \tag{12}
\end{aligned}$$

where $F_D^{(3)}$ is the Lauricella function [16, eq. (15)].

equiprobable, and correlated binary signals with noncoherent detection is given by

3.4. *Noncoherent Detection of Equiprobable Correlated Binary Signals and $\pi/4$ DQPSK*. The average SEP of equal energy,

$$\begin{aligned}
P_e &= \frac{1}{2\pi} \int_0^\pi M_\gamma \left[\frac{(b^2 - a^2)^2}{2(a^2 + b^2) - 4ab \cos \theta} \right] d\theta \\
&= \frac{\sqrt{\pi}}{2\Gamma(m_s) \Gamma(m_r)} \sum_{k_1=0}^{\lfloor m_s \rfloor} \sum_{k_2=0}^{\lfloor m_r \rfloor} \sum_{l=0}^{\lfloor (m_s - k_1 + m_r - k_2)/2 \rfloor} \frac{(\lfloor m_s \rfloor + k_1)! (\lfloor m_r \rfloor + k_2)! \beta_s^{m_s/2 - 1/4 - k_1/2}}{4^{k_1 + k_2} k_1! k_2! (\lfloor m_s \rfloor - k_1)! (\lfloor m_r \rfloor - k_2)!} \times \frac{\beta_r^{m_r/2 - 1/4 - k_2/2} (m_s - k_1 + m_r - k_2)!}{2^{2l} l! (m_s - k_1 + m_r - k_2 - 2l)!} \\
& \times \Gamma(m_s + m_r - k_1 - k_2 - l - \frac{1}{2}) \times (-1)^l \left(\sqrt{\beta_s} + \sqrt{\beta_r} \right)^{m_s - k_1 + m_r - k_2 - 2l} \\
& \times \left(\beta_s + \beta_r + 2\sqrt{\beta_s \beta_r} + \frac{(b-a)^2}{2} \right)^{1/2 + k_1 + k_2 + l - m_s - m_r} \\
& \times F_1 \left[\frac{1}{2}, m_s + m_r - k_1 - k_2 - l - \frac{1}{2}, \frac{1}{2} + k_1 + k_2 + l - m_s - m_r; 1; \frac{8ab (\sqrt{\beta_s} + \sqrt{\beta_r})^2}{(a+b)^2 (2(\sqrt{\beta_s} + \sqrt{\beta_r})^2 + (b-a)^2)}, \frac{4ab}{(a+b)^2} \right], \tag{13}
\end{aligned}$$

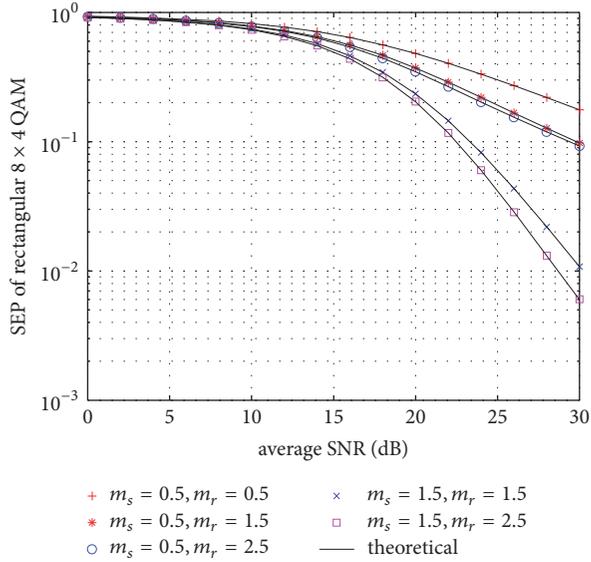
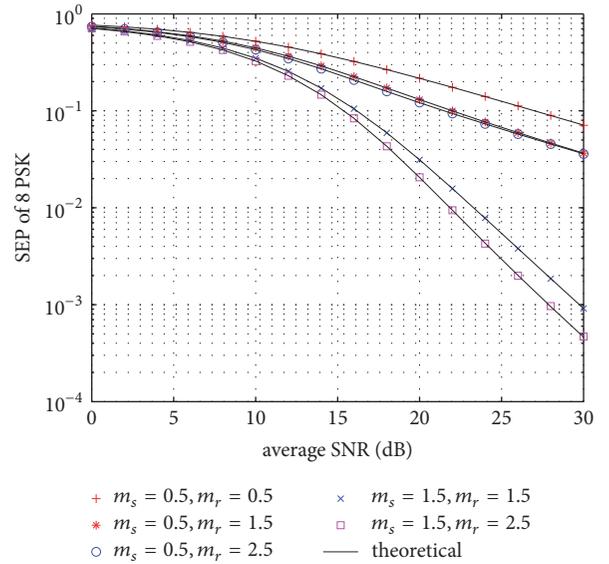

 FIGURE 1: Comparison of SEP for 8×4 QAM.


FIGURE 2: Comparison of SEP for 8 PSK.

where

$$a = \sqrt{\frac{1 - \sqrt{1 - |\rho|^2}}{2}}, \quad (14)$$

$$b = \sqrt{\frac{1 + \sqrt{1 - |\rho|^2}}{2}} \quad (15)$$

and $0 \leq |\rho| \leq 1$ is the magnitude of the cross correlation coefficient between the two signals. When $a = \sqrt{2 - \sqrt{2}}$ and $b = \sqrt{2 + \sqrt{2}}$, (13) corresponds to $\pi/4$ DQPSK with gray coding.

4. Simulation Results

In this section, the simulation results of the error probability for QAM, MPSK, MDPSK, and DQPSK are evaluated. The average SNR per symbol is defined as $P_s/N_0 = P_r/N_0$, where P_s and P_r represent transmit power of the source and the relay, respectively. N_0 is the noise variances. The channel gain is normalized to unit. The theoretical results highly agree with the simulations for 8×4 QAM and integer plus one-half m . The same coincidence can be deduced for 8 PSK, 8 DPSK, and DQPSK modulation constellations and fading parameters m . This demonstrates the accuracy and validity of the proposed formula.

Figures 1–4 show that the diversity gain is an increasing function of the fading parameter m . For example, when the SEP of 8 PSK is 0.1, the diversity gain of the case of $m_s = 0.5$, $m_r = 0.5$ is achieved about 4 dB compared with the case of $m_s = 0.5$, $m_r = 1.5$. Moreover, the diversity gain increased to about 10 dB when $m_s = 1.5$, $m_r = 1.5$. Similar observations can be found in 8 DPSK modulation in Figure 3. The diversity order is dominated by the worse link between $S \rightarrow R$ and $R \rightarrow$

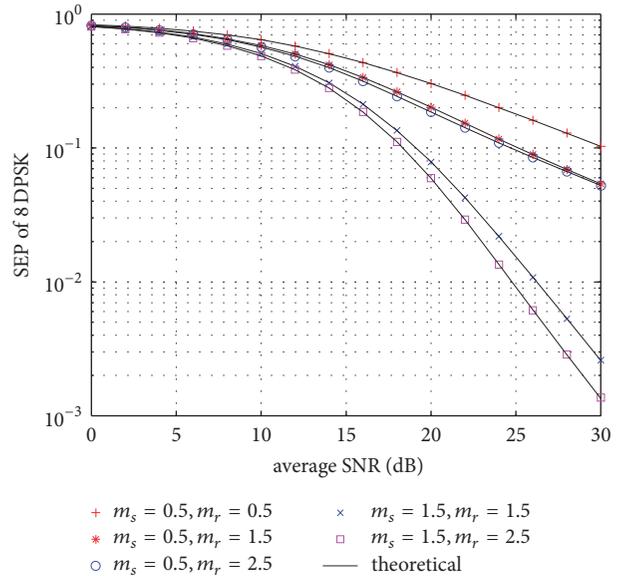


FIGURE 3: Comparison of SEP for 8 DPSK.

D. For example, let us observe two cases of $m_s = 0.5$, $m_r = 0.5$ and $m_s = 0.5$, $m_r = 1.5$. From Figure 2, when $m_s = m_r = 0.5$, in the high SNR region, the SEP is 0.089645629941062 at average SNR = 28 dB and 0.070958408816342 at average SNR = 30 dB. This implies that the diversity gain is

$$\frac{10 \lg(0.089645629941062/0.070958408816342)}{30 - 28} \quad (16)$$

$$= 0.5076 \approx \min(m_s, m_r).$$

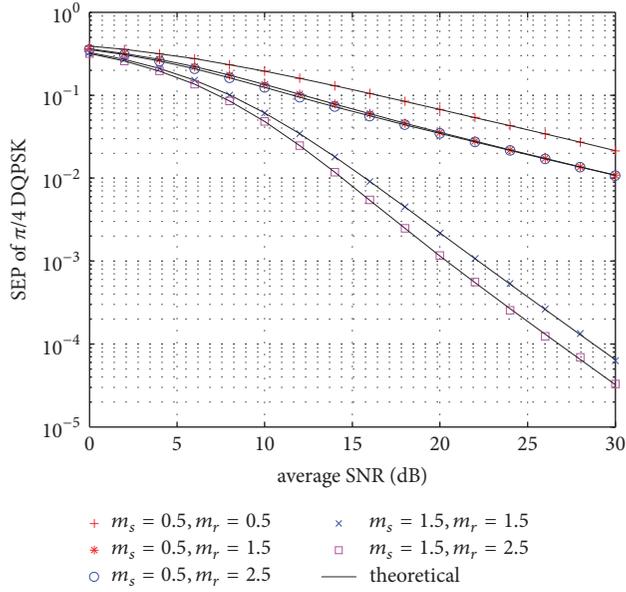
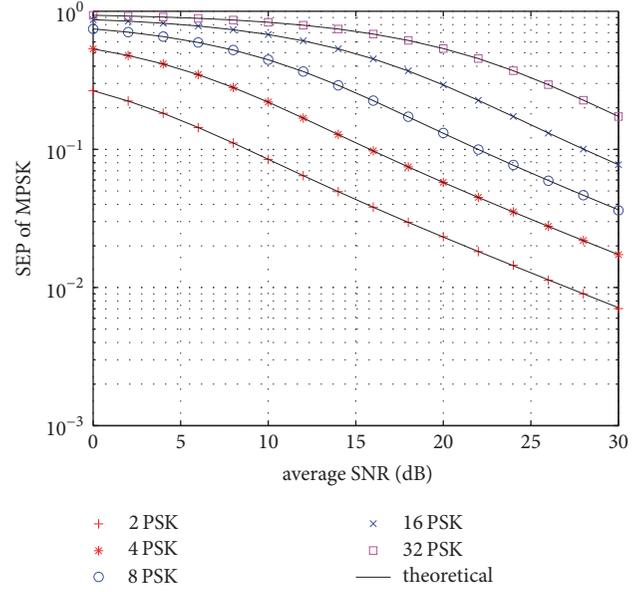
FIGURE 4: Comparison of SEP for $\pi/4$ DQPSK.

FIGURE 6: Comparison of SEP for MPSK.

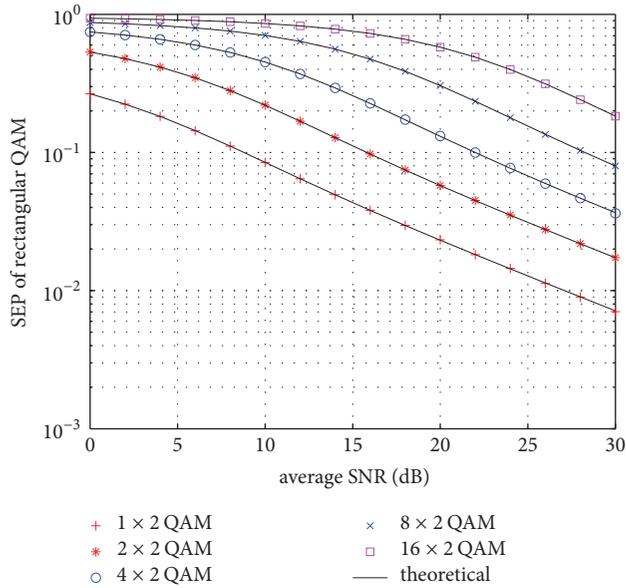


FIGURE 5: Comparison of SEP for MQAM.

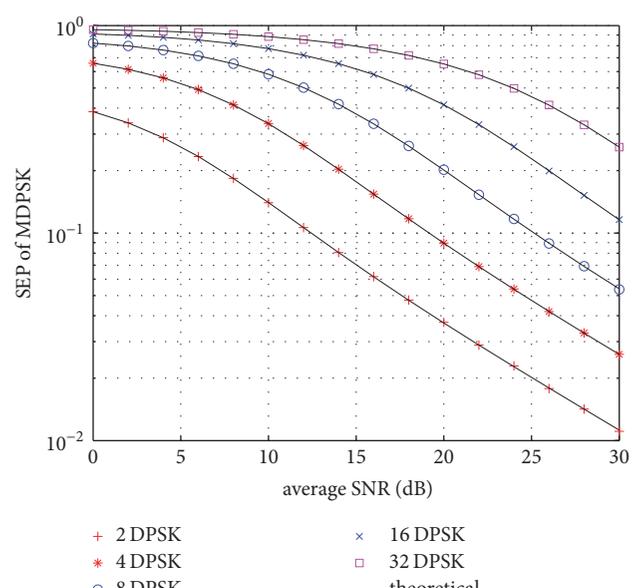


FIGURE 7: Comparison of SEP for MDPSK.

While for $m_s = 0.5, m_r = 1.5$, the SEP is 0.046485872872970 at 28 dB and 0.036583504432831 at 30 dB. The diversity gain becomes

$$\frac{10 \lg (0.046485872872970 / 0.036583504432831)}{30 - 28} \quad (17)$$

$$= 0.5202 \approx \min(m_s, m_r).$$

Although the parameters $m_r = \{0.5, 1.5\}$ are different, the results are the same. So they achieve the same diversity order.

Figures 5–7 show the average SEP of M -ary modulation schemes, where $m_s = 0.5, m_r = 1.5$. The average SEP increases with the increase of M because the minimum

distance between symbols becomes smaller. But the slopes of the SEP curves are nearly the same, implying the same diversity gain. The influence of cross correlation coefficient on SEP is drawn in Figure 8. When the correlation between two signals is small, the SEP is relatively small.

5. Conclusion

In this paper, we study the SEP of QAM, MPSK, MDPSK, and DQPSK modulation in cooperative AF system. Exact closed-form expressions for average SEP are obtained over independent Nakagami- m fading channels with integer plus

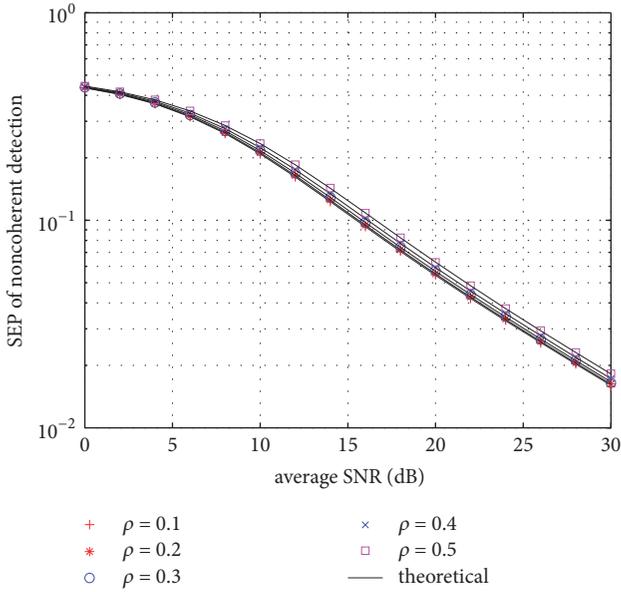


FIGURE 8: Comparison of SEP for correlated binary signals with noncoherent detection.

one-half m . Simulation results agree well with the theoretical analysis.

Mathematical Notations

- $f_{\gamma}(\cdot)$: Probability density function of the variable γ
- $M_{\gamma}(\cdot)$: Moment generating function of the variable γ
- $\lfloor \cdot \rfloor$: Floor function
- $Q(\cdot)$: Gaussian Q function
- ${}_2F_1$: Gauss hypergeometric function
- F_1 : Appell hypergeometric function
- $F_D^{(n)}$: Lauricella function.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant nos. 61761030, 61661028, 61701214, 61461030, and 61661032), by the China Postdoctoral Science Foundation (Grant no. 2017M622103), and by the Young Natural Science Foundation of Jiangxi Province (Grants nos. 20171BAB212002 and 20161BAB212038).

References

[1] Y. G. Kim and N. C. Beaulieu, "Exact closed-form solutions for the BEP of decode-and-forward cooperative systems in nakagami- m fading channels," *IEEE Transactions on Communications*, vol. 59, no. 9, pp. 2355–2361, 2011.

[2] Y. G. Kim and N. C. Beaulieu, "SEP of Decode-and-forward cooperative systems with relay selection in nakagami- m fading channels," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 5, pp. 1882–1894, 2015.

[3] J. Zhang, T. Zhang, J. Huang, and R. Yuan, "ABEP of amplify-and-forward cooperation in Nakagami- m fading channels with arbitrary m ," *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4445–4449, 2009.

[4] H. A. Suraweera and G. K. Karagiannidis, "Closed-form error analysis of the non-identical Nakagami- m relay fading channel," *IEEE Communications Letters*, vol. 12, no. 4, pp. 259–261, 2008.

[5] S. N. Datta and S. Chakrabarti, "Unified error analysis of dual-hop relay link in Nakagami- m fading channels," *IEEE Communications Letters*, vol. 14, no. 10, pp. 897–899, 2010.

[6] N. C. Beaulieu and Y. Chen, "An accurate approximation to the average error probability of cooperative diversity in nakagami- m fading," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2707–2711, 2010.

[7] N. Yang, M. ElKashlan, J. Yuan, and T. Shen, "On the ser of fixed gain amplify-and-forward relaying with beamforming in Nakagami- m fading," *IEEE Communications Letters*, vol. 14, no. 10, pp. 942–944, 2010.

[8] T. Q. Duong, H.-J. Zepernick, and V. N. Q. Bao, "Symbol error probability of hop-by-hop beamforming in Nakagami- m fading," *IEEE Electronics Letters*, vol. 45, no. 20, pp. 1042–1044, 2009.

[9] S. N. Datta, S. Chakrabarti, and R. Roy, "Comprehensive error performance analysis of distributed selection combining with multi-antenna amplify-and-forward relay over Nakagami- m fading channels," *IEEE Electronics Letters*, vol. 46, no. 22, pp. 1523–1525, 2010.

[10] F. Gong, J. Ge, and N. Zhang, "SER analysis of the mobile-relay-based M2M communication over double nakagami- m fading channels," *IEEE Communications Letters*, vol. 15, no. 1, pp. 34–36, 2011.

[11] T. Lu, J. Ge, Y. Yang, and Y. Gao, "On bit error performance of full-rate signal space cooperative communication over Nakagami- m fading channels," *IEEE Communications Letters*, vol. 16, no. 8, pp. 1224–1227, 2012.

[12] I. Trigui, S. Affes, and A. Stephenne, "Closed-form error analysis of variable-gain multihop systems in nakagami- m fading channels," *IEEE Transactions on Communications*, vol. 59, no. 8, pp. 2285–2295, 2011.

[13] L.-L. Yang and H.-H. Chen, "Error probability of digital communications using relay diversity over Nakagami- m fading channels," *IEEE Transactions on Wireless Communications*, vol. 7, no. 5, pp. 1806–1811, 2008.

[14] J. Yang, P. Fan, T. Q. Duong, and X. Lei, "Exact performance of two-way af relaying in nakagami- m fading environment," *IEEE Transactions on Wireless Communications*, vol. 10, no. 3, pp. 980–987, 2011.

[15] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products*, Academic Press, 7th edition, 2007.

[16] S. Sreng, B. Escrig, and M.-L. Boucheret, "Exact symbol error probability of hybrid/integrated satellite-terrestrial cooperative network," *IEEE Transactions on Wireless Communications*, vol. 12, no. 3, pp. 1310–1319, 2013.

[17] M. K. Fikadu, P. C. Sofotasios, S. Muhaidat, Q. Cui, G. K. Karagiannidis, and M. Valkama, "Error rate and power allocation analysis of regenerative networks over generalized fading channels," *IEEE Transactions on Communications*, vol. 64, no. 4, pp. 1751–1768, 2016.