Single and Multiple User Pair Cooperation Schemes with Delay Issues

by

Moyuan Chen

B.Sc., Beijing University of Posts and Telecommunications, 2009

A Thesis Submitted in Partial Fulfillment of the

Requirements for the Degree of

MASTER OF APPLIED SCIENCE

in the Department of Electrical and Computer Engineering

Single and Multiple User Pair Cooperation Schemes with Delay Issues

by

Moyuan Chen

B.Sc., Beijing University of Posts and Telecommunications, 2009

Supervisory Committee

Dr. Xiaodai Dong, Supervisor

(Department of Electrical and Computer Engineering)

Dr. T. Aaron Gulliver, Departmental Member

(Department of Electrical and Computer Engineering)

Supervisory Committee

Dr. Xiaodai Dong, Supervisor
(Department of Electrical and Computer Engineering)

Dr. T. Aaron Gulliver, Departmental Member
(Department of Electrical and Computer Engineering)

ABSTRACT

Cooperative communication is a promising technique to provide spatial diversity in a virtual multi-input and multi-output (MIMO) manner. However, as application evolves toward a more practical situation, realistic constraints and issues such as channel state information (CSI) assumption must be accounted when developing appropriate cooperative schemes. In this thesis, we have addressed delay related problems in both single user pair cooperation (SUPC) and multiple user pair cooperation (MUPC) networks. In SUPC, realizing that the outdated CSI caused by delay between relay selection instant and transmission instant can impair diversity order severely, we propose an opportunistic multiple relay selection (MRS) scheme to achieve desired diversity order and combat the variation of the wireless environment. On the other hand, for multiple user pairs cooperation (MUPC), we start from one of the notable work, two hop opportunistic relaying (THOR), and analyze its the delay related problems. We propose an opportunistic pair scheduling (OPS) scheme which can get rid of the buffer requirement at the relay nodes of THOR and incurs no loss in terms of throughput scaling. Furthermore, we extend OPS to a general scheduling

scheme, L scheduling, which can achieve controllable throughput-and-delay trande-offs.

Contents

List of Tables

List of Figures

List of Abbreviations

AaF	amplify and forward
APR	all participating relaying
AWGN	additive white Gaussian noise
BER	bit error probability
BHM	best harmonic mean
BW CS	best worse channel selection
CDF	cumulative distribution function
CSI	channel state information
CTS	clear-to-send
DaF	decode and forward
DS	decoding subset
GSC	generalized-selection-combining
i.i.d.	independent and identically distributed
i.n.d.	independent and nonidentically distributed
LSD	least start-up delay
MGF	moment generating function
MIMO	multiple-input and multiple-output
MORS	modified ORS
MRC	maximum ratio combining

MRS	multiple relay selection
MUPC	multiple user pair cooperation
N+NT-ORS	N plus normalized threshold opportunistic relay selection
OFDMA	orthogonal frequency division multiple access
OPS	opportunistic pair scheduling
ORS	opportunistic relay selection
OT-MRS	output-threshold MRS
PDF	probability density function
RS	relay selection
RTS	request-to-send
SINR	signal-to-interference-plus-noise ratio
SNR	signal-to-noise ratio
SRS	single relay selection
SUPC	single user pair cooperation
THOR	two-hop opportunistic relaying

Acknowledgement

First and foremost, my utmost gratitude to my supervisor Dr. Xiaodai Dong, whose sincerity and encouragement I will never forget. This thesis would not have been possible without her strong enthusiasm, constant motivation, invaluable guidance, and ample support. One simply could not wish for a better or friendlier supervisor.

I would like to express my sincere gratitude to my committee member Dr. T. Aaron Gulliver for his insightful guidance and constructive comments, and to Dr. Daniela Constantinescu for being as the external examiner.

Besides the supervisory committee members for my thesis, I would like to express my gratitude to Dr. Wu-Sheng Lu, Dr. Lin Cai, Dr. Antoniou, Dr. Kui Wu for their guidance and help through graduate courses, which equipped me with solid foundation in both theory and practice.

Here I would like to offer my deepest thanks to Yang, my girlfriend. Being one of the most important persons in my life, Yang has accompanied me through my entire master study. It would not have been possible for me to overcome so many difficulties in this two years study without her love and encouragement.

I also would like to express my gratitude to my dear group colleagues and friends in Victoria for their presences and help in both study and life. They are Ted C.-K. Liu, Yi Shi, Zhuangzhuang Tian, Wei Xu, Lebing Liu, Yuzhe Yao, Youjun Fan, Guowei Zhang, Shuai He, Xue Dong, Congzhi Liu, Chenyuan Wang, Qingzhong Li,

Tong Xue, Biao Yu, Binyan Zhao, Teng Ge, Lan Zhao, Xi Tu, Dan Li, Ping Li, Jun Zhu, Bojiang Ma.

Most importantly, this work would not have been possible without my parents. Without your love, patience, encouragement, understanding and care, I would not have been where I am today.

Dedication

To my dearest parents

Chapter 1

Introduction

In wireless networks, it is well known that signal fading arising from multipath propagation can be mitigated through the use of diversity [39]. The multiple-input and multiple-output (MIMO) diversity techniques are particularly attractive for its significant improvement to information rate and transmission reliability [16, 55]. However, high cost and complicate implementation issues bring challenges to MIMO systems. Cooperative communication has been demonstrated to be an effective way to combat wireless fading by providing spatial diversity without the need of multi-antenna configurations [28, 42]. When designing a feasible cooperative network, two aspects are especially important.

- Channel State Information (CSI). The availability of CSI is always an essential issue in designing a practical wireless network. For a static channel CSI at receiver side is typically assumed, since it is fairly easy to obtain the channel gains through the pilot sequence sent form the transmitter for channel estimation. However, to obtain CSI at transmitter side, it requires feedback mechanisms to send back the CSI from the receiver to the transmitter via a feedback path. Depending on system complexity and the functions of transmitter and receiver

nodes, different CSI assumptions are made.

- Centralized or distributed. When considering cellular network, it is commonly assumed that the nodes in the network are coordinated by central control. Thus, it is possible to perform optimization and other techniques that requires global CSI and high processing complexity. While in wireless ad hoc networks, usually the nodes have reduced functions and are distributed deployed. As a result, simple and efficient cooperation schemes with the use of only local CSI are desired.

In the following two sections, we will introduce background and related literature for the thesis. There exists numerous work on cooperative communications, studying various wireless setups. The review is by no means exhaustive. Instead, we start from two most common scenarios, i.e., single source and destination pair cooperative networks and multiple source and destination pairs, and try to characterize the different system behaviors and the metrics of interest of both system setups. Moreover, we focus on the two aspects we mentioned above when introducing the related works to distinguish them from each other and identify appropriate environment for each scheme. In both scenarios, multiple relays are available to assist the transmission between source nodes and destination nodes, via either amplify and forward (AaF) or decode and forward (DaF) cooperative protocol. The single source and destination pair cooperative networks are reviewed in Sec. 1.1, with emphasis on different relay selection schemes, while multiple source and destination pairs cooperative networks are reviewed in Sec. 1.2 focusing on network capacity.

1.1 Single User Pair Cooperation

We denote a source and destination pair by a user pair. In this section, we consider single user pair cooperation (SUPC). To date, there has been a large body of literature developing and studying cooperative schemes to fully exploit multi-user diversity when employing multiple relays in SUPC. In a cooperative network, multiple relays work as intermediate conduits to enable different spatial paths for relaying information. Laneman and Wornell in their seminal work [29] proposed the use of all available relays to perform space-time coded cooperative diversity protocols, which is henceforth called all-participating relaying (APR). However, with the use of multiple relays in orthogonal time slots, the overall system rate will decrease as the number of relays increases. In this sense, although the APR achieves full spatial diversity, it is spectrally inefficient. The recognition of the low efficiency of APR inspired a line of work that focuses on developing appropriate relay selection (RS) schemes that instruct not all the relays but a subset of relays to cooperate [3–5, 22, 24, 33, 41, 58]. According to the number of relays involved in the cooperation, the RS schemes can generally be classified into two categories: single-RS (SRS) and multiple-RS (MRS) schemes [24].

Among the notable works on the SRS schemes [3–5, 24, 41, 58], opportunistic relay selection (ORS) proposed by Bletsas *et. al.* [3–5] stands out in developing efficient relay selection schemes in both AaF and DaF, as well as achieving full spatial diversity in a distributed fashion without the requirement of global CSI. The key idea of ORS is to select and use the best relay in forwarding information from a set of available relays. The selection scheme is realized by a timer at each relay. By setting a timer whose length is inversely proportional to the estimated metric involving only local CSI, the relay is selected if its timer expires first. Because of this distributed timer-based architecture, ORS requires no network topology information, i.e., location of each

nodes, and is only based on local measurements of the instantaneous CSI[1]. The ORS is initially proposed in [4] with the emphasis on describing the scheme and addressing practical issues such as synchronization and protocol implementations. It is later extended to [5], in which the scheme is fully developed in both reactive and proactive manners and proved therein that selecting the relay with the best instantaneous CSI is outage optimal in both DaF and AaF modes.

Following the ORS, several other SRS schemes emerged in the same spirit as ORS but with different choice of relay selection criteria. The optimal SRS scheme that maximizes the end-to-end signal-to-noise ratio (SNR) while achieving full diversity with the highest throughput is examined in [24, 58]. A distributed nearest-neighbor protocol for RS is proposed in [41], where the user selects a neighboring node as the relay based on its proximity to the source node. The best worse channel selection is proposed in [4,45], where for each relay, the worse channel between the source-to-relay ($S \to R$) and the relay-to-destination ($R \to D$) links is recognized as its bottleneck channel, and among all the relays, the one whose bottleneck channel is the best gains the permission to forward. A derivative of the above methods is used for AaF relay systems in [4] and [40] where the relay with the best harmonic mean of the $S \to R$ and the $R \to D$ links is selected. Except the distributed nearest-neighbor protocol, it has been shown in [24] that the SRS schemes achieve full diversity under perfect CSI.

Despite its popularity in relay networks, SRS nonetheless suffers performance loss since it does not fully exploit spatial diversity. In contrast, multiple relay selection (MRS) not only fully utilizes available spatial diversity but it can also incorporate additional constraints in its design to make the problem formulation more realistic. For that, several MS schemes based on adaptive power control methods have been

[1]In this thesis, we use the term instantaneous CSI interchangeably with perfect CSI to denote the same meaning.

studied [25, 30, 32, 33]. In [33], a selection scheme of bit error probability (BER) minimization under total energy constraints is proposed. Refs. [30] and [32] consider MRS based on short-term aggregated relay power constraints with AaF relays. In [25], multiple relays based beamforming technique is proposed with AaF relaying under short-term power constraints on each node.

Although both beamforming and optimization techniques are possible strategies in MRS, they often entail excessive processing complexity and communication overhead. Moreover, arbitrary power adjustments are hard to manage and require more advanced nodes. As such, several researchers have investigated a more practical and easy-to-implement setup, where each relay either cooperates with its maximum power or keep silent [2,22–24]. In [24], Jing *et. al.* proposed a multiple relay selection scheme which maximizes the received SNR by exhaustive search. Unfortunately, the optimal MRS can not be found by linear algorithms so several suboptimal MRS schemes are proposed in [24] as substitutes. The generalized-selection-combining (GSC)-based MRS scheme developed in [22, 23] for both AaF and DaF relaying is another MRS that utilizes relays based on an on-or-off pattern, in which the top N relays are selected. Recently, Amarasuriya *et. al.* [2] proposed an output-threshold MRS (OT-MRS) scheme which selects a subset of relays that provide adequate combining SNR.

Nevertheless, one problem which is often overlooked in the aforementioned literatures is that, in practice, the available CSI at hand may not be instantaneous but outdated. The outdated CSI phenomena may occur when there exists significant time delay between the selection and data transmission instants [48]. The quality of the outdated CSI depends on the correlation coefficient ρ between the instantaneous CSI and the outdated CSI with zero being uncorrelated. As a result, when the user is in motion with respect to (w.r.t.) the relay, the CSI used for RS becomes an outdated version of the CSI observed at data transmission. For instance, given a 2 ms delay

between RS instant and transmission instant at a center frequency of 900 MHz, ρ can vary from unity down to zero. Thereby, outdated CSI occurs almost in all scenarios when the user is in motion or when non-fixed relays are used. There have been emerging attentions on studying cooperative networks with the presence of outdated CSI. The impact of outdated CSI on the performance of SRS scheme is investigated for both DaF and AaF relaying systems in [48] and [34], respectively. It is stated in [48] that the diversity order of the original ORS system with DaF reduces to unity under outdated CSI, and similar behavior is also observed for AaF in [34]. However, exactly how to mitigate the effect of outdated CSI is not addressed in both cases. Furthermore, a thorough study on the impact of outdated CSI on all existing SRS schemes is lacking. In [49], Vicario *et. al.* proposed a modified ORS with SNR estimation. It first generates an estimate of the current instantaneous CSI based on the outdated CSI and then selects the relay with the best estimated CSI. In [56], it is shown that an opportunistic scheme with outdated CSI may cause system performance loss as compared to a generalization of hybrid-automatic repeat request.

1.2 Multiple User Pair Cooperation

In this section, we take one step further by changing the network setup from SUPC to multiple user pair cooperation (MUPC), in which multiple source nodes transmit messages to their associated destination nodes. Stimulated by the fact that communication devices are becoming increasingly pervasive, there has been significant and increasing interest in MUPC in recent years. Such networks are distinguished from conventional infrastructure networks for their distributed and unsupervised nature, which have sparkled theoretic interests in characterizing the fundamental capacity limits in the form of capacity scaling laws, as well as developing appropriate architec-

tures and practical communication schemes that seek to approach these performance boundaries [6, 11, 12, 14, 20, 21, 35, 38, 54].

Gupta and Kumar [21] spearheaded this line of research by considering a communication network where n source nodes transmit data to their designated destination nodes through a shared wireless medium. They showed that in the limiting region, as the number of nodes n grows sufficiently large, $\Theta(\sqrt{n/\log n})$ transmitters can talk simultaneously to randomly chosen receivers. They have also derived a theoretical throughput upper bound of $\Theta(\sqrt{n})$, achieved with optimal node locations, optimally assigned traffic patterns, and optimal range chosen for each transmission. This performance gap was later closed by Franceschetti *et al.* [14] via percolation theory. Nevertheless, the throughput scaling of $\Theta(\sqrt{n})$ is still a disheartening result since it suggests that as the number of node n goes to infinity, the per source-destination (S-D) pair throughput will necessarily diminish to zero. It is not long for Grossglauser and Tse [20] to realize that the fundamental limitation of the Gupta-Kumar network model lies in the lack of node mobility. By allowing the nodes to move across the deployment area and split their packet streams to as many different nodes as possible, they demonstrate the up-to-now best aggregated throughput scaling of $\Theta(n)$, even when nodes are restricted to move along one-dimensional paths [12].

The Gupta-Kumar model postulates a *dense* network configuration, where the total area is fixed and the density of nodes increases [38]. This assumption has led to sublinear scaling of the system throughput, since as growing number of S-D pairs are packed into the area, inter-node interference becomes the ultimate performance bottleneck. The recognition of the interference-limited nature of Gupta-Kumar's setup has inspired a line of research that focuses on developing appropriate architectures and protocols that manage interference efficiently. As such, the level of CSI knowledge proves to be an essential factor that affects the capability of handling interference

significantly [6, 38]. With *global CSI*, meaning that the channel knowledge of the whole network is available *a priori* to all the nodes, Özgür *et al.* [38] described a hierarchical cooperation scheme by allowing both source nodes and destination nodes to cooperate in *clusters* and form a *virtual* MIMO antenna arrays. The mutually interfering signals are therefore turned into useful ones, yielding a linear throughput scaling of $\Theta(n)$. Cadambe and Jafar [6] introduced the *interference alignment* technique whereby through a centralized transmitting signal vector design, the interfering signals remain distinguishable at the intended receivers which constitutes a linear scalable network.

The assumption of global CSI and centralized coordination entails excessive overhead that drastically affects the effective throughput [38]. As such, several researchers have investigated the performance of practical schemes that operate on *limited CSI knowledge*. In accordance with the recent interests in relay-assisted communication [28, 42], a wireless network structure with two-hop relaying emerges as a common problem setup. Dana and Hassibi [11] considered the communication of n S-D pairs aided by k AaF relays [28]. Each relay is assumed to have full knowledge of its own *backward* (sources-to-relay) channel and *forward* (relay-to-destinations) channel (but not the channel knowledge pertain to other relays), so that the relays can perform *distributed beamforming*. They demonstrated in this case the attainment of a system throughput scaling of $\Theta(n)$ with the support of $k = \Theta(n^2)$ nodes in the network functioning as relays. Since the number of relays k dominates the size of the network in the limiting region, equivalently, it amounts to a system throughput of $\Theta(\sqrt{k})$. Morgenshtern and Bölcskei [35] built upon Dana and Hassibi's model but instead assume each relay has access to only the backward and forward channel knowledge of a subset of S-D pairs. Not surprisingly, the relaxation in the level of CSI knowledge reduced the throughput from $\Theta(\sqrt{k})$ to $\Theta(\sqrt[3]{k})$.

Among the notable works on relay-assisted networks, the two-hop opportunistic relaying (THOR) scheme proposed by Cui *et al.* [9, 10] stands out in understanding the throughput limit of two-hop DaF networks [28], as well as in achieving optimal throughput scaling with overhead-inexpensive schemes. THOR was initially proposed in [9] for a Rayleigh fading environment and later extended in [10] to incorporate a general fading model with diverse system settings, where the key idea is to schedule at each hop only the subset of nodes that can benefit from *multiuser diversity gain* [26]. Specifically, THOR regards the two-hop wireless network as a cascade of two (isolated) single-hop networks and applies the classic opportunistic scheduling technique to each hop individually. Each receiver has full knowledge of the backward channel and feeds back only the "best" connection index to the corresponding transmitter. Cui *et al.* showed that as long as the number of relays k scales no faster than $\Theta(\log n)$, THOR is equally capable of restoring the two-hop interference channel into k orthogonal parallel channels, as that of the centralized counterpart with full cooperation and global CSI. Notwithstanding the enlightening throughput results, THOR has delay issues as confirmed in a recent study by Wang *et al.* in [50]. They studied THOR's throughput-delay tradeoff performance and demonstrated an upper-bound of the end-to-end delay of $O(n)$. A better upper bound $O\left(n/\sqrt{\log n}\right)$ was also found achievable through a similar redundant scheduling procedure in the spirit of [37]. In fact, the associated delay issue along with opportunistic scheduling schemes are well-recognized [20]. It is generally agreed that capacity improvements by exploiting opportunism affects communication delays considerably [31]. Accordingly, there have been several recent studies that dwell upon the relationship between the achievable throughput and the packet delay in large wireless networks [15, 37, 47, 50]. Neely and Modiano [37] showed that the delay of Grossglauer-Tse's model is upper-bounded by $O(n)$. They pointed out that letting each user send redundant packets along multiple paths

to the destination can improve the delay performance. Gamal *et al.* [15] cast the Gupta-Kumar fixed network model and Grossglauer-Tse mobility model into a general framework as two extreme cases, the throughput-delay tradeoff of which were shown to be $\Theta(n)$ and $\Theta\left(\sqrt{n}v(n)\right)$, respectively, where $v(n)$ denotes the velocity of the mobile nodes. In order to achieve the optimal delay for any given throughput, they proposed schemes that vary in the number of hops, transmission range and the degree of node mobility. In light of the delay bound $O(n)$, Toumpis and Goldsmith [47] developed a scheme to achieve $\Theta\left(n^{(d-1)/2}(\log n)^{-5/2}\right)$ throughput scaling under a delay profile of $O(n^d)$ with $0 < d < 1$.

1.3 Contributions

So far, we reviewed cooperative communication networks from two classifications, SUPC and MUPC, respectively. In SUPC, significant progress has been made in terms of designing simple and efficient relay selection schemes achieving high transmission performance dedicated to the only user pair involved in the communication. On the other hand, in MUPC, the emphasis is placed on investigating the fundamental capacity limit of the whole network and the throughput scaling law, in the presence of interference from other user pairs' communication.

The review of literature of both SUPC and MUPC revealed delay related issues. For SUPC, the multi-user diversity order of traditional ORS decreases to 1 with outdated CSI, which is caused by the delay between the selection instant and the transmission instant. For MUPC, one of the recent outstanding work THOR achieves throughput scaling linear to the number of relays with sacrifice on delay aspects, including packet delay due to the use of buffers at the relay nodes and system startup delay.

In this thesis, we continue the study on developing simple and efficient cooperative schemes for both SUPC and MUPC. The contributions of the thesis can be summarized as follows.

For SUPC, we first outline several SRS schemes including ORS which achieve full diversity when using perfect CSI, and prove that the diversity order of these schemes reduces to unity if outdated CSI exists. To deal with the performance deterioration due to the outdated CSI phenomena, we propose a new multiple relay selection scheme for both AaF and DaF relay systems, namely the N plus normalized threshold opportunistic relay selection (N+NT-ORS). The performance of the scheme is analyzed theoretically in terms of outage probability, and its asymptotic diversity order is also examined. In addition, we outline a distributed N+NT-ORS selection protocol which makes no assumption of the global CSI at each relay. Numerical results confirm the analysis and show the advantage of the N+NT-ORS scheme against existing counterparts in highly dynamic networks.

For MUPC, by selecting THOR as our starting point, we explore the possibility of achieving *controllable* delay-throughput tradeoff without sacrificing *linear* throughput scaling. We begin with an opportunistic pair scheduling scheme (OPS) by restricting the relays to schedule only matched source-destination pairs over two hops. This scheme is found to maintain the throughput linearity yet reduce the end-to-end delay to the minimum. We then develop this scheme further into a general L-scheduling scheme that captures THOR and OPS as two extreme cases and achieves arbitrary throughput-delay tradeoffs in a controlled manner. Through adjusting a design parameter L, desired tradeoff can be achieved catering to applications with diverse delay and throughput requirements.

1.4 Thesis Outline

The rest of the thesis is organized as follows:

- **Chapter 2** studies relay selection schemes in cooperative communications in the presence of outdated CSI. Thereafter, observing the fact that the outdated CSI impairs the diversity order of existing single relay selection schemes, we propose a multiple relay selection scheme which can obtain desired diversity order even when using outdated CSI. We compare our scheme to its predecessors in terms of outage probability, diversity order and robustness to the variation of the wireless environment in dynamic networks.

- **Chapter 3** investigates throughput scaling law in wireless ad hoc networks. Driven by the fact that the delay issue of THOR is caused by the independent scheduling of two transmission hops, we propose a pair scheduling scheme which can avoid the use of buffer at relay nodes and bears no loss in terms of throughput scaling compared to THOR. Closed-form throughput lower bounds and scaling laws are derived under Nakagami fading environment. Furthermore, we propose a general opportunistic scheme that incorporates OPS and THOR as special cases.

- **Chapter 4** concludes the thesis and discusses possible future work.

Notation : We write $X \sim \Gamma(m, 1/m)$ to indicate a random variable (RV) X that follows the Gamma distribution with shape m and scale $1/m$. For a RV X with cumulative distribution function $F(x)$, we use $\overline{F}(x)$ to denote the corresponding complementary cumulative distribution function or the tail distribution. The cardinality of the set $\mathcal{X}$ is denoted by $|\mathcal{X}|$. $\mathbb{Z}^+$ denotes the set of natural numbers (excluding zero). For two functions $f(n)$ and $g(n)$, $f(n) = O\left(g(n)\right)$ means

that $\lim_{n\to\infty} |f(n)/g(n)| < \infty$, and $f(n) = \Omega(g(n))$ means that $g(n) = O(f(n))$. We let $f(n) = o(g(n))$ denote $\lim_{n\to\infty} |f(n)/g(n)| = 0$ and $f(n) = \Theta(g(n))$ denote $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$.

Chapter 2

A New Multiple Relay Selection in SUPC with Outdated CSI

2.1 Background

As discussed in the introduction chapter, in order to combat the severe diversity loss due to outdated CSI, appropriate MRS schemes are required. However, to the best knowledge of the author, there has not been any work studying MRS schemes with outdated CSI. In light of the existing RS schemes, we realize that a distributed RS scheme that is independent of global CSI while being both easy to implement and manage is always favorable. For that we ask the following: Is it possible to keep the distributed timer structure of the ORS for an MRS scheme so that global CSI is not required at each relay? Can the MRS scheme achieve adequate cooperative diversity in the presence of outdated CSI? Can the MRS scheme be robust against a highly mobile environment[1] without having to frequently adjust the number of participating relays? These questions are addressed in this chapter.

[1]We define mobile environment in terms of either node mobility, i.e., high node mobility means ρ changes quickly over time, or network size, i.e., highly mobile network size means nodes arriving and exiting frequently within the network. Please refer to Secs. 2.6.1 and 2.6.2 for more information.

The remainder of this chapter is organized as follows: Section 2.2 describes our system model, Section 2.3 analyzes the impact of outdated CSI on several SRS schemes. Our MRS scheme is proposed in Section 2.4, together with the performance analysis. We present the distributed selection protocol for the N+NT-ORS in Section 2.5. Finally, Section 2.6 presents the simulation results and concluding remarks are given in Section 2.7.

2.2 System Model

We consider a cooperative network with one source node (S), one destination (D) and K relays (R) as depicted in Fig. 2.1. We assume that S communicates with

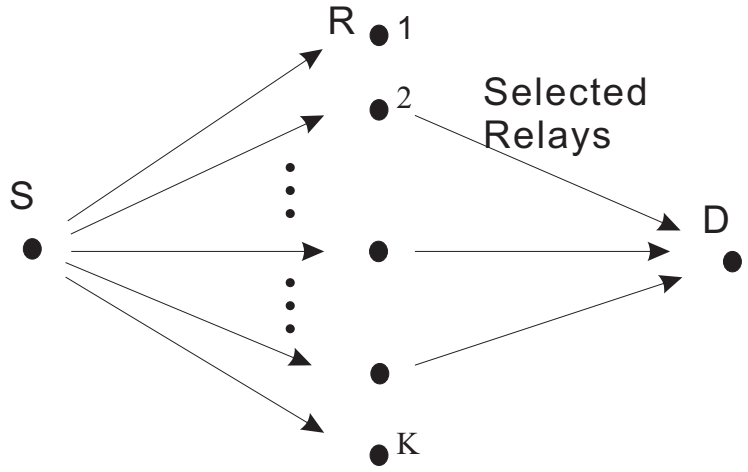


Figure 2.1: System Model of SUPC.

D in half-duplex dual-hop mode via K relays. For the ease of analysis we assume that there is no direct link between S and D due to direct line-of-sight blockage. All nodes are equipped with single antenna. Denote the received signal in an arbitrary

link $A \to B$ between two nodes A and B as:

$$r_B = h_{A,B} x_A + n_B \,,$$

where $x_A \in \mathbb{C}$ is the transmitted symbol from node A with power $P_A = \mathbb{E}[|x_A|^2]$[2], $n_B \in \mathbb{C}$ is the additive white Gaussian noise (AWGN) at node B with double-sided spectrum density $N_0/2$ and $h_{A,B} \in \mathbb{C}$ is the channel gain between the link $A \to B$. Denote $\gamma_{A,B} = P_A |h_{A,B}|^2 / N_0$ as the instantaneous SNR in the link $A \to B$ in a given time slot and $\bar{\gamma}_{A,B} = P_A \Omega_{A,B} / N_0$ as its long-term average. For ease of analysis, two assumptions are made: 1) equal power allocation for each transmitting node, i.e., $P_A = P$; 2) independent and identically distributed (i.i.d.) and symmetric $S \to R$ and $R \to D$ channels[3] with normalized Rayleigh fading, i.e., $\Omega_{A,B} = \Omega$, $h_{A,B} \sim \mathcal{CN}(0,1)$. Extension to unequal power allocation, independent and non-identically distributed and asymmetric channels with any other channel statistics in analysis can be made following a similar approach presented herein. Under these assumptions, $\bar{\gamma} = P/N_0$ denotes the average per-hop SNR. Moreover, we assume the transmitting relays are perfectly synchronized.

Throughout this chapter, both AaF and DaF protocol are considered as relaying strategies. Let xSRS denotes a particular SRS scheme[4], the rest of this section generalizes to how xSRS would work in both AaF and DaF relay systems.

[2] $\mathbb{E}[\cdot]$ denotes the expectation operation and $|x_A|^2$ denotes the squared magnitude for complex number x_A.

[3] This assumption is only imposed for the ease of analysis herein. We will compare the performance of different RS schemes under independent but non-identically distributed channels via extensive simulation in Sec. 2.6.

[4] The SRS schemes will be described in more detail in Sec. 2.3.

2.2.1 AaF Relaying

In a cooperative network with AaF strategy, each relay normalizes the signal received from the source during the first phase then orthogonally retransmits to the destination in the second phase. The end-to-end SNR of an $S \to R \to D$ link is well known and can be expressed as [24, 57, 58]

$$\gamma_{S,k,D} = \frac{\gamma_{S,k}\gamma_{k,D}}{\gamma_{S,k} + \gamma_{k,D} + 1}\,,\ k = 1, \cdots, K\,, \tag{2.1}$$

where $\gamma_{S,k} = P|h_{S,k}|^2/N_0$ and $\gamma_{k,D} = P|h_{k,D}|^2/N_0$.

The selection scheme is performed among all K relays with respect to their end-to-end SNR. Define $\mathcal{G}_{\text{AaF}}^{x\text{SRS}}(k)$ as the function of interest for the xSRS scheme in AaF relaying. By ordering all the relays in terms of this function value, the xSRS scheme selects the "best" relay with the largest function value as

$$b_{\text{AaF}}^{x\text{SRS}} = \arg \max_{1,\cdots,K} \mathcal{G}_{\text{AaF}}^{x\text{SRS}}(k)\,, \tag{2.2}$$

where $b_{\text{AaF}}^{x\text{SRS}}$ is the selected relay.

2.2.2 DaF Relaying

For the case of DaF, only the relays that successfully receive the message during the first phase are eligible for selection. These relays form the decoding subset $\mathcal{DS}$[5], as defined by

$$\mathcal{DS} = \{k\colon\ \log_2(1+\gamma_{S,k}) \geq 2R\} = \left\{k\colon\ \gamma_{S,k} \geq 2^{2\mathcal{R}} - 1\right\}\,,$$

[5]We let $|\mathcal{DS}|$ denotes the cardinality of the set $\mathcal{DS}$.

where $\mathcal{R}$ is the end-to-end spectral efficiency resulted from the 2-hop system requiring two phases. Relay k is assumed to decode successfully if no outage happens during the first phase.

Then, the selection schemes is carried out on the candidate relays in the decoding subset with respect to their $R \to D$ SNRs, i.e.,

$$b_{\mathrm{DaF}}^{x\mathrm{SRS}} = \arg \max_{k \in \mathcal{DS}} \mathcal{G}_{\mathrm{DaF}}^{x\mathrm{SRS}}(k)\,. \tag{2.3}$$

2.3 Existing SRS and Their Diversity Order with Outdated CSI

In this section, we first review several SRS schemes that achieve full diversity order with perfect CSI. After describing the outdated CSI model, we establish a theorem on the diversity order of these schemes under the influence of outdated CSI.

2.3.1 Opportunistic Relay Selection

In [5], the ORS is proposed which chooses the "best" relay that maximizes either the instantaneous end-to-end SNR amongst $S \to R \to D$ links for AaF relaying or the instantaneous one-hop SNR amongst $R \to D$ links for DaF relaying. Therefore, the ORS scheme is described as

$$\mathcal{G}_{\mathrm{AaF}}^{\mathrm{ORS}}(k) = \gamma_{S,k,D}\,, \tag{2.4}$$

$$\mathcal{G}_{\mathrm{DaF}}^{\mathrm{ORS}}(k) = \gamma_{k,D}\,. \tag{2.5}$$

The ORS is equivalent to the SRS schemes proposed in [57, 58]. The result in this literature shows that this scheme achieves the full diversity order of K when using instantaneous CSI.

2.3.2 Modified ORS with SNR Estimation

In [49], the conventional ORS in DaF systems is shown to be inferior when using outdated CSI. The authors proposed an alternative based on channel estimation. It generates an estimate of the instantaneous SNR $\gamma_{k,D}$ using the outdated SNR $\hat{\gamma}_{k,D}$ and the coefficient ρ as

$$\mathcal{E}(\gamma_{k,D}|\hat{\gamma}_{k,D},\bar{\gamma}) = \rho^2\hat{\gamma}_{k,D} + (1+\rho^2)\bar{\gamma}\,, \tag{2.6}$$

where $\mathcal{E}(\cdot)$ is the estimator function. Therefore, the selection function for this modified ORS (MORS) scheme is

$$\mathcal{G}_{\mathrm{DaF}}^{\mathrm{MORS}}(k) = \mathcal{E}(\gamma_{k,D}|\hat{\gamma}_{k,D},\bar{\gamma})\,. \tag{2.7}$$

Note that this MORS scheme proposed in [49] only considers DaF systems.

2.3.3 Best Worse Channel Selection

The best worse (BW) channel selection (CS) is proposed in [4, 45], where the relay with the largest of the worst channels from either $S \to R$ or $R \to D$ links gains the permission to retransmit. The selection function is

$$\mathcal{G}_{\mathrm{AaF}}^{\mathrm{BW}}(k) = \min(\gamma_{S,k},\gamma_{k,D})\,. \tag{2.8}$$

While ORS aims to maximize the received SNR, the BW CS considers the balance of the two channels for each relay. Ref. [4] has shown that when the CSI available is instantaneous the BW CS can achieve full diversity.

2.3.4 Best Harmonic Mean (BHM) Selection

The selection method used in [4] and [40], again for AaF relay systems, chooses the relay with the best harmonic mean of the $S \to R$ and the $R \to D$ links. The selection function is

$$\mathcal{G}_{\text{AaF}}^{\text{BHM}}(k) = \frac{1}{\gamma_{S,k}^{-1} + \gamma_{k,D}^{-1}} \,. \tag{2.9}$$

Approximated symbol error rate analysis of this scheme is performed in [40] using an upper bound on the end-to-end SNR. This result shows that the best harmonic mean selection achieves the full diversity of K when instantaneous CSI is available.

2.3.5 Outdated CSI and its Impact on Diversity Order

Due to the possible dynamic movements in between the nodes, the CSI in the selection instant may possibly be outdated, thus it is necessary to study the impact of the outdated CSI on different relay selection schemes. Define $\hat{h}_{A,B}$ as a delayed version of the instantaneous CSI for the link $A \to B$. Note that the outdated CSI $\hat{h}_{A,B}$ follows the same distribution as the instantaneous one, i.e., $\hat{h}_{A,B} \sim \mathcal{CN}(0,1)$. The instantaneous and outdated CSIs are jointly Gaussian and $h_{A,B}$ conditioned on $\hat{h}_{A,B}$ follows a Gaussian distribution [48]:

$$h_{A,B}|\hat{h}_{A,B} \sim \mathcal{CN}\left(\rho\hat{h}_{A,B}, 1-\rho^2\right) \,,$$

where ρ is the correlation coefficient between $\hat{h}_{A,B}$ and $h_{A,B}$. Throughout this chapter, Jakes' model is adopted to represent the outdated CSI and thus $\rho = J_0(2\pi f_{d_k} T_{D_k})$, where f_{d_k} stands for Doppler frequency, T_{D_k} is the delay between the selection instant and the transmission instant, and $J_0(\cdot)$ denotes the zero-order Bessel function of the first kind. Thereby, from the expression, it can be observed that when T_{D_k} is given, Doppler frequency $f_{d_k} = v f_c / c$ determines the value of ρ, where v is the relative

velocity between the relay and the user, f_c is the carrier frequency, c is the velocity of light. Therefore, from the outdated CSI $\hat{h}_{A,B}$, the delayed SNR is $\hat{\gamma}_{A,B} = P|\hat{h}_{A,B}|^2/N_0$. One can easily show that the instantaneous SNR, $\gamma_{A,B}$, conditioned on its delayed version $\hat{\gamma}_{A,B}$, follows a non-central chi-square distribution with 2 degrees of freedom, whose probability density function (PDF) is expressed as [48]:

$$f_{\gamma_{A,B}|\hat{\gamma}_{A,B}}\left(\gamma_{A,B}|\hat{\gamma}_{A,B}\right) = \frac{1}{\bar{\gamma}(1-\rho^2)} e^{\frac{-(\gamma_{A,B}+\rho^2\hat{\gamma}_{A,B})}{\bar{\gamma}(1-\rho^2)}} \cdot I_0\left(\frac{2\sqrt{\rho^2\gamma_{A,B}\hat{\gamma}_{A,B}}}{\bar{\gamma}(1-\rho^2)}\right), \qquad (2.10)$$

where $I_0(\cdot)$ stands for the zero-order modified Bessel function of the first kind.

The following theorem characterize the impact of outdated CSI on the above mentioned SRS schemes.

Theorem 1. *The diversity order of the ORS, MORS, BW, BHM selection schemes with outdated CSI and* $0 \leq \rho < 1$ *equals to* 1.

Proof: See Appendix A.

Note that although MORS is shown to outperform the conventional ORS in the presence of outdated CSI [49], the diversity order remains 1. In other words, the detrimental impact of the outdated CSI on the ORS scheme can not be mitigated by using MORS.

2.4 Proposed Multiple Relay Selection Scheme

As outlined in the previous section, irrespective to the selection criteria the diversity order of the SRS schemes is guaranteed to reduce to unity under outdated CSI. Therefore, in this section, we propose and analyze a MRS scheme to overcome this performance deterioration.

2.4.1 N Plus Normalized Threshold Opportunistic Relay Selection (N+NT-ORS)

With the outdated CSI from $S \to R$ and $R \to D$ links available at the k-th relay, our proposed scheme first computes either the end-to-end SNR $\hat{\gamma}_{S,k,D}$ for AaF or the one-hop $R \to D$ SNR $\hat{\gamma}_{k,D}$ for DaF. Then, the MRS scheme arranges all of the computed SNRs from the candidate set in descending order, where the candidate set is defined as either all K relays for AaF or relays in the decoding subset of $S \to R$ links for DaF. Thereafter, it opportunistically selects N ($N \leq K$ or L, where L is the size of the decoding subset for DaF) best relays from this ordered candidate set. Finally, the ratio of the SNR on the remaining candidate relays to that of the N-th highest SNR relay are tested against a normalized threshold $\mu \in [0, 1]$ and only those relays passing this test are selected in addition to the N best relays. It is possible that the size of the decoding subset in DaF relaying is smaller or equal to N. In that case, all relays in the decoding subset are selected. Due to the SNR normalization and a comparison to the threshold μ combined together with the ORS, this selection scheme is named as the N+NT-ORS MRS scheme.

The simple goal of this selection scheme is to ensure that a proper number of relays are selected. If one simply selects N best relays without the use of μ, whose AaF and DaF cases are equivalent to the GSC-based MRS scheme from [22, 23][6], it is still quite difficult to determine N especially in a dynamic cooperative network where either D is moving w.r.t. R or the number of R varies. In that case, the number N may be chosen to be too small leading to inadequate performance or too large resulting in high power consumption. On the contrary, if relay selection is operated only according to μ (i.e., $N = 1$), in the same spirit as the NT-GSC proposed by Sulyman *et. al.* [46] which addresses the shortcomings of the fixed diversity branch combining of the conventional

[6]Note in [22, 23], the outdated CSI phenomena and its impact on performance is not considered.

GSC receiver [1, 13] in non-relay communication systems, its diversity order is only one according to the analysis later in this chapter. Our N+NT-ORS is inspired by the N+NT-GSC proposed by Xiao *et. al.* [53] which combines the conventional GSC and NT-GSC in non-relay communication systems. The rest of the section presents the performance analysis of the proposed scheme in both AaF and DaF relaying.

2.4.2 Amplify-and-forward Relaying

For AaF relaying, the outdated end-to-end SNR of the $S \to k \to R$ link is given by

$$\hat{\gamma}_{S,k,D} = \frac{\hat{\gamma}_{S,k}\hat{\gamma}_{k,D}}{\hat{\gamma}_{S,k} + \hat{\gamma}_{k,D} + 1}\,. \tag{2.11}$$

For tractability of subsequent analysis, a well known tight upper bound for $\hat{\gamma}_{S,k,D}$, which is also exploited in [22], is written as

$$\hat{\gamma}_{S,k,D} \leq \hat{\gamma}_k = \min(\hat{\gamma}_{S,k}, \hat{\gamma}_{k,D}),\ k = 1, \cdots, K\,, \tag{2.12}$$

One can easily show that the PDF of $\hat{\gamma}_k$ is $f_{\hat{\gamma}_k}(\gamma) = (1/\bar{\gamma}_k)e^{-\gamma/\bar{\gamma}_k}$, where $\bar{\gamma}_k = \bar{\gamma}_{S,k}\bar{\gamma}_{k,D}/(\bar{\gamma}_{S,k} + \bar{\gamma}_{k,D}) = \bar{\gamma}/2$. Of all independent individual end-to-end SNR upper-bounds, $\hat{\gamma}_1, \hat{\gamma}_2, \cdots, \hat{\gamma}_K$, the N-th largest among them, denoted by $\hat{\gamma}^{(N)}$, is chosen to be the anchor element. The received sum SNR of the AaF N+NT-ORS scheme is $\gamma_{\text{ub}} = \sum_{k=1}^{K} T(\hat{\gamma}_k)$, where $T(\hat{\gamma}_k)$ is given by

$$T(\hat{\gamma}_k) = \begin{cases} 0, & \hat{\gamma}_k < \mu\hat{\gamma}^{(N)} \\ \gamma_k, & \hat{\gamma}_k \geq \mu\hat{\gamma}^{(N)} \end{cases}\,. \tag{2.13}$$

Theorem 2. *The MGF of γ_{ub} in AaF relaying with outdated CSI can be computed as*

$$\Phi_{\gamma_{ub}}^{AaF}(s) = K\binom{K-1}{N-1}\sum_{k=0}^{K-N}\sum_{q=0}^{K-N-k}\sum_{m=0}^{k}\left(\frac{1}{1-\frac{s\bar{\gamma}}{2}}\right)^{K-1-k}\frac{(K-N)!(-1)^{q+m}}{(K-N-k-q)!m!q!(k-m)!}\frac{1}{H(s)}, \tag{2.14}$$

where

$$H(s) = \left(1-\frac{s\bar{\gamma}}{2}\right)\left[(K-N-k-q)\mu+(q+N)\right]+m\mu\left[1-\frac{s\bar{\gamma}}{2}(1-\rho^2)\right].$$

Proof: See Appendix B.

The number of selected relays J is a discrete random variable with range $1 < J < K$. It indicates the tradeoffs amongst power consumption, complexity and performance. Hence, it is necessary to analyze J.

Theorem 3. *The average number of the selected relays $\bar{J}$ of N+NT-ORS in AaF relaying is*

$$\bar{J}_{N+NT\text{-}ORS}^{AaF} = K\sum_{m=0}^{K-N-1}\frac{(K-N-1)!(K-1)!(-1)^m}{(K-N-1-m)!(K-N-1)!(N-1)!m!}\cdot\frac{1}{N+m+\mu}. \tag{2.15}$$

Proof: See Appendix C. It is worth noting that $\bar{J}_{N+\text{NT-ORS}}^{\text{AaF}}$ is independent of both γ and ρ.

Outage Probability

The system outage probability is defined as the probability that the mutual information I between the source and the destination is lower than a threshold, which is the target data rate $\mathcal{R}$. The mutual information for AaF N+NT-ORS is given by $I = 1/(J+1)\log_2(1+\gamma)$, where J is the number of selected relays and γ is the

end-to-end SNR [17]. Therefore, the outage probability is given by

$$P_{\text{out}}^{\text{AaF}}(J) = \Pr(I < \mathcal{R}) = \Pr\left(\gamma < 2^{(J+1)\mathcal{R}} - 1\right) . \tag{2.16}$$

Since J is a discrete random variable, the outage probability can be obtained in terms of a long term average $\mathbb{E}_J\left[P_{\text{out}}^{\text{AaF}}(J)\right] = \sum_{j=1}^{K} \Pr(J = j) P_{\text{out}}^{\text{AaF}}(J)$. Given that the average number of selected relays is independent of γ, we obtain

$$P_{\text{out}}^{\text{AaF}} = \mathbb{E}_J\left[P_{\text{out}}^{\text{AaF}}(J)\right] = \Pr\left(\gamma < 2^{(\mathbb{E}[J]+1)\mathcal{R}} - 1\right) = \Pr\left(\gamma < 2^{(\bar{J}+1)\mathcal{R}} - 1\right) . \tag{2.17}$$

There exists a number of ways to evaluate the outage probability with given MGF. One widely adopted method, which is extensively used in [22, 23], is that first rewrite the MGF using partial fraction which is in the general form of $(1 + As)^{-m}$ and then take the inverse Laplace transform of the MGF using the fact that $\tilde{\mathcal{S}}\{(1 + As)^{-m}\} = \frac{1}{(m-1)!A^m} x^{m-1} e^{-x/A}$. As such, the exact closed form pdf and cdf expressions can be obtained, which can be further exploited to derive BER, ergodic capacity, etc. On the other hand, Ko *et. al.* proposed a numerical technique in [27] to approximate outage probability with given MGF. Due to the fact that such method enjoys low computational complexity and high approximation accuracy and the outage probability is adequate to characterize the performance of our scheme, especially in terms of diversity order, we adopt Ko's method to calculate the outage probability. Particularly, if the MGF is known as $\mathcal{M}_\gamma(s)$, we have

$$\Pr(\text{outage}) = \frac{2^{-Q} e^{A/2}}{\gamma_{th}} \sum_{q=0}^{Q} \binom{Q}{q} \sum_{n=0}^{M+q} \frac{(-1)^n}{\beta_n} \cdot \mathcal{R}\left\{ \frac{\mathcal{M}_\gamma\left(-\frac{A+2\pi jn}{2\gamma_{th}}\right)}{\frac{A+2\pi jn}{2\gamma_{th}}} \right\} + E(A, M, Q) , \tag{2.18}$$

where the parameters A, M and Q are chosen to let the overall error term $E(A, M, Q)$

to be negligible compared to outage probability [27] and $\gamma_{th} = 2^{(\bar{J}+1)\mathcal{R}} - 1$.

By setting $\mathcal{M}_{\gamma}(s) = \Phi^{\text{AaF}}_{\gamma_{\text{ub}}}(s)$, the total system outage probability using N+NT-ORS scheme in AaF relay system is obtained.

Asymptotic Diversity Order

We adopt the definition of diversity order as described in [59]: $d = \lim_{\bar{\gamma}\to\infty} - \log(P_{\text{out}})/\log(\bar{\gamma})$.

Proposition 1. *The asymptotic diversity order of the AaF relay system using the N+NT-ORS scheme with outdated CSI and $0 \leq \rho < 1$ is N for high SNR.*

Proof: See Appendix D.

Remark 1. *We can see from the description of the proposed scheme in Sec. 2.4.1 that if $N = 1$, the N+NT-ORS transforms to the NT-ORS where only the value of μ controls the number of selected relays. In this case, the diversity order decreases to 1 as a special case of Proposition 1 with $N = 1$. Therefore, the value of N is crucial to improve the diversity order. It can also be seen that if $N > 1, \mu = 1$, the N+NT-ORS transforms to the GSC-based MRS [22]. Moreover, if $N = 1, \mu = 1$, the N+NT-ORS transforms to ORS and the diversity order of ORS in AaF with outdated CSI is 1 as a special case of Proposition 1 with $N = 1$.*

2.4.3 Decode-and-forward Relaying

For the DaF protocol, the decoding relays are those with the mutual information from the source in the first phase larger than a threshold $2^{(Z+1)\mathcal{R}} - 1$ where Z is the number of time slots used in the second phase. For the ease of analysis, we assume all selected relays retransmit via orthogonal frequency division multiplexing access (OFDMA) so that the entirety of the second phase transmission only occupies one time slot, i.e., $Z = 1$.

Define $\mathcal{DS}^L$ as the set of all decoding subsets with L relays, and $\mathcal{DS}_p^L$ as the p-th element of the set. The received sum SNR of DaF N+NT-ORS scheme with L relays is $\gamma = \sum_{l=1}^{L} U(\hat{\gamma}_l)$, where $U(\hat{\gamma}_l)$ is the indicator function written in a manner similar to (2.13), but $\hat{\gamma}_l$ is now replaced by $\hat{\gamma}_{l,D}$ the SNR of l-th $R \to D$ link, $l = 1, \ldots, K$, and $\hat{\gamma}^{(N)}$ is the N-th largest $\hat{\gamma}_l$ in the $\mathcal{DS}^L$.

The CDF of the SNR of N+NT-ORS using DaF protocol is computed as [48]

$$P_{out}(y) = \sum_{L=0}^{K} \sum_{p=1}^{|\mathcal{DS}^L|} \Pr(\text{outage}|\mathcal{DS}_p^L) \Pr(\mathcal{DS}_p^L) \tag{2.19}$$

$$= \sum_{L=0}^{K} \Pr(\text{outage}|\, |\mathcal{DS}| = L) \Pr(|\mathcal{DS}| = L)\,, \tag{2.20}$$

where $y = 2^{2R} - 1$, $\Pr(\text{outage}|\mathcal{DS}_p^L)$ is the probability of outage conditioned on the decoding subset $\mathcal{DS}_p^L$, $\Pr(\mathcal{DS}_p^L)$ is the probability of the occurance of that subset, and $\Pr(|\mathcal{DS}| = L)$ is the probability that the decoding subset has L relays. When $L = 0$, the system outage probability is the probability that the first hop is in outage. Given the assumption that all the $S \to R$ channels are i.i.d., $\Pr(\mathcal{DS}_p^L)$ are the same for any p with a given value of L and so is $\Pr(\text{outage}|\mathcal{DS}_p^L)$, which leads (2.19) to (2.20).

Under Rayleigh fading, one can obtain $\Pr(|\mathcal{DS}| = L)$ as follows

$$\Pr(|\mathcal{DS}| = L) = \binom{K}{L} \left(1 - e^{-\frac{y}{\bar{\gamma}}}\right)^{K-L} e^{-\frac{yL}{\bar{\gamma}}}\,. \tag{2.21}$$

Now we try to solve $\Pr(\text{outage}|\, |\mathcal{DS}| = L)$. If the size of the decoding subset, L, is smaller or equal to N, maximum ratio combining (MRC) is done for *all* the relays in the decoding subset and hence, $\Pr(\text{outage}|\, |\mathcal{DS}| = L \leq N)$ is given by [44]

$$\Pr(\text{outage}|\, |\mathcal{DS}| = L \leq N) = 1 - e^{-\gamma_{th}/\bar{\gamma}} \sum_{k=1}^{L} \frac{(\gamma_{th}/\bar{\gamma})^{k-1}}{(k-1)!}\,, \tag{2.22}$$

where $\gamma_{th} = y = 2^{2R} - 1$. If $L > N$, N+NT-ORS is applied. In the rest of this section, we derive the MGF for $\Pr(\text{outage}|\,|\mathcal{DS}| = L > N)$, then apply the relationship between MGF and outage probability to obtain the latter [27]. With $\Pr(\text{outage}|\,|\mathcal{DS}| = L)$ available, we can compute the total system outage probability.

The analysis of the MGF of γ for $\Pr(\text{outage}|\,|\mathcal{DS}| = L > N)$ follows the same procedure as the analysis in the AaF case as the random variables in the summations for both cases are all exponentially distributed. Therefore, the MGF of γ for $\Pr(\text{outage}|\,|\mathcal{DS}| = L > N)$ can be obtained by replacing K and $\bar{\gamma}/2$ in (2.14) by L and $\bar{\gamma}$.

We again utilize the numerical technique proposed in [27] to approximate the outage probability using MGF as

$$\Pr(\text{outage}|\,|\mathcal{DS}| = L > N) = \frac{2^{-Q}e^{A/2}}{\gamma_{th}} \sum_{q=0}^{Q} \binom{Q}{q} \sum_{n=0}^{M+q} \frac{(-1)^n}{\beta_n} \cdot \mathcal{R}\left\{ \frac{\mathcal{M}_\gamma\left(-\frac{A+2\pi jn}{2\gamma_{th}}\right)}{\frac{A+2\pi jn}{2\gamma_{th}}} \right\} + E(A, M, Q)\,, \tag{2.23}$$

where $\gamma_{th} = 2^{2R} - 1$.

By setting $\mathcal{M}_\gamma(s) = \Phi_{N\text{+NT-ORS}}(s)$, $\Pr(\text{outage}|\,|\mathcal{DS}| = L > N)$ is obtained. Therefore, the total system outage probability using N+NT-ORS scheme in DaF relaying can be computed by (2.20).

The average number of selected relays of N+NT-ORS in DaF is given in the following theorem.

Theorem 4. *The average number of the selected relays $\bar{J}$ of N+NT-ORS in DaF relay systems is*

$$\bar{J}^{DaF}_{N+NT\text{-}ORS} = \sum_{L=0}^{K} \bar{j}(L) \Pr(|DS| = L)\,, \tag{2.24}$$

where

$$\bar{j}(L) = \begin{cases} L\,, & L \leq N \\ L\sum_{m=0}^{L-N-1} \frac{(L-N-1)!(L-1)!(-1)^m}{(L-N-1-m)!(L-N-1)!(N-1)!m!} \cdot \frac{1}{N+m+\mu}\,, & L > N \end{cases} . \tag{2.25}$$

Proof. Let variable $\bar{j}(L)$ be the average number of selected relays when the size of the decoding subset is L. When $L \leq N$, all relays in the subset are selected, i.e., $\bar{j}(L) = L$. When $L > N$, we can view $\bar{j}(L)$ as the average number of selected relays in AaF relaying except the total number of relays is not K but L. Replacing K by L in (2.15), Theorem 4 holds. □

Asymptotic Diversity Analysis

Proposition 2. *The asymptotic diversity order of the DaF relay systems using the N+NT-ORS scheme with outdated CSI and* $0 \leq \rho < 1$ *is N for high SNR.*

Proof: See Appendix E.

Remark 2. *For DaF relaying, it can be easily seen that ORS is once again a special case of N+NT-ORS when* $N = 1$ *and* $\mu = 1$*. Similarly, Proposition 2 can be carried over to describe that the diversity order of ORS decreases to* 1 *with outdated CSI in DaF relay systems. Similar transformation to GSC-based DaF [22][7] can also be carried out by setting* $N > 1$ *and* $\mu = 1$ *in our N+NT-ORS. Same as the case in AaF relaying, we see that N determines the diversity order for N+NT-ORS with DaF relaying.*

[7]Note in [22], outdated CSI is not considered.

2.5 Implementation Issues

In [3–5], a method of distributed timers is proposed to carry out the ORS without requiring global CSI at each relay. For practical implementation, we list below the procedures for a possible distributed RS protocol of the N+NT-ORS:

1. Each relay k acquires its own CSI $\hat{h}_{S,k}$ and $\hat{h}_{k,D}$ by listening to the Request-to-Send (RTS) signal from S and the Clear-to-Send (CTS) signals from D. Notice that the CSI available is outdated. Relay k then calculates the outdated SNR $\hat{\gamma}_{S,k,D}$ or $\hat{\gamma}_{k,D}$, depending on AaF or DaF protocol the system is using. Upon receiving the RTS and the CTS, all the candidate relays[8] start a synchronized timer T_{sync} and the kth relay starts a timer T_k. For AaF, $T_k \propto 1/\hat{\gamma}_{S,k,D}$, $\forall k$. For DaF, $T_k \propto 1/\hat{\gamma}_{k,D}$, $\forall k$.

2. The timer of the "best" relay $1'$[9] expires first and a flag packet is transmitted to notify the rest of the network. Then relay $1'$ goes to the temporary sleep mode. The length of the flag packet D_{f_k} is proportional to the relay's timer T_k[10], but much shorter (e.g., $D_{f_k} = T_k/100$, where the denominator is a proportional parameter λ, which equals 100 in this example.). Notice that this proportional parameter λ is set for all the relays prior to RS.

3. Before relay k's timer T_k expires, relay k updates its number of received flag packets j_k when it hears a flag packet from another relay ($j_k = j_k + 1$). Once relay k's timer T_k expires, it computes the N-th best relay's CSI using the length of the N-th flag packet and the proportion parameter λ (i.e., $T_{N'} = D_{f_{N'}}/\lambda$). If $j_k < N$, it sends its flag packet and goes to the temporary sleep mode.

[8] Again, the candidate relays are all K relays for AaF relaying while for DaF relaying, they are the relays in the decoding subset.

[9] Let variable k' denotes the index for the k-th best relay with respect to the outdated SNR.

[10] We inherits the timer structure of the ORS scheme which uses the flag length to represent the channel rather than include the CSI in the flag packet.

1.	Set parameters N, μ, λ and T_{sync}. Set the number of the flag packets that a relay has received $j_k = 0$ for relay k, $\forall k$.
2.	Upon receiving the RTS and the CTS, relay k starts a timer T_{sync} and its own timer T_k, where $T_k \propto 1/\hat{\gamma}_{S,k,D}$, $\forall k$ for AaF and $T_k \propto 1/\hat{\gamma}_{k,D}$, $\forall k$ for DaF.
3.	Before T_k expires, $j_k = j_k + 1$ if relay k hears a flag packet from another relay.
4.	When T_k expires, relay k computes $T_{N'}$ from $D_{f_{N'}}$, then if $j_k < N$, relay k sends its flag packet of duration D_{f_k} and goes to the sleep mode, where $D_{f_k} = \lambda T_k$; elseif $T_k \leq T_{N'}/\mu$ relay k sends its flag packet and goes to the sleep mode; else relay k directly goes to the sleep mode.
5.	When T_{sync} expires, relays that have sent flag packets start to transmit.

Table 2.1: Distributed relay selection protocol for N+NT-ORS

Otherwise, relay k then compares the value of its own timer T_k to the value $T_{N'}/\mu$. If $T_k \leq T_{N'}/\mu$, it sends out its flag packet and goes to the temporary sleep mode. Otherwise, it directly goes to the temporary sleep mode.

4. When T_{sync} expires, all the "sleeping" relays which have sent out flag packets (i.e., selected relays) wake up and start to receive data for retransmission. For AaF relaying, the data is forwarded sequentially. The transmission sequence is known for each selected relay from the number of the flag packets it has received, i.e., if relay k has received j_k flag packets, it is the $(j_k + 1)$-th relay to transmit. For DaF relaying, all relays forward the data simultaneously in an OFDMA manner.

The protocol described above can be summarized in Table 2.1. The length of synchronized timer T_{sync} is relative to parameters N and μ for the N+NT-ORS. Intuitively, the larger the delay between the selection instant and data transmission instant, the more severe outdated CSI at the relay side. In other words, to exploit

the diversity of the N+NT-ORS, we may choose a larger N and a smaller μ leading to a large delay. In this case, the total number of selected relays J may be large, and accordingly a relatively longer timer T_{sync} is needed.

The conventional ORS greatly simplifies the system synchronization requirement. As a result of the fact that the proposed N+NT-ORS inherits the timer configuration of the conventional ORS, it enjoys not only being performed in a distributed fashion, but also the relaxed requirement on synchronization. Specifically, for the above described N+NT-ORS protocol to perform properly, two assumptions are necessary. First, at the relay selection stage, all the relays need to start the timer T_{sync} simultaneously, which is a common assumption for most single and multiple relay selections including conventional ORS. Secondly, at the data transmission stage, all relays maintain time slot configurations for the sequential transmission in AaF, which is similar to TDMA.

For the distributed protocol of the conventional ORS [5], the probability that any two relays have their timers expire simultaneously is analyzed in [3] and [4]. Same problem exists for the distributed protocol in this paper. There are several parameters affecting this probability: the propagation delays of $D \rightarrow R$ and $R \rightarrow R$ links, radio listen-to-transmit switch time and the duration of flag packet. Due to an architecture that is similar to the distributed timer structure of ORS, the analysis to the problems above is similar to that of ORS. Interested readers are advised to refer to [5]. However, the analysis of this probability is beyond the scope of this thesis and is left for future research.

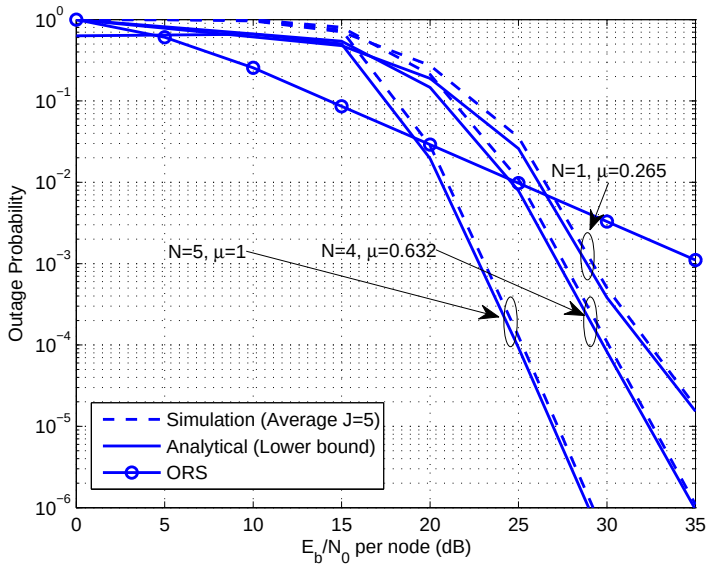
2.6 Numerical Results

We verify the closed-form expressions for outage probability by comparing them with simulation results. We consider a scenario with $K = 10$ relays, a target data rate $R = 1$ bit/s/Hz, and i.i.d. channels unless otherwise specified. We further assume that the distributed N+NT-ORS protocol outlined in Sec. 2.5 can be performed without error.

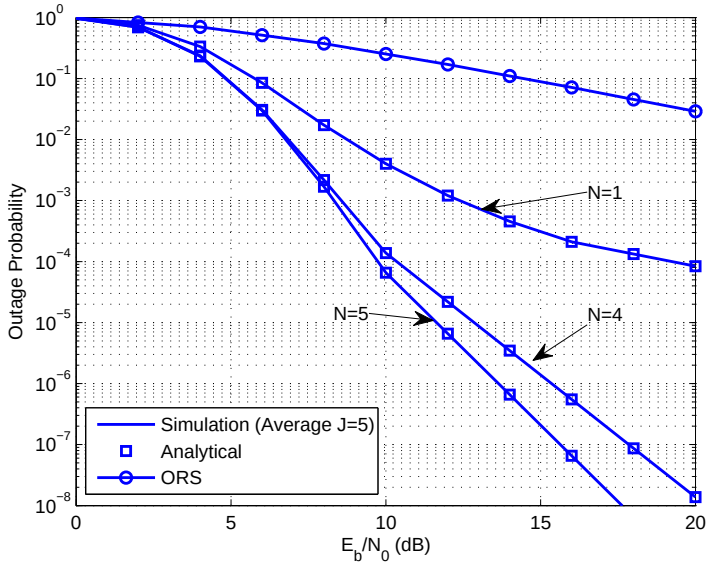
Figs. 2.2 and 2.2 plot the outage probability of N+NT-ORS in AaF and DaF relaying, respectively, for different combinations of N and μ. For fair comparison, they are all under the same total power constraint. In particular, the combinations of N and μ are selected such that the average number of selected relays $\bar{J}$ remains constant for both figures, i.e., expressions (2.15) and (2.24)[11]. For AaF, in the low SNR region, the outage probability of MRS does not decrease rapidly. This is because although the total SNR at the destination is higher using MRS, it is insufficient to compensate for the rate degradation introduced by using multiple orthogonal time slots; whereas, in the high SNR region, the effect of the total SNR increase becomes dominant and as a result diversity order of N can be visualized. Please note that the exact diversity order as stated in Proposition 1 will not occur for AaF until at a sufficiently high SNR, and we have only shown its result for up to 35 dB in Fig. 2.2. For DaF, due to the use of OFDMA, only two time slots are needed. Therefore, the N diversity order of the outage performance is clearly seen in Fig. 2.2.

We compare the outage probability of the proposed N+NT-ORS scheme with that of several other RS schemes, including: conventional ORS [4, 5], GSC-based MRS [22, 23], random MRS and APR [58], in Figs. 2.3 and 2.3 for AaF and DaF

[11] As discussed in Sec. 2.4, $\bar{J}$ is independent of both ρ and $\bar{\gamma}$ for AaF, so the value of μ is fixed for one combination over different SNRs. In the contrast, $\bar{J}$ is only independent of ρ for DaF. Therefore, in Fig. 2.2, the value of μ is not shown for any combination since it is varying over different SNRs to maintain a constant $\bar{J}$.



(a) AaF



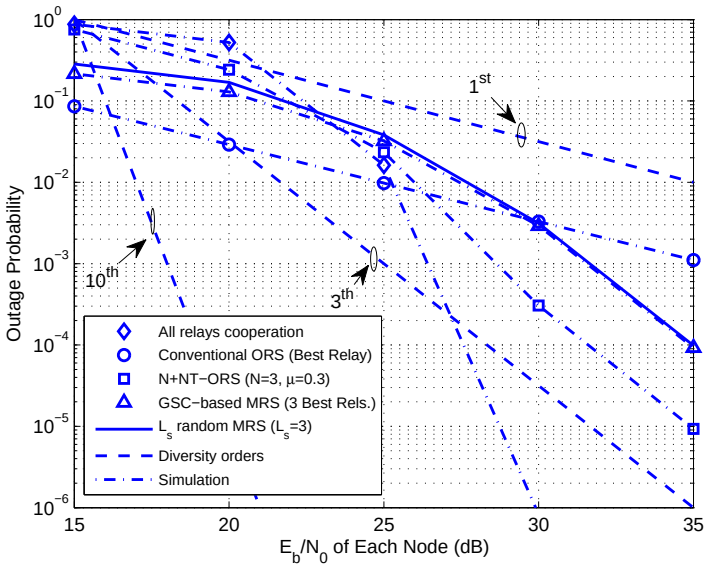
(b) DaF

Figure 2.2: Outage probability of N+NT-ORS with 5 average selected relays with $K = 10$ and $\rho = 0.1$ for (a) AaF and (b) DaF relaying protocols.

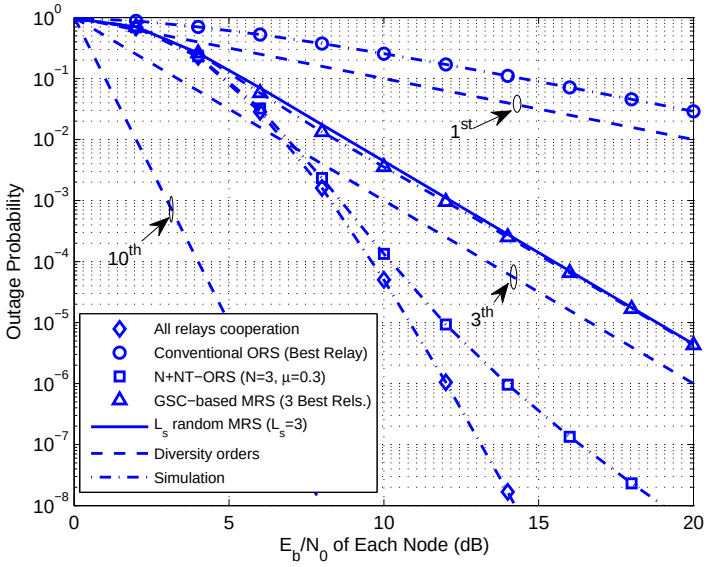
relaying protocols, respectively. The random MRS scheme is defined as one which randomly selects up to L_s relays for relaying. The APR is included as a benchmark since it results in the best performance but is difficult to implement in practice due to high total power consumption and the need of node coordination [4, 5, 58]. From the figures, several observations can be made both in the high SNR region of AaF and in the whole SNR range of DaF as follows: 1) As expected, the diversity order of the conventional ORS with outdated CSI becomes unity as investigated in [48]; 2) The outage of GSC-based MRS is close to random MRS because ρ is 0.1; 3) N+NT-ORS with $N = 3, \mu = 0.3$ performs better than GSC-based MRS with 3 best relays selected, since with this specific μ the average number of selected relays $\bar{J}$ for N+NT-ORS is larger than 3, indicating a trade-off between performance and number of cooperative relays under outdated CSI; 4) the outage probability decays at a rate $N = 3$, which matches the diversity order analysis of Secs. 2.4.2 and 2.4.3. Similar to the description of Fig. 2.2 above, the AaF RS scheme that uses the most relays would have the worst outage performance in the low SNR region since its data rate is severely impaired by the use of orthogonal time slots. This impairment is overcome at the high SNR region by the high total sum SNR.

The average number of selected relays is plotted in Fig. 2.4 versus the values of μ. As can be seen in the figure, while the value of N in N+NT-ORS decides the diversity order, the value of μ controls the number of relay being selected. Together, these two parameters results in trade-offs among performance and total power consumption. Moreover, the average number of selected relays is independent of the average SNR E_b/N_0 for AaF but depends on it for DaF.

In a dynamic environment where node mobility may vary, it is important to study its impact on the performance of different RS schemes. In the next two subsections, we will examine the effect of both a varying network size as well as mobile nodes in

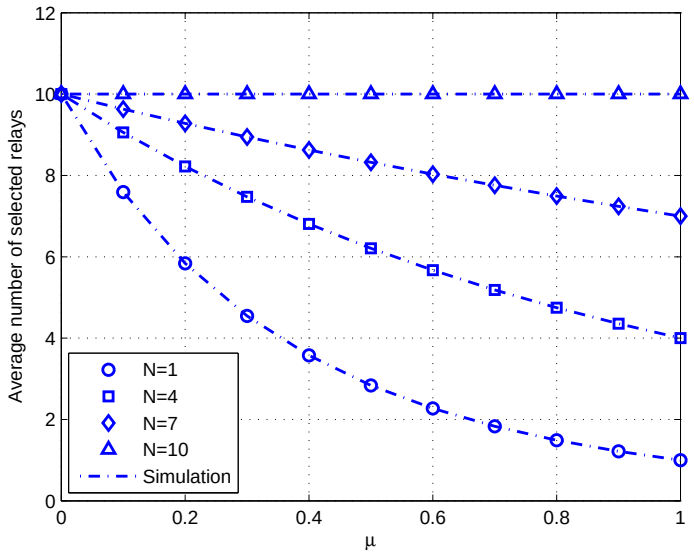


(a) AaF

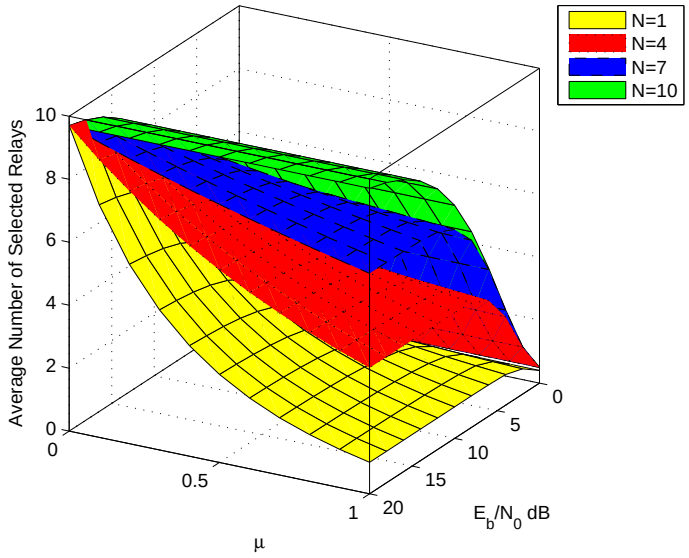


(b) DaF

Figure 2.3: Comparison of the outage probability of different RS schemes with $K = 10$, $\rho = 0.1$ for (a) AaF and (b) DaF relaying protocols.



(a) AaF



(b) DaF

Figure 2.4: Average number of selected relays as a function of μ.

a dynamic network. A varying network size may arise when the S and D nodes are deployed under different applications at different time instances; whereas, nodes may become mobile w.r.t. fixed position relays resulting in network mobility.

2.6.1 Dynamic Network Size

To first model a network with varying size, we assume that the total number of relays follows the discrete uniform distribution in the interval $[\bar{K} - n/2, \bar{K} - n/2 + 1, \ldots, \bar{K} + n/2]$, with mean value $\bar{K} = 20$ and the interval of total number of relays $n = \{0, 4, 8, 12, 16, 20\}$. Since the variance of the discrete uniform distribution is $(n^2 - 1)/12$, large n indicates the high variability of the network size. Moreover, we consider both i.i.d. and independent but non-identically distributed (i.n.d.) channels, where for i.i.d. channels the received average SNRs are 10 dB for all relays, and for i.n.d. channels the maximum average SNR is 10 dB and the average SNRs received from two adjacent relays decreases by 1 dB. The rationale behind this consideration is due to the fact that a dynamic network with varying size usually leads to different nodes with different channel statistics, i.e., different average SNRs.

A comparison between the outage probability of N+NT-ORS and other DaF MRS schemes with both i.i.d. and i.n.d. channels as a function of n in the above network model is shown in Fig. 2.5 with $\rho = 0.1$. Three observations are worth noting: 1) N+NT-ORS shows the best robustness for both i.i.d. and i.n.d. channels; 2) MRS schemes exhibit more variation on outage performance under i.n.d. channel conditions; 3) APR performs the best for all values of n since it employs all the relays. However, the use of APR in practice is undesirable due to its high power consumption and the needs of complicated node coordinations [4, 5, 58]. The simulation demonstrates that our N+NT-ORS scheme utilizes a proper number of relays to cooperate without the need of adjusting N and μ. If GSC-based MRS needs to maintain a

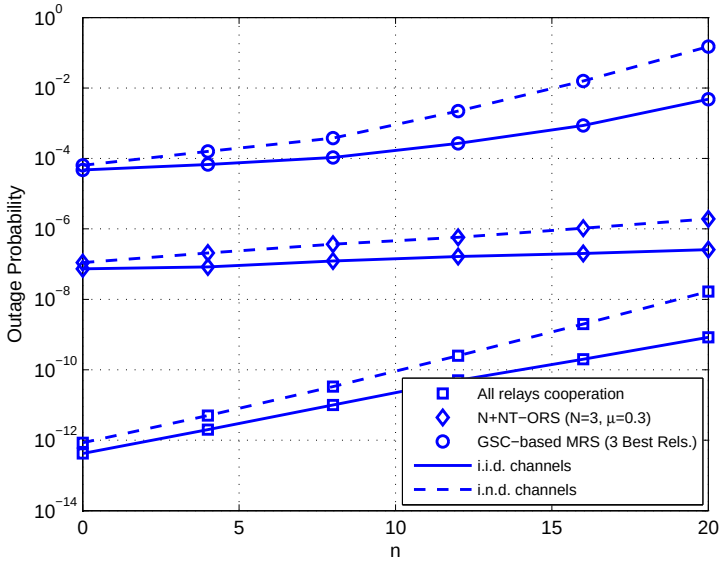


Figure 2.5: Comparison of the outage probability of different DaF MRS schemes in a dynamic network with varying network size for both i.i.d. and i.n.d. channels, $\bar{K} = 20$, and $\rho = 0.1$.

constant outage performance, it would have to frequently decide on the appropriate number of relays to cooperate due to the frequently changing network size. For example, as shown in Fig. 2.5, when n is high, the outage performance of GSC-based MRS deteriorates more severely than N+NT-ORS, which necessitates an adjustment to N. Similar behavior can be seen for AaF relaying, but is not shown due to space limitation.

2.6.2 Dynamic Network Mobility

We compare the proposed scheme with other RS schemes in a network with high mobility in which the correlation coefficient ρ between the instantaneous CSI and the outdated CSI varies between 0 and 1. Fig. 2.6 depicts the outage probability of N+NT-ORS, conventional ORS [4, 5], GSC-based MRS [22, 23], random MRS and APR [58] versus ρ in DaF relaying. The influence on the outage performance of con-

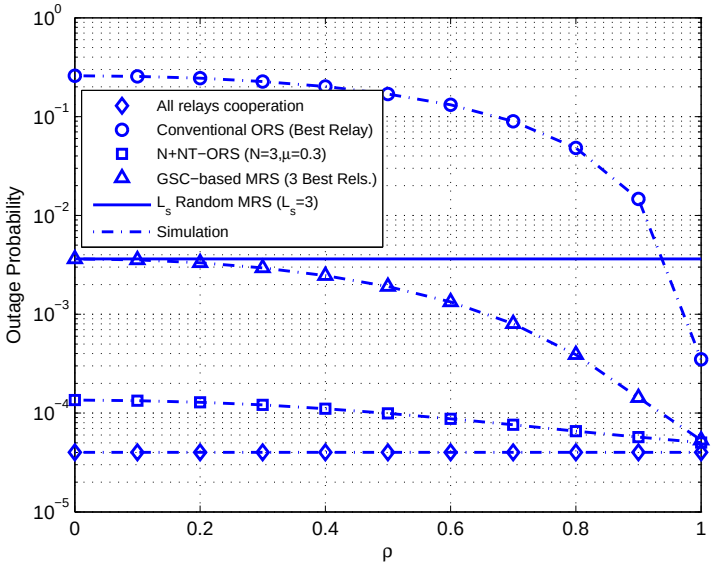


Figure 2.6: Outage probability as a function of ρ for several DaF RS schemes with $K = 10$ and $E_b/N_0 = 10$ dB.

ventional ORS is most evident as ρ decreases. Once again, N+NT-ORS outperforms GSC-based MRS in all ranges of ρ. Although the outage behavior of both random MRS and APR show no dependence to the value of ρ, these RS schemes are in fact inferior. With $N = 3$ and $\mu = 0.3$, the average number of selected relays is about 5. We see that with much less than 10 relays selected, we achieves significant performance similar to APR. When comparing with N+NT-ORS, random MRS is more ineffective at dealing with outdated CSI, e.g., Fig. 2.6, and APR necessitates higher total power because all relays are involved in transmission as well as complex node coordination [4,5,58]. Similar behavior can be seen in AaF relaying, but is not shown due to space limitation.

2.7 Conclusions

We now summarize the contribution of the work in this chapter as follows:

- We prove that the *diversity order* of all the existing SRS schemes, i.e., schemes from [4, 5, 40, 45, 49, 57, 58], which can achieve full diversity using perfect CSI decreases to unity in the presence of outdated CSI. Henceforth, we demonstrated that it is not possible to combat outdated CSI by SRS schemes.

- Based on the GSC-based MRS schemes, we propose a novel MRS scheme, the N plus normalized threshold selection based on ORS (N+NT-ORS) in both AaF and DaF one way relay networks to offer reliable outage performance under outdated CSI by determining the number of relays to cooperate. From the extensive simulation results, we show that N+NT-ORS outperforms the existing MRS scheme, especially in a wireless environment with high mobility in terms of both node mobility and network size.

- For the N+NT-ORS, we derive the closed form expressions for outage probability, from which we show that the asymptotic diversity order equals to N at high SNR[12].

- Based on the distributed timers in [3–5], we outline a distributed RS protocol for the N+NT-ORS which does not require the collection and use of global CSI at each relay. Unlike other authors in the description of their MRS schemes, we present the details about how the distributed timer protocol can be realized in our scheme.

[12] As a byproduct, the impact of outdated CSI on GSC-based MRS is also examined since it can be casted as a special case of our proposed MRS scheme.

Chapter 3

Achieving Linear Throughput Scaling in MUPC with Controllable Delay-Throughput Tradeoffs

3.1 Background

As we introduced in Chapter 1, for MUPC, one of the recent notable work, THOR, achieves enlightening throughput results but suffers from delay related problems. In this chapter, we continue on developing decentralized schemes for large two-hop DaF wireless networks, with an emphasis on achieving *linear* throughput scaling with *controllable* delay-throughput tradeoff performance. An in-depth delay analysis of Cui's THOR scheme [9] reveals a number of practical issues, which turns out as direct consequences of the *independent* opportunistic scheduling over two hops. Consider first the scenario when relay r connects to source i in the first hop, but is then

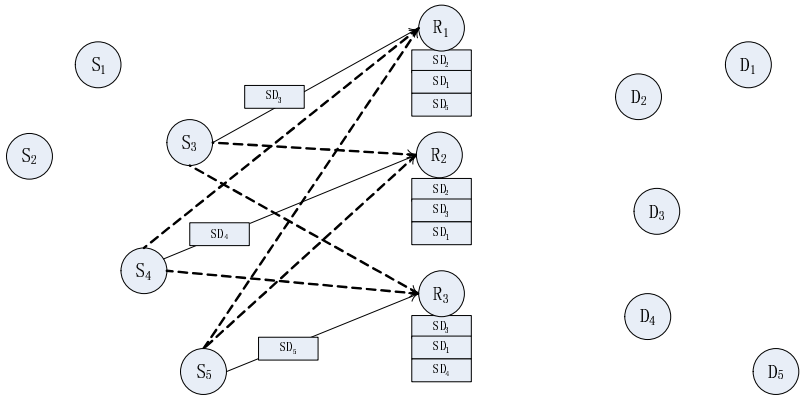
scheduled by destination j at the second hop. This mismatch between the source node and the destination node occurs frequently, since usually the packets have to be buffered at the relays until the forwarding channel become strong relative to other destinations. Therefore, the relays need to have buffers, the sizes of which can grow large as the number of S-D pairs increases. Furthermore, since the packets from one particular source can travel through different routes by way of different relays, they are likely to be disordered when arriving at the destinations. In the worse case that a preamble packet arrives last in order among all packets of a whole frame, the destination node is kept waiting for an entire communication session. The decoding delay of recovering the whole frame is likely intolerable for delay-sensitive applications (e.g., wireless data streaming services). Finally, in order for THOR to be practically functional, each relay needs to have information packets to be delivered to all the destinations buffered in their respective queues. This requirement translates into an additional *start-up time period* for the whole network to "warm up" and reach steady states. In light of the above, we pose the following questions. Is it possible to achieve linear throughput scaling in the number of relays k without requiring buffers at relays, so that no packet disorder at the destination occurs and no start-up time period are required? If such a scheme exists, how fast can k grow while still retaining throughput linearity? What is the throughput-delay tradeoff performance of such a scheme? Are we able to achieve the tradeoff curve boundaries through simple parametrization design? These questions are addressed in this chapter.

The remainder of the chapter is organized as follows. In Section 3.2, we describe the basic system model and propose an opportunistic pair scheduling scheme (OPS), followed by a complete throughput analysis in Section 3.3, where closed-form throughput lower bounds and scaling laws are derived under Nakagami fading environment. Next, in Section 3.4, we propose a general opportunistic scheme that incorporates

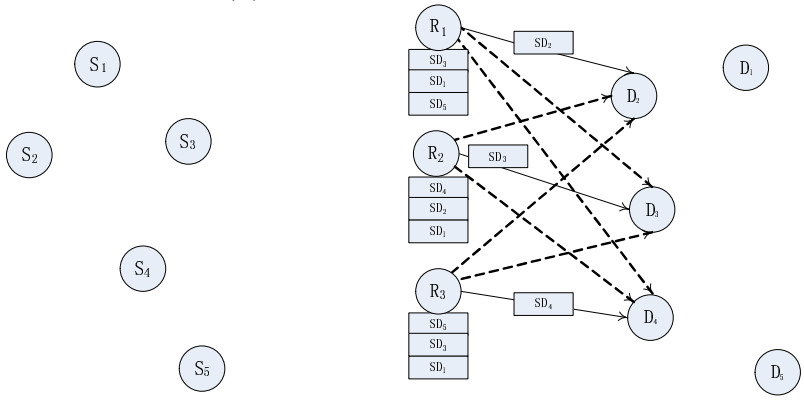
OPS and THOR as special cases. The delay performance is investigated in Section 3.5. Section 3.6 presents some simulation results. Finally, concluding remarks are drawn in Section 3.7. The detailed proofs and mathematical derivations of the lemmas and theorems can be found in the appendices.

3.2 System Model and Pair Scheduling

Consider a wireless network where n S-D pairs and k relays are densely populated in a fixed area and operate in the same frequency band, as depicted in Fig. 3.1. The communication between each S-D pair is impaired by fading and is possible only through intermediate relaying nodes. The relays are assumed to work in half-duplex mode and employ a two-hop decode-and-forward communication protocol [28]. In the first hop of the communication, a subset of sources are selected and scheduled by the relays for concurrent transmissions. Due to the broadcasting nature of the wireless medium, each relay receives the desired signal from a pre-notified source node, superimposed with the interfering signals that were intended for other relays. It then attempts to decode the desired signal by treating all the other signals as additive noise. The difference with Cui's THOR model [9, 10] lies in the second hop. Depending on hardware constraints and/or application requirements, each relay either forwards the decoded packet *instantly* to the corresponding destination node, which eliminates the buffer requirement, or *stores* the decoded packet in a queue and forward the packet pertain to the *best* destination node as indicated by a feedback message. In the latter case, the set of destinations scheduled for packet reception does not necessarily match the set of destinations associated with the source set in the first hop, which entails the necessity of packet storage and communication delays. In this section, we restrict our attention to the former case by letting relays serve



(a) First hop transmission



(b) Second hop transmission

Figure 3.1: A two-hop network with $n = 5$ S-D pairs and $k = 3$ relays. Circles denote source, relay and destination nodes. Rectangles denote data for specified S-D pairs and the rectangles at each relay indicates that these packets are buffered at the relay. For example, rectangle S-D_2 denotes the information transmitted by S_2 to destination D_2. (a) In the first hop, source nodes $3, 4, 5$ transmit to the relays. (b) In the second hop, the relays transmit to the destination nodes $2, 3, 4$. Each relay maintains a buffer to store the data packets from the source nodes. Solid lines indicate desired signal transmissions and dashed lines indicate interference signals.

only matched S-D pairs. In other words, once relay r serves source i in the first hop, it will continue to forward packets originated from source i to destination i in the second hop. This restriction will be relaxed in Section 3.4.

Throughout the chapter, the general channel model of Nakagami fading is considered, which represents a wide range of realistic multi-path channels [36]. The Nakagami fading parameter m signifies the fading severity and smaller values of m represent more fading in the channel. It includes Rayleigh fading ($m = 1$) as a special case. Although Nakagami and Rician distributions are not identical, one can approximate Rician distribution with Nakagami distribution through a one-to-one parameter mapping, which is generally known to be accurate around their means. We assume that the channel coefficients from source i to relay r, denoted by $h_{i,r}$ ($\forall i = 1, ..., n$ and $r = 1, ..., k$) are independent and identically distributed (i.i.d.) and follow Nakagami fading model with parameter m. Accordingly, the channel power gains are known to follow Gamma distributions, i.e., $\alpha_{i,r} = |h_{i,r}|^2 \sim \Gamma(m, 1/m)$. Likewise, the channel power gains from relay r to destination j, denoted by $\beta_{r,j}$ ($\forall r = 1, ..., k$ and $j = 1, ..., n$), are i.i.d. and $\beta_{r,j} \sim \Gamma(m, 1/m)$. The channel gains $\alpha_{i,r}$ and $\beta_{r,j}$ are assumed to be quasi-static, which remain unchanged during a complete two-hop transmission session but vary independently in different sessions. We assume that, at each hop, the receivers have full knowledge on the backward channel but not on the forward channel, i.e., the CSIR assumption, while the transmitters have access to only an integer value, fed back from the receivers to indicate a node chosen for transmission. The assumption on the channel knowledge entails acceptable signaling overhead and is deemed reasonable for decentralized implementation.

We are ready to describe the opportunistic pair scheduling scheme (OPS). Let us begin with the second hop. By the CSIR assumption, the relays have no knowledge on their forward channels to the destinations, and thus the destinations are responsible

for making scheduling decisions in the second hop. For this purpose, we let all relay nodes transmit training sequences simultaneously with fixed power $P_R = P$. Assuming that the additive noise at all the receivers are i.i.d. and follow the complex normal distribution $\mathcal{CN}(0,1)$, destination j can compute k signal-to-interference-plus-noise ratio (SINR) by assuming that relay r is the desired sender and the messages from the other relays are interference as

$$\mathsf{SINR}_{r,j}^{\mathsf{H2}} = \frac{\beta_{r,j}}{\frac{1}{P} + \sum_{1\le t\le k, t\neq r} \beta_{t,j}}, \qquad r = 1, \ldots, k. \tag{3.1}$$

It is not difficult to show that there is at most one "good" relay r for any destination j such that $\mathsf{SINR}_{r,j}^{\mathsf{H2}} \geq 1$ for all relays $1 \leq r \leq k$ [9, Lemma 2]. Therefore, if destination j identifies one such good relay r, it will feed back the corresponding relay index to notify relay r being chosen for transmission. Otherwise, destination node j does not feed back any information. For each relay r, denote by $\mathcal{D}_r$ the set of destination nodes that have chosen relay r for transmission in the second hop. Depending on specific channel conditions, the candidate set $\mathcal{D}_r$ can be an empty set (receives no feedback), a singleton (one feedback), or a set with multiple elements (multiple feedbacks). Accordingly, relay r will remain silent in case of no feedback, schedule the corresponding S-D pair in case of one feedback, and schedule the S-D pair with the *strongest* source-to-relay (S-R) channel in case of multiple feedbacks, i.e., $i_r = j_r = \arg\max_{i\in\mathcal{D}_r} \alpha_{i,r}$, where i_r and j_r denote the source node and destination node scheduled by relay r respectively. In the latter two cases, relay r needs to feedback the source index i_r to initiate a two-hop transmission session at the beginning of the transmission block, which incurs a single integer overhead per relay node. The scheduled source nodes, constituting a set $\mathcal{S}_r \subset \{1, \ldots, n\}$, then transmit simultaneously with constant power P and fixed data rate of 1 bit/s/Hz[1]. Relay r can correctly decode source i_r's message

[1]Generalization to communication rate other than 1 bit/s/Hz is straightforward.

if the corresponding SINR is greater or equal to 1, i.e.,

$$\mathsf{SINR}_{i_r,r}^{\mathsf{H1}} = \frac{\alpha_{i_r,r}}{\frac{1}{P} + \sum_{t\in\mathcal{S}_r, t\neq i_r} \alpha_{t,r}} \geq 1, \qquad r = 1, \ldots, k\,. \tag{3.2}$$

The pair scheduling schedules only matched S-D pairs in two hops. It improves the THOR scheme in three aspects: 1) no buffers are required at the relays; 2) communication delay between each S-D pair is maintained at the minimum level; and 3) no packet arrives in disorder at the destination. Nevertheless the improvement on delay comes at the price of reduced multiuser diversity (in the first hop) and thus lowered end-to-end throughput. To be specific, while the relays are matched to the destinations within the subset $\mathcal{D}_r$ in the second hop, respectively, the scheduled source nodes are not necessarily among the strongest n source-to-relay connections. While it seems clear that the loss of multiuser diversity will affect the throughput of each connection negatively and results in lower system throughput, this observation is not straightforward when the interference coupling among the transceiver pairs are taken into consideration. For example, even though relay r is not connected to the best possible source node, the aggregated interference power from all the other source nodes also differs from the case when the sources and relays are matched with the best connections. Instead of digging into the complicated shifted interference patterns, we pose a fundamentally important question: Can OPS maintain a linear throughput scaling in the number of relays k? If yes, how fast can k grow while still retaining throughput linearity? What is the impact of Nakagami fading on the throughput scaling laws? These questions are answered in the following section.

3.3 Throughput Scaling: Pair Scheduling

In this section, we study the throughput scaling laws for the pair scheduling scheme in the regime of large number of S-D pairs, n, but for small and finite number of relays, k. This consideration is deemed in line with the current wireless communication landscape, which portrays an tremendous increase in the number of wireless terminals of various types, yet only a small increase in the number of wireless infrastructure (relay) supports. Emphasis is placed on showing the achievability of linear throughput scaling and the corresponding affordable scaling speed of the number of relays. Specifically, we show the throughput of both hops achieve an average throughput of k bit/s/Hz, yielding a $k/2$ bit/s/Hz end-to-end throughput for the complete two-hop transmission. Compared with the information-theoretic upper bound of $k/2 \cdot \log\log n$ bit/s/Hz, achieved with full relay cooperation and global CSI, the pair scheduling maintains the linearity of the throughput in the number of relay nodes, along with minimum-level signaling overhead and communication delay. We start by analyzing the throughput achieved on the second hop, then, building upon these results, we derive the throughput for the first hop.

3.3.1 The Second Hop: Relays to Destinations

The reason we treat the throughput analysis of the second hop before that of the first hop is in accordance with the pair scheduling scheme: The set of source nodes scheduled is determined by the way the relays matched to the destination nodes in the second hop. In this section, we shall develop an exact expression for the sum-rate of all the relay-to-destination links. When the connection between any relay-destination link is impaired by Nakagami fading, we have

Lemma 1. *For any k, n, $m \geq 0.5$, and P, the throughput of OPS in the second hop*

is given by

$$R_2 = k\Big(1 - \big(F(1)\big)^n\Big), \tag{3.3}$$

where $F(x)$ is given by

$$F(x) = \int_0^x \frac{U\big(-m, 1-km\,; m(1+\tau)/P\big)\tau^{m-1}e^{-m\tau/P}}{\Gamma(m)(1+\tau)^{km}}\,d\tau, \tag{3.4}$$

and $U(a,b\,;x)$ is the Tricomi confluent hypergeometric function [18, 9.210.2], the integral representation of which is given by $U(a,b\,;x) = 1/\Gamma(a)\int_0^\infty e^{-xt}t^{a-1}(1+t)^{b-a-1}dt$ [18, 9.211.4].

Proof: The second hop throughput extends Cui's result on Rayleigh fading to the general case of Nakagami fading. It can be verified that (3.3) reduced to [9, eqn. (8)] under Rayleigh fading. Based on the observation proved in [9, Lemma 2] that there exists at most one *good* relay for any destination node (those that can support the required SINR larger than one). Calculating the system sum rate amounts to summing up the total number of relays that finds at least one good destination to support the required 1-bit/s/Hz rate.

According to OPS, all the relay nodes transmit training sequences simultaneously with fixed power $P_R = P$. Given the i.i.d. channel model and Nakagami fading, the SINR measured at the destination is given by the general form (cf. (3.1))

$$\mathsf{SINR}_{r,j}^{\mathsf{H2}} = \frac{\Gamma(m, 1/m)}{1/P + \Gamma((k-1)m, 1/m)}. \tag{3.5}$$

Following similar steps in the proof of [43, Lemma 2], the probability distribution

function (pdf) of $\mathsf{SINR}_{r,j}^{\mathsf{H2}}$ can be obtained as

$$f(x) = \frac{U\big(-m, 1-km\,; m(1+x)/P\big) x^{m-1} e^{-mx/P}}{\Gamma(m)(1+x)^{km}}. \tag{3.6}$$

where $U(a, b\,; x)$ is the Tricomi confluent hypergeometric function. Accordingly, the CDF of $\mathsf{SINR}_{r,j}^{\mathsf{H2}}$, $F(x)$, is given by

$$F(x) = \int_0^x \frac{U\big(-m, 1-km\,; m(1+\tau)/P\big) \tau^{m-1} e^{-m\tau/P}}{\Gamma(m)(1+\tau)^{km}}\, d\tau. \tag{3.7}$$

Next, since each destination can have at most one good relay as its desired sender [9, Lemma 2], the system sum rate equals the number of relays that are scheduled by destinations. Mathematically, the throughput can be expressed as

$$\begin{aligned} R_2 &= \sum_{r=1}^{k} \Pr\left[\text{relay } r \text{ receives at least one feedback}\right] \cdot 1 \\ &= \sum_{r=1}^{k} \Big(1 - \Pr\left[\text{relay } r \text{ receives no feedback}\right]\Big) \\ &= \sum_{r=1}^{k} \left(1 - \Pr\left[\mathsf{SINR}_{r,j}^{\mathsf{H2}} \le 1, \forall j\right]\right) = k\Big(1 - \big(F(1)\big)^n\Big). \end{aligned} \tag{3.8}$$

■

Since $0 < F(1) < 1$ for fixed k, it follows that $R_2 \to k$ as $n \to \infty$, which establishes linear throughput scaling for the second hop.

3.3.2 The First Hop: Sources to Relays

We now study the sum-rate of the source-relay links. After the second hop feedback phase, each relay r maintains a candidate destination set $\mathcal{D}_r$ consisting of all the *qualified* destination indices that can support the required rate of 1 bit/s/Hz. Denote by $X_1, X_2, ..., X_{|\mathcal{D}_r|}$ the corresponding $|\mathcal{D}_r|$ S-R channel gains. Then, according to

the OPS scheme, each relay r simply serves the S-D pair that has the strongest S-R channel gain, i.e., $r_i = \arg\max_{i\in\mathcal{D}_r} X_i$. Analyzing the throughput meets two obstacles. The first obstacle lies in the fact that the total number of selected sources $|\mathcal{S}_r|$ can be smaller than the number of relays k, which occurs when multiple relay select the same source node. To overcome this difficult, we adopt a similar technique as in [9] by resorting to a lower bound on the sum rate, which considers only the case when k distinct source nodes are scheduled. The bounding technique is verified to be applicable to the general case of Nakagami fading as well. The second obstacle lies in the randomness of the candidate destination sets $\mathcal{D}_r$, the cardinality of which is a random variable (RV) varying between 0 and n. Characterizing the exact average throughput requires going through each possible case of $\mathcal{D}_r$, which is mathematically involved. Nevertheless, in the limiting region of $n \to \infty$, we can approximate $|\mathcal{D}_r|$ by its mean value. The techniques employed above are the key in demonstrating the linear throughput scaling of the first hop.

To begin with, we denote the event of having k distinct selected source nodes by N_k. The throughput of the first hop, R_1, is lower bounded by

$$R_1 \geq \Pr[N_k] \sum_{r=1}^{k} \Pr[\mathsf{SINR}_{i_r,r}^{\mathsf{H1}} \geq 1] \cdot 1 = k \Pr[N_k] \Pr[\mathsf{SINR}^{\mathsf{H1}} \geq 1], \tag{3.9}$$

where the equality follows from the i.i.d. channel model and $\mathsf{SINR}^{\mathsf{H1}}$ denotes the SINRs at all relay nodes in the first hop. By symmetry, one can readily obtain that $\Pr[N_k] = n(n-1)\cdots(n-k+1)/n^k$. Then, we focus on the $\Pr[\mathsf{SINR}^{\mathsf{H1}} \geq 1]$ term. According to the scheduling scheme and SINR expression (3.2), $\mathsf{SINR}^{\mathsf{H1}}$ can be expressed in the form of $\frac{X^{(0)}}{1/P+Y}$ (We add a superscript to X to differentiate from several similar notations appeared later), where $X^{(0)} := \max\{X_1, ..., X_{|\mathcal{D}_r|}\}$ and $Y := \sum_{i\in\mathcal{K}'} X_i$, where $\mathcal{K}'$ is any randomly selected $(k-1)$ element subset of $\{1, \ldots, |\mathcal{D}_r|\} \setminus$

$\{j : X_j = X^{(0)}\}$. To calculate $\Pr\left[\frac{X^{(0)}}{1/P+Y} \geq 1\right]$, we need to characterize the statistic behavior of $X^{(0)}$ and Y.

Let us study first the aggregated interference term Y. Conditioned on the event that none of the $k-1$ element is the maximum among $|\mathcal{D}_r|$ channel strengths, each interference term in Y is no longer a Gamma RV and they are not independent of each other in general. Nevertheless, it can be shown as in [9, Appendix A] that the error of approximating Y by the sum of $k-1$ independent Gamma RVs diminishes to zero as $n \to \infty$. Aided by this approximation, Y can be treated as a Gamma RV with scale $1/m$ and shape $(k-1)m$, the PDF and CDF of which are given by, respectively,

$$f_Y(y) = m^{(k-1)m} e^{-my} \frac{y^{(k-1)m-1}}{\Gamma((k-1)m)}, \tag{3.10}$$

$$F_Y(y) = 1 - \frac{\gamma((k-1)m, my)}{\Gamma((k-1)m)}, \tag{3.11}$$

where $\gamma(s,x) = \int_0^x t^{s-1}e^{-t}dt$ is the lower incomplete Gamma function [18, 8.350.1] and $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$ is the Gamma function [18, 8.310.1] . For the case of Nakagami fading, the CDF of $X^{(0)}$ (the maximum among $\bar{n}$ Gamma RVs) is

$$F_{X^{(0)}}(x) = \left(1 - m^m e^{-mx} \frac{x^{m-1}}{\Gamma(m)}\right)^{\bar{n}}. \tag{3.12}$$

Please note that the CDF expression (3.12) is still a RV due to the randomness in $\bar{n}$, which varies from 1 to n. It is mathematically involved to obtain an exact expression of $F_{X^{(0)}}(x)$ by averaging out all the possible values of $|\mathcal{D}_r|$. Nevertheless, in the limiting region of $n \to \infty$, the value of $\bar{n}$ can be approximated by its mean

$$\bar{n} = n \Pr\left[\mathsf{SINR}_{r,j}^{\mathsf{H2}} \geq 1\right] = n\overline{F}(1). \tag{3.13}$$

Here the notation $\overline{F}(x)$ means the complementary cumulative distribution function

of the corresponding RV. This completes the study of the statistics of both $X^{(0)}$ and Y. To compute $\Pr\left[\mathsf{SINR}^{\mathsf{H1}} \geq 1\right]$, we still need the following lemma. The proof is relegated to Appendix F.

Lemma 2. *Suppose $X^{(0)}$ is the maximum among $\bar{n}$ Gamma distributed RVs. For $s_0 = \frac{1}{m}\log\bar{n} + \frac{m-2}{m}\log\log\bar{n}$, we have $\Pr\left[X^{(0)} \leq s_0\right] \to 0$ as $\bar{n} \to \infty$.*

With Lemma 2, it is not difficult to obtain a similar probability lower bound as [9, Eqns. (4),(5)] by introducing a real auxiliary variable s $(0 < s \leq s_0)$

$$\Pr\left[\frac{X^{(0)}}{1/P + Y} \geq 1\right] \geq \overline{F}_{X^{(0)}}(s) F_Y\left(s - \frac{1}{P}\right). \tag{3.14}$$

Plugging (3.11) and (3.12) into (3.14), we have

Lemma 3. *For any P, k, $m \geq 0.5$, and $0 < s \leq s_0$, the throughput of OPS in the first hop is lower-bounded by*

$$R_1 \geq k\frac{n(n-1)\cdots(n-k+1)}{n^k}\left[1 - \left(1 - m^m e^{-ms}\frac{s^{m-1}}{\Gamma(m)}\right)^{n\overline{F}(1)}\right] \times \left(1 - \frac{\gamma((k-1)m, m\,(s-1/P))}{\Gamma((k-1)m)}\right) \text{ as } n \to \infty, \tag{3.15}$$

where $F(x)$ is given by (3.4).

Lemma 3 implies that with fixed k, all the terms $\frac{n(n-1)\cdots(n-k+1)}{n^k}$, $\left[1 - \left(1 - m^m e^{-ms}\frac{s^{m-1}}{\Gamma(m)}\right)^{\bar{n}}\right]$ and $\left(1 - \frac{\gamma((k-1)m, m(s-1/P))}{\Gamma((k-1)m)}\right)$ approach 1 as $n \to \infty$. Thus, the throughput in the first hop scales linearly in the number of relays k. With the results in both Lemma 1 and Lemma 3, the overall two-hop system throughput can be readily shown as follows.

Theorem 5. *OPS achieves a system throughput of $k/2$ bits/s/Hz under nakagami fading environment with fixed k and $n \to \infty$.*

3.3.3 How Fast Can k Grow?

The growth of the number of relays k has two conflicting effects: increased multiplexing gain (more S-D pairs can transmit simultaneously) and exacerbated multiuser interference. Answering the question of how fast can k grow amounts to finding the optimal k^* that maximizes the system throughput. Cui *et al.* answered this question in [9, Section IV] for the THOR scheme under Rayleigh fading and showed that k cannot grow faster than $\Theta(\log n)$ in order to maintain a linearly scalable network. In this section, we characterize the affordable growing speed of k for OPS under Nakagami fading. The main results are summarized as follows. The detailed proof is deferred to Appendix G.

Theorem 6. *For Rayleigh fading, OPS achieves a maximum throughput of* $\frac{\log 2n}{4(1+\log 2)}$ *and a throughput scaling of* $\Theta(\log n)$. *For Nakagami fading with parameter* $m > 1$ *taking integer values, OPS achieves a maximum throughput of* $-\frac{1}{4m}\mathcal{W}_{-1}\left(\frac{-c}{n}\right)$ *and a throughput scaling of* $\Theta\left(-\mathcal{W}_{-1}(-n^{-\frac{1}{m-1}})\right)$ *where* c *is a constant for fixed transmission power* P *and Nakagami fading parameter* m*, given by*

$$c = \frac{2^m}{m(1-m+m\log 2)} \exp\left\{\frac{-2^m e^{m/P} + m^2(1+2P+m(1+P)(\log 2-1))}{mP(1-m+m\log 2)}\right\}. \tag{3.16}$$

□

The following result is obtained as a byproduct of Theorem 6.

Theorem 7. *THOR achieves a maximal throughput of* $\frac{1}{4m}\log n$ *and a throughput scaling of* $\Theta(\log n)$ *under Nakagami fading with fading parameter* m. □

The proof of the above theorem can be verified by the following fact. THOR exploits multi-user diversity among all n users, while OPS exploits multi-user diversity

among all $\bar{n}$ users. Therefore, substituting $\bar{n}$ by n in the proof of Theorem 6 leads to Theorem 7.

Remark 3. *Theorem 6 and Theorem 7 establish the throughput scaling behaviors of OPS and THOR under Nakagami fading channels, respectively. Theorem 7 indicates that Nakagami fading has no effect on the throughput scaling laws for THOR. In other words, THOR maintains the same scaling speed of* $\Theta(\log n)$ *regardless of the specific value of the Nakagami fading parameter* m*. The maximal achievable throughput is, however, penalized by a multiplicative term of* $1/4m$*, which decreases the system throughput rapidly as* m *increases. Intuitively, as the fading parameter* m *increases, the tail of the CDF of the corresponding SINR becomes thinner, and the less likely there is a* qualified *transmitter-receiver connection. In other words, increasing* m *reduces channel fluctuations and the corresponding multiuser diversity gain. The case of OPS is more complicated since the throughput of OPS is governed by the negative branch of the Lambert W function* $\mathcal{W}_{-1}$*. Resorting to a recent inequality result on the Lambert W function [8], we can tightly lower and upper bound* $\mathcal{W}_{-1}$ *by*

$$\log n^{m-1} < -\mathcal{W}_{-1}(-n^{-\frac{1}{m-1}}) < \log n^{m-1} + \log\big((m-1)\log n\big). \tag{3.17}$$

It is then not difficult to show that

$$-\mathcal{W}_{-1}(-n^{-\frac{1}{m-1}}) = \Theta(\log n), \quad m > 1, m \in \mathbb{Z}^+, \tag{3.18}$$

which implies that OPS achieves the same scaling speed as THOR and is immune to the effect of Nakagami fading as well. Finally, it is noteworthy to point out that although a general throughput upper bound of $O(\sqrt{n})$ *has been derived for THOR under a class of fading distributions with finite mean and variance [10, Theorem 1] (including the case of Namagami fading), the exact throughput scaling behavior can be*

Table 3.1: The maximum supportable number of relays k that maintains throughput linearity for OPS and THOR

	$n = 1200$		$n = 2400$		$n = 4800$		$n = 9600$	
m	OPS	THOR	OPS	THOR	OPS	THOR	OPS	THOR
1	4.538	7.090	4.947	7.783	5.357	8.476	5.766	9.170
2	2.765	3.545	2.991	3.892	3.216	4.238	3.440	4.585
3	2.067	2.263	2.233	2.594	2.396	2.825	2.557	3.057
4	1.664	1.773	1.800	1.946	1.934	2.119	2.064	2.292

drastically different from this upper bound under specific channel models [10, Remark 1]. Therefore, the throughput scaling laws need to be studied on a case-by-case basis, which justifies the significance of Theorem 7.

Remark 4. *To illustrate the effect of Nakagami fading on the system throughout, we summarize in Table 3.1 the maximum supportable number of relays k^* that maintains throughput linearity for both OPS and THOR schemes when the fading parameter $m = \{1, 2, 3, 4\}$ and the number of S-D pairs $n = \{1200, 2400, 4800, 9600\}$. Clearly, there is a one-to-one correspondence between k^* and the maximum system throughput. As demonstrated in Appendix G, the analytical expressions of k^* equal the maximum achievable throughput multiplied by the constant 4, which are given by*

$$k^* = \begin{cases} \frac{1}{m} \log n & \text{for THOR} \\ \frac{1}{m} \log\big(n\overline{F}(1)\big) & \text{for OPS} \end{cases}. \tag{3.19}$$

This comparative result precisely captures the throughput loss of OPS in trading system throughput for the minimum delay. In order for the relays to schedule only matched S-D pairs, OPS can only exploit multiuser diversity from the candidate source set with only $n\overline{F}(1)$ source nodes, in comparison to the whole source set with n source nodes for THOR. The shrink in the candidate source set directly affects the multiuser diversity

gain achievable over the first hop and thereupon the end-to-end system throughput, as visualized by the penalty term $0 < \overline{F}(1) < 1$ *in* (3.19). *As can be observed from Table 3.1, the throughput degradation is more salient for smaller value of* m*, since as* m *increases, the performance of both schemes are limited by the reduced multiuser diversity gain (the pre-log* $1/m$ *factor), and the differences between them become negligible. We emphasize here that OPS achieves the same throughput scaling of* $\Theta(\log n)$ *as THOR and thus incurs no loss in large system limits. More importantly, while THOR is prone to select non-associated S-D pairs with large system delay, OPS requires no buffer at the relays and incurs no packet disorder at the destination, and zero start-up delay for the whole system.*

3.4 The L-scheduling Scheme

In the previous section, we proposed a simple matched opportunistic pair scheduling scheme. While THOR fully harnesses multiuser diversity in attaining the highest throughput, the proposed OPS scheme heads another extreme in trading throughput for the minimum communication delay. Through a comparative study, we have demonstrated that OPS incurs no loss in terms of throughput scaling laws, but is penalized in the total number of supported concurrent transmissions. For instance, in a Rayleigh fading environment, THOR can serve up to $\log n$ S-D pairs for simultaneous transmissions, which reduces to $\frac{\log 2n}{1+\log 2}$ for OPS. The aim of this section is to propose a general opportunistic scheme that captures THOR as one extreme point on the delay-throughput boundaries and OPS as another. This is done by introducing a design parameter L, whereby one can easily control the tradeoff between system throughput and delay, thus achieving the desired tradeoff point catering to different applications.

The L-scheduling scheme works as follows. After receiving the feedback messages from the destinations in the second hop, each relay r maintain a candidate set $\mathcal{D}_r$ listing all qualified destinations that can support the required communication rate. In the L-scheduling scheme, we let each relay r maintain a *candidate source* set $\mathcal{S}_r$ as well. This set is constructed by ranking the S-R channel gain in a non-increasing order, i.e., $\alpha_{i_r^{(1)}} \geq \alpha_{i_r^{(2)}} \geq \cdots \geq \alpha_{i_r^{(n)}}$, and then take the source indices corresponding to the first L elements into $\mathcal{S}_r$, i.e., $\mathcal{S}_r = \{i_r^{(1)}, i_r^{(2)}, ..., i_r^{(L)}\}$. Define the intersection between the two candidate sets as a matching set $\mathcal{M}_r$, where $\mathcal{M}_r = \mathcal{S}_r \cap \mathcal{D}_r$. When the intersection set is non-empty ($\mathcal{M}_r \neq \emptyset$), the S-D pair in $\mathcal{M}_r$ with the strongest S-R channel gain will be scheduled for transmission, i.e., $i_r = j_r = \arg\max_{i \in \mathcal{M}_r} \alpha_{i,r}$. This scenario is analogous to the OPS scheme described in Section 3.2. On the other hand, when relay r identifies no matched S-D pair ($\mathcal{M}_r = \emptyset$), it will schedule the best source node $i_r = i_r^{(1)} = \arg\max_{i \in \mathcal{S}_r} \alpha_{i,r}$ and randomly selects one destination from $\mathcal{D}_r$ to serve in the second hop. This scenario is analogous to the THOR scheme. Consequently, by varying the value of L between 0 and n (and thus the cardinality of the candidate source set), one can achieve different tradeoff points on the performance boundaries. In summary, the L-scheduling scheme is a general opportunistic scheduling scheme that captures THOR at one extreme and the pair scheduling scheme at the other (with L equal to 0 and n, respectively). More discussions on this point can be found in Remark 6. We now proceed to throughput scaling analysis.

3.4.1 Throughput Scaling

The second-hop throughput behavior of L-scheduling is the same as that of OPS, since the scheduled destination node set remains unchanged in the second hop. To obtain the first-hop throughput behavior, we apply the same technique as in Section 3.3.2 by approximating the size of $\mathcal{D}_r$ by its mean, i.e., $|\mathcal{D}_r| = \bar{n}$. Accordingly, the probability

that any S-D pair falls into the intersection set $\mathcal{M}_r$ is given by $p = |\mathcal{D}_r| \cdot |\mathcal{S}_r|/n^2 = \bar{n}L/n^2$. Denote $q = \min\{L, \bar{n}\}$. Thus, the number of matched S-D pairs $|\mathcal{M}_r|$ varies from 0 to q. By the law of total probability, $\Pr\left[\mathsf{SINR}^{\mathsf{H1}} \geq 1\right]$ can be expressed as

$$\Pr\left[\mathsf{SINR}^{\mathsf{H1}} \geq 1\right] = \sum_{i=0}^{q} \Pr\left\{|\mathcal{M}_r| = i\right\} \Pr\left\{\mathsf{SINR}^{\mathsf{H1}} \geq 1 \middle| \, |\mathcal{M}_r| = i\right\}. \tag{3.20}$$

We then compute $\Pr\left\{\mathsf{SINR}^{\mathsf{H1}} \geq 1 \middle| \, |\mathcal{M}_r| = i\right\}$ by considering $\mathcal{M}_r = \emptyset$ and $\mathcal{M}_r \neq \emptyset$, respectively.

$\mathcal{M}_r = \emptyset$

In this case, relay r finds no matched source node for any destination in $\mathcal{D}_r$. Therefore, the L-scheduling scheme reduces to THOR. Consequently, $\mathsf{SINR}^{\mathsf{H1}}$ can be written in the form of $\frac{X^{(1)}}{1/P+Y}$, with $X^{(1)}$ being the maximum value among n (not $\bar{n}$) Gamma distributed RVs. Exploiting the same technique used to obtain (3.14), we have

$$\Pr\left\{\mathsf{SINR}^{\mathsf{H1}} \geq 1 \middle| \, |\mathcal{M}_r| = 0\right\} \geq \left[1 - \left(1 - m^m e^{-ms} \frac{s^{m-1}}{\Gamma(m)}\right)^n\right] \times \left(1 - \frac{\gamma((k-1)m, m\,(s - 1/P))}{\Gamma((k-1)m)}\right), \tag{3.21}$$

where $s \in (0, s_1]$ and $s_1 = \frac{1}{m}\log n + \frac{m-2}{m}\log\log n$.

$\mathcal{M}_r \neq \emptyset$

In this case, relay r successfully identifies matched S-D pair to serve and thus $\mathsf{SINR}^{\mathsf{H1}}$ is expressed in the form of $\frac{X^{(2)}}{1/P+Y}$. Note here $X^{(2)}$ is the maximum variable among neither n nor $\bar{n}$ Gamma RVs, but among the $|\mathcal{M}_r| = i$ RVs, conditioned on that these i RVs are drawn from the L largest samples of n Gamma RVs. The distribution of $X^{(2)}$ is given by $F_{X^{(2)}}(x, i) = (F_W(x))^i$, where W is a RV drawn from the L largest samples of n Gamma RVs. We then derive the following lemma which characterizes the asymptotic behavior of $X^{(2)}$, the proof of which is given by Appendix H.

Lemma 4. *Let $X_1, \ldots, X_n$ be n $\Gamma(m, 1/m)$ RVs. Arrange $(X_1, \ldots, X_n)$ in an increasing order of magnitude and let $X_{1:n} \leq \cdots \leq X_{n:n}$. Thus, the first L statistics are $X_{n-L+1:n}, \ldots, X_{n:n}$. Suppose $X^{(2)}$ is the maximum among i variables out from $X_{n-L+1:n}, \ldots, X_{n:n}$, where $i = 1, \ldots, q$ with $q = \min\{L, \bar{n}\}$. It follows that $F_{X^{(2)}}(s_2) \to 0$ as $n \to \infty$ for $s_0 \leq s_2 \leq s_1$, where $s_0 = \frac{1}{m}\log\bar{n} + \frac{m-2}{m}\log\log\bar{n}$ and $s_1 = \frac{1}{m}\log n + \frac{m-2}{m}\log\log n$.*

With Lemma 4, we can again exploit the same technique in obtaining (3.14) to get the lower bound

$$\Pr\left\{\mathsf{SINR}^{\mathsf{H1}} \geq 1 \middle| \, |\mathcal{M}_r| = i\right\} \geq \overline{F}_{X^{(2)}}(s_2) F_Y\left(s_2 - \frac{1}{P}\right). \tag{3.22}$$

By substituting (3.21) and (3.22) into (3.20), we have

Lemma 5. *The throughput of the L-scheduling scheme in the first hop is lower-bounded by*

$$\begin{aligned} R_1 \geq & k\frac{n(n-1)\cdots(n-k+1)}{n^k} \\ & \times \left\{(1-p)^n\left[1 - \left(1 - m^m e^{-ms_1}\frac{s_1^{m-1}}{\Gamma(m)}\right)^n\right] \times \left(1 - \frac{\gamma((k-1)m, m\,(s_1 - 1/P))}{\Gamma((k-1)m)}\right)\right. \\ & \left. + \sum_{i=1}^{q}\binom{n}{i}p^i(1-p)^{n-i}\left(1 - F_{X^{(2)}}(s_2, i)\right)\left(1 - \frac{\gamma((k-1)m, m\,(s_2 - 1/P))}{\Gamma((k-1)m)}\right)\right\}, \end{aligned} \tag{3.23}$$

where $s_0 \leq s_2 \leq s_1$, $\bar{n} = n\overline{F}(1)$, $q = \min\{\bar{n}, L\}$, $p = L\bar{n}/n^2$ and

$$F_{X^{(2)}}(s_2, i) = \left(\frac{1}{L}\sum_{j=0}^{L}\left[\left(1 - \frac{\gamma(m, ms_2)}{\Gamma(m)}\right)^{n-j}\left(\frac{\gamma(m, ms_2)}{\Gamma(m)}\right)^j\right]\right)^i. \tag{3.24}$$

With the second-hop throughput expression (3.8), we then have the following theorem.

Theorem 8. *L-scheduling achieves at least the same throughput scaling as that of OPS, which is* $\Theta(\log n)$ *for Rayleigh fading and* $\Theta\big(-\mathcal{W}_{-1}(-n^{-\frac{1}{m-1}})\big)$ *for Nakagami fading with* $m > 1, m \in \mathbb{Z}^+$.

Proof: From the proof of Theorem 6, which can be found in Appendix G, we know that the first additive term in the lower-bound expression (3.23) scales the same as that of THOR, while the second maintains the same scaling as that of OPS. Hence, the dominant scaling behavior of L-scheduling depends on their respective probability weight, i.e., $(1-p)^n$ for the first term and $\binom{n}{i}p^i(1-p)^{n-i}$ for the second. Since the affordable growing speed of k for OPS is slower than that of THOR, we conclude that L-scheduling achieves at least the same throughput scaling as that of OPS. ■

Remark 5. *Theorem 8 reveals an interesting insight in the throughput scaling of L-scheduling. A portion of the throughput increases linearly in* k *until* $\frac{1}{m}\log n\overline{F}(1)$*, and begins to decrease after this point. On the other hand, the rest of the throughput maintains linearity in a much wider range as* k *varies between* 1 *and* $\frac{1}{m}\log n$*. Therefore, the end-to-end throughput of L-scheduling is guaranteed to increase linearly in* k *till* $\frac{1}{m}\log n\overline{F}(1)$ *and decrease rapidly after* $\frac{1}{m}\log n$*. When* k *falls in between these two points, the throughput is composed of two weighted contributions from two non-dominant throughput terms, which levels off at some value with slight fluctuations (cf. Fig. 3.3). We point out that (3.23) is a throughput scaling lower bound of L-scheduling and is insufficient to characterize the exact throughput scaling law. Nevertheless, being lower bounded by OPS, the throughput of L-scheduling is guaranteed to scale as fast as that of OPS and thus incurs no loss in comparison to THOR as well.*

Remark 6. *It is important at this point to emphasize the differences between "throughput scaling" and "maximum achievable throughput". As we have shown, although the*

throughput scaling of the general L-scheduling scheme is lower bounded by that of OPS, as L increases from 0 to n, the maximum number of relays k^* that maintains linearity shifts from $\frac{1}{m}\log n$ to $\frac{1}{m}\log n\overline{F}(1)$. In other words, although no loss is incurred on how the system throughput scales as the network grows in size, the maximum achievable throughput does decrease due to the reduced number of concurrent transmissions supported. In the next section, we shall show that increasing L also has the dual effect of improving the system delay performance. Therefore, the parameter L plays a dual role of controlling the maximum achievable throughput as well as the system delay. Through properly designing L, one can achieve the desired delay-throughput tradeoff catering to different application requirements.

3.5 Delay Analysis

There has been a line of work concentrating on the delay-throughput tradeoff of wireless networks, as reviewed in Section 3.1. Here we follow the delay analysis methodology used in [37] for mobile ad-hoc networks [20], which has also been applied in [50] to examine the delay performance of THOR scheme. We shall first derive a closed-form end-to-end delay upper bound for the general L-scheduling scheme, whereby the delay performance of THOR and OPS are readily available by setting $L = 0$ and $L = n$, respectively. Next, we study and compare the least start-up delay (LSD) across different schemes, followed by defining a delay-throughput tradeoff performance metric and discussion some of its properties.

3.5.1 End-to-End Delay

In this section we study the end-to-end delay performance of L-scheduling. The delay performance of OPS is readily available by setting $L = n$. We assume that each

source node generates data traffic to be sent to its corresponding destination node. The total end-to-end delay consists of the delays experienced over two hops. The packet transmission in both hops can be modeled by queues. Take the ith S-D pair for example. In the first hop, S_i maintains a queue for the buffered data packets, while in the second hop, each relay maintains a queue that contains data packets from all the source nodes. It is assumed the arrival process is a Poisson process with arrival rate λ, which is commonly adopted for queuing analysis in wireless networks [15,37,47]. The service opportunity arises at a certain probability and accordingly has a Bernounlli distribution with an average rate μ. For an S-D pair, say the i-th S-D pair, there are two modes regarding the way it is scheduled by relay r, $1 \leq r \leq K$: 1) the i-th S-D pair is scheduled by relay r according to OPS; 2) the source node and destination node are scheduled independently according to THOR. We assume that the i-th S-D pair is scheduled by mode 1 for j relays and by mode 2 for $K-j$ relays, where $0 \leq j \leq K$. For different value of j, we analyze the behavior of the queues in Appendix I, leading to the following theorem on the delay performance of the proposed schemes.

Theorem 9. *The delay of L-scheduling is upper bounded by* $D_1 \leq \left(\frac{n}{k} - L\overline{F}(1)\right)\frac{2-\rho}{2(1-\rho)}$ *for the first hop, and by* $D_2 = n - kL\overline{F}(1)$ *for the second hop. Consequently, the end-to-end delay for the whole system is upper-bounded by*

$$\begin{aligned} D = D_1 + D_2 &\leq n - kL\overline{F}(1) + \left(\frac{n}{k} - L\overline{F}(1)\right)\frac{2-\rho}{2(1-\rho)} \\ &= n + \frac{n}{k}\frac{2-\rho}{2(1-\rho)} - \left(k + \frac{2-\rho}{2(1-\rho)}\right)\overline{F}(1)\cdot L\,, \end{aligned} \tag{3.25}$$

where $\rho = \lambda/\mu_0$ *and* $\mu_0 = 1/n$. □

As can be observed from (3.25), for any fixed system profile, n, k, m and P, and queueing profile ρ, the system end-to-end delay upper bound decreases linearly in L in the speed of $\left(k + \frac{2-\rho}{2(1-\rho)}\right)\overline{F}(1)$. The following proposition follows immediately.

Proposition 3. *L-scheduling reduces end-to-end delay linearly as L increases.*

In the preceding section we have shown that as L increases, it becomes increasingly difficult to maintain throughput linearity. The results in Theorem 9 indicates that the penalty on the achievable throughput is accompanied by an improvement in system delay performance. Again, two extreme cases worth being highlighted to provide insights in the role of L. When $L = 0$, L-scheduling reduces to THOR scheme. In this case, the maximum number of relays can grow as large as $\log n$ ensuring the throughput scales linearly in k. However, the system delay behaves the worse for all possible values of L, i.e., $D \leq n + \frac{n}{k}\frac{2-\rho}{2(1-\rho)}$. When $L = n$, L-scheduling reduces to OPS and the system achieves the minimum end-to-end delay of $n + \frac{n}{k}\frac{2-\rho}{2(1-\rho)} - n \cdot \left(k + \frac{2-\rho}{2(1-\rho)}\right)\overline{F}(1)$. We point out that increasing L only brings a linear decrease in the system end-to-end delay, which is still in the order of n. This in turn verifies the results in [50]. We comment also that for fixed L, increasing m decreases $\overline{F}(1)$ and thus negatively affects the the slope of delay reduction $\left(k + \frac{2-\rho}{2(1-\rho)}\right)\overline{F}(1)$. In other words, the effect of delay improvement is less significant for larger values of m.

3.5.2 The Least Start-up Delay

In order for THOR to be practically functional, each relay needs to have information packets to be delivered to all the destinations buffered in their queues, respectively. This requirement translates into a start-up time period for the whole network to "warm up" and reach steady states. In this subsection, we quantify this delay for L-scheduling. Again, the respective performance of THOR and OPS can be obtained setting $L = 0$ and $L = n$.

We denote the time duration of one-hop transmission by t_0. It is straightforward to show that the LSD of THOR is nt_0. For L-scheduling, if the intersection set $\mathcal{M}_r$ for relay r is empty, the LSD corresponds to that of THOR, nt_0. On the other hand,

if $\mathcal{M}_r$ is non-empty, the LSD is zero since an S-D pair will be scheduled and no data needs to be buffered in advance. Therefore, the LSD of L-scheduling can be obtained as

$$\begin{aligned}\mathsf{LSD} &= \Pr\{\mathcal{M}_r = \emptyset\} \cdot nt_0 + \Pr\{\mathcal{M}_r \neq \emptyset\} \cdot 0 \\ &= (1-p)^n \cdot nt_0 \\ &= \left(1 - \frac{L}{n}\overline{F}(1)\right)^n \cdot nt_0 \to e^{-\overline{F}(1)L} \cdot nt_0\,, \end{aligned} \tag{3.26}$$

where the approximation follows from the property of exponential functions. The following proposition ensues immediately.

Proposition 4. *L-scheduling reduces system least start-up delay exponentially in L.*

3.5.3 Delay-throughput Tradeoff

To characterize the delay-throughput tradeoff, we follow the convention and study the following performance metric

$$\mathcal{T} = \frac{\min(R_1, R_2)}{2D}, \tag{3.27}$$

which represents the achievable system throughput per unit delay. With the closed-form performance bounds derived in Theorem 8 and Theorem 9, a lower bound on the delay-throughput tradeoff of L-scheduling can be obtained as

$$\mathcal{T}(L) = \frac{R_1(L)}{2\,D(L)}, \tag{3.28}$$

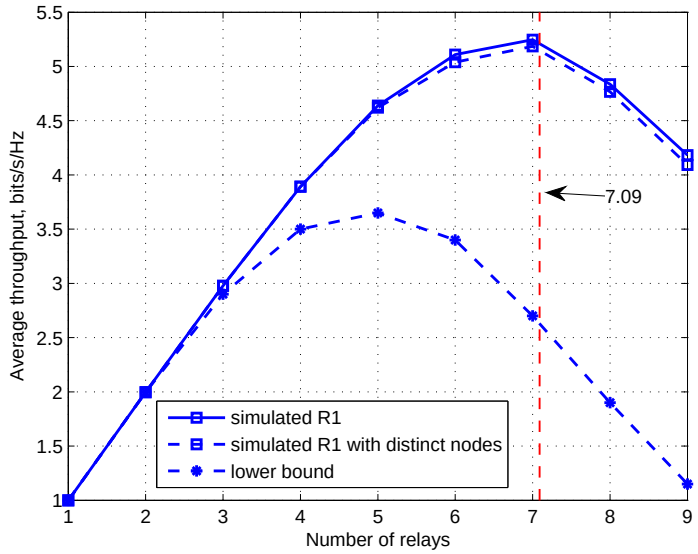
where R_1 is given by (3.23) and D is given by (3.25). We note that for a given system profile, the tradeoff expression (3.28) is a function of a single variable L defined over the domain of $[0, n]$. Since deriving the optimized L^* in closed-form is mathematically

involved, we resort to numerical optimization methods and some results are presented in the forthcoming section.

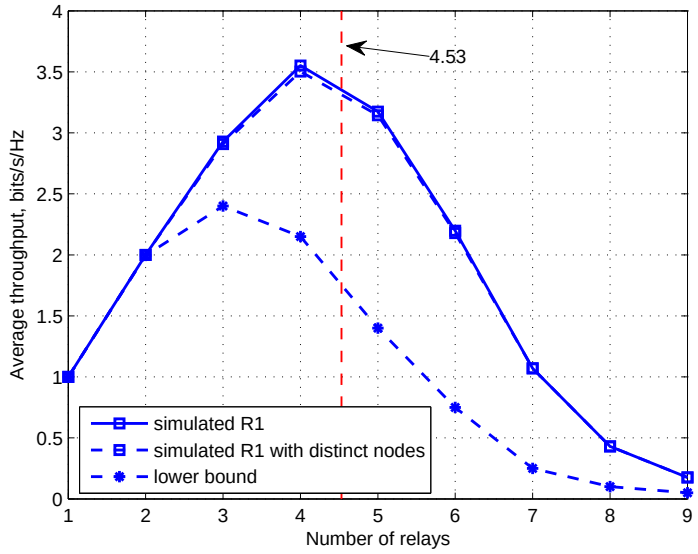
3.6 Simulation Results

In this section, we provide some numerical examples to illustrate 1) the delay-throughout tradeoff of the proposed schemes in comparison to THOR and 2) The impact of Nakagami fading on this tradeoff. Throughout these examples, the SNR is set at 10 dB ($P = 10$ dB) for both hops.

We show in Fig. 3.2 the first-hop average throughput R_1 of OPS versus THOR for a network with k relays and $n = 1200$ S-D pairs under Rayleigh fading. The simulated curves were obtained by averaging over 2000 channel realizations. The "simulated R_1" curve refers to the average throughput using all assignments of source nodes, while the "simulated R_1 with distinct nodes" curve corresponds to only assignments of distinct source nodes. These two curves match each other very well, which establishes the fact that considering only distinct source node assignments leads to a tight lower bound for large n. The lower bound curves in both subfigures were obtained by numerically evaluating (3.23) for $L = 0$ and $L = n$, respectively. In particular, the curves shown in Fig. 3.2(a) for THOR match the results in [9, Fig. 2] very well. Observe also the trend that as the number of relays k increases, the system can support more concurrent transmitting S-D pairs leading to a linear throughput increase. However, as the number of relays exceeds a threshold, the throughput gain from multiuser diversity cannot compensate the harmful multiuser interference, and thus the system throughput begins to decline monotonically. This effect is consistent with our theoretical results (3.19) in Section 3.3 with $\log 1200 = 7.09$ and $\frac{\log 2n}{1+\log 2} = 4.53$, respectively.

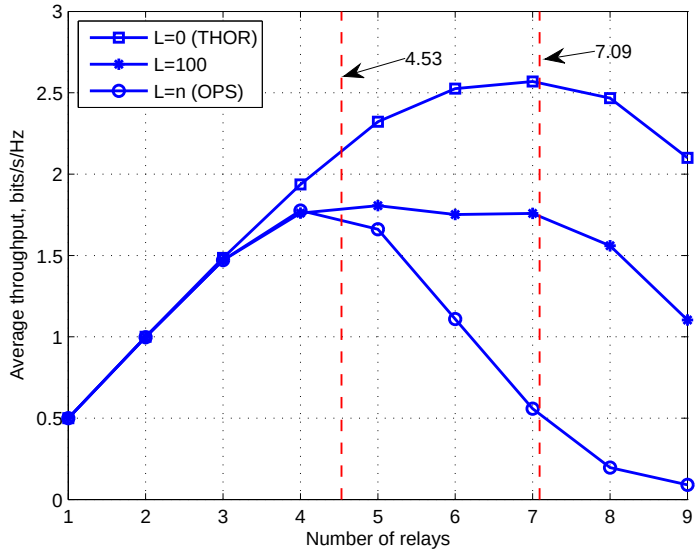


(a) THOR

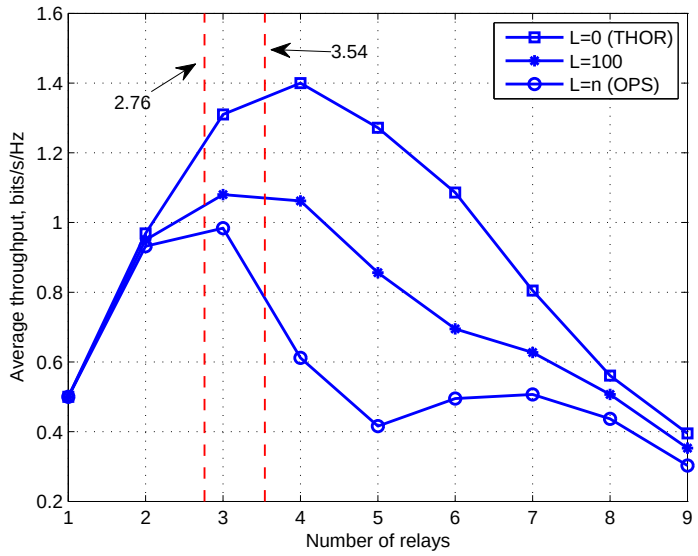


(b) OPS

Figure 3.2: First-hop average throughput R_1 as a function of the number of relays k for $n = 1200$ S-D pairs under Rayleigh fading. (a) $L = 0$, i.e., Cui's THOR scheme; (b) $L = 1200$, i.e., the proposed OPS scheme. The square solid line refers to the average throughput using all assignments of source nodes, while the square dashed line refers to the average throughput using only assignments of distinct source nodes. The star dashed line represents the theoretical lower bound (3.23). The vertical dash line refers to the maximum theoretical value of k to maintain the linear throughput in k.



(a) $m = 1$, Rayleigh fading

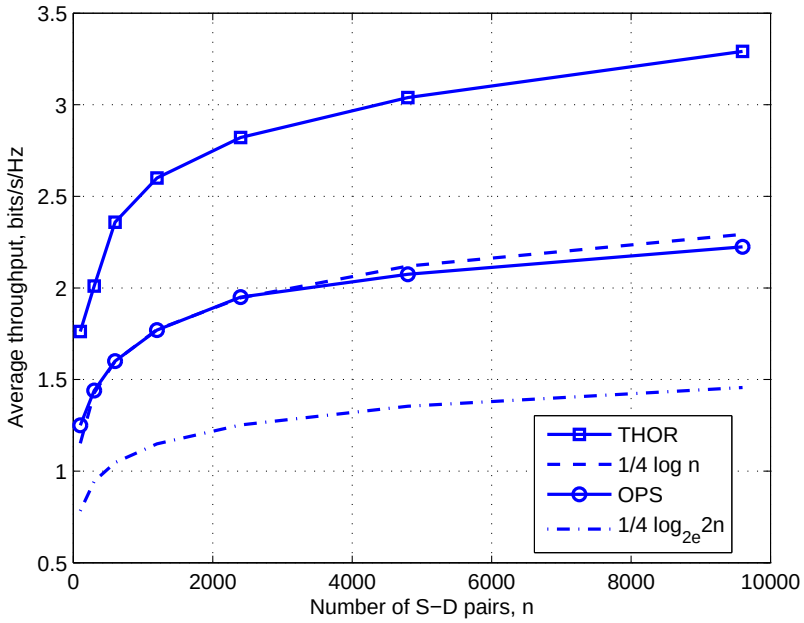


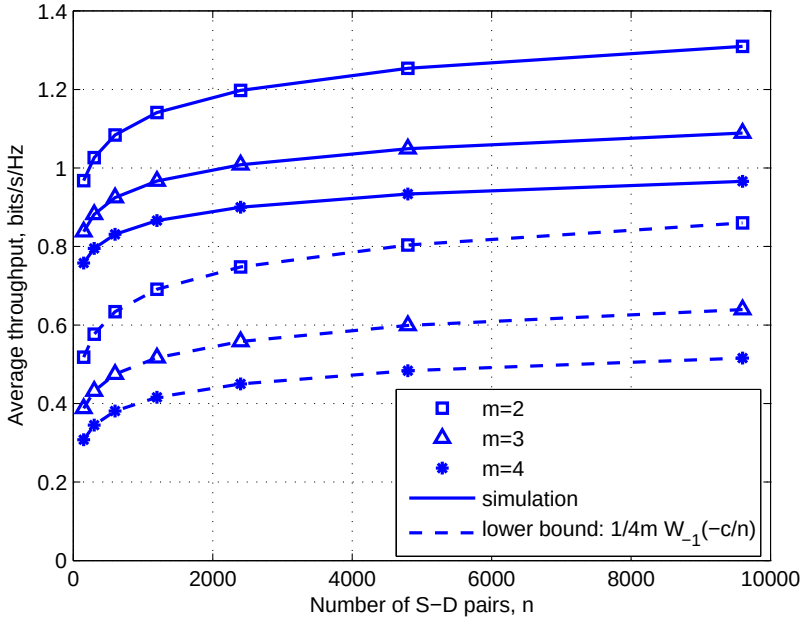
(b) $m = 2$

Figure 3.3: System average throughput as a function of the number of relays k for $n = 1200$ S-D pair. (a) Rayleigh fading; (b) Nakagami fading with $m = 2$. The square solid line refers to Cui's THOR scheme, the circle dashed line refers to the proposed OPS scheme, and the star dashed line refers to the proposed L-scheduling scheme with $L = 100$. The vertical dash line refers to the maximum theoretical value of k to maintain the linear throughput in k.

Fig. 3.3 illustrates the impact of Nakagami fading on the system throughput. For the two subfigures Fig. 3.3(a) and Fig. 3.3(b), the Nakagami fading parameter m takes the values of $m = 1$ and $m = 2$, respectively. Observe that as the fading parameter m increases, the system can support less number of concurrent transmissions and the throughput decreases accordingly. In particular, the maximum number of relays that maintains throughput linearity drops from 4.53 to 2.76 for OPS and from 7.09 to 3.54 for THOR as m increases from 1 to 2, which again verifies our theoretical analysis in (3.19). This is because, high values of m represent high line-of-sight (LOS) components in the channel between the transmitter and receiver. Consequently, as the fluctuations of the signal strength reduce, the probability of identifying qualified destination nodes with SINR larger than 1 (the value of $\overline{F}(1)$) decreases as well, which leads to lower system throughput. Besides, for L-scheduling with $L \in (0, n)$, the system throughput can fluctuate when k falls in $\left(\frac{1}{m}\log n\overline{F}(1), \frac{1}{m}\log n\right)$, as shown in Fig. 3.3(a) for the case of $L = 100$. This is because, in this case, the system throughput consists of two weighted non-dominant throughput contributions from THOR and OPS, respectively, resulting in possible throughput fluctuations. Despite of the unpredictable behavior of L-scheduling in this region of k, we can observe that it at least guarantees the same linear throughput increasing region as that of OPS.

Although Nakagami fading affects the system throughout negatively, Fig. 3.4 shows that it has no impact on the throughput scaling laws. Note that here for different numbers of S-D pairs, the number of relays k is selected to be the optimal value that attains the highest throughput. The lower bound is given by Theorem 6. Observe that regardless of the value of m, the throughputs of all the schemes scale equally fast as the number of S-D pairs n increases. This confirms our theoretical result in (3.18), implying that the proposed general L-scheduling scheme incurs no loss in regard to throughput scaling laws.





(b) $m = \{2, 3, 4\}$

Figure 3.4: System average throughput as a function of the number of S-D pairs n and for optimized number of relays k^*. (a) THOR and OPS under Rayleigh fading. (b) OPS with Nakagami fading $m = \{2, 3, 4\}$. The lower bound curves are given by Theorem 6 and Theorem 7.

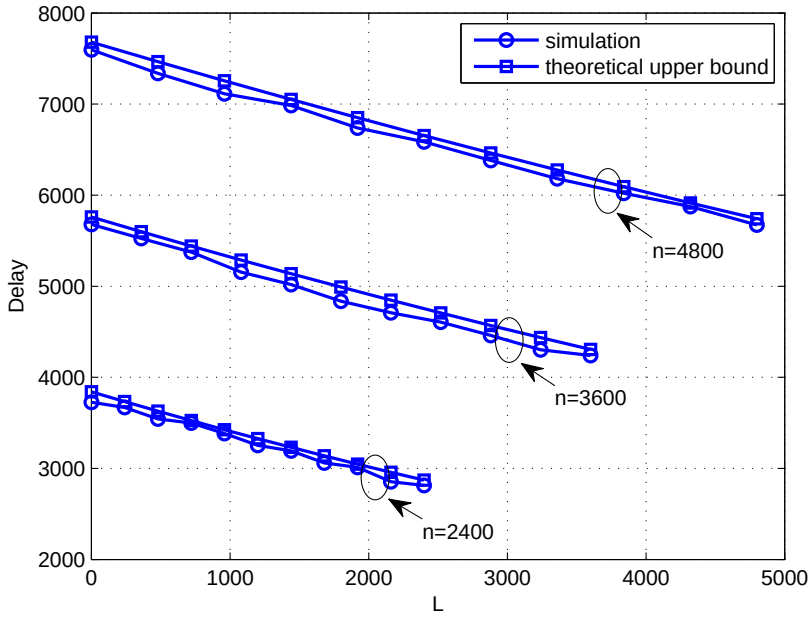


Figure 3.5: System end-to-end delay of the L scheduling as a function of L for $n = 1500$ S-D pairs under Rayleigh fading. Queueing profile $\rho = 0.8$ and number of relays $k = 5$. The square line refers to the theoretical delay upper bound given by (3.25) and the circle line refers to the simulated system delay.

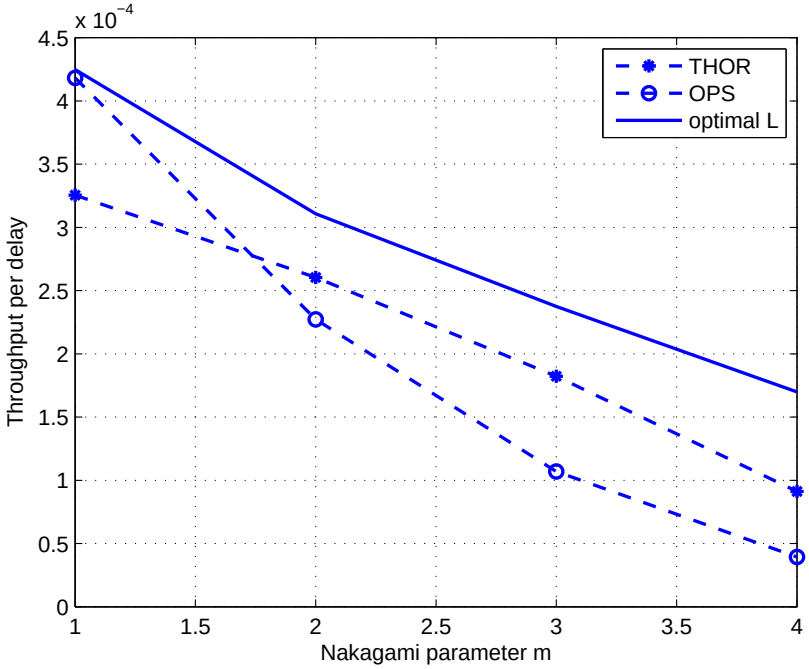


Figure 3.6: Delay-throughput tradeoff as a function of the Nakagami fading parameter m. Number of S-D pairs $n = 4800$, number of relays $k = 5$, and queueing profile $\rho = 0.8$. The solid line refers to the delay-throughput tradeoff with optimized L^*, the star dashed line refers to the tradeoff of THOR, and the circle dashed line refers to the tradeoff of OPS.

Table 3.2: Throughput scaling comparison under Nakagami fading with parameter m

	Throughput ($m = 1$)	Throughput ($m > 1, m \in \mathbb{Z}^+$)
THOR [9]	$\Theta(\log n)$	$\Theta(\log n)$
OPS	$\Theta(\log n)$	$\Theta\big(-\mathcal{W}_{-1}(-n^{-\frac{1}{m-1}})\big)$
L-scheduling	$\Theta(\log n)$	$\Theta\big(-\mathcal{W}_{-1}(-n^{-\frac{1}{m-1}})\big)$

Fig. 3.5 demonstrates the delay performance of the proposed scheme versus the value of L for an ad-hoc network with $n = \{2400, 3600, 4800\}$ S-D pairs. It is observed that this delay bound is linearly decreasing in L for all values of n, consistent with our analysis in (3.25). For a wireless system that prioritizes throughput, one is advised to select smaller L to fully harness the benefit of multiuser diversity. On the other hand, for delay-sensitive applications, large L should be adopted with some penalty on the system throughput. This tradeoff has been further illustrated in Fig. 3.6 for THOR, OPS, as well as L-scheduling with optimized L^* (obtained numerically by maximizing the throughput per delay expression in (3.27)) in different fading environments. Three observations are noteworthy relative to Fig. 3.6. First, OPS approaches the optimal throughput-delay tradeoff under Rayleigh fading, but becomes inefficient as m increases. Second, the throughput-delay tradeoff with optimized L^* outperforms both OPS and THOR schemes. The improvement is more significant for OPS under large m. Third, the tradeoff curve decreases as m increases. In a nutshell, increasing the Nakagami fading parameter m has the dual effect of decreasing system throughput and increasing delay simultaneously, which contributes negatively to system throughput-delay tradeoffs.

3.7 Conclusions

Our main contributions in this chapter are summarized as follows. By fully exploiting multiuser diversity over two hops, THOR [9] achieves an order-optimal throughput of $\Theta(\log n)$ under Rayleigh fading, at the expense of large communication and decoding delay [9, 50]. The objective of our work is to seek for alterative opportunistic scheduling schemes that can achieve balanced throughput-delay tradeoffs. We start by considering an opportunistic pair scheduling scheme which restricts the relays to serve only *matched* S-D pairs. In comparison to THOR that sacrifices delay for the maximum throughput, OPS heads another extreme and trades throughput for the minimum system delay. By selecting THOR as our starting point, we have conducted a comparative theoretical study in a comprehensive manner. Specifically, we have obtained accurate throughput lower bound and scaling law for both OPS and THOR under Nakagami fading, thereby generalizing Cui's results to the case of Nakagami fading. The obtained throughput scaling laws are summarized in Table 3.2. Some new physical insights are

- **Nakagami fading has no effect on the system throughput scaling laws**. While this statement is clear for THOR (cf. Table 3.3), which achieves a maximum throughput scaling of $\Theta(\log n)$ regardless of the value of the Nakagami fading parameter m, the throughput scaling of OPS is more complicated. According to Table 3.2, the throughput behavior of OPS is governed by the negative branch of the Lambert W function[2] $\mathcal{W}_{-1}$ [7], with m taking positive integer values large than 1. Resorting to a recent inequality result on the Lambert W function [8], we show that $\Theta(-\mathcal{W}_{-1}(-n^{-\frac{1}{m-1}})) = \Theta(\log n)$, $\forall m \in \mathbb{Z}^+$, which

[2]The Lambert W function $\mathcal{W}(x)$, also known as the Omega function or product logarithm, is a set of functions defined by the solution to the transcendental equation $\mathcal{W}(x)e^{\mathcal{W}(x)} = x$ for $x > -e^{-1}$. We focus on the lower branch (or the negative branch) which has $\mathcal{W}(x) \leq -1$, denoted by $\mathcal{W}_{-1}(x)$. For $x \in (-e^{-1}, 0^-)$, the value of $\mathcal{W}_{-1}(x)$ decreases from $\mathcal{W}_{-1}(-e^{-1}) = -1$ to $\mathcal{W}_{-1}(0^-) = -\infty$.

implies that Nakagami fading has no impact on the throughput scaling laws of all the schemes.

- **OPS and L-scheduling incur no throughput scaling loss in comparison to THOR**. This statement follows directly from (3.18) which indicates that all the schemes achieve the same throughput scaling of $\Theta(\log n)$. Consequently, the proposed schemes incur no loss in terms of how throughput scales when the number of S-D pairs grows large. Moreover, they all operate on the same level of CSI knowledge (perfect CSI at the receiver side (CSIR) and an indexed-value integer feedback at the transmitter side), yet OPS outshines in terms of buffer requirement, communication delay and decoding delay.

- **L-scheduling achieves the largest throughput per unit delay with optimized L**. Although OPS achieves the same throughput scaling as that of THOR, it is disadvantageous in the achievable throughput due to the lost of multiuser diversity in the first hop. For this reason, L-scheduling serves as an adaptive scheme that allows network designers to achieve desired throughput-delay tradeoffs by fine tuning L, thereby catering to diverse application requirements. As L increases, delay improves and the achievable throughput degrades. THOR and OPS correspond to two extreme cases of L-scheduling with L equal to 0 and n, respectively. It turns out that one can further optimize L to achieve the largest throughput per unit delay of all the schemes (see Fig. 3.6).

In addition to the results on throughput scaling laws, we have derived also a closed-form delay performance bound for the L-scheduling scheme. This performance bound is valid for any fading scenario with $F(x)$ being the corresponding CDF of the SINR measured at a destination node. The delay bounds of THOR and OPS are readily available by setting L to 0 and n, respectively. The results are summarized

Table 3.3: Delay comparison under Nakagami fading with parameter m

	End-to-End Delay	Least Start-up Delay
THOR [9]	$n + \frac{n}{k}\frac{2-\rho}{2(1-\rho)}$	nt_0
OPS	$n + \frac{n}{k}\frac{2-\rho}{2(1-\rho)} - \left(k + \frac{2-\rho}{2(1-\rho)}\right)\overline{F}(1)\,n$	0
L-scheduling	$n + \frac{n}{k}\frac{2-\rho}{2(1-\rho)} - \left(k + \frac{2-\rho}{2(1-\rho)}\right)\overline{F}(1)L$	$e^{-\overline{F}(1)L} \cdot nt_0$

in Table 3.3. A few key points to note are

- **L-scheduling reduces end-to-end delay linearly**. Observe from Table 3.3 that the end-to-end delay of L-scheduling decreases linearly in L. In particular, for a fixed L, the slope of delay reduction is proportional to the tail distribution $\overline{F}(x)$ evaluated at $x = 1$. The larger the Nakagami fading parameter m is, the thinner the tail and the less likely there is a *qualified* transmitter-receiver connection[3]. In other words, increasing m has the negative effect of reducing channel fluctuations and thus the amount of multiuser diversity gain. Additionally, it is worth noting that OPS has the minimum end-to-end delay and start-up delay among all the schemes.

- **L-scheduling reduces least start-up delay exponentially**. In order for THOR to work in practice, every relay node must have information packets to be delivered to all the destinations stored locally. This necessitates what we call a start-up time period for the whole network. We show that the least start-up delay for THOR is nt_0 which grows unbounded as n approaches the limit. In contrast, OPS schedules only matched S-D pairs and thus has *zero* start-up delay. L-scheduling, on the other hand, reduces the start-up delay by a multiplicative term $e^{-\overline{F}(1)L}$, which decays exponentially fast and approaches zero for reasonably large L. The role of Nakagami fading played in start-up delay is the similar as that in the end-to-end delay.

[3]A qualified transmitter-receiver connection refers to the transceiver pair that can achieve an SINR larger than a threshold to support a successful transmission.

Chapter 4

Conclusions

The very feature of cooperative communications that it can provide spatial diversity in a virtual MIMO manner is highly favorable in applications where small size user end devices are used. However, as application evolves toward a more practical situation, realistic constraints and issues such as CSI assumption and delay must be accounted when developing appropriate cooperative schemes. In this thesis, we have addressed delay related problems in both SUPC and MUPC networks.

In Chapter 2, we have studied the impact of outdated CSI, caused by the delay between the selection instant and the transmission instant, on the existing SRSs. All these SRSs achieves full diversity order with perfect CSI, but surprisingly, their diversity order all reduce to 1 with outdated CSI. We then proposed an MRS which can function in a similar way to ORS in a completely distributed manner. Such scheme achieves desired diversity order and has been demonstrated to be robust to the variation of wireless environment.

As addressed in Chapter 3, we investigate the possibility of achieving linear throughput scaling in a two-hop wireless ad-hoc network with arbitrary throughput-delay tradeoffs. Our results reveal that instead of matching the best transceiver pairs

in each hop separately (THOR), one can exploit multiuser diversity in a joint manner by restricting relays to serve only matched S-D pairs (OPS), without compromising the (linear) throughput scaling law yet reducing delay to the minimum (cf. Theorem 6). This somewhat surprising result propels us to further develop a general opportunistic scheduling scheme (L-scheduling) that incorporates THOR and OPS as two extreme cases, and most importantly, achieves arbitrary performance points along the throughput-delay tradeoff boundaries. Delay performance bounds of all the schemes have been derived in closed-form (cf. Theorem 9). In regard to fading, we show that Nakagami fading has no impact on throughput scaling laws, but degrades throughput and delay in general as the fading parameter m increases. Finally, we emphasize that all the proposed schemes allow decentralized operation and achieve the throughput scaling upper bound with affordable overheads.

There are several limitations of this thesis which can be topics of future work. First, as discussed in Section 2.5, we assumed that synchronization is perfect for our proposed MRS. However, in practice, we have to consider the influence of different level of synchronization when implementing MRS schemes. Secondly, although the N+NT-ORS scheme provides desired diversity order, how to determine the necessary diversity order and the predetermined threshold μ is worth further investigation. One possibility is to use optimization technique to model the problem with appropriate total or individual power constraints. However, we would like to stress that it is very likely that the optimization approach will break the distributed fashion of the scheme. Thirdly, as we mentioned in Section 3.3.3, the throughput scaling of THOR should be studied on a case-by-case basis for different fading types. Likewise, future effort can be placed on analyzing the proposed OPS and L scheduling under other fading environments.

Bibliography

[1] M.-S. Alouini and M. K. Simon. Performance of coherent receivers with hybrid SC/MRC over Nakagami-m fading channels. *IEEE Trans. Veh. Technol.*, 48:1155–1164, July 1999. 23

[2] G. Amarasuriya, M. Ardakani, and C. Tellambura. Adaptive multiple relay selection scheme for cooperative wireless networks. In *Proc. of IEEE WCNC*, pages 1–6, April 2010. 5

[3] A. Bletsas. *Intelligent antenna sharing in cooperative diversity wireless networks*. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, September 2005. 3, 30, 32, 42

[4] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman. A simple cooperative diversity method based on network path selection. *IEEE J. Sel. Areas Commun.*, 24(3):659–672, March 2006. 3, 4, 19, 20, 30, 32, 33, 35, 38, 40, 41, 42, 91

[5] A. Bletsas, H. Shin, and M. Z. Win. Cooperative communications with outage-optimal opportunistic relaying. *IEEE Trans. Wireless Commun.*, 6(9):3450–3460, September 2007. 3, 4, 18, 30, 32, 33, 35, 38, 40, 41, 42

[6] V. R. Cadambe and S. A. Jafar. Interference alignment and degrees of freedom of the k-user interference channel. *IEEE Trans. Inf. Theory*, 54(8):3425–3441, August 2008. 7, 8

[7] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the lambert w function. *Adv. Comput. Math.*, 5(4):329–359, March 1996. 76

[8] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. Approximation of the lambert w function. *RGMIA Research Report Collection*, 8(4), Article 12, 2005. 57, 76

[9] S. Cui, A. M. Haimovich, O. Somekh, and H. V. Poor. Opportunistic relaying in wireless networks. *IEEE Trans. Inf. Theory*, 55(11):5121–5137, November 2009. 9, 43, 45, 48, 51, 52, 53, 54, 55, 56, 68, 75, 76, 78, 100

[10] S. Cui, A. M. Haimovich, O. Somekh, and H. V. Poor. Throughput scaling of wireless networks with random connections. *IEEE Trans. Inf. Theory*, 56(8):3793–3806, July 2010. 9, 45, 57, 58

[11] A. F. Dana and B. Hassibi. On the power efficiency of sensory and ad hoc wireless networks. *IEEE Trans. Inf. Theory*, 52(7):2890–2914, July 2006. 7, 8

[12] S. N. Diggavi, M. Grossglauser, and D. N. C. Tse. Even one-dimensional mobility increases the capacity of wireless networks. *IEEE Trans. Inf. Theory*, 51(11):3947–3954, November 2005. 7

[13] T. Eng, N. Kong, and L. B. Milstein. Comparison of diversity combining techniques for Rayleigh-fading channels. *IEEE Trans. Commun.*, 44:1117–1129, March 1996. 23

[14] M. Franceschetti, O. Dousse, D. N. C. Tse, and P. Thiran. Closing the gap in the capacity of wireless networks via percolation theory. *IEEE Trans. Inf. Theory*, 53(3):1009–1018, March 2007. 7

[15] A. El Gamal, J. Mammen, B. Prabhakar, and D. Shah. Throughput delay tradeoff in wireless networks. In *IEEE INFOCOM*, November 2004. 9, 10, 65

[16] A. Goldsmith, S. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE J. Sel. Areas Commun.*, 21(5):684–702, June 2003. 1

[17] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, 2005. 25

[18] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic Press, San Diego, CA, 6th edition, 2000. 51, 54, 93, 98

[19] D. Gross and C. M. Harris. *Fundamentals of queuing theory*. Wiley, John and Sons, 3rd edition, 1998. 106

[20] M. Grossglauser and D. N. C. Tse. Mobility increases the capacity of ad hoc wireless networks. *IEEE/ACM Trans. Netw.*, 10(4):477–486, August 2002. 7, 9, 64

[21] P. Gupta and P. R. Kumar. The capacity of wireless networks. *IEEE Trans. Inf. Theory*, 46(2):388–404, March 2000. 7

[22] S. S. Ikki and M. H. Ahmed. Performance analysis of generalized selection combining for amplify-and-forward cooperative diversity networks. In *Proc. of IEEE ICC*, June 2009. 3, 5, 22, 23, 25, 26, 29, 33, 40, 91

[23] S. S. Ikki and M. H. Ahmed. Performance analysis of generalized selection combining for decode-and-forward cooperative diversity networks. In *Proc. of IEEE VTC*, October 2010. 5, 22, 25, 33, 40

[24] Y. Jing and H. Jafarkhani. Single and multiple relay selection schemes and their diversity orders. In *Proc. of IEEE ICC Workshops*, pages 349–353, May 2008. 3, 4, 5, 17

[25] Yindi Jing and H. Jafarkhani. network beamforming using relays with perfect channel information. *IEEE Trans. Inf. Theory*, 55(6):2499–2517, May 2009. 5

[26] R. Knopp and P. Humblet. Information capacity and power control in single cell multiuser communications. In *Proc. of IEEE ICC*, volume 1, pages 331–335, June 1995. 9

[27] Y.-C. Ko, M.-S. Alouini, and M. K. Simon. Outage probability of diversity systems over generalized fading channels. *IEEE Trans. Commun.*, 48(11):1783–1787, 2000. 25, 26, 28

[28] J. N. Laneman, D. Tse, and G. W. Wornell. Cooperative diversity in wireless networks: efficient protocols and outage behaviour. *IEEE Trans. Inf. Theory*, 50(12):3062–3080, December 2004. 1, 8, 9, 45

[29] J. N. Laneman, D. N. C. Tse, and G. W. Wornell. Distributed space-time-coded protocols for exploiting cooperative diversity in wireless netowrks. *IEEE Trans. Inf. Theory*, 49(10):2415–2425, October 2004. 3

[30] P. Larsson and H. Rong. Large-scale cooperative relay network with optimal coherent combining under aggregate relay power cosntraints. In *Proc. of Working Group 4, WWRF8 Meeting*, February 2004. 5

[31] Xiaojun Lin and Ness Shroff. On the fundamental relationship between the achievable capacity and delay in mobile wireless networks. *Adv. Pervasive Computing and Networking*, pages 17–55, July 2005. 9

[32] I. Maric and R. D. Yates. Bandwidth and power allocation for cooperative strategies in Gaussian relay networks. In *Asilomar Conf. Signals, Syst., Computers*, November 2004. 5

[33] D. S. Michalopoulos, G. K. Karagiannidis, T. A. Tsiftsis, and R. K. Mallik. An optimized user selection method for cooperative diversity systems. In *Proc. of IEEE Globecom*, Nov./Dec. 2006. 3, 5

[34] D. S. Michalopoulos, H. A. Suraweera, G. K. Karagiannidis, and R. Schober. Amplify-and-Forward relay selection with outdated channel state information. In *Proc. of IEEE Globecom*, December 2010. 6

[35] V. I. Morgenshtern and H. Bölcskei. Crystallization in large wireless networks. *IEEE Trans. Inf. Theory*, 53(10):3319–3349, October 2007. 7, 8

[36] M. Nakagami. The m-distribution - a general formula of intensity distribution of rapid fading. *Statistical Methods in Radio Wave Propagation*, pages 3–36, (W. G. Hoffman ed.), Pergamon Press, 1960. 47

[37] M. J. Neely and E. Modiano. Capacity and delay tradeoffs for ad hoc mobile networks. *IEEE Trans. Inf. Theory*, 51(6):1917–1937, June 2005. 9, 64, 65

[38] Özgür, O. A. Leveque, and D. N. C. Tse. Hierarchical cooperation achieves optimal capacity scaling in ad hoc networks. *IEEE Trans. Inf. Theory*, 53(10):3549–3572, October 2007. 7, 8

[39] J. G. Proakis. *Digital Communications*. McGraw-Hill, Inc, New York, 4th edition, 2001. 1

[40] A. Ribeiro, X. Cai, and G. B. Giannakis. Symbol error probabilities for general cooperative links. *IEEE Trans. Wireless Commun.*, 4:1264–1273, May 2005. 4, 20, 41, 91

[41] A. K. Sadek, Z. Han, and K. J. R. Liu. A distributed relay-assignment algorithm for cooperative communications in wireless networks. In *Proc. of IEEE ICC*, 2006. 3, 4

[42] A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity, Part I and II. *IEEE Trans. Commun.*, 51(11):1927–1948, November 2003. 1, 8

[43] M. Sharif and B. Hassibi. On the capacity of the MIMO broadcast channels with partial side information. *IEEE Trans. Inf. Theory*, 51(2):506–522, February 2005. 51, 98, 99

[44] M. K. Simon and M-S. Alouini. *Digital Communications over Fading Channels*. John Wiley & Sons, Hoboken, NJ, second edition, 2005. 27, 96

[45] V. Sreng, H. Yanikomeroglu, and D. D. Falconer. Relay selection strategies in cellular networks with peer-to-peer relaying. In *Proc. of IEEE VTC*, 2003. 4, 19, 41

[46] A. I. Sulyman and M. Kousa. Bit error rate performance of a generalized diversity selection combining scheme in Nakagami fading channels. In *Proc. of IEEE WCNC*, sep. 2000. 22

[47] S. Toumpis and A. Goldsmith. Large wireless networks under fading mobility, and delay constraints. In *IEEE INFOCOM*, November 2004. 9, 10, 65

[48] J. L. Vicario, A. Bel, J. A. Lopez-Salcedo, and G. Seco. Opportunistic relay selection with outdated CSI: Outage probability and diversity analysis. *IEEE Trans. Wireless Commun.*, 8(6):2872–2876, June 2009. 5, 6, 20, 21, 27, 35, 90, 95

[49] J.L. Vicario, A. Bel, A. Morell, and G. Seco-Granados. A robust relay selection strategy for cooperative systems with outdated CSI. In *Proc. of IEEE VTC*, 2009. 6, 19, 21, 41, 90

[50] Y. Wang, S. Cui, and A. M. Haimovich. Delay-throughput trade-off with opportunistic relaying in wireless networks. In *Proc. of IEEE Globecom*, December 2011. 9, 64, 66, 76

[51] Z. Wang and G. B. Giannakis. A simple and general parameterization quantifying performance in fading channels. *IEEE Trans. Commun.*, 51:1389–1398, August 2003. 95, 96

[52] R. W. Wolff. *Stochastic modeling and the theory of queues*. Prentice-Hall, 3rd edition, 1989. 108

[53] L. Xiao and X. Dong. Unified analysis of generalized selection combining with normalized threshold test per branch. *IEEE Trans. Wireless Commun.*, 5(8):2153–2163, September 2006. 23, 92

[54] F. Xue and P. R. Kumar. Scaling laws for ad hoc wireless networks: An information theoretic approach. *Foundations and Trends in Networking*, 1(2):145–270, 2006. 7

[55] W. Zhang, X. Xiang-Gen, and K. B. Letaief. Space-time/frequency coding for mimo-ofdm in next generation broadband wireless systems. *IEEE Trans. Wireless Commun.*, 14(3):32–43, June 2007. 1

[56] B. Zhao and M. C. Valent. Practical relay networks: a generalization of hybrid-ARQ. *IEEE J. Sel. Areas Commun.*, 23(1):7–18, January 2005. 6

[57] Y. Zhao, R. Adve, and T. J. Lim. Symbol error rate of selection amplify-and-forward relay systems. *IEEE Trans. Commun.*, 10:757–759, November 2006. 17, 18, 41

[58] Y. Zhao, R. Adve, and T. J. Lim. Improving amplify-and-forward relay networks: Optimal power allocation versus selection. *IEEE Trans. Wireless Commun.*, 6(8):3114–3123, August 2007. 3, 4, 17, 18, 33, 35, 38, 40, 41

[59] L. Zheng and D. Tse. Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels. *IEEE Trans. Inf. Theory*, 49(5):1073–1096, May 2003. 26

Appendix A

Proof of Theorem 1

Proof of Unit Diversity Order for ORS

To prove the unit diversity order of ORS with outdated CSI, we first point out that according to our proposed N+NT-ORS, we can easily see that N+NT-ORS can be transformed to ORS when $N = 1$ and $\rho = 1$. Then, according to Proposition 1 in Sec. 2.4.2 and Proposition 2 in 2.4.3, the diversity order of N+NT-ORS equals to N under outdated CSI for both AaF and DaF systems. Consequently, the diversity order of ORS is equal to 1 with outdated CSI for both AaF and DaF systems. It is worth noting that the unit diversity order of ORS with outdated CSI is also proven in [48, Corollay 1] for DaF but not for AaF.

Proof of Unit Diversity Order for MORS

The MORS generates an estimate of the instantaneous CSI for each relay and selects the relay with the best estimated CSI [49] that maximizes the estimator metric $\mathcal{E}(\gamma_{k,D}|\hat{\gamma}_{k,D}, \bar{\gamma}) = \rho^2\hat{\gamma}_{k,D} + (1+\rho^2)\bar{\gamma}$. Notice that ρ and $\bar{\gamma}$ are constant and it assumes i.i.d. channels in [49]. Thus, according to the assumption of $0 < \rho < 1$, it can be

readily seen that selecting the relay with the best estimated CSI is equivalent to ORS which selects the relay with the best outdated CSI. Therefore, Proposition 1 stands for MORS and the diversity order of MORS reduces to 1.

Proof of Unit Diversity Order for BW CS

In the presence of outdated CSI, the selection function for the BW CS is $\mathcal{G}_{\text{AaF}}^{\text{BW}}(k) = \min(\hat{\gamma}_{S,k}, \hat{\gamma}_{k,D})$. Let $\hat{\gamma}_k = \min(\hat{\gamma}_{S,k}, \hat{\gamma}_{k,D})$. Therefore, the BW CS is equal to selecting the best relay with respect to $\hat{\gamma}_k$. In the process of obtaining Proposition 1, we have shown that when the resulting SNR of each relay equals to $\hat{\gamma}_k$, the diversity of N+NT-ORS under AaF relaying equals to N. Letting $N = 1$, we are selecting a single relay with the best $\hat{\gamma}_k$, which is identical to BW CS. Therefore, according to Proposition 1, BW CS achieves unit diversity order.

Proof of Unit Diversity Order for BHM Selection

The selection function for the best harmonic mean selection is $\mathcal{G}_{\text{AaF}}^{\text{BHM}}(k) = \frac{1}{\gamma_{S,k}^{-1}+\gamma_{k,D}^{-1}}$ [4, 40]. The following well-known inequality is again exploited, i.e. [22],

$$\frac{1}{\gamma_{S,k}^{-1} + \gamma_{k,D}^{-1}} = \frac{\gamma_{S,k}\gamma_{k,D}}{\gamma_{S,k} + \gamma_{k,D}} \leq \gamma_k = \min\left(\gamma_{S,k}, \gamma_{k,D}\right) . \tag{A.1}$$

The rest of the proof follows that of Appendix A-C.

Appendix B

Proof of Theorem 2

In order to obtain the MGF of $\gamma_{\rm ub}$ using N+NT-ORS, we exploit a probability space partition approach proposed in [53]. Define a set $\mathcal{A}_i$, $i = 1, \cdots, K$, as the collection of all K-tuples $(\hat{\gamma}_1, \hat{\gamma}_2, \cdots, \hat{\gamma}_K)$ where $\hat{\gamma}_i$ is the N-th largest, i.e. $\hat{\gamma}^{(N)}$. It is given by

$$\mathcal{A}_i := \left\{(\hat{\gamma}_1, \hat{\gamma}_2, \cdots, \hat{\gamma}_K) \,|\, \exists\, |I| = N-1, \text{ s.t. } k \in I \Leftrightarrow \hat{\gamma}_k > \hat{\gamma}_i \text{ or } (\hat{\gamma}_k = \hat{\gamma}_i \text{ and } k > i)\right\}, \tag{B.1}$$

where $I \subset \{1, 2, \cdots, K\} - \{i\}$ is an index set which contains the indices of those relays that have individual end-to-end SNR upperbound larger than $\hat{\gamma}_i$.

Following the analysis in [53], the MGF of the SNR $\gamma_{\rm ub}$ of the N+NT-ORS scheme with outdated CSI in AaF relaying can be expressed as

$$\Phi_{\gamma_{\rm ub}}^{\rm AaF}(s) = \sum_{i=1}^{K} \int_0^\infty \sum_{\substack{\text{all I} \\ |I|=N-1}} \underbrace{\int_0^\infty \cdots \int_0^\infty}_{K} \underbrace{\int_0^{\hat{\gamma}_i} \cdots \int_0^{\hat{\gamma}_i}}_{K-N} \underbrace{\int_{\hat{\gamma}_i}^\infty \cdots \int_{\hat{\gamma}_i}^\infty}_{N-1} \prod_{k=1}^{K} \left[e^{sT(\hat{\gamma}_k)} f_k(\gamma_k|\hat{\gamma}_k) f_k(\hat{\gamma}_k)\right]$$

$$d\hat{\gamma}_{k_1} \cdots d\hat{\gamma}_{k_{N-1}} d\hat{\gamma}_{k'_1} \cdots d\hat{\gamma}_{k'_{K-N}} d\gamma_1 \cdots d\gamma_K d\hat{\gamma}_i\,, \tag{B.2}$$

where $k_l \in I$ is the index of a relay with end-to-end SNR upperbound larger than $\hat{\gamma}_i$ and $k'_l \in \{1, 2, \cdots, K\} - \{i\} - I$ is the index of a relay with end-to-end SNR

upperbound less than $\hat{\gamma}_i$. The number of all possible I's is $\binom{K-1}{N-1}$. Substituting (2.10) and (2.13) into (B.2), and using [18, (6.614.3), (9.220.2) and (9.215.1)], Theorem 2 is obtained.

Appendix C

Proof of Theorem 3

The proof follows that of Theorem 2 except that the testing function $T(\hat{\gamma}_k)$ is replaced by the on-off indicator $W(\hat{\gamma}_k)$ as

$$W(\hat{\gamma}_k) = \begin{cases} 0, & \hat{\gamma}_k < \mu\hat{\gamma}^{(N)} \\ 1, & \hat{\gamma}_k \geq \mu\hat{\gamma}^{(N)} \end{cases} . \tag{C.1}$$

Appendix D

Proof of Proposition 1

For high SNR $\bar{\gamma} \gg 1$, we can obtain an approximation of $\Phi_{\gamma_{\mathrm{ub}}}^{\mathrm{AaF}}(s)$ as

$$\Phi_{\gamma_{\mathrm{ub}}}^{\mathrm{AaF}}(s) = K\binom{K-1}{N-1}\sum_{k=0}^{K-N}\sum_{q=0}^{K-N-k}\sum_{m=0}^{k}\left(\frac{1}{-s\bar{\gamma}/2}\right)^{K-1-k}\frac{(K-N)!(-1)^{q+m}}{(K-N-k-q)!m!q!(k-m)!} \frac{1}{\left(-\frac{s\bar{\gamma}}{2}\right)\left[(K-N-k-q)\mu+(q+N)\right]+m\mu\left[-\frac{s\bar{\gamma}}{2}(1-\rho^2)\right]}. \tag{D.1}$$

In [51, Proposition 3], it is stated that the diversity order is equal to the order of the MGF's pole. From (D.1), it can be seen that the asymptotic diversity order is N for both cases $\rho \to 1$ and $\rho \to 0$. Applying the following equation used in [48], i.e.,

$$d_{\rho\to 0} \leq d_{\rho'} \leq d_{\rho\to 1}\,, \tag{D.2}$$

where ρ' is an arbitrary value of ρ, Proposition 1 is obtained.

Appendix E

Proof of Proposition 2

For high SNR $\bar{\gamma} \gg 1$, we can obtain an approximation of $\Phi_\gamma(\gamma|\,|\mathcal{DS}| = L > N)$ as shown in (D.1) through replacing K by L.

By applying [51, Proposition 3], it is known that the asymptotic diversity order of $F_\gamma(\gamma|\,|\mathcal{DS}| = L > N) = \Pr(\text{outage}||\mathcal{DS}| = L > N)$ is N for both cases $\rho \to 1$ and $\rho \to 0$. To facilitate the analysis, we represent $\Pr(\text{outage}||\mathcal{DS}| = L > N)$ as $G_1(\rho)(y/\bar{\gamma})^N + o((y/\bar{\gamma})^N)$, where $G_1(\rho)$ is a scaling factor that depends on ρ and $y = 2^{2R}-1$. Similarly, $\Pr(\text{outage}||\mathcal{DS}| = L \leq N)$ can be represented as $G_2(\rho)(y/\bar{\gamma})^L + o((y/\bar{\gamma})^L)$ since maximal ratio combining reaches full diversity [44], where $G_2(\rho)$ is a scaling factor that depends on ρ.

Now, the outage probability of N+NT-ORS in DaF relaying can be expressed as

$$P_{out}^{\text{DaF}}(y) = \left(\frac{y}{\bar{\gamma}}\right)^K + \sum_{L=1}^{N}\left[G_2(\rho)\binom{N}{L}\left(\frac{y}{\bar{\gamma}}\right)^N\left(1-\frac{y}{\bar{\gamma}}\right)^L + o\left(\left(\frac{y}{\bar{\gamma}}\right)^N\right)\right] + \sum_{L=N+1}^{K}\left[G_1(\rho)\binom{K}{L}\left(\frac{y}{\bar{\gamma}}\right)^{K-L+N}\left(1-\frac{y}{\bar{\gamma}}\right)^L + o\left(\left(\frac{y}{\bar{\gamma}}\right)^{K-L+N}\right)\right]. \tag{E.1}$$

In deriving (E.1), the following equality is exploited

$$\left(1-e^{-\frac{y}{\bar{\gamma}}}\right)^K = \left(\frac{y}{\bar{\gamma}}\right)^K + o\left(\left(\frac{1}{\bar{\gamma}}\right)^K\right) .$$

By considering that $1-\frac{y}{\bar{\gamma}} \approx 1$, it shows that the first summation in (D.1) depends on $(1/\bar{\gamma})^N$ and the predominant term of the second summation, which is the term $L = K$, also depends on $(1/\bar{\gamma})^N$. Therefore, the asymptotic diversity order is N for both cases $\rho \to 1$ and $\rho \to 0$. By applying (D.2), Proposition 2 holds.

Appendix F

Proof of Lemma 2

In this appendix, we prove a property of the maximum among $\bar{n}$ RVs $\{x_1, x_2, \ldots, x_{\bar{n}}\}$ stated in Lemma 2. Let $f_X(x)$ and $F_X(x)$ be the pdf and cdf of x_i, the growth function is defined as $g_X(x) = \frac{1-F_X(x)}{f_X(x)}$ [43]. For the ease of illustration, we define μ_n to be the unique solution to $1 - F_X(\mu_n) = \frac{1}{n}$ and restate Corollary A.1 in [43] below

Corollary 1. *If* $\mu_n = O(\log n)$ *and* $g(x)$ *satisfies* $\lim_{x\to\infty} g(x) = c > 0$ *and* $g^{(m)}(\mu_n) = O(1/\mu_n^m)$*, then*

$$\Pr\left[\mu_n - c\log\log n \le \max x_i \le \mu_n + c\log\log n\right] \ge 1 - O\left(\frac{1}{\log n}\right). \tag{F.1}$$

In what follows, we show that x_i, a Gamma distributed RV $\Gamma(m, 1/m)$, meets all the three conditions stated in the above corollary. We start by showing $\mu_n = O(\log n)$. To find μ_n, we apply the asymptotic expansion of the incomplete Gamma function given by [18, 8.257], and we have

$$1 - F_X(\mu_n) = \gamma(m, m\mu_n) = (m\mu_n)^{m-1} e^{-m\mu_n}(1 + O(1/\mu_n)) = \frac{1}{n}, \tag{F.2}$$

based on which, μ_n can be obtained as

$$\mu_n = \frac{1}{m}\log n + \frac{(m-1)}{m}\log\log n + O(\log\log\log n), \tag{F.3}$$

implying $\mu_n = O(\log n)$.

Next, the growth function $g_X(x)$ of a Gamma distributed RV can be expressed as

$$g_X(x) = \frac{1 - F_X(x)}{f_X(x)} = \frac{\Gamma(m, mx)}{m^m x^{m-1} e^{-mx}} \to \frac{1}{m} \text{ as } x \to \infty \tag{F.4}$$

where we have used the following property of Gamma function

$$\frac{\Gamma(s, x)}{x^{s-1}e^{-x}} \to 1 \text{ as } x \to \infty, \tag{F.5}$$

implying $g_X^{(m)}(\mu_n) = 0 = O(1/\mu_n^m)$. Therefore, from Corollary A.1 in [43], we have

$$\Pr\left\{\max x_i \leq \frac{1}{m}\log n + \frac{m-2}{m}\log\log n\right\} \to 0 \text{ as } n \to \infty. \tag{F.6}$$

The lemma is established.

Appendix G

Proof of Theorem 6

In this appendix, we prove the throughput scaling law results established given by Theorem 6 in Section 3.3.3. The proof follows similar steps as that of [9, Section IV] and consists of three basic steps. We first show that in order for the first hop to maintain linear throughput increase, the maximum number of relays cannot be greater than k^*. Next, we demonstrate that the corresponding limitation placed on the second hop is loosened with a value larger than k^*. Finally, since the end-to-end throughout for DF systems are governed by $\frac{1}{2}\min\{R_1, R_2\}$, we conclude that the system throughput is limited by the first hop and the maximum number of concurrent transmission supported is therefore k^*.

Let us begin by deriving k^* in the first hop. We know from Lemma 2 that for all $s \in \big(0, \frac{1}{m}\log\bar{n}$ $+ \frac{m-2}{m}\log\log\bar{n}\big]$, both terms $\frac{n(n-1)\cdots(n-k+1)}{n^k}$ and $\left[1-\left(1-m^m e^{-ms}\frac{s^{m-1}}{\Gamma(m)}\right)^{\bar{n}}\right]$ corresponding to $\overline{F}_{X^{(0)}}(x)$ in (3.15) approach 1 as $n \to \infty$. Therefore, as long as the other term $F_Y(s-\frac{1}{P_R}) = \Theta(1)$, linear throughput k can be guaranteed for the first hop. To establish the property of $F_Y(s-\frac{1}{P_R})$, let us consider the exemplary case of $s = \frac{1}{m}\log\bar{n} + \frac{m-2}{m}\log\log\bar{n}$ and assume for the moment that $k = \frac{1}{m}\log\bar{n}$. We then

have

$$F_Y\left(s_0-\frac{1}{P}\right)=F_Y\left(\frac{1}{m}\log\bar{n}+\frac{m-2}{m}\log\log\bar{n}-\frac{1}{P}\right)$$
$$\approx F_Y\left(\frac{1}{m}\log\bar{n}\right)=\frac{1}{2}, \tag{G.1}$$

where the approximation assumes $1/P$ can be neglected which is valid when a typical amount of power is supplied ($P>1$), and the last equality follows from a combination of the Central Limit Theorem, whereby the interference term Y in (3.11) can be approximated by a Gaussian RV with with mean and variance equal to $\frac{1}{m}\log\bar{n}$, and the symmetry of the Gaussian distribution. Therefore, we show that as long as $k=\frac{1}{m}\log\bar{n}$, OPS achieves a linear throughput of $k/2$ in the first hop. For positive integer value of m, the closed-form solution to the equation $k=\frac{1}{m}\log\bar{n}$, denoted by k^*, can be derived as

$$k^*=\begin{cases}\frac{\log 2n}{1+\log 2}, & m=1\\ -\frac{1}{m}\mathcal{W}_{-1}\left(\frac{-c}{n}\right), & m>1, m\in\mathbb{Z}^+\end{cases}, \tag{G.2}$$

where c is a constant for fixed transmission power P and Nakagami fading parameter m, given by

$$c=\frac{2^m}{m(1-m+m\log 2)}\exp\left\{\frac{-2^m e^{m/P}+m^2(1+2P+m(1+P)(\log 2-1))}{mP(1-m+m\log 2)}\right\}. \tag{G.3}$$

In step 2, we show that the maximum achievable second-hop throughput is larger than that of the first hop. We rewrite the throughput expression derived in (3.8) for the ease of presentation

$$R_2=k\Big(1-\big(F(1)\big)^n\Big). \tag{G.4}$$

Consider then the exemplary case of $k=\frac{\frac{1}{m}\log n+\frac{m-2}{m}m-\frac{1}{P}}{\log 2}+1$. After some algebra, we

have

$$\overline{F}(1) = \frac{1}{(2\log 2)^{m-1}\Gamma(m)}\frac{\log n}{n}. \tag{G.5}$$

Let $d = \frac{1}{(2\log 2)^{m-1}\Gamma(m)}$. Then we obtain

$$\begin{aligned}(F(1))^n &= \left(1 - d\frac{\log n}{n}\right)^n = e^{n\log(1-d\frac{\log n}{n})} \\ &= e^{-d\log n + O\left(\frac{\log^2 n}{n}\right)} \\ &= O\left(\frac{1}{n^d}\right). \end{aligned} \tag{G.6}$$

From the expression of $F(x)$ given in (3.4), it can be observed that $F(1)$ is an increasing function of k. Thus, when $k \leq \frac{\frac{1}{m}\log n + \frac{m-2}{m}\log\log n - \frac{1}{P}}{\log 2} + 1$, we have $(F(1))^n \to 0$ as $n \to \infty$, which implies from (G.4) that $R_2 \to k$. In other words, the second-hop throughput R_2 maintains linearity in k when $k \leq \frac{\frac{1}{m}\log n + \frac{m-2}{m}\log\log n - \frac{1}{P}}{\log 2} + 1$.

Finally, we prove $k^* < \frac{\frac{1}{m}\log n + \frac{m-2}{m}\log\log n - \frac{1}{P}}{\log 2} + 1$. Given that k^* is the solution to the equation $k = \frac{1}{m}\log\Big(n\big(1 - F(1)\big)\Big)$. With a large n and a typical P ($P > 1$), we have

$$\begin{aligned} k^* &= \frac{1}{m}\log\Big(n\big(1 - F(1)\big)\Big) \\ &\leq \frac{1}{m}\log n \\ &\leq \frac{\frac{1}{m}\log n}{\log 2} + 1 \qquad \text{(G.7)} \\ &\approx \frac{\frac{1}{m}\log n + \frac{m-2}{m}\log\log n - \frac{1}{P}}{\log 2} + 1, \qquad \text{(G.8)} \end{aligned}$$

which establishes the fact that the maximum achievable system throughput is limited by the first hop.

Finally, the maximum achievable system throughput is given by

$$R = \frac{1}{2}\min\{R_1, R_2\} = \frac{k^*}{4} = \begin{cases} \frac{\log 2n}{4(1+\log 2)}, & m = 1 \\ -\frac{1}{4m}\mathcal{W}_{-1}\left(\frac{-c}{n}\right), & m > 1, m \in \mathbb{Z}^+ \end{cases}. \tag{G.9}$$

Correspondingly, we can obtain the throughput scaling. For Rayleigh fading, OPS achieves a throughput scaling of $\Theta(\log n)$. For Nakagami fading with parameter $m > 1$ taking integer values, OPS achieves a throughput scaling of $\Theta\big(-\mathcal{W}_{-1}(-n^{-\frac{1}{m-1}})\big)$.

Appendix H

Proof of Lemma 4

In this appendix, we prove an order statistics result for the maximum of L out of n Gamma RVs given by Lemma 4 in Section 3.4.1. From Section 3.4, we know $F_{X^{(2)}}(x) = (F_W(x))^i$ where

$$F_W(x) = \frac{1}{L}\sum_{j=0}^{L}\left[\left(1-\frac{\gamma(m,mx)}{\Gamma(m)}\right)^{n-j}\left(\frac{\gamma(m,mx)}{\Gamma(m)}\right)^{j}\right]. \tag{H.1}$$

To prove this lemma, it amounts to show that the critical value s_2 for $F_W(s_2) \to 0$ locates in between s_0 and s_1, i.e., $s_0 \le s_2 \le s_1$, since $F_{X^{(2)}}(s_2) \to 0$ can be readily obtained given $F_W(s_2) \to 0$.

We first prove $s_2 \le s_1$. From Appendix A, we know that $F_{X^{(1)}}(s_1) = \left(1 - m^m e^{-ms_1}\frac{s_1^{m-1}}{\Gamma(m)}\right)^n \to 0$ with $s_1 = \frac{1}{m}\log n + \frac{m-2}{m}\log\log n$ and $F_{X^{(0)}}(s_0) = \left(1 - m^m e^{-ms_0}\frac{s_0^{m-1}}{\Gamma(m)}\right)^{\bar{n}} \to 0$ with $s_0 = \frac{1}{m}\log \bar{n} + \frac{m-2}{m}\log\log \bar{n}$.

$$\begin{aligned} F_W(s_2) &= \frac{1}{L}\left[F_{X_{n-L+1:n}}(s_2) + F_{X_{n-L+2:n}}(s_2)\cdots + F_{X_{n:n}}(s_2)\right] \\ &\ge \frac{1}{L}\left[F_{X_{n:n}}(s_2) + F_{X_{n:n}}(s_2)\cdots + F_{X_{n:n}}(s_2)\right] \\ &= F_{X_{n:n}}(s_2) = F_{X^{(1)}}(s_2). \end{aligned} \tag{H.2}$$

Thus, from $F_{X^{(1)}}(s_2) \leq F_W(s_2) \to 0$ and $F_{X^{(1)}}(s_1) \to 0$, it follows $s_2 \leq s_1$.

Now, we prove $s_2 \geq s_0$.

$$\begin{aligned} F_W(s_2) &= \frac{1}{L}\left[F_{X_{n-L+1:n}}(s_2) + F_{X_{n-L+2:n}}(s_2) \cdots + F_{X_{n:n}}(s_2)\right] \\ &\leq \frac{1}{L}\left[F_{X_{n-L+1:n}}(s_2) + F_{X_{n-L+1:n}}(s_2) \cdots + F_{X_{n-L+1:n}}(s_2)\right] \\ &= F_{X_{n-L+1:n}}(s_2)\,. \end{aligned} \tag{H.3}$$

We now focus on proving $F_{X_{n-L+1:n}}(s_2) \leq F_{X^{(0)}}(s_2)$.

$$\begin{aligned} \frac{F_{X_{n-L+1:n}}(s_2)}{F_{X^{(0)}}(s_2)} &= \frac{F_X^{n-L+1}(s_2)\left(1-F_X(s_2)\right)^{L-1}}{F_X^{\bar{n}}(s_2)} \\ &= \frac{\left(1-F_X(s_2)\right)^{L-1}}{F_X^{L-1}(s_2)} \cdot \frac{F_X^{n}(s_2)}{F_X^{\bar{n}}(s_2)} \\ &\leq \left(\frac{\left(1-F_X(s_2)\right)}{F_X(s_2)}\right)^{L-1} \cdot F_X^{n-\bar{n}}(s_2) \\ &< \left(\frac{\left(1-F_X(s_2)\right)}{F_X(s_2)}\right)^{L-1}\,. \end{aligned} \tag{H.4}$$

As defined in Appendix A, $f_X(x)$ and $F_X(x)$ are the pdf and cdf of x_i, a Gamma distributed RV $\Gamma(m, 1/m)$.

Thus, if $\frac{(1-F_X(s_2))}{F_X(s_2)} \leq 1$, we have $F_{X_{n-L+1:n}}(s_2) \leq F_{X^{(0)}}(s_2)$.

$$\begin{aligned} \frac{\left(1-F_X(s_2)\right)}{F_X(s_2)} \leq 1 &\Leftrightarrow F_X(s_2) \geq \frac{1}{2} \\ &\Leftrightarrow s_2 \geq \log 2\,. \end{aligned} \tag{H.5}$$

The above condition $s_2 \geq \log 2$ holds for most cases. Therefore, $F_{X_{n-L+1:n}}(s_2) \leq F_{X^{(0)}}(s_2)$. And from (H.3), we have $F_W(s_2) \leq F_{X^{(0)}}(s_2)$. Together with the fact that $F_W(s_2) \to 0$ and $F_{X^{(0)}}(s_0) \to 0$, we obtain $s_2 \geq s_0$. This completes the proof.

Appendix I

Proof of Theorem 9

In this appendix, we derive the closed-form delay expression for L-scheduling given by Theorem 9 in Section 3.5.1. We first derive a delay upper bound for the first hop. From Section 3.4, we know that source node i is scheduled by j relays with probability 1 and by $k-j$ relays with probability $1/n$, implying a service rate of $\mu_j = j + (k-j)/n$. Accordingly, the source node is modeled as a M/G/1 queue with Poisson arrival rate λ and Bernoulli service rate μ_j. Define $\rho = \lambda/\mu_0$, and assume $0 < \rho < 1$. Thus $\rho_j = \lambda/\mu_j = \lambda/\left(j + (k-j)/n\right)$, $\rho_0 = \rho$, and $0 < \rho_j < 1$. Applying results for the M/G/1 queue [19, p.212], the average number of packets in the queue is given by

$$\begin{aligned} M_{H1} &= \left(\frac{\rho_j}{1-\rho_j}\right)\left[1 - \frac{\rho_j}{2}(1-\mu_j^2\sigma_j^2)\right] \\ &= \frac{\rho_j}{1-\rho_j} + \frac{\rho_j^2\mu_j^3}{2(1-\rho_j)} - \frac{\rho_j^2}{2(1-\rho_j)} - \frac{\rho_j^2\mu_j^4}{2(1-\rho_j)}, \end{aligned} \tag{I.1}$$

where the second equality follows from the property of the Bernoulli distribution, i.e., $\sigma_j^2 = \mu_j - \mu_j^2$. By Little's Law, we have

$$D_1(j) = \frac{M_{H1}}{\lambda} \leq \frac{1}{\mu_j} \frac{2-\rho_j}{2(1-\rho_j)}. \tag{I.2}$$

Therefore, the delay in the first hop is upper-bounded by

$$D_1 \leq \sum_{j=0}^{k} \binom{k}{j} (1-p)^{k-j} \frac{1}{n^j} \frac{1}{j+(k-j)/n} \frac{2-\rho_j}{2(1-\rho_j)}. \tag{I.3}$$

The case of $j = 0$ corresponds to the event that the i-th S-D pair is scheduled independently by all the relays. As can be readily observed from (I.3), this case contributes the largest amount of delay and therefore dominates all the other terms. Therefore, we approximate the upper bound by this dominant term where $j = 0$[1]. Hence, we have

$$\begin{aligned} D_1 &\leq (1-p)^k \frac{n}{k} \frac{2-\rho}{2(1-\rho)} \\ &\leq (1-kp) \frac{n}{k} \frac{2-\rho}{2(1-\rho)} \\ &= \left(\frac{n}{k} - L\overline{F}(1)\right) \frac{2-\rho}{2(1-\rho)}. \end{aligned} \tag{I.4}$$

The last equality is due to $p = \frac{\bar{n}L}{n^2} = \frac{LF(1)}{n}$.

We now derive the delay upper bound for the second hop. The queue in each relay can be modeled as a G/G/1 queue. The arrival rate is $\tilde{\lambda} = \lambda/k$ and the average service rate is $\tilde{\mu} = (j + (k-j)/n)/k$. Let σ_1^2 and σ_2^2 be the variances of arrival distribution and service distribution, respectively. According to Kingman's upper

[1] Although it is not mathematically strict to approximate an upper bound by one of the involving terms, such approximation is reasonable since the total number of terms k is mostly at the magnitude of $\log n$ and is verified by extensive simulations.

bound for G/G/1 queues [52, p.476], we have

$$\begin{aligned} D_2(j) &\leq \frac{\tilde{\lambda}^2\left(\sigma_1^2+\sigma_2^2\right)}{2\left(1-\rho_j^2\right)}+\frac{1}{\tilde{\mu}} \\ &= \frac{1}{\tilde{\mu}}+\frac{\tilde{\mu}^2\rho_j(1+\rho_j)}{2(1-\rho_j)}-\frac{\tilde{\mu}^3\rho_j^2(1+\rho_j^2)}{2(1-\rho_j)} \end{aligned} \tag{I.5}$$

Following the same steps in obtaining (I.4), the delay in the second hop can be expressed as

$$D_2 \leq n - kL\overline{F}(1)\,. \tag{I.6}$$

This completes the proof.