

**OVERHEAD, UNCERTAINTY, AND INTERFERENCE IN WIRELESS
NETWORKS**

by
Hao Feng

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Electrical and Computer Engineering

Fall 2014

© 2014 Hao Feng
All Rights Reserved

UMI Number: 3685109

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI 3685109

Published by ProQuest LLC (2015). Copyright in the Dissertation held by the Author.

Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against unauthorized copying under Title 17, United States Code



ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 - 1346

**OVERHEAD, UNCERTAINTY, AND INTERFERENCE IN WIRELESS
NETWORKS**

by

Hao Feng

Approved: _____

Kenneth E. Barner, Ph.D.
Chair of the Department of Electrical and Computer Engineering

Approved: _____

Babatunde Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: _____

James G. Richards, Ph.D.
Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Leonard J. Cimini, Jr., Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Kenneth E. Barner, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Javier Garcia-Frias, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Chien-Chung Shen, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Steven Weber, Ph.D.

Member of dissertation committee

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the support of many people. I would like to express my sincere appreciation to those who have provided a lot of help in carrying out my research and made my graduate study the most rewarding experience in my life.

I would like to start by thanking my advisor, Prof. Leonard J. Cimini, Jr., for his invaluable support and endless encouragement throughout my graduate career. Prof. Cimini has been an example of an excellent professor, wonderful advisor, and an efficient manager, and I am grateful for having the chance to be his student. During the years of my Ph.D. studies, his deep knowledge and insight to the research helped me tremendously in developing precious research skills in approaching and analyzing complicated problems. He has also shown great care and concern for his students, and has always helped them advance and become better in their personal and professional lives.

I would like to thank my defense committee members, Dr. Kenneth E. Barner, Dr. Javier Garcia-Frias, Dr. Chien-Chung Shen, and Dr. Steven Weber, both for laying the technical foundation several years ago for me to pursue this dissertation and for taking the time to serve on my committee. Without their knowledge and assistance, this dissertation would not have been successful.

A special thank is due to the friends who I have enjoyed working and developing friendships with during my time at University of Delaware. I cannot name all of them; however, I have to mention a few: Qi Wang, Hongzheng Wang, Yao Xiao, Yang Guan, Gubong Lim, Chenzi Jiang, Bohan Zhang, Guangyi Liu, Li Li, Haina Ye, Chengming Zhou, and Jing Li. Their helpful discussions and friendship have greatly facilitated my study and research.

Lastly, and most importantly, I would like to thank my family, Guanghui Feng and Yanfei Zhang, for their selfless love and unconditional support throughout my life, especially in the past few years when I was pursuing my PhD overseas. I dedicate this dissertation to them.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xi
GLOSSARY	xvi
ABSTRACT	xix
 Chapter	
1 OVERVIEW	1
1.1 Background and Motivation	1
1.1.1 Overhead	1
1.1.2 Uncertainty	3
1.1.3 Interference	4
1.2 Dissertation Outline	5
2 TWO-HOP DECODE-AND-FORWARD COOPERATIVE RELAYING NETWORKS	9
2.1 Introduction	9
2.2 Decentralized Cooperative Strategy with Low Overhead	11
2.2.1 Optimization Formulation	13
2.2.2 Optimality of M -group	16
2.2.3 Simulation Results	20
2.3 Overhead-Performance Tradeoff for Best Relay Selection	24
2.3.1 Rate Distortion Formulation	26
2.3.2 Overhead-Performance Tradeoff	29

2.3.3	Simulation Results	34
2.4	Spectral Efficiency of Centralized and Decentralized Cooperative Networks with Relay Selection	39
2.4.1	Centralized Feedback-Based Relay Selection (FBRS)	40
2.4.2	Decentralized Timer-Based Relay Selection (TBRS)	42
2.4.3	Simulation Results	48
2.5	Summary	53
3	MULTI-USER POISSON NETWORK WITH COOPERATION	55
3.1	System Model	57
3.1.1	Non-Cooperative Strategy	58
3.1.2	Cooperative Strategy	59
3.2	Performance Analysis	62
3.2.1	Non-Cooperative Strategy	62
3.2.2	Cooperative Strategy	63
3.2.3	To Cooperate or Not to Cooperate	70
3.3	Simulation Results	74
3.3.1	Validation of Analysis	74
3.3.2	To Cooperate or Not to Cooperate	75
3.4	Summary	81
4	MULTI-HOP LINEAR WIRELESS NETWORKS	84
4.1	Multi-hop Linear Network with Randomly Located Nodes	84
4.1.1	System Model	86
4.1.2	Analysis	88
4.1.3	Simulation Results	91
4.2	Relay Deployment in Multi-Hop Linear Network with Cooperation	95
4.2.1	System Model	96
4.2.2	Optimum Relay Deployment Strategy	98

4.2.3	Simulation Results	103
4.3	Summary	106
5	MULTI-USER DOWNLINK NETWORKS	109
5.1	Introduction and System Model	109
5.2	Quantization Techniques for Downlink MU-MIMO	113
5.2.1	Quantization Methods	115
5.2.2	Simulation Results	122
5.3	Overhead-Performance Tradeoff for Downlink MU-MIMO Systems	127
5.4	Summary	133
6	HETEROGENEOUS NETWORKS	134
6.1	Energy-Efficient User Pairing	135
6.2	Robust User Pairing	139
6.3	Summary	145
7	CONTRIBUTIONS AND FUTURE WORK	146
7.1	Contributions	146
7.2	Future Works	148
	BIBLIOGRAPHY	151

LIST OF TABLES

2.1	Estimated $R(D)$ and corresponding RMSE	35
5.1	Comparisons between VQ and SCQ in terms of the number of feedback bits, search complexity, and storage requirements.	119
5.2	Classic rate distortion results versus our approach	128
6.1	Simulation Parameters	143

LIST OF FIGURES

1.1	Illustration of overhead-aware design for wireless networks.	3
1.2	Illustration of uncertainty-aware design for wireless networks.	4
2.1	A two-hop decode-and-forward cooperative relaying network.	10
2.2	A two-hop decode-and-forward cooperative relaying network. Potential relay nodes are randomly distributed and M -group STBC is adopted.	12
2.3	Outage probability as a function of the number of nodes in the first group. Only the effect of Rayleigh fading is considered. Two different SNR margins are used. The randomness in the first hop is ignored.	21
2.4	Outage probability as a function of the transmit power. Only the effect of Rayleigh fading is considered.	22
2.5	CDF of the gap between the average received power for the first and second groups. The effect of path loss is considered. The randomness in the first hop is ignored.	23
2.6	Outage probability as a function of the transmit power. The effect of path loss is included.	24
2.7	A two-hop decode-and-forward cooperative relaying network consisting of a source, a destination, and K potential relay nodes, from which the best relay is selected.	25
2.8	Simulation and Gaussian-fitting approximations for $R(D)$	35
2.9	The lower bound on capacity loss D_{\min} when no overhead is allowed.	36
2.10	The capacity loss ratio $\rho(\gamma, K)$ when no overhead is allowed.	37

2.11	Data rate (performance) versus required information (overhead). ($B = 10$ MHz, 64 subcarriers)	37
2.12	Data rate (performance) versus required information (overhead). ($B = 10$ MHz, 64 subcarriers)	38
2.13	A timing diagram for FBRS. Each relay node sends a pilot symbol such that the destination can estimate the channels. Then, the destination broadcasts the relay selection results via a feedback packet. Afterwards, the selected relay starts data transmission. . .	41
2.14	A timing diagram for TBRS. The destination broadcasts a pilot symbol such that all nodes can estimate the channels. Each node sets a timer. Once the first timer expires, the corresponding node sends a pilot symbol to ensure that the destination can estimate the channel. Then, it starts data transmission, and all other nodes back off. . . .	43
2.15	t_{sel}/λ versus the number of potential relay nodes K	46
2.16	Validation of empirical approximations (2.65) and (2.66).	49
2.17	Normalized spectral efficiency as a function of the number of relays K	51
2.18	Normalized spectral efficiency as a function of $\tau = T_c/t_g$	51
2.19	Operating regimes for centralized and decentralized relay selection schemes for different values of K and τ	52
3.1	A large-scale wireless network with cooperative relays.	57
3.2	Outage probability versus the average received signal power ($d = 50$ m, $\alpha = 4$, $\lambda = 10^{-6}$, $K = K_v = k$).	76
3.3	Outage probability versus the average received signal power. K and K_v follow a Poisson distribution with mean κ ($d = 50$ m, $\alpha = 4$, $\lambda = 10^{-6}$).	76
3.4	Outage probability versus the intensity of interfering sources λ ($d = 50$ m, $\alpha = 4$, $K = K_v = 8$).	78
3.5	Asymptotic behavior of the success probability as the source-destination distance d increases ($\alpha = 4$, $\lambda = 10^{-5}$, $K = K_v = 8$).	78

3.6	Asymptotic behavior of the success probability as the intensity of interfering sources λ increases ($d = 50$ m, $\alpha = 4$, $K = K_v = 8$). . . .	79
3.7	Outage probability versus the intensity of the interfering sources λ . K and K_v follow a Poisson distribution with mean 16 ($d = 50$ m, $\alpha = 4$). . . .	80
3.8	Threshold for cooperation β^* versus the path loss exponent α . In the case with a random number of relays, K and K_v follow a Poisson distribution with mean 16.	80
3.9	Asymptotic behavior of the success probability as the source-destination distance d increases. $\alpha = 4$, $\lambda = 10^{-5}$, K and K_v follow a Poisson distribution with mean 16.	81
3.10	Asymptotic behavior of the success probability as the intensity of the interfering sources λ increases. $d = 50$ m, $\alpha = 4$, K and K_v follow a Poisson distribution with mean 16.	82
4.1	Linear network model with equidistant relay nodes.	85
4.2	Random shift model: linear network with randomly located relay nodes.	86
4.3	Illustration of a practical multi-hop wireless network	91
4.4	CDF of the maximized distance d_{\max} for a realistic system and the random shift model. ($\mathcal{R}_c = d_s = 100$ m, $\mathcal{R}_d = 30$ m, $N = 4$, $\delta = 0.2$)	92
4.5	Lower bound and an approximation for the spectral efficiency using the random shift model (10^6 trials, $d_s = 100$ m, $\alpha = 4$, $\delta = 1$).	93
4.6	Spectral efficiency as a function of the number of hops for different values of average received SNR and shift range. (10^6 trials, $d_s = 100$ m, $\alpha = 4$).	94
4.7	Optimum number of hops N^* (10^6 trials, $d_s = 100$ m, $\alpha = 4$, $\delta = 0.25$).	94
4.8	Multi-hop wireless ad hoc network. The route contains a source-destination pair and several relay clusters. The nodes in each cluster are selected from the potential relay nodes (small circles) according to a given routing protocol.	95

4.9	Linear network model with cooperative relays.	97
4.10	Optimum number of hops for different values of K (optimal selection), $\eta = 2$ bps/Hz, $\gamma_d = 4$ dB.	104
4.11	Optimum number of hops for different values of K (ad hoc selection), $\eta = 2$ bps/Hz, $\gamma_d = 4$ dB.	105
4.12	Optimum number of hops for different η (optimal and ad hoc selection schemes). $K = 4$, $\alpha = 4$, $\gamma_d = 4$ dB.	106
4.13	Optimum number of relays for different values of N (optimal selection), $\gamma_d = 4$ dB.	107
4.14	Optimum number of relays for different values of N (ad hoc selection), $\gamma_d = 4$ dB.	107
5.1	The multi-user MIMO broadcast channel with L users. The AP/BS is equipped with N_t antennas and each user has N_r antennas. We assume $N_t = LN_r$	110
5.2	Analytical approximation and simulation results of the sum rate achieved by SCQ.	123
5.3	Sum rate comparison between VQ and SCQ to achieve 1% rate loss with respect to the performance with perfect CSI at 20 dB SNR. $N_t = 4$, $N_r = 1$ and $L = 4$	124
5.4	Feedback overhead for the 3-dB performance gap versus SNR. $N_t = 4$, $N_r = 2$ and $L = 2$	125
5.5	Net capacity versus the overhead time ratio for different values of ρ . The carrier frequency $f_c = 5$ GHz, velocity $v = 25$ m/s, and doppler frequency $f_d = 500$ Hz. $N_t = 4$, $N_r = 2$, $L = 2$ and SNR = 20 dB.	126
5.6	Net capacity versus the velocity of the user v . $N_t = 4$, $N_r = 2$, $L = 2$ and SNR = 20 dB.	127
5.7	Rate distortion bounds versus vector quantization for different values of transmit power. $N_t = 2$, $N_r = 1$, $L = 2$. The channel magnitudes are assumed to be fixed as 1 for both users.	132

6.1	An example of user pairing scenario with cooperative communications and P2P cooperation.	136
6.2	Performance of user pairing with Gaussian uncertainty.	144
6.3	Performance of user pairing with a polyhedron uncertainty set. . . .	144
6.4	Performance of user pairing with an ellipsoid uncertainty set.	145

GLOSSARY

$ \cdot $	Absolute value of a complex number
$\Pr\{\cdot\}$	Probability of a event
\mathbb{E}	Expectation of a random variable
$\lfloor \cdot \rfloor$	Largest integer that is not greater than the operand
\mathbf{I}	Identity matrix
\mathbb{R}	Set of real numbers
\mathbb{Z}^+	Set of positive integers
$\ \cdot\ $	2-norm of a vector/matrix
$\ \cdot\ _{\text{F}}$	Frobenius norm of a vector/matrix
$\log(\cdot)$	Base-2 logarithm operation
$\ln(\cdot)$	Base- e logarithm operation
$(\cdot)^{\text{T}}$	Transpose of a matrix
$(\cdot)^{\text{H}}$	Hermitian (conjugate transpose) of a matrix
$\det(\cdot)$	Determinant of a square matrix
$\text{tr}(\cdot)$	Trace of a square matrix
$\Re(\cdot)$	Real part of a complex number
$\Im(\cdot)$	Imaginary part of a complex number

3GPP	3rd Generation Partnership Project
AP	Access point
AWGN	Additive white Gaussian noise
BD	Block diagonalization
BS	Base station
CDF	Cumulative distribution function
CSI	Channel state information
dB	Decibel
DF	Decode-and-forward
i.i.d.	Independent and identically distributed
LHS	Left-hand-side of an equality/inequality
LTE-A	Long Term Evolution-Advanced
MAC	Media access control
MANET	Mobile ad hoc network
MIMO	Multiple-input multiple-output
MU	Multi-user
NP	Nondeterministic polynomial time
OFDM	Orthogonal frequency division multiplexing
PN	Pseudo-random noise
RHS	Right-hand-side of an equality/inequality
RMSE	Root-mean-square error
SINR	Signal-to-interference-plus-noise ratio
SIR	Signal-to-interference ratio
SNR	Signal-to-noise ratio
SQ	Scalar quantization
SQP	Sequential quadratic programming

STBC	Space-time block coding
SU	Single-user
VQ	Vector quantization
SCQ	Sparse coding quantization
WLAN	Wireless local area network
WSN	Wireless sensor network

ABSTRACT

In general, the performance of many wireless systems is approaching the fundamental limits on transmission capacity. For example, current commercial wireless standards such as 3GPP LTE-A and IEEE 802.11ac have a near-optimal physical layer. In order to meet the ever growing demand for capacity, other directions for improving network performance must be found.

In most existing research on wireless networks, overhead, the “non-data” portion including coordination, control signaling and other costs of serving different purposes, is assumed to be negligible. However, the final application throughput could be much lower than the theoretical bounds as a result of overhead, especially in large and dynamic networks. Therefore, it is critical to quantitatively analyze the overhead in wireless networks, which could provide clear insights on the performance in practical systems and could help to identify opportunities for improvements in their designs. Surprisingly, the fundamental limits on overhead are largely unknown, and the framework needed to design overhead-aware systems has not been adequately investigated.

In addition, interference is one of the main performance-limiting factors in most future wireless applications. Conventional “interference avoidance” techniques might not be feasible because the degrees of freedom (for example, bandwidth, number of orthogonal codes, and time) might be limited. Although the interference can be mitigated quite efficiently with centralized control, existing approaches are usually very sensitive to channel uncertainties; if the knowledge of the channel state information is imperfect, the system performance could be severely degraded. Also, collecting accurate information incurs a significant amount of overhead due to the time-varying nature of the wireless medium. Thus, it is imperative to jointly consider overhead, uncertainty, and interference.

In this dissertation, we investigate practical and overhead-aware designs that can achieve better performance in a realistic networking context. We start with a simple, single-user, two-hop cooperative relaying network model. For this model, we first prove that M -group cooperation is the optimal distributed space-time block coding strategy when neither central control nor inter-relay communications is permitted. Then, we consider the relay selection problem where a small and acceptable amount of overhead is allowed. The tradeoff between the feedback overhead and the performance is investigated via rate distortion theory. Compared to existing research, which is usually highly dependent on the specific implementation approaches, the analysis presented here addresses the fundamental tradeoff of a general network. Using our theoretical results, we also compare practical centralized and decentralized relay selection schemes in terms of spectral efficiency.

Then, interference-limited networks with multiple concurrent transmissions are studied. We analyze and compare the performance of cooperative and non-cooperative schemes. Although cooperation among relay nodes increases the reliability of point-to-point transmission, it also produces a higher level of interference and degrades the overall performance of a multi-user network. The tradeoff between cooperative gain and the additional interference is investigated, and a criterion which determines whether we should cooperative or not is derived.

We next focus on multi-hop linear networks, which have one or more intermediate nodes along the path that receive and forward information via wireless links. Instead of assuming equal hop distances, we propose a novel model that permits randomness in the node locations, and then we determine the optimum number of hops for maximizing the end-to-end spectral efficiency. Then, for a multi-hop linear network with cooperative relays, a relay deployment strategy is proposed and studied.

After that, for downlink multi-user networks, we present a novel quantization technique, sparse coding quantization (SCQ), which is an extension of classic vector quantization (VQ) and provides a balance between performance and complexity. In

particular, the computational complexity of conventional VQ can be significantly reduced by applying SCQ, with a negligible reduction in performance. Comparisons among different quantization techniques are also provided. Beside considering specific quantization schemes, we also study the overhead-performance tradeoff for general MU-MIMO systems by applying a rate distortion framework.

Finally, we investigate robust a user pairing problem for a heterogeneous network in the presence of channel uncertainty. Different definitions of robustness and uncertainty are considered to formulate the corresponding optimization problems. We develop an algorithm that is robust to uncertainty in channel measurement and thereby performs well in practical systems. Simulation results validate the robustness of the proposed method.

Chapter 1

OVERVIEW

1.1 Background and Motivation

Wireless communications and networking has brought revolutionary changes to the way people communicate, work, and entertain. Today, there are about 7 billion mobile subscribers worldwide [1], which is equivalent to 95.5 percent of the world population; and, the number of smartphone users is expected to exceed 1.75 billion by the end of 2014 [2]. This proliferation of mobile devices is driving the tremendous growth in mobile traffic demands.

To a large extent, the performance of many wireless systems is approaching the fundamental limits on transmission capacity. Current commercial wireless standards such as 3GPP Long Term Evolution-Advanced (LTE-A)[3] and IEEE 802.11ac [4] have a near-optimal physical layer; this performance is achieved by using a combination of modern techniques including Orthogonal Frequency Division Multiplexing (OFDM) [5, 6], Multiple-input Multiple-output (MIMO) [7–9], and turbo codes [10, 11], among others. In order to meet the ever growing demand for capacity, other directions for improving network performance must be found.

1.1.1 Overhead

In most existing research, overhead, the “non-data” portion including coordination, control signaling and other costs of serving different purposes, is assumed to be negligible. However, the final application throughput could be much lower than the theoretical bounds as a result of overhead, especially in large and dynamic networks [12–14].

Even for wired networks, such as Ethernet over twisted pair copper wire, a significant portion of the overall Internet traffic is due to the overhead in different protocols [15]. The additional complexities of the wireless medium, such as its broadcasting nature and varying channel characteristics, require additional overhead for coordinating transmissions and resolving conflicts between users. Applications that require wireless communications are also more dynamic, with mobile users frequently joining and leaving the network. The mobility and dynamic network topology incur extra overhead requirements for synchronization and routing. Moreover, for wireless transmissions, reliable estimates of the channel state information (CSI) are required at both the transmitter and the receiver to approach the theoretical performance limits. To acquire this information, a significant amount of overhead, such as training sequences and CSI feedback signals, could be incurred.

Fig. 1.1 is meant to illustrate the importance of overhead-aware designs for wireless networks. Both the theoretical bounds and the actual performance are sketched. According to Fig. 1.1, when no or low overhead is allowed, the performance is usually poor since the advanced technologies for handling channel uncertainties cannot be applied. On the other hand, the theoretical optimum performance can be achieved, but usually at the expense of a significant amount of overhead. In this case, the actual performance might also be very poor since a large part of the resources might be occupied by the overhead. Therefore, it is critical to analyze fundamental limits on the overhead in wireless networks, which could provide clear insights on the performance in practical systems and could help to identify opportunities for improvements in their design. Surprisingly, the fundamental limits on overhead are largely unknown, and the framework needed to design overhead-aware systems has not been adequately investigated.

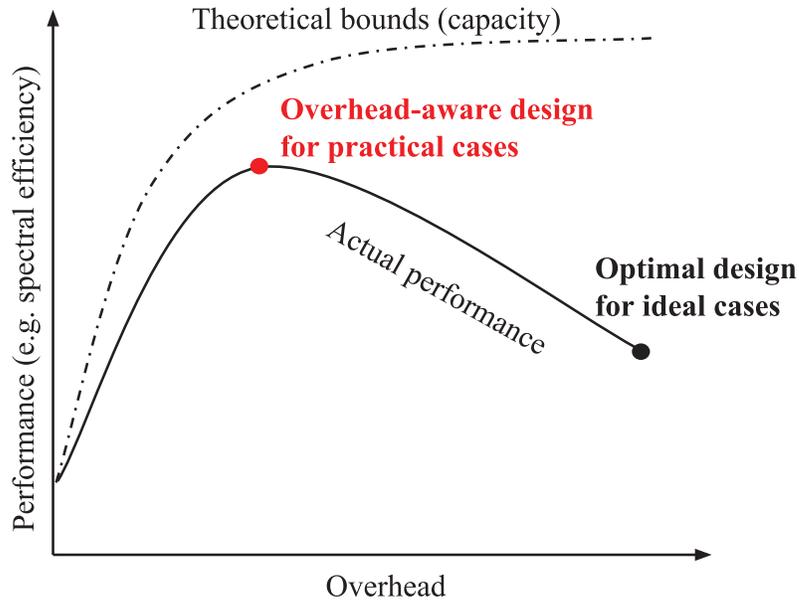


Figure 1.1: Illustration of overhead-aware design for wireless networks.

1.1.2 Uncertainty

The performance of wireless networks relies on the model of the underlying physical layer. The model should be as accurate as necessary and as simple as possible. However, due to practical constraints, only imperfect characterizations of the real system are available. For example, some properties of real systems are usually not modeled explicitly to ensure an analytically tractable framework. Also, time-variant parameters, which should be accurately acquired, are generally not known perfectly because of the non-deterministic nature of wireless communications and networking (for example, channel uncertainty, node mobility, random network topology).

Conventional approaches are designed assuming that the model is correct and its parameters are perfectly estimated. The uncertainties of the model are usually ignored and the estimated parameters are applied as if they were error-free. If the uncertainties are negligible, the conventional approaches yield satisfactory results. However, the performance of wireless networks could be significantly degraded due to the unavoidable errors, as illustrated in Fig. 1.2.

Alternatively, a robust design [16, 17] aims at minimizing the performance loss

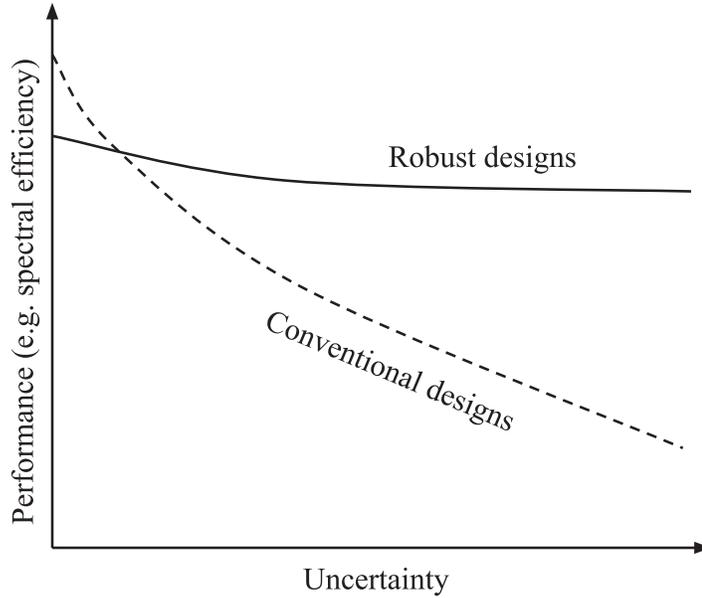


Figure 1.2: Illustration of uncertainty-aware design for wireless networks.

due to model errors and uncertainties. As we can see from Fig. 1.2, robust designs guarantee reasonable system performance even under large uncertainty, while sacrificing some performance with respect to the ideal design. Achieving a balance between robustness and performance is a fundamental challenge. Improved techniques for estimation and prediction of the channel parameters together with robust optimization at the physical layer can contribute to designing efficient and practical wireless networks.

1.1.3 Interference

Interference is one of the main performance-limiting factors in most future wireless applications. Since the degrees of freedom, for example, bandwidth, number of orthogonal codes, and time, are usually limited, “interference avoidance” techniques, which attempt to avoid collisions by transmitting data in different orthogonal channels, might be impossible. In other words, some transmissions will inevitably occur at the same time in the same frequency band, separated only in space, and the signals from many undesired or interfering transmitters add to the desired transmitter’s signal at a

receiver. This interference can be mitigated quite efficiently in systems with centralized control. For example, a base station (BS) or access point (AP) can coordinate the channelization and the power/interference levels of the individual terminals. Alternatively, sophisticated multi-user detection or interference cancellation schemes could be employed. However, many emerging classes of wireless systems, such as ad hoc and sensor networks, do not permit centralized control, requiring a more distributed resource allocation. In this case, the interference is not tightly controllable and a degradation in the system performance is inevitable.

Note that these performance-limiting factors are not independent. Typically, centralized approaches for managing interference are very sensitive to the uncertainties; if the knowledge of the channel information is imperfect, the system performance will be severely affected. Due to the time-varying nature of the wireless medium, collecting accurate information incurs a significant amount of overhead. Hence, it is imperative for us to jointly consider overhead, uncertainty, and interference.

1.2 Dissertation Outline

In this dissertation, instead of focusing on optimal designs for idealized scenarios, we investigate practical and overhead-aware designs that can achieve better performance in a realistic networking context. We start with a simple, single-user, two-hop cooperative relaying network model. For this model, we present a low-overhead design and preliminary results on the overhead-performance tradeoff. Then, we expand the analysis to more complicated models such as multi-user, two-hop, Poisson networks and single-user, multi-hop, linear networks. Our contributions include proposing practical overhead-aware designs with low overhead, quantifying the amount of overhead for given systems, analyzing the theoretical limits on overhead for a given transmission strategy, and optimizing the system performance taking into consideration the different types of overhead.

In Chapter 2, we first formulate the design of protocols for a single-user, two-hop, cooperative network as an optimization problem with an overhead constraint.

By solving the proposed problem, we prove that M -group cooperation is the optimal distributed space-time block coding (STBC) strategy when neither central control nor inter-relay communications is permitted, i.e., CSI at the transmitter is not available. The optimality of M -group holds in not only an ideal Rayleigh fading environment but also in more realistic scenarios where path loss is included. Then, we consider the case where a small and acceptable amount of overhead is allowed. In particular, we assume that the destination collects the CSI for all potential relays and then selects the node with highest channel gain. The tradeoff between the feedback overhead and the performance of this relay selection scheme is investigated via rate distortion theory. Compared to existing research, which is usually highly dependent on the specific implementation approaches, the analysis presented here addresses the fundamental questions: (1) How much extra information is required? and (2) What is the optimal tradeoff for general selective relaying networks? The theoretical analysis for the rate distortion function and its asymptotic properties are presented. Finally, we compare the spectral efficiency of centralized and decentralized cooperative communication systems. The impact of overhead is included in our analysis. We show that the centralized scheme usually achieves a higher spectral efficiency than the decentralized scheme if the number of nodes is small and/or the channel is static. Although the decentralized scheme significantly reduces the amount of overhead, it suffers from unavoidable performance loss. For given system environments, criteria for determining which scheme should be applied is also provided.

In Chapter 3, we consider a large-scale and dynamic network with multiple concurrent transmissions. As shown in Chapter 2, the reliability of point-to-point transmission can be significantly improved by sharing each cooperative node's antenna to form a virtual antenna array. However, cooperation among different nodes may produce a higher level of interference and degrade the overall performance of a multi-user network. Therefore, for practical environments, it is essential to investigate the tradeoff between cooperative gain and the additional interference. By analyzing and comparing the outage performance of non-cooperative and cooperative strategies, we

derive a criterion for deciding whether we should cooperate or not. We show that a cooperative strategy is preferred for sparse networks. As the wireless networks become more dense, the benefit of cooperation is eventually eliminated by the excessive amount of interference, which implies that non-cooperative strategies should be used.

In Chapter 4, instead of focusing on two-hop transmissions, we investigate multi-hop networks. A linear network model, which is a commonly used model to approximate practical networks, is adopted. We first propose a novel linear network model to characterize the randomness in the node location. Unlike most previous network models, which usually assume that the nodes are equidistant, we consider a linear network with randomly located nodes. The randomness in the distance is analyzed and validated to be a reasonable approximation to reality. The optimum number of hops and the spectral efficiency of the proposed model are studied. In addition, we provide performance analyses for the linear multi-hop network with cooperative relays. Using an outage analysis, we derive the optimum relay cluster locations which minimize the end-to-end outage probability. Further, we consider the required cooperation overhead by using the overhead-performance tradeoff analysis in Chapter 2. A large number of relays could lead to a performance loss because of the extra overhead costs in implementing cooperation.

In Chapter 5, a novel quantization method, sparse coding quantization (SCQ), is proposed for downlink, multiuser, multiple-input multiple-output (MU-MIMO) systems. Compared to conventional vector quantization (VQ), in which the original channel information is represented by a vector codeword, SCQ utilizes a linear combination of several codewords rather than a single one to represent the channel matrix. We show that the proposed technique can achieve the same sum rate performance as VQ at a reasonable cost in feedback overhead. Thus, SCQ is more practical because it significantly reduces the time and storage complexity for generating, searching and storing the codebook. Also, comparisons among the different quantization techniques are provided. The required feedback bits for a specified rate loss are quantified, as well as the complexity for each technique. Furthermore, the net capacity, which incorporates

the effect of the overhead, is studied. Both analytical and simulation results reveal the advantages and drawbacks of each quantization method and demonstrate under what conditions to use one of them rather than the other. On the other hand, instead of focusing on specific quantization techniques, we also apply the rate distortion framework discussed in Chapter 2 to MU-MIMO systems. The tradeoff between the rate loss due to CSI quantization and the amount of feedback overhead has been quantified and investigated.

In Chapter 6, we focus on heterogeneous networks, which can be exploited to improve cellular system performance. Specifically, the use of multiple cells can significantly increase the system throughput and reduce power consumption. However, complete and perfect information is required to coordinate transmissions and to allocate the available resources to different users. In practical systems, the performance of traditional algorithms degrades if some assumptions are incorrect or imprecise. In this chapter, we propose a robust user pairing algorithm which jointly exploits cooperative communications and peer-to-peer streaming. By evaluating the performance under different definitions of uncertainty, we show that the proposed algorithm outperforms the traditional optimal solution in the presence of imperfect CSI.

Finally, in Chapter 7, we summarize our contributions and describe interesting open problems for future research.

Chapter 2

TWO-HOP DECODE-AND-FORWARD COOPERATIVE RELAYING NETWORKS

2.1 Introduction

Cooperative relaying in wireless networks has been a very active area of research in recent years. In contrast to direct transmission, cooperative relaying can exploit the must-needed diversity inherent in multiple spatially distributed wireless links [18–20], and thereby improve the reliability of desired transmissions.

In this chapter, a decode-and-forward two-hop cooperative relaying network is investigated. All the nodes in the network are assumed to have a single antenna and be capable of only half-duplex transmission. The entire network is synchronized in both time and frequency.¹ The direct link between the source and the destination is assumed to be very weak due to the presence of deep fading or shadowing, as is typically the case. As illustrated in Fig. 2.1, there are multiple potential relay nodes in the network which can help the transmission. We assume that the potential relay nodes are randomly distributed in the network, and the spatial interference caused by other concurrent data transmissions is ignored. The channels are assumed to follow a quasi-static flat Rayleigh fading model.

The transmission process can be described as follows. In the first hop, the source node broadcasts its signal s and each node receives an impaired version

$$y_i = h_{sr_i} \sqrt{P_t} s + n_i, \quad (2.1)$$

¹ Synchronization is critical to the performance of cooperative networks. In [21, 22], it has been shown that the network performance will significantly degrade without perfect synchronization. The implementation of synchronization can be achieved by using a global clock [23], training symbols [24], or PN sequences [25].

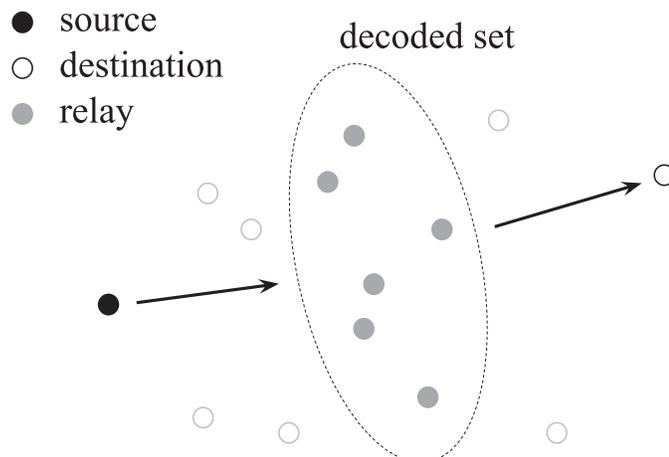


Figure 2.1: A two-hop decode-and-forward cooperative relaying network.

where P_t represents the transmit power of the source, h_{sr_i} is the channel from the source to the i th relay, and n_i denotes the white Gaussian noise at the relays. The noise power P_N is assumed to be the same for all receivers. As long as the instantaneous SNR at a node is higher than a pre-determined threshold γ_{th} , i.e.,

$$\frac{P_t |h_{sr_i}|^2}{P_N} \geq \gamma_{\text{th}}, \quad (2.2)$$

we assume that the node can successfully decode s ; such a node is called a decoded node. The set of decoded nodes is called the decoded set and denoted by \mathcal{D} . It is clear that the number of decoded nodes is upper bounded by the total number of nodes. Notice that \mathcal{D} is a random set and the number of decoded nodes K is thus a random variable, both varying with the fading channels in the first hop. For the sake of simplicity, we will temporarily ignore the randomness of K and consider it only in the simulations.

In the second hop, the decoded set, consisting of all decoded nodes, can form a

virtual antenna array where each node emulates one antenna of a multiple-antenna array. Significant spatial diversity gain can be achieved by applying well-known multiple-antenna techniques, such as cooperative beamforming [26], STBC [27–34], relay selection [35–39], and so on. The end-to-end performance, which depends on the transmission strategy used in the second hop, will be discussed in the following sections.

The main difference between cooperative relaying and multiple-antenna transmission is that the relay nodes in a cooperative network are distributed in space. So, obtaining and exploiting CSI is an additional challenge. Implementing a practical cooperative network is non-trivial because it usually requires central control or full inter-relay communications. In other words, a significant overhead will be incurred, and the net performance gain might be significantly reduced. Therefore, designing a decentralized cooperative strategy with low overhead is crucial for real-world applications. Investigating the theoretical limits on overhead for a given transmission strategy is another important task that will also be covered in this chapter.

2.2 Decentralized Cooperative Strategy with Low Overhead

STBC has been shown to be a promising technique for achieving the available diversity benefits of cooperation. Although conventional STBC schemes [27, 28] usually require a central controller for coordinating the transmission, decentralized STBC [30–34] has been investigated for eliminating the coordination overhead. In [30], the authors propose a decentralized STBC scheme called M -group for cooperative networks. With M -group, the nodes which have already successfully decoded the information from the source node will independently and randomly divide themselves into M groups; each group then emulates one antenna of an M -antenna system. By applying a pre-determined STBC scheme, cooperation among the relays, and the resulting performance gain, can be obtained. It has been shown that M -group can achieve almost the same outage performance as a true multiple-antenna STBC system, but with very low complexity. Moreover, M -group requires the least amount of overhead compared with other decentralized STBC schemes, requiring neither inter-relay

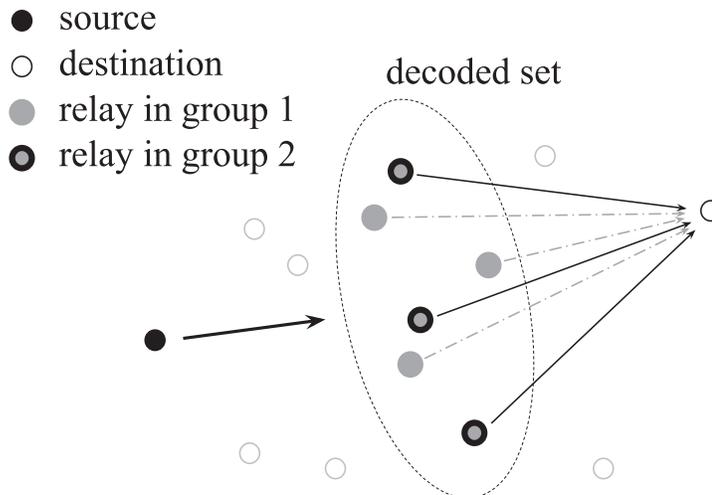


Figure 2.2: A two-hop decode-and-forward cooperative relaying network. Potential relay nodes are randomly distributed and M -group STBC is adopted.

communications nor central control.

As shown in Fig. 2.2, the decoded set \mathcal{D} is divided into M groups $\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_M$, and a pre-determined orthogonal STBC scheme designed for an M -antenna system is adopted for the transmission in the second hop. We assume that the STBC matrix, denoted by $\mathbf{X}(s) = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_M]$, is known to all nodes in the network. Each group \mathcal{G}_i then emulates an antenna of the underlying M -antenna system, i.e., all relays in \mathcal{G}_i transmit the same column \mathbf{x}_i to the destination. After that, the destination can extract the desired signal s using a decoding rule for orthogonal STBC [28]. The end-to-end outage performance depends on how the M groups are constructed.

In this section, we prove the optimality of M -group decentralized STBC. By solving an optimization problem minimizing the outage probability of the system, we can show that M -group is the optimal transmission strategy under the condition that *neither inter-relay communications nor central control is permitted*. The robustness of M -group will also be discussed; this justifies the use of M -group in practical cooperative wireless networks.

2.2.1 Optimization Formulation

There are two basic challenges in a decentralized STBC system: (i) How to design a proper space-time code? and (ii) How to perform the STBC using the available decoded set? Given the dynamic and distributed nature of a wireless network, these two challenges are difficult without central control or full inter-relay negotiations, which would incur a significant amount of overhead. To reduce the overhead, we need to relax the design problem. The first relaxation is to fix the underlying STBC scheme *a priori*, which means that not only the coding scheme but also the size of the underlying STBC matrix are pre-specified and known to all nodes in the network. This addresses the first challenge.

To solve the second challenge, as discussed above, we can divide the decoded set into M groups and let each group emulate an antenna. Then, the design problem is simplified to the problem of grouping these decoded nodes into M groups such that the inter-relay communications and feedback (negotiation overhead) is less than some given tolerance.

Recall that the number of nodes in the decoded set is K . Define a $K \times K$ binary “overhead” matrix $\mathbf{\Omega}$ with

$$\Omega_{i,j} = \begin{cases} 1, & \text{If the } i\text{th decoded node is permitted to} \\ & \text{communicate with the } j\text{th node } (i \neq j) \\ 0, & \text{otherwise.} \end{cases} \quad (2.3)$$

Note that the diagonal elements of $\mathbf{\Omega}$ are always set to 0. The number of 1’s in $\mathbf{\Omega}$ can be used as a measure of the overhead incurred by the inter-relay communications. Assuming all nodes in the decoded set relay the source information, when inter-relay communications is prohibited, the number of 1’s in $\mathbf{\Omega}$ reaches its minimum value 0. Furthermore, since $\mathbf{\Omega}$ is a binary matrix, the number of 1’s is equal to the square of its Frobenius norm. In the following, we will use the Frobenius norm $\|\mathbf{\Omega}\|_F$ of $\mathbf{\Omega}$ to represent the overhead. It can be seen that $\|\mathbf{\Omega}\|_F \geq 0$ where equality holds when no inter-relay communications is permitted.

Without loss of generality, we suppose that the M groups transmit s to the destination using an orthogonal STBC scheme [40]. Since we have assumed that the noise power is the same for all links, we can set it to 1 to simplify the notation. The SNR at the destination becomes

$$\text{SNR} = \sum_{i=1}^M \left| \sum_{j \in \mathcal{G}_i} \sqrt{P_j} h_{r_j d} \right|^2, \quad (2.4)$$

where P_j represents the transmit power of the j th node in the i th group \mathcal{G}_i and $h_{r_j d}$ is the channel from the corresponding relay to the destination. The outage probability is defined as

$$p_{\text{out}} = \Pr\{\text{SNR} \leq \gamma_{\text{th}}\}. \quad (2.5)$$

Then, the problem can be formulated as

$$\begin{aligned} & \min_{\mathcal{G}_1, \mathcal{G}_2, \dots, \mathcal{G}_M} p_{\text{out}} \\ \text{s.t.} \quad & \|\mathbf{\Omega}\|_F \leq \Omega_{\text{th}} \\ & \bigcup_{i=1}^M \mathcal{G}_i = \mathcal{D} \\ & \mathcal{G}_i \neq \emptyset, \quad \forall i \\ & \mathcal{G}_i \cap \mathcal{G}_j = \emptyset, \forall i \neq j \end{aligned} \quad (2.6)$$

where Ω_{th} is the overhead tolerance.

Note that (2.6) is a combinatorial optimization problem and difficult to solve [30]. Instead, we examine an extreme case where no inter-relay communications is permitted. We can represent this extreme case by setting the overhead tolerance $\Omega_{\text{th}} = 0$. In other words, we assume that only the destination knows the CSI on the channels from the relays to the destination, and the estimated CSI will not be fed back to the relays.

Using (2.4), the received signal power contributed by the i th group is

$$P_s^{(i)} = \left| \sum_{j \in \mathcal{G}_i} \sqrt{P_j} h_{r_j d} \right|^2. \quad (2.7)$$

Since the $h_{r_j d}$'s are i.i.d. complex Gaussian random variables, each $P_s^{(i)}$ is a random variable with an exponential distribution and can be parametrized by

$$\lambda_i = \frac{1}{\mathbb{E}[P_s^{(i)}]} = \frac{1}{\mathbb{E}\left[\left|\sum_{j \in \mathcal{G}_i} \sqrt{P_j} h_{r_j d}\right|^2\right]}, \quad (2.8)$$

where \mathbb{E} is the expectation operator, and λ_i represents the reciprocal of the average received power from the i th group \mathcal{G}_i . Since no inter-relay communications is allowed, we cannot apply power allocation or any relay selection technique. So we assume that all the nodes in the decoded set \mathcal{D} transmit with the same power P_t . We use the notation v to denote the ratio of γ_{th} , the SNR threshold, to the transmit power P_t , i.e., $v \triangleq \gamma_{\text{th}}/P_t$. Note that the noise power is assumed to be one without loss of generality. Then, λ_i is the reciprocal of the number of nodes in the i th group \mathcal{G}_i multiplied by P_t ,

$$\lambda_i = \frac{1}{P_t |\mathcal{G}_i|}, \quad (2.9)$$

where $|\mathcal{G}_i|$ denotes the cardinality of the i th group. Let $K_i \triangleq |\mathcal{G}_i|$. By using characteristic-function and partial-fraction techniques, we can derive a closed-form expression for the outage probability for the case $\lambda_i \neq \lambda_j, \forall i \neq j$,

$$\begin{aligned} p_{\text{out}} &= \Pr\left\{\sum_{i=1}^M P_s^{(i)} \leq \gamma_{\text{th}}\right\} \\ &= \sum_{i=1}^M (1 - e^{-\gamma_{\text{th}} \lambda_i}) \prod_{j \neq i} \frac{1}{1 - \frac{\lambda_i}{\lambda_j}} \\ &= \sum_{i=1}^M \left(1 - e^{-\frac{v}{K_i}}\right) \prod_{j \neq i} \frac{1}{1 - \frac{K_j}{K_i}}. \end{aligned} \quad (2.10)$$

Although this expression is not valid when $\lambda_i = \lambda_j$, we can show that p_{out} is a continuous and second-order differentiable function of the λ_i 's (or K_i 's). In particular, if

$\lambda_i = \lambda_j = \lambda, \forall i, j$, the closed-form expression for the outage probability is

$$\begin{aligned}
p_{\text{out}} &= \Pr \left\{ \sum_{i=1}^M P_s^{(i)} \leq \gamma_{\text{th}} \right\} \\
&= 1 - \sum_{i=0}^{M-1} \frac{e^{-\gamma_{\text{th}}\lambda} (\gamma_{\text{th}}\lambda)^i}{i!} \\
&= 1 - e^{-\frac{vM}{K}} \sum_{i=0}^{M-1} \frac{1}{i!} \left(\frac{vM}{K} \right)^i.
\end{aligned} \tag{2.11}$$

The specialized optimization problem, i.e., (2.6) with $\Omega_{\text{th}} = 0$, can be reformulated as

$$\begin{aligned}
&\min_{K_1, K_2, \dots, K_M} p_{\text{out}}(K_1, K_2, \dots, K_M) \\
&\text{s.t.} \quad K_1 + K_2 + \dots + K_M = K \\
&\quad \quad K_i \in \{0, 1, \dots, K\}
\end{aligned} \tag{2.12}$$

In order to make this NP-hard problem tractable, we also relax the integer constraint, giving

$$\begin{aligned}
&\min_{K_1, K_2, \dots, K_M} p_{\text{out}}(K_1, K_2, \dots, K_M) \\
&\text{s.t.} \quad K_1 + K_2 + \dots + K_M = K \\
&\quad \quad K_i \in \mathbb{R}, \quad K_i \geq 0
\end{aligned} \tag{2.13}$$

2.2.2 Optimality of M -group

If a uniform grouping is the optimal solution to (2.13), it is then clear that we cannot do better than letting the decoded nodes independently and uniformly decide which group they belong to. In the following, we prove this conjecture and show that a uniform grouping is indeed the optimal strategy.

Theorem 2.1. A uniform grouping, i.e., $K_i = K/M, \forall i \in \{1, 2, \dots, M\}$, is the optimal solution to (2.13).

Proof. We use mathematical induction to prove the theorem.

Basis: In the simplest case $M = 2$,

$$p_{\text{out}}(K_1, K_2) = \begin{cases} 1 - \frac{K_1}{K_1 - K_2} e^{-\frac{v}{K_1}} - \frac{K_2}{K_2 - K_1} e^{-\frac{v}{K_2}} & \text{if } K_1 \neq K_2 \\ 1 - e^{-\frac{2v}{K}} - \frac{2v}{K} e^{-\frac{2v}{K}} & \text{if } K_1 = K_2 \end{cases}. \tag{2.14}$$

Since $p_{\text{out}}(K_1, K_2)$ is a continuous and second-order differentiable function, we only need to consider the general expression for $K_1 \neq K_2$. Consider the Lagrange multiplier

$$\mathcal{L}(K_1, K_2, \lambda) = 1 - \frac{K_1}{K_1 - K_2} e^{-\frac{v}{K_1}} - \frac{K_2}{K_2 - K_1} e^{-\frac{v}{K_2}} + \lambda(K_1 + K_2 - K). \quad (2.15)$$

Then, the corresponding Lagrange conditions are

$$\begin{cases} \frac{K_2}{(K_1 - K_2)^2} \left(e^{-\frac{v}{K_1}} - e^{-\frac{v}{K_2}} \right) - \frac{v}{K_1(K_1 - K_2)} e^{-\frac{v}{K_1}} + \lambda = 0 \\ \frac{K_1}{(K_1 - K_2)^2} \left(e^{-\frac{v}{K_2}} - e^{-\frac{v}{K_1}} \right) - \frac{v}{K_2(K_2 - K_1)} e^{-\frac{v}{K_2}} + \lambda = 0 \\ K_1 + K_2 = K \end{cases} \quad (2.16)$$

which can be rewritten as

$$\frac{K_1 + K_2}{K_1 - K_2} = \frac{v(K_1 - K_2) \left(\frac{1}{K_1} e^{-\frac{v}{K_1}} + \frac{1}{K_2} e^{-\frac{v}{K_2}} \right)}{e^{-\frac{v}{K_1}} - e^{-\frac{v}{K_2}}}. \quad (2.17)$$

Let $\Delta K = K_1 - K_2$. Notice that $K_1 + K_2 = K$, then (2.17) is equivalent to the following equation

$$K \left(e^{-\frac{2v}{K+\Delta K}} - e^{-\frac{2v}{K-\Delta K}} \right) - v \left(\frac{2\Delta K}{K+\Delta K} e^{-\frac{2v}{K+\Delta K}} + \frac{2\Delta K}{K-\Delta K} e^{-\frac{2v}{K-\Delta K}} \right) = 0. \quad (2.18)$$

Obviously, (2.18) holds when $\Delta K = 0$. By calculating the derivatives, it is easy to show that the left-hand-side of (2.18) is a strictly increasing function of ΔK . Hence $\Delta K = 0$ (or $K_1 = K_2 = K/2$) is the unique solution of (2.17). Since we know that the boundary is the worst case (when $K_1 = 0$ or $K_2 = 0$, no diversity gain is obtained), the global minimum is achieved by $K_1 = K_2 = K/2$.

Inductive step: When $M > 2$, assume the inductive hypothesis holds for $M = M'$. For $M = M' + 1$, we have

$$\begin{aligned} p_{\text{out}}(K_1, K_2, \dots, K_{M'+1}) &= \Pr \left\{ \sum_{i=1}^{M'+1} P_s^{(i)} \leq \gamma_{\text{th}} \right\} \\ &= \int_{\sum_{i=1}^{M'+1} P_s^{(i)} \leq \gamma_{\text{th}}} p(P_s^{(1)}, P_s^{(2)}, \dots, P_s^{M'+1}) dP_s^{(1)} dP_s^{(2)} \dots dP_s^{(M'+1)} \\ &= \int_0^{\gamma_{\text{th}}} \Pr \left\{ \sum_{i=1}^{M'} P_s^{(i)} \leq \gamma_{\text{th}} - P_s^{(M'+1)} \mid P_s^{(M'+1)} \right\} p(P_s^{(M'+1)}) dP_s^{(M'+1)} \end{aligned}$$

$$\begin{aligned}
&\stackrel{(a)}{\geq} \int_0^{\gamma_{\text{th}}} \left[1 - \sum_{i=0}^{M'-1} \frac{e^{-\frac{M'(\gamma_{\text{th}} - P_s^{(M'+1)})}{P_t(K - K_{M'+1})}} \left(\frac{M'(\gamma_{\text{th}} - P_s^{(M'+1)})}{P_t(K - K_{M'+1})} \right)^i}{i!} \right] \frac{1}{P_t K_{M'+1}} e^{-\frac{P_s^{(M'+1)}}{P_t K_{M'+1}}} dP_s^{(M'+1)} \\
&\stackrel{(*)}{=} \int_0^v \left[1 - \sum_{i=0}^{M'-1} \frac{1}{i!} e^{-\frac{M'(v - P_s)}{K - K_{M'+1}}} \left(\frac{M'(v - P_s)}{K - K_{M'+1}} \right)^i \right] \frac{1}{K_{M'+1}} e^{-\frac{P_s}{K_{M'+1}}} dP_s \\
&= 1 - e^{-\frac{v}{K_{M'+1}}} - \\
&\frac{1}{K_{M'+1}} \sum_{i=0}^{M'-1} \frac{1}{i!} \left(\frac{M'}{K - K_{M'+1}} \right)^i e^{-\frac{M'v}{K - K_{M'+1}}} \int_0^v (v - y)^i e^{\left(\frac{M'}{K - K_{M'+1}} - \frac{1}{K_{M'+1}} \right) y} dy \\
&\stackrel{(b)}{\geq} 1 - e^{-\frac{v(M'+1)}{K}} \sum_{i=0}^{M'} \frac{1}{i!} \left(\frac{v(M'+1)}{K} \right)^i, \tag{2.19}
\end{aligned}$$

where $Y_{M'+1}/P$ is replaced by y in step (*).

The expression in the last line of (2.19) is the outage probability with all K_i 's being equal, given by (2.11). In order to prove that the hypothesis still holds for $M = M' + 1$, we only need to show that the minimum value is actually achievable. The equality of step (a) holds if and only if the parameters of $Y_1, \dots, Y_{M'}$ are the same, and the equality of step (b) holds if and only if $\frac{M'}{K - K_{M'+1}} = \frac{1}{K_{M'+1}}$, i.e., the minimum value can be achieved when $K_1 = K_2 = \dots = K_{M'+1}$. Hence, the hypothesis also holds when $M = M' + 1$.

By mathematical induction, the statement holds for all $M \geq 2$ and the theorem is proved. \blacksquare

Theorem 2.1 indicates that the optimal grouping strategy is to allocate the decoded nodes across the decoded set \mathcal{D} as uniformly as possible. In [30], a similar problem is considered, but it only shows that uniform allocation can minimize an upper bound of the outage probability. Since no inter-relay communications is allowed, the only way to implement a uniform grouping is to let the decoded nodes independently and uniformly choose the group they belong to. In other words, M -group is the optimal strategy under the constraint that no inter-relay communications is permitted when the channels $h_{r_j d}$'s are i.i.d. complex Gaussian random variables.

Remark 2.1. In general, K/M might not be an integer, which means that we need to round the fractional solution to obtain an integer-valued feasible solution. This also implies that M should not exceed K . Since the gap between the optimal solutions to the original integer program and the relaxed optimization problem is well understood in the literature, the rounding process will not be discussed here. The simulation results in Section 2.2.3 also show that the gap is negligible. Alternatively, if we use a greedy algorithm to solve the integer program (2.12), we should first allocate $K_i = \lfloor \frac{K}{M} \rfloor$ decoded nodes to each group. The remaining $K - MK_i$ decoded nodes can be randomly put into these groups.

In a more general case where the distances from the relays to the destination are not the same and the effect of path loss is also considered, the distribution of $P_s^{(i)}$ in (2.7) is affected not only by the number of nodes K_i in \mathcal{G}_i but also by the distances to the destination.

Let d_j denote the distance from the j th decoded node in the i th group \mathcal{G}_i to the destination and $\ell(d_j)$ represent the path loss on this link, where $\ell(d_j)$ is a continuous, positive, non-increasing function of d_j . Then, $P_s^{(i)}$ has an exponential distribution parametrized by

$$\tilde{\lambda}_i = \frac{1}{P \sum_{j \in \mathcal{G}_i} \ell(d_j)} \quad (2.20)$$

which is the reciprocal of the average total received power from the i th group \mathcal{G}_i when path loss is considered. Following the proof of Theorem 2.1, we can easily show that the approach utilized above is still applicable.

Corollary 2.1. A grouping strategy with $\tilde{\lambda}_i = \tilde{\lambda}_j, \forall i, j \in \{1, 2, \dots, M\}$ is the optimal solution for Problem (2.6) with $\Omega_{\text{th}} = 0$ when the distances from the decoded nodes to the destination are not the same and path loss is considered.

Corollary 2.1 shows that the optimal grouping strategy is to allocate the decoded nodes such that the average received power from each group is equal. However, this is impossible without central control or full inter-relay communications because we would

need to know the exact power gain for each link to obtain the optimal performance. M -group can, nevertheless, achieve a near-optimal performance without incurring extra overhead.

2.2.3 Simulation Results

In this section, we present results to verify the analysis. A two-hop cooperative network with 16 potential relays is considered. All relays are uniformly distributed in a square area, and the source and destination nodes are placed at opposite corners. We consider the path loss model in [40]

$$\ell(d) = \text{PL}_0 \left(\frac{d}{d_0} \right)^{-\alpha}, \quad (2.21)$$

where PL_0 is a constant that depends on the antenna characteristics and the average channel attenuation, d_0 is the reference distance for the antenna far field, and α is the path loss exponent. Because of scattering phenomena in the antenna near field, the model (2.21) is generally valid only at transmission distances $d > d_0$. We assume $\text{PL}_0 = 1$, $d > d_0 = 1$ and $\alpha = 4$, since different values for these parameters will not affect our analysis and simulation results. The largest possible path loss, i.e., the one from one corner to its opposite corner, is set to 86 dB. The source and the relays use the same transmit power. The Alamouti STBC scheme [27] is adopted to implement a two-column M -group decentralized STBC.

Depending on whether the effect of path loss is considered, we evaluate two different scenarios. In Figs. 2.3 and 2.4, only the effect of Rayleigh fading is considered. In Fig. 2.3, we choose the transmit power so that the average received SNR is 15 dB or 20 dB above the SNR threshold γ_{th} , i.e., the received SNR margin is 15 dB and 20 dB, respectively. In Fig. 2.4, the received SNR margin changes over a large range from 0 dB to 30 dB. By contrast, in Figs. 2.5 and 2.6, we present the outage performance in the presence of path loss. In Fig. 2.5, the transmit power is set so that the average received SNR at the destination contributed by a node close to the source is about 4 dB above the SNR threshold γ_{th} . In Fig. 2.6, the transmit power is chosen to let the

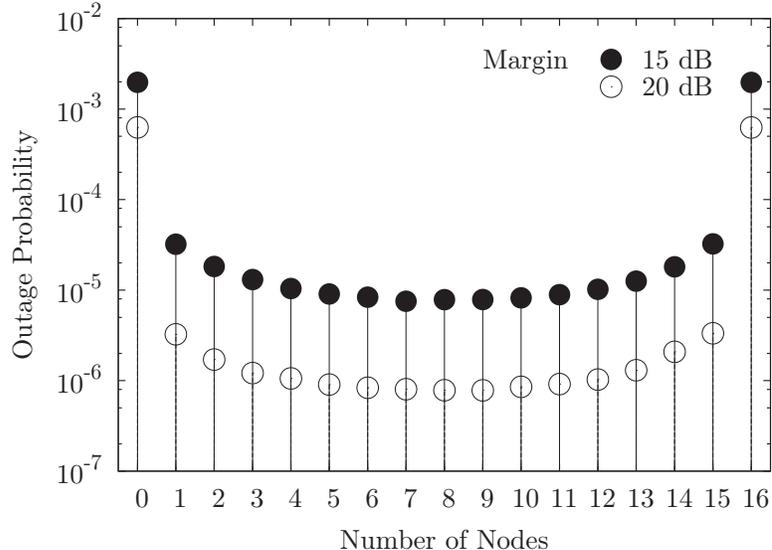


Figure 2.3: Outage probability as a function of the number of nodes in the first group. Only the effect of Rayleigh fading is considered. Two different SNR margins are used. The randomness in the first hop is ignored.

average received SNR margin caused by the node near the source change over a range from -36 to -6 dB.

In Fig. 2.3, the outage probability is plotted as a function of the number of decoded nodes in the first group. The randomness in the first hop is ignored, i.e., the transmissions between the source and the relays are assumed to be perfect and all 16 potential relays are selected as decoded nodes. The trends shown are very similar for the two values of transmit power. It can be seen that the outage probability always achieves the minimum value when the two groups have the same number of nodes, as expected. In addition, we can see that the outage performance changes only slightly if both groups are not empty; this is a good indication that M -group is robust against the uncertainty introduced by the random assignment.

Next, we examine how the outage performance varies as the transmit power increases. The outage probability curve is plotted as a function of the transmit power for several different scenarios. The curves labeled “Perfect 1st hop” are obtained by assuming all 16 potential relays are selected as decoded nodes. By contrast, the curves

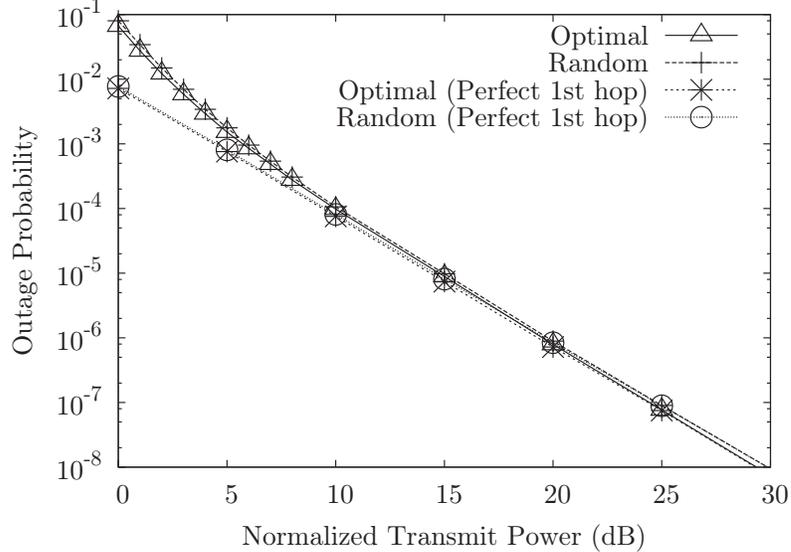


Figure 2.4: Outage probability as a function of the transmit power. Only the effect of Rayleigh fading is considered.

without such labels represent the cases in which the randomness in the first hop is also considered. As the transmit power increases, the outage performance increases with a slope of one decade every 5 dB. The comparison between the optimal grouping strategy (uniform group) and the random one (M -group) is shown. As expected, the optimal one has better outage performance, but the gap is negligible. It can also be observed that the randomness in the first hop is more significant when the transmit power is low. In the low-transmit-power regime, the decoded set \mathcal{D} is usually not full, and increasing the transmit power can effectively increase the decoded set, thus improving the outage performance. When the transmit power is high, however, the decoded set would be full almost all the time and the randomness in the first hop is negligible. Thus, the curves for the cases with and without a perfect first hop coincide.

We now evaluate the outage performance in the presence of path loss. The randomness of the relay locations is also considered. Recall that the path loss on the links from the decoded nodes to the destination is denoted by $\ell(d_j)$. The optimal grouping strategy, as in (2.20), is to divide the decoded set into two parts with equal sums of $\ell(d_j)$'s rather than equal numbers of nodes. Mathematically, finding a partition

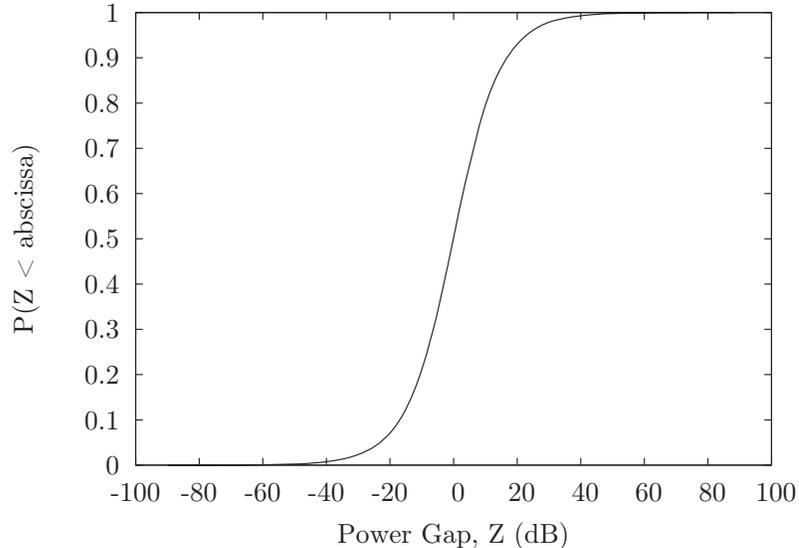


Figure 2.5: CDF of the gap between the average received power for the first and second groups. The effect of path loss is considered. The randomness in the first hop is ignored.

over the set of $\ell(d_j)$'s with the equal (or, as close as possible) sum is an NP-complete problem [41]. Fortunately, there are several available algorithms to achieve this [41].

First, we examine the effectiveness of M -group by evaluating the statistical properties of the gap between the average received power from the two randomly divided groups. The power gap is defined as the ratio of the average received power from the second group to that from the first group. Since all the relays use the same transmit power, the exact value of transmit power does not affect the statistics of the power gap. The randomness of the relay locations is taken into account. Assuming a perfect first hop, the cumulative distribution function (CDF) of the power gap is plotted in Fig. 2.5. It can be seen that the power gap varies around 0 dB, with a transition phase of about 40 dB. In other words, the two groups generated by M -group usually have a non-negligible difference in the average received power, which in turn leads to a performance degradation.

Second, the relationship between the outage probability and the transmit power is illustrated in Fig. 2.6. It can be seen that the gap between the optimal strategy and M -group is not negligible. This is consistent with the observation from Fig. 2.5.

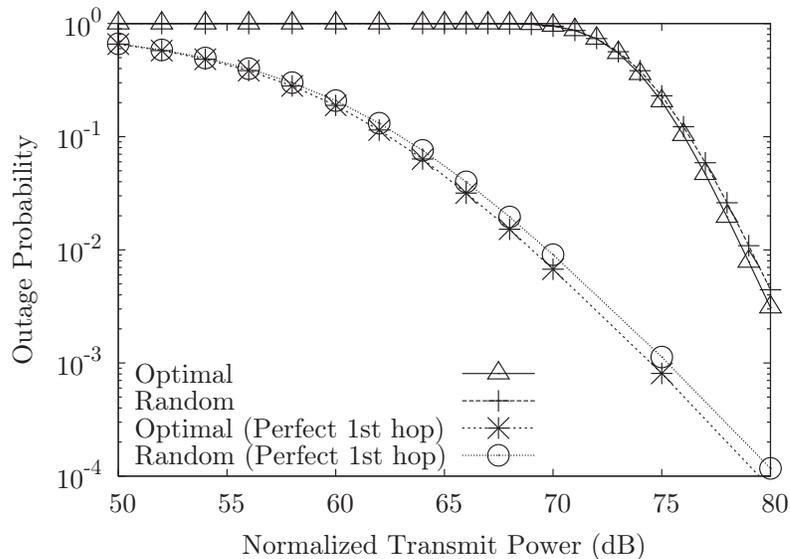


Figure 2.6: Outage probability as a function of the transmit power. The effect of path loss is included.

In other words, such a coin-flip-driven strategy may not always be satisfying, and a penalty in outage performance is inevitable to remove the overhead incurred by inter-relay negotiations. Intuitively, we are inclined to believe that a small and acceptable amount of overhead might improve the performance significantly. It can also be seen that the outage with a non-perfect first hop is 1 when the transmit power is less than 70 dB ². This is because no decoded nodes can be found at all in the first hop when the transmit power is not high enough.

2.3 Overhead-Performance Tradeoff for Best Relay Selection

Relay selection is another attractive technique for realizing the benefits of cooperation and enhancing the network performance [35]. After the source broadcasts its data, instead of using all the decoded nodes, the data is transmitted to the destination using the “best” relay in the decoded set.

² This value is relatively high since we assume that the noise power is 1. A practical example is to assume that the noise spectral density is -174 dBm/Hz and the bandwidth is 20 MHz. In this case, the noise power is -141 dB, and the outage for the first hop is 1 when the transmit power is less than -71 dB.

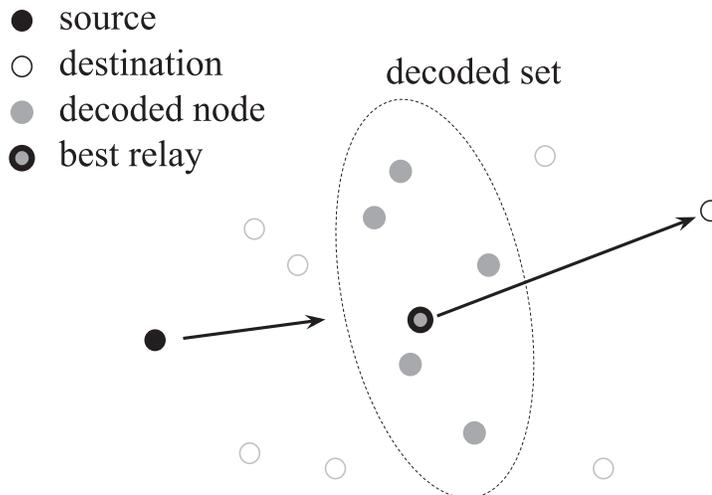


Figure 2.7: A two-hop decode-and-forward cooperative relaying network consisting of a source, a destination, and K potential relay nodes, from which the best relay is selected.

The first hop of the cooperative protocol remains the same as that introduced in Section 2.1. In the second hop, the relay which has the highest channel power gain to the destination will be selected, as illustrated in Fig. 2.7. Since only one node transmits in the second stage, it is reasonable to assume that the selected relay node always uses transmit power P_t . Let $h_{r_i,d}$ represent the channel coefficient from the i th relay to the destination and $g_i = |h_{r_i,d}|^2$ represent the corresponding channel power gain. Then, the “best” relay is the j th relay where

$$j = \operatorname{argmax}_{i=1,2,\dots,K} g_i. \quad (2.22)$$

The selected relay then transmits the source information to the destination.

In most existing papers, relay selection is assumed to be perfect and the overhead required is ignored. In practice, the implementation of cooperation among relay nodes incurs overhead including control signals and inter-node communication. It is intuitively obvious that better performance can be achieved if more CSI is available at the transmitter. However, the overhead required for obtaining and transmitting the CSI could be significant enough that the performance gain of using cooperation

remains uncertain. Thus, the overhead-performance tradeoff is an important issue for investigation.

Distributed relay selection algorithms which only require local information have been investigated in recent papers. In [36, 37], a distributed mechanism is described that is based on back-off timers to reduce the overhead for relay selection. Another approach, time-slotted splitting, is proposed in [38] to quantify the amount of overhead needed for the selection process. The overhead-performance tradeoff for the proposed time-slotted splitting algorithm is discussed in [39]. However, the results in [36–39] are highly dependent on the specific implementations. The fundamental questions remain unanswered: How much overhead is required for effective relay selection? What is the optimal tradeoff for general selective relaying networks?

In this section, we reformulate the overhead-performance tradeoff for selective relaying networks as an optimization problem by applying a rate distortion approach. It is straightforward to use the destination as a central controller that collects CSI and selects the relays with the highest channel gain. The destination needs to utilize the feedback channel to notify the selected relay. Our main contribution is to theoretically determine how many feedback bits are required for effective relay selection. By solving the proposed rate distortion problem, we obtain the minimum required amount of overhead. Note that, unlike [36–39], we are not specifying the implementations such as timer-based best select or the splitting algorithm.

2.3.1 Rate Distortion Formulation

The pioneering work in [42] provides a novel information-theoretic approach to study the overhead-performance tradeoff through rate distortion theory. The primary purpose of rate distortion theory is to find the optimal compression rate under a given distortion measure for lossy source coding problems. In [42], rate distortion theory is extended to analyze the minimum amount of overhead to be transmitted under a given time delay constraint. Several recent papers have also focused on this approach. In [43], the overhead required for transmitting traffic information for centralized scheduling

in a multiple access system is investigated. Rate distortion theory is used in [44] to derive a lower bound on the overhead for a specified location estimation error in geographical routing. The tradeoff between the routing overhead for link state updating and the maximum achievable transport capacity is studied in [45]. In [46], the overhead-performance tradeoff for a beamforming system with partial CSI has been quantified.

Here, we use the extended rate distortion theory to study the impact of overhead on the performance of cooperative relaying networks. In [43], the concept of distortion in conventional rate distortion theory is interpreted to be the difference between the optimal and actual network performance. Then, the minimum amount of overhead required for relay selection can be found by minimizing the mutual information between the correct and incorrect decision policies.

Define $\mathbf{u} = [u_1, u_2, \dots, u_K]$ as the best-select decision policy, i.e.,

$$u_i = \begin{cases} 1, & g_i = \max g_k \\ 0, & \text{otherwise} \end{cases}. \quad (2.23)$$

Obviously, $\Pr\{u_i = 1\} = \frac{1}{K}$ since all g_i have the same distribution. Suppose $\hat{\mathbf{u}}$ is the actual decision policy. If $\hat{\mathbf{u}} = \mathbf{u}$, i.e., the correct decision has been made based on perfect knowledge of the CSI, then the optimal performance can be achieved.

In reality, $\hat{\mathbf{u}} \neq \mathbf{u}$ since the actual decision policy is usually based on imprecise CSI. Suppose $\Phi = [\phi_{ij}]_{K \times K}$ is the transition probability matrix between \mathbf{u} and $\hat{\mathbf{u}}$, i.e.,

$$\phi_{ij} = \Pr\{\hat{u}_i = 1 | u_j = 1\}, \quad (2.24)$$

and $\Pr\{\hat{u}_i = 1\} = \sum_{j=1}^K \Pr\{\hat{u}_i = 1 | u_j = 1\} \Pr\{u_j = 1\} = \frac{1}{K} \sum_{j=1}^K \phi_{ij}$. Then, the mutual information between \mathbf{u} and $\hat{\mathbf{u}}$, which characterizes the amount of overhead required, is defined by

$$\mathcal{I}(\mathbf{u}, \hat{\mathbf{u}}) = H(\hat{\mathbf{u}}) - H(\hat{\mathbf{u}}|\mathbf{u}), \quad (2.25)$$

where [47]

$$\begin{aligned} \mathbf{H}(\hat{\mathbf{u}}) &= \mathbf{H} \left(\frac{1}{K} \sum_{j=1}^K \phi_{1j}, \frac{1}{K} \sum_{j=1}^K \phi_{2j}, \dots, \frac{1}{K} \sum_{j=1}^K \phi_{Kj} \right) \\ &= -\frac{1}{K} \sum_{i=1}^K \left[\left(\sum_{j=1}^K \phi_{ij} \right) \log \left(\frac{1}{K} \sum_{j=1}^K \phi_{ij} \right) \right], \end{aligned} \quad (2.26)$$

and

$$\mathbf{H}(\hat{\mathbf{u}}|\mathbf{u}) = \frac{1}{K} \sum_{i=1}^K \mathbf{H} \left(\frac{\phi_{1i}}{\sum_{j=1}^K \phi_{ji}}, \dots, \frac{\phi_{Ki}}{\sum_{j=1}^K \phi_{ji}} \right). \quad (2.27)$$

To describe the performance degradation when $\mathbf{u} \neq \hat{\mathbf{u}}$, we choose the ergodic capacity [40] as the performance measure when the decision policy is $\hat{\mathbf{u}}$. Let $\mathbf{g} = [g_1, \dots, g_K]$ be the channel gain vector. The ergodic capacity can be written as

$$\mathbf{C}(\hat{\mathbf{u}}) = \int_0^\infty \log(1 + \gamma \mathbf{g}^H \hat{\mathbf{u}}) f(\mathbf{g}) d\mathbf{g}, \quad (2.28)$$

where $\gamma = P_t/P_N$ is the average received SNR. Then, the performance degradation is $\Delta(\mathbf{u}, \hat{\mathbf{u}}) = \mathbf{C}(\mathbf{u}) - \mathbf{C}(\hat{\mathbf{u}})$. We can rewrite the distortion metric $\Delta(\mathbf{u}, \hat{\mathbf{u}})$ as Δ_{ij} , where $u_i = 1$ and $\hat{u}_j = 1$. Obviously, $\Delta_{ij} \geq 0$, and the equality holds if and only if $i = j$. Therefore,

$$\mathbb{E}[\Delta(\mathbf{u}, \hat{\mathbf{u}})] = \sum_{i=1}^K \sum_{j=1}^K \phi_{ij} \Delta_{ij} \leq \Delta_{\mathbf{c}} \left(1 - \frac{1}{K} \sum_{i=1}^K \phi_{ii} \right), \quad (2.29)$$

where $\Delta_{\mathbf{c}}$ is the supremum of the average capacity loss. For simplicity, we use this supremum to relax the inequality constraint in the rate distortion formulation. The following analysis and simulation results verify that the relaxation has minimal impact. With this simplification, the rate distortion function can be written as

$$\begin{aligned} R(D) &= \min_{\Phi} \mathcal{I}(\mathbf{u}, \hat{\mathbf{u}}) \\ \text{s.t. } &\Delta_{\mathbf{c}} \left(1 - \frac{1}{K} \sum_{i=1}^K \phi_{ii} \right) \leq D, \\ &\sum_{j=1}^K \phi_{ij} = 1, \quad \forall i = 1, 2, \dots, K. \end{aligned} \quad (2.30)$$

The problem in (2.30) is similar to the optimization problem proposed in [46]. However, we formulate a discrete and multivariate optimization problem; in [46], a continuous but univariate problem is considered. In addition, different scenarios are considered here.

2.3.2 Overhead-Performance Tradeoff

Since the average rate loss Δ_{C} in (2.30) is neither concave nor convex, the analytical solution is intractable. Here, we first consider two representative cases for analysis: the “best” and the “worst.” The “best” means $D = 0$, i.e., the optimal performance is achieved. In this case, we want to investigate the amount of overhead required to guarantee the optimal performance. The “worst” means $R(D) = 0$, i.e., no overhead is allowed. In this case, we want to derive a lower bound on the performance loss due to incorrect selection. Then, we propose a closed-form approximation for the rate distortion function by using an independence assumption.

Best Case: Lower Bound on Overhead

It is easy to show that $R(0) = \log K$. Because $D = 0$ and $\phi_{ij}\Delta_{ij} \geq 0$, the average distortion must satisfy

$$\mathbb{E}[\Delta(\mathbf{u}, \hat{\mathbf{u}})] = \sum_{i=1}^K \sum_{j=1}^K \phi_{ij}\Delta_{ij} \leq 0 \Rightarrow \phi_{ij}\Delta_{ij} = 0 \quad \forall i \text{ and } j, \quad (2.31)$$

which implies that $\phi_{ij} = 0$ for $i \neq j$. Note that since $\sum_{j=1}^K \phi_{ij} = 1$, the only feasible solution for (2.30) is the transition matrix $\Phi = \mathbf{I}$. In this case, $\hat{\mathbf{u}} = \mathbf{u}$ since, to achieve the optimal performance, no error in the selection policy is allowed. Hence, $H(\hat{\mathbf{u}}|\mathbf{u}) = 0$, and $R(D) = H(\hat{\mathbf{u}}) = \log K$.

Intuitively, $\log K$ is the number of information bits to transmit the index i , $1 \leq i \leq K$. Since it is equally likely for each relay to be the best relay, the entropy (uncertainty) of the index for the best relay is $\log K$. In other words, an overhead of at least $\log K$ bits is required to guarantee the optimal performance is achieved.

Worst Case: Lower Bound on Performance Loss

To analyze the minimum performance loss D_{\min} subject to $R(D) = 0$, we use the fact that $R(D) = \min \mathcal{I}(\mathbf{u}; \hat{\mathbf{u}}) = 0$ which indicates that \mathbf{u} and $\hat{\mathbf{u}}$ are independent. In other words, the elements ϕ_{ij} in the transition matrix Φ are

$$\phi_{ij} = \Pr\{\hat{u}_i = 1 | u_j = 1\} = \Pr\{\hat{u}_i = 1\}. \quad (2.32)$$

An intuitive explanation for (2.32) is that the selection policy must be random when we do not have any prior knowledge about the CSI or about the optimal selection policy at the transmitter. Since all the channels are i.i.d., the random selection criterion should be equally likely for any potential relay node. Then, (2.28) can be rewritten as

$$\begin{aligned} \mathbb{E}[C(\hat{\mathbf{u}})] &= \int_0^\infty \log(1 + \gamma g) f(g) dg \\ &= \log(e) \cdot e^{1/\gamma} \Gamma(0, 1/\gamma), \end{aligned} \quad (2.33)$$

where

$$\Gamma(m, x) = \int_x^\infty t^{m-1} e^{-t} dt \quad (2.34)$$

is the incomplete gamma function.

Therefore, the inequality constraint in (2.30) becomes

$$\mathbb{E}[\Delta(\mathbf{u}, \hat{\mathbf{u}})] \leq \Delta_{\text{c}} \left(1 - \frac{1}{K}\right) \leq D. \quad (2.35)$$

The equalities hold if and only if $p_{ij} = \frac{1}{K}$ for any $i, j = 1, 2, \dots, K$. Thus, D_{\min} is proportional to the average capacity loss Δ_{c} . The following theorem provides a closed-form expression for D_{\min} .

Theorem 2.2. When $R(D) = 0$,

$$\frac{D_{\min}}{\log(e)} = \frac{K-1}{K} \Delta_{\text{c}} \text{ bps/Hz}, \quad (2.36)$$

where

$$\Delta_{\text{c}} = \left[\sum_{k=1}^K (-1)^{k-1} \binom{K}{k} e^{k/\gamma} \Gamma(0, k/\gamma) - e^{1/\gamma} \Gamma(0, 1/\gamma) \right], \quad (2.37)$$

and $\binom{K}{k}$ is the binomial coefficient indexed by K and k .

Proof. We derive the ergodic capacity $\mathbb{E}[\mathbf{C}(\mathbf{u})]$ for best-select relaying with K decoded nodes as

$$\begin{aligned}
\frac{\mathbb{E}[\mathbf{C}(\mathbf{u})]}{\log(e)} &= \int_0^\infty \ln(1 + \gamma x) K (1 - e^{-x})^{K-1} e^{-x} dx \\
&= \int_0^\infty \ln(1 + \gamma x) d[(1 - e^{-x})^K - 1] \\
&= \lim_{R \rightarrow \infty} [(1 - e^{-R})^K - 1] \ln(1 + \gamma R) - \int_0^\infty [(1 - e^{-x})^K - 1] \cdot \frac{\gamma}{1 + \gamma x} dx.
\end{aligned} \tag{2.38}$$

Note that $(1 - e^{-x})^K - 1 = \sum_{k=1}^K (-1)^k \binom{K}{k} e^{-kx}$, which leads to $\lim_{R \rightarrow \infty} [(1 - e^{-R})^K - 1] \ln(1 + \gamma R) = 0$. Thus,

$$\begin{aligned}
\frac{\mathbb{E}[\mathbf{C}(\mathbf{u})]}{\log(e)} &= - \sum_{k=1}^K \int_0^\infty (-1)^k \binom{K}{k} e^{-kx} \frac{\gamma}{1 + \gamma x} dx \\
&= \sum_{k=1}^K (-1)^{k-1} \binom{K}{k} \gamma \int_0^\infty \frac{e^{-kx}}{1 + \gamma x} dx \\
&= \sum_{k=1}^K (-1)^{k-1} \binom{K}{k} e^{k/\gamma} \Gamma(0, k/\gamma).
\end{aligned} \tag{2.39}$$

Theorem 2.2 is proved by combining (2.33), (2.35) and (2.39). ■

The closed-form expression in (2.36), however, does not give much insight. To better understand this problem, we consider another important metric, the ratio between the capacity loss and the capacity. The capacity loss ratio is defined as

$$\rho(\gamma, K) = \frac{D_{\min}}{\mathbb{E}[\mathbf{C}(\mathbf{u})]} = \left(1 - \frac{1}{K}\right) \left(1 - \frac{\mathbb{E}[\mathbf{C}(\hat{\mathbf{u}})]}{\mathbb{E}[\mathbf{C}(\mathbf{u})]}\right). \tag{2.40}$$

Next, we provide some asymptotic properties of $\rho(\gamma, K)$.

Theorem 2.3. (i) As the average SNR γ approaches infinity, the capacity loss ratio $\rho(\gamma, K) \rightarrow 0$. (ii) As γ approaches 0,

$$\rho(\gamma, K) \rightarrow \left(1 - \frac{1}{K}\right) \left(1 - \frac{1}{H_K}\right), \tag{2.41}$$

where H_K is the K -th harmonic number defined by $H_K = \sum_{k=1}^K \frac{1}{k} = \int_0^1 \frac{1-x^K}{1-x} dx$.

Proof. (i) According to (2.33) and (2.39),

$$\frac{\mathbb{E}[\mathbf{C}(\mathbf{u})]}{\mathbb{E}[\mathbf{C}(\widehat{\mathbf{u}})]} = \sum_{k=1}^K (-1)^{k-1} \binom{K}{k} \psi_k(\gamma), \quad (2.42)$$

where

$$\psi_k(\gamma) = \frac{\int_0^\infty \frac{e^{-kx}}{1+\gamma x} dx}{\int_0^\infty \frac{e^{-x}}{1+\gamma x} dx} = \frac{\int_0^\infty \frac{e^{-ky/\gamma}}{1+y} dy}{\int_0^\infty \frac{e^{-y/\gamma}}{1+y} dy}. \quad (2.43)$$

Notice that $\int_0^\infty \frac{e^{-kx}}{1+\gamma x} dx \leq \int_0^\infty \frac{e^{-x}}{1+\gamma x} dx$, which indicates that if $\lim_{\gamma \rightarrow \infty} \psi_k(\gamma)$ exists, then $0 \leq \lim_{\gamma \rightarrow \infty} \psi_k(\gamma) \leq 1$. For any $k = 1, 2, \dots, K$, we have

$$\begin{aligned} \lim_{\gamma \rightarrow \infty} \psi_k(\gamma) &\stackrel{(a)}{=} \lim_{\gamma \rightarrow \infty} \frac{\int_0^\infty \frac{ky/\gamma^2 e^{-ky/\gamma}}{1+y} dy}{\int_0^\infty \frac{y/\gamma^2 e^{-y/\gamma}}{1+y} dy} \\ &= \lim_{\gamma \rightarrow \infty} \frac{k \int_0^\infty e^{-ky/\gamma} dy - k \int_0^\infty \frac{e^{-ky/\gamma}}{1+y} dy}{\int_0^\infty e^{-y/\gamma} dy - \int_0^\infty \frac{e^{-y/\gamma}}{1+y} dy} \\ &\stackrel{(*)}{=} \frac{1 - k \lim_{\gamma \rightarrow \infty} \int_0^\infty \frac{e^{-kx}}{1+\gamma x} dx}{1 - \lim_{\gamma \rightarrow \infty} \int_0^\infty \frac{e^{-x}}{1+\gamma x} dx} \stackrel{(b)}{=} 1, \end{aligned}$$

where y/γ is replaced by x in step (*). Equality (a) follows from L'Hospital's Rule. Although $\lim_{\gamma \rightarrow \infty} \int_0^\infty \frac{e^{-kx}}{1+\gamma x} dx$ and $\lim_{\gamma \rightarrow \infty} \int_0^\infty \frac{e^{-x}}{1+\gamma x} dx$ are not Riemann integrable, by using the Fatou-Lebesgue theorem [48], we can exchange the order of the limitation operation and the integral operation giving (b). Using (2.42) and (2.44),

$$\lim_{\gamma \rightarrow \infty} \frac{\mathbb{E}[\mathbf{C}(\mathbf{u})]}{\mathbb{E}[\mathbf{C}(\widehat{\mathbf{u}})]} = - \sum_{k=0}^K \binom{K}{k} 1^{K-k} (-1)^k + 1 = 1. \quad (2.44)$$

Hence, $\lim_{\gamma \rightarrow \infty} \rho(\gamma, K) = 0$.

(ii) Since $\lim_{\gamma \rightarrow 0} \int_0^\infty \frac{e^{-kx}}{1+\gamma x} dx$ and $\lim_{\gamma \rightarrow 0} \int_0^\infty \frac{e^{-x}}{1+\gamma x} dx$ are Riemann integrable, we have

$$\lim_{\gamma \rightarrow 0} \psi_k(\gamma) = \lim_{\gamma \rightarrow 0} \frac{\int_0^\infty \frac{e^{-kx}}{1+\gamma x} dx}{\int_0^\infty \frac{e^{-x}}{1+\gamma x} dx} = \frac{\int_0^\infty e^{-kx} dx}{\int_0^\infty e^{-x} dx} = \frac{1}{k}. \quad (2.45)$$

The combinatorial expression $\sum_{k=1}^K (-1)^{k-1} \frac{1}{k} \binom{K}{k}$ is equivalent to the K -th harmonic number H_K , which proves (2.41). ■

Notice that $H_K \rightarrow \infty$ as $K \rightarrow \infty$, which means that, in the low-SNR regime, the capacity loss approaches the capacity when K is large. However, most of the time, $\rho(0, K)$ is not large since $H_K = \mathcal{O}(\log K)$, which indicates that the capacity loss ratio for the low-SNR regime is tolerable in practice.

At first glance, it might seem that (2.41) is counterintuitive. To better explain (2.41), we can view the ratio $\rho(\gamma, K)$ as the difference between the slopes of two capacity curves. From an information-theoretic point of view, the capacity can be described through two key parameters: the minimum required energy per information bit and the asymptotic slope, especially in the low-SNR regime. In [49], it is shown that the value of the asymptotic slope varies when the knowledge of the CSI at the transmitter changes.

A Closed-Form Approximation

Here, we relax the original problem to a convex optimization problem by using an independence assumption. Suppose that the j th node has the highest channel power gain. We assume that $|h_{r_i d}|^2 - |h_{r_j d}|^2, i = 1, 2, \dots, K, i \neq j$ are i.i.d. random variables. Under this assumption, the following theorem provides a closed-form expression for the rate distortion function.

Theorem 2.4. The rate distortion function can be found as

$$R(D) = \log K - \frac{D}{\Delta_C} \log(K - 1) - \mathcal{H}\left(\frac{D}{\Delta_C}\right), \quad (2.46)$$

where $\mathcal{H}(x)$ is defined by

$$\mathcal{H}(x) = \begin{cases} -x \log x - (1 - x) \log(1 - x), & 0 < x < 1, \\ 0, & x = 0, 1 \end{cases} \quad (2.47)$$

and Δ_C is given in (2.37).

Proof. Given the independence assumption, we can transform the rate distortion problem (2.30) to the following optimization problem

$$\begin{aligned}
R(D) = \min_{\Phi} \quad & \mathcal{I}(\mathbf{u}; \hat{\mathbf{u}}) \\
\text{s.t.} \quad & \sum_{i=1}^K \phi_{ii} \geq KD \left(1 - \frac{D}{\Delta_c}\right), \\
& \sum_{j=1}^K \phi_{ij} = 1, \forall i = 1, 2, \dots, K.
\end{aligned} \tag{2.48}$$

According to the Karush-Kuhn-Tucker (KKT) conditions, we can solve (2.48) and obtain the closed-form solution

$$\Phi = \begin{bmatrix} 1 - \frac{D}{\Delta_c} & \frac{D}{(K-1)\Delta_c} & \cdots & \frac{D}{(K-1)\Delta_c} \\ \frac{D}{(K-1)\Delta_c} & 1 - \frac{D}{\Delta_c} & \cdots & \frac{D}{(K-1)\Delta_c} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{D}{(K-1)\Delta_c} & \frac{D}{(K-1)\Delta_c} & \cdots & 1 - \frac{D}{\Delta_c} \end{bmatrix}, \tag{2.49}$$

The rate distortion function $R(D)$ can then be obtained by combining (2.48) and (2.49). ■

In (2.46), we have $R(0) = \log K$, which implies that the minimum required number of feedback bits for perfect relay selection is $\log K$ bits. In particular, we can encode the index of K nodes by using $\log K$ bits if K is known *a priori*. On the other hand, when $R(D) = 0$, $D = \frac{K-1}{K}\Delta_c$ is the performance loss if there is no feedback. These results match our analysis for the two extreme cases.

2.3.3 Simulation Results

In this section, we present results to verify the analytical derivations. Since the objective function and the constraints in (2.30) are twice continuously differentiable, we can use sequential quadratic programming (SQP) [50] to numerically solve it. As a simple example, we choose $K = 2$; the transmit power is chosen so that the average received SNR is between 0 and 15 dB. In Fig. 2.8, the resulting rate distortion function $R(D)$, which represents the minimum amount of overhead, is plotted as a function of the performance gap D (in bps/Hz).

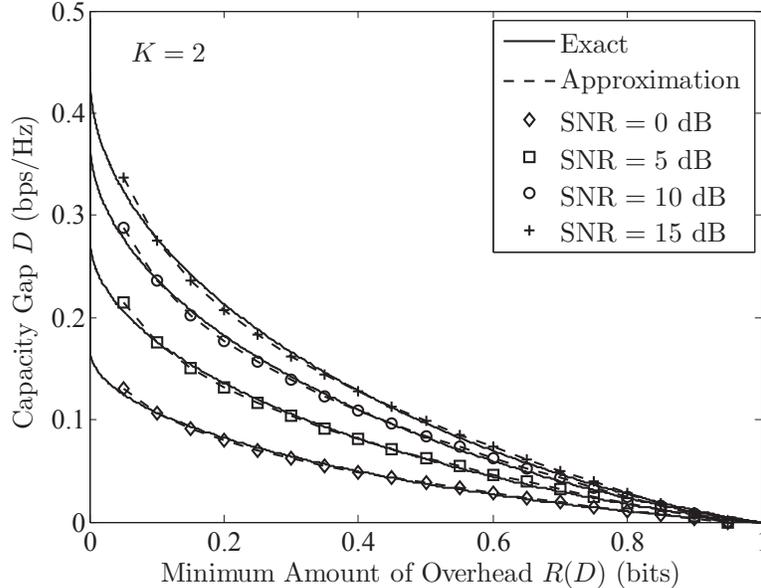


Figure 2.8: Simulation and Gaussian-fitting approximations for $R(D)$.

We can also use the following expression, called the Gaussian-fitting approximation [51], to evaluate $R(D)$,

$$R(D) \approx c_1 \exp\left(-\frac{(D + c_2)^2}{c_3}\right). \quad (2.50)$$

The parameters c_1 , c_2 and c_3 , which depend on the average SNR γ and the number of relay nodes K , can be estimated by using numerical methods. Using this approximation, we can easily obtain near-optimal lower bounds for $R(D)$ and D for any given system environment.

The approximation is suitable since the differences between the estimated values and the true values are negligible, which is also observed in Fig. 2.8. Table I shows the root-mean-square error (RMSE) of the estimated $R(D)$ for different γ .

γ	c_1	c_2	c_3	RMSE
0 dB	2.182	0.1152	0.0159	0.0021
5 dB	2.151	0.1874	0.0429	0.0035
10 dB	2.147	0.2509	0.0774	0.0031
15 dB	2.135	0.2953	0.1063	0.0036

Table 2.1: Estimated $R(D)$ and corresponding RMSE

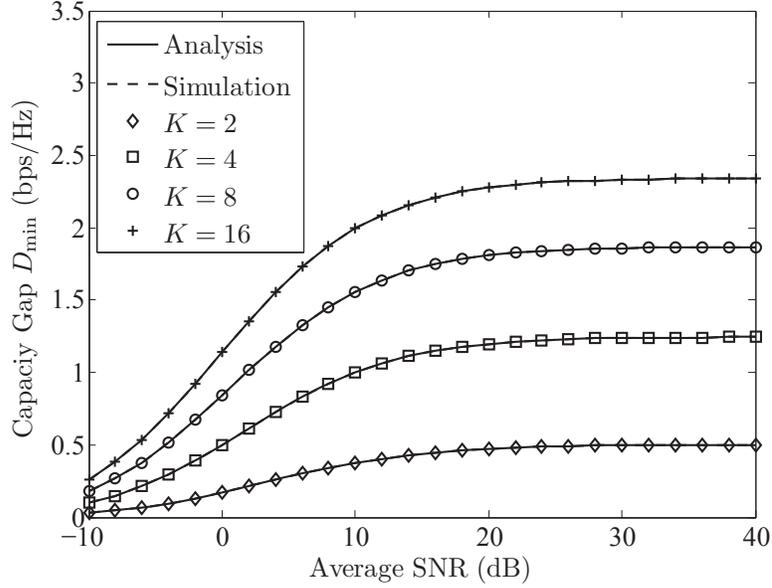


Figure 2.9: The lower bound on capacity loss D_{\min} when no overhead is allowed.

According to Fig. 2.8, we see that only one extra information bit is required to achieve the optimal performance for this simple example, which verifies our analysis for $R(0)$. In a real system, the receiver only needs to broadcast one bit of information to announce the index of the node which has the best channel power gain. Depending on the quality of the feedback channel, we can specify the achievable code rate and the practical cost for this feedback signal. The overhead for the receiver to obtain the estimates of these channel gains must also be computed in practice.

Figs. 2.9 and 2.10 provide the lower bound on the capacity loss D_{\min} and the capacity loss ratio ρ , respectively, for the case where no overhead is allowed ($R(D) = 0$). In the simulation, we assume $R(D) \approx 0$ if $R(D) \leq 10^{-10}$ to obtain the corresponding D_{\min} . According to the results, as expected, D_{\min} increases when we have more relay nodes since the selection gain is high for larger K . On the other hand, ρ will decrease as we increase the transmit power P_t , which indicates that the impact of selection gain becomes negligible in the high-SNR regime, also as expected.

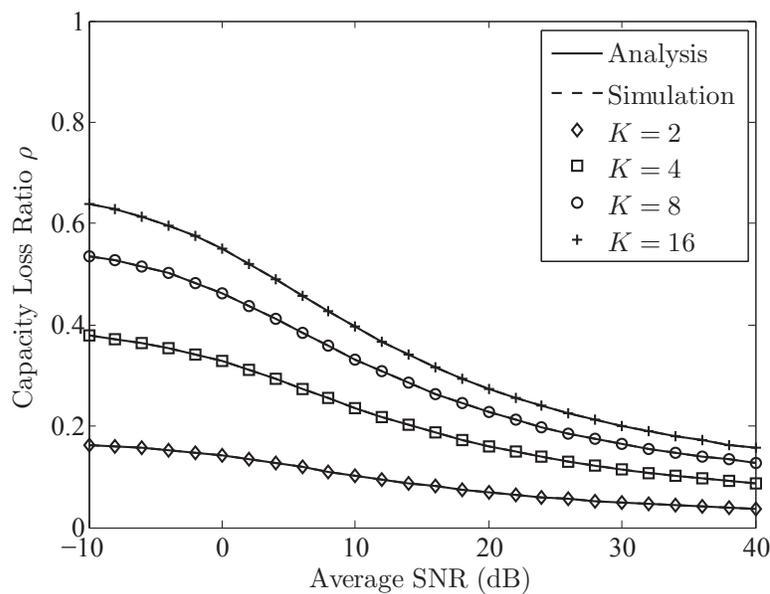


Figure 2.10: The capacity loss ratio $\rho(\gamma, K)$ when no overhead is allowed.

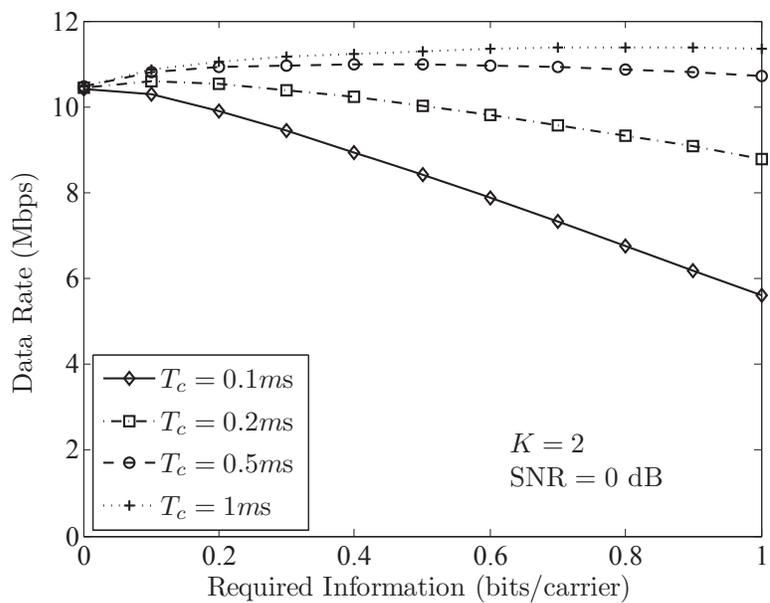


Figure 2.11: Data rate (performance) versus required information (overhead). ($B = 10$ MHz, 64 subcarriers)

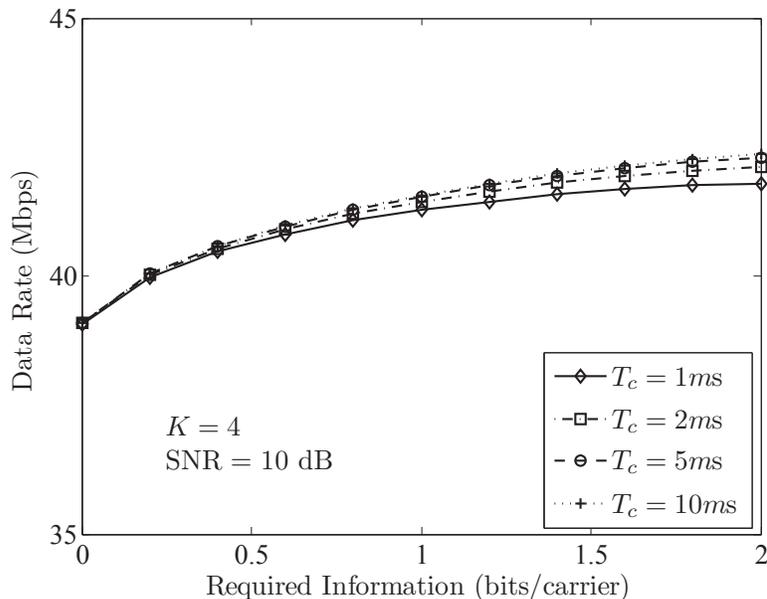


Figure 2.12: Data rate (performance) versus required information (overhead). ($B = 10$ MHz, 64 subcarriers)

Overhead-Performance Tradeoff for a Practical System

Now, we examine the overhead-performance tradeoff in a selective OFDM relaying system [52] consisting of K relays and one source-destination pair. We do not consider the time delay and bandwidth consumption that is needed for practical relay selection.

The tradeoff between data rate (performance) and required number of information bits (overhead) is plotted in Figs. 2.11 and 2.12. Assume that the node mobility and carrier frequency are such that the channel coherence time T_c is between 0.1 and 10 msec [53]. The bandwidth B is chosen to be 10 MHz, which is divided into 64 subcarriers. We set the length of each time slot to be $0.1T_c$, and in each time slot and each subcarrier, $R(D)$ bits is consumed to achieve capacity $C(\mathbf{u}) - D$ bps/Hz. In Fig. 2.11, we have two relays in the system, and the transmit power is chosen such that the average received SNR is 0 dB. In Fig. 2.12, there are four relays and the transmit power is chosen so that the average received SNR is 10 dB. Clearly, the extra overhead occupies a large part of the resource in the low-SNR regime with a small coherence time, and

more overhead does not lead to better performance, as we can see in Fig. 2.11. On the other hand, when the node mobility is low (large coherence time) and the SNR is high, the impact of $R(0)$ information bits becomes negligible. Note from Fig. 2.12 that we can always get a gain in performance with more overhead. Notice, also, that the data rate is not as sensitive to the coherence time when $K = 4$ and SNR = 10 dB.

2.4 Spectral Efficiency of Centralized and Decentralized Cooperative Networks with Relay Selection

As we mentioned, an essential issue for designing a relay selection scheme is how to select the “best” relay out of all available cooperating nodes. A straightforward approach is to use the destination as a central controller that selects the best relay. In Section 2.3, we introduced this centralized feedback-based relay selection scheme and quantified the feedback overhead required for a given performance loss. The amount of overhead incurred by channel estimation, however, was not studied.

The centralized implementation of relay selection usually involves extra control signals and inter-node communications, which can degrade the overall system performance, especially for systems with a large number of relays. Decentralized selection mechanisms, which only require local information, have been investigated for overcoming these implementation issues. A distributed mechanism, which is based on back-off timers to reduce the required overhead for relay selection, is described in [36,37,54,55]. Another approach, called the time-slotted splitting algorithm, is proposed in [38,39,56] to characterize the amount of overhead needed for a specified selection process.

Intuitively, decentralized relay-selection schemes incur much less overhead than centralized schemes, but at the cost of performance loss. In a decentralized implementation, such as timer-based relay selection, transmissions may fail due to an unsuccessful relay-selection process. By contrast, centralized schemes usually achieve optimal performance, but with a significant amount of overhead. Thus, it is valuable to study which scheme to choose for a given transmission environments.

In this section, we follow the same system model discussed in Section 2.3. Overhead analyses for different approaches of relay selection are first provided. By comparing the optimum spectral efficiency for centralized and decentralized schemes, the criteria for deciding which scheme should be applied for a given environment is then investigated. We show that, if the number of relays is small or the channel is relatively static, the effect of overhead for centralized schemes becomes negligible, implying that centralized schemes should be employed. On the other hand, if there are many available relays or the channel changes rapidly, the decentralized schemes outperform the centralized approaches due to the reduced amount of overhead.

We choose the spectral efficiency [14] as our performance metric to compare the centralized and decentralized schemes. It is defined as the successfully delivered bits from the source to the destination per channel use,

$$\eta = \frac{RT - T_o}{2T}, \quad (2.51)$$

where R is the transmission rate, T_o is the expected time consumption for overhead, T is the total transmission time, and the factor of $1/2$ comes from the two-stage transmission. Note that R can be the ergodic capacity (for variable-rate transmission) or the outage capacity (for fixed-rate transmission) [40].

Here, we investigate the spectral efficiency of two different schemes: centralized (feedback-based) and decentralized (timer-based) relay selection. Different types of overhead including channel estimation and relay selection are discussed.

2.4.1 Centralized Feedback-Based Relay Selection (FBRS)

It is straightforward to use the destination as a central controller that selects the relay with the highest channel gain. In order to collect global channel information, each relay node has to first send pilot symbols to the destination. By using pilot symbols which are known at all nodes in the network, the destination can obtain accurate channel estimates. In [57], it has been shown that the optimum length of the pilot symbols which maximizes the capacity is equal to the number of antennas. Therefore,

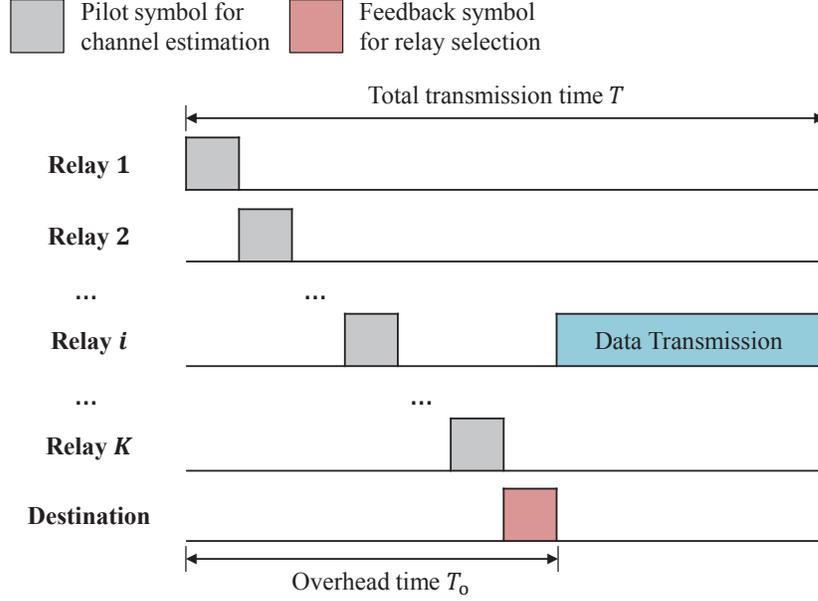


Figure 2.13: A timing diagram for FBRs. Each relay node sends a pilot symbol such that the destination can estimate the channels. Then, the destination broadcasts the relay selection results via a feedback packet. Afterwards, the selected relay starts data transmission.

we assume that the destination can obtain perfect knowledge of the channel power gain $|h_{r_i,d}|^2$ after the i th node sends a pilot symbol. In other words, the minimum time consumption for the channel estimation overhead is Kt_s , where K is the number of potential relay nodes, and t_s is the symbol duration.

The destination can then utilize the feedback channel to notify the selected relay. The time consumption for feedback, t_f , is determined by the size of the feedback packets and the feedback rate R_{fb} . We can easily obtain closed-form expressions for t_f based on our analysis in Section 2.3. As illustrated in Fig. 2.13, the time consumption for overhead is

$$T_o = Kt_s + t_f. \quad (2.52)$$

According to (2.46), we can choose an appropriate value of Δ for variable-rate transmission to maximize the spectral efficiency defined in (2.51). Here, we focus on fixed-rate transmission and assume that $\Delta = 0$ for simplicity. In this case, the spectral

efficiency is

$$\eta_{\text{FBRS}} = \frac{R_d(1 - p_{\text{out}})T - Kt_s - t_f}{2T}, \quad (2.53)$$

where R_d is the fixed transmission rate and p_{out} is the outage probability for relay selection, given as [20]

$$\begin{aligned} p_{\text{out}} &= \Pr \left\{ \log_2 \left(1 + \frac{P_t}{P_n} \max_i |h_{r_i d}|^2 \right) < R_d \right\} \\ &= \left(1 - e^{-\frac{P_n}{P_t}(2^{R_d} - 1)} \right)^N. \end{aligned} \quad (2.54)$$

2.4.2 Decentralized Timer-Based Relay Selection (TBRS)

TBRS is first proposed in [36] and has been shown to perform well. A timing diagram for TBRS is shown in Fig. 2.14. At first, the destination sends a pilot symbol to all nodes. Exploiting the reciprocity of wireless links, all relay nodes are able to estimate their own channel power gains. Based on this local channel information, every relay node sets up an individual timer so that the best node will have the shortest timer. Once the first timer expires (at the best node), the node starts its data transmission, and the other nodes back off. Consequently, TBRS always ensures that the timer of a node with a larger metric (for example, channel power gain) expires earlier than that of a node with a smaller metric.

Ideally, the best relay can always be selected successfully; however, in practice, the selection process may fail due to collisions. As discussed in [14, 36], due to the different propagation delays and the processing delay to prepare the information packet, it is possible that, before overhearing the transmission from the best relay, the timers at the other relays have already expired. In this case, more than one relay will rebroadcast the signal; a collision occurs and the destination cannot decode the message correctly (note that no capture effect is taken into account). Therefore, to avoid collisions and achieve a successful selection, the second minimum timer must be larger than the sum of the minimum timer and a guard interval, t_g , the value of which depends on the capabilities of the system [36].

The value of t_g , which depends on system capabilities, can be viewed as a pre-defined parameter. As discussed in [36], the guard interval t_g includes the propagation delay, the processing delay, synchronization offset, and the transmission symbol length. Typically, the delays and offset are negligible compared to the symbol length. Therefore, without loss of generality, we assume that t_g also represents the symbol length.

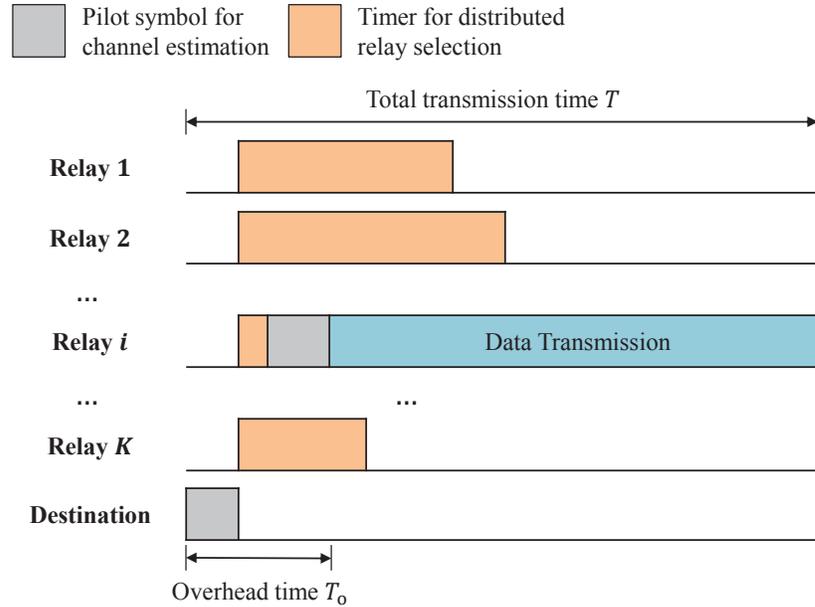


Figure 2.14: A timing diagram for TBRS. The destination broadcasts a pilot symbol such that all nodes can estimate the channels. Each node sets a timer. Once the first timer expires, the corresponding node sends a pilot symbol to ensure that the destination can estimate the channel. Then, it starts data transmission, and all other nodes back off.

Inverse Timer

In [36], an inverse timer is adopted, i.e., an individual timer is set as inversely proportional to the channel power gain, that is, $t_i = \frac{\lambda}{|h_{r_i d}|^2}$ (λ is a constant system parameter). We denote the ordered sequence of timers as $t_{(1)} \leq t_{(2)} \leq \dots \leq t_{(K)}$.

Then, a collision occurs if $t_{(2)} < t_{(1)} + t_g$ and the collision probability, p_{coll} , is [14]

$$\begin{aligned} p_{\text{coll}} &= Pr(t_{(2)} < t_{(1)} + t_g) \\ &= 1 - K(K-1) \int_{t_g}^{\infty} \frac{f(x)F(x-t_g)}{(1-F(x))^{2-K}} dx \end{aligned} \quad (2.55)$$

where

$$F(x) = Pr\{t_i \leq x\} = Pr\left\{\frac{\lambda}{|h_{r_i d}|^2} \leq x\right\} = \begin{cases} e^{-\frac{\lambda}{x}} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (2.56)$$

is the CDF of t_i , and

$$f(x) = \frac{dF(x)}{dx} = \begin{cases} \lambda x^{-2} e^{-\frac{\lambda}{x}} & x > 0 \\ 0 & x \leq 0 \end{cases} \quad (2.57)$$

is the corresponding probability density function (PDF).

As for the selection overhead in TBRS, in addition to the possible collisions, a duration t_{sel} is consumed by the selection process, during which there is no data transmission. According to [14], the expected selection time t_{sel} is the expectation of the minimum timer, $t_{(1)}$. Given K , the CDF of $t_{(1)}$ is

$$\begin{aligned} Pr(t_{(1)} \leq x) &= 1 - Pr(t_{(1)} > x) = 1 - \prod_i Pr(t_i > x) \\ &= 1 - (1 - F(x))^K = 1 - (1 - e^{-\frac{\lambda}{x}})^K \end{aligned} \quad (2.58)$$

Then, the PDF is

$$f(x)_{t_{(1)}} = K(1 - e^{-\frac{\lambda}{x}})^{K-1} e^{-\frac{\lambda}{x}} \lambda x^{-2} \quad (2.59)$$

Therefore, the expected selection time t_{sel} is

$$t_{\text{sel}} = \mathbb{E}(t_{(1)}) = K\lambda \int_0^{\infty} \frac{e^{-\frac{\lambda}{x}}}{x} (1 - e^{-\frac{\lambda}{x}})^{K-1} dx, \quad (2.60)$$

where $\mathbb{E}(\cdot)$ represents expectation. The following theorem provides a closed-form expression for t_{sel} .

Theorem 2.5. The expected selection time

$$t_{\text{sel}} = \lambda \sum_{i=0}^{K-1} \binom{K}{i} (-1)^{K-i} (K-i) \ln(K-i), \quad (2.61)$$

where $\binom{K}{i}$ is the binomial coefficient indexed by K and i .

Proof. We can prove the theorem by rewriting (2.60) as

$$\begin{aligned} t_{\text{sel}} &= K\lambda \int_0^\infty \frac{1}{x} e^{-\frac{\lambda}{x}} \left(1 - e^{-\frac{\lambda}{x}}\right)^{K-1} dx \\ &\stackrel{(a)}{=} K\lambda \int_0^\infty \frac{1}{x} e^{-\frac{\lambda}{x}} \sum_{i=0}^{K-1} \binom{K-1}{i} (-1)^{K-1-i} e^{-\frac{\lambda(K-1-i)}{x}} dx \\ &\stackrel{(b)}{=} \lambda \sum_{i=0}^{K-1} \binom{K}{i} (-1)^{K-1-i} \int_0^\infty \frac{1}{x} e^{-\frac{\lambda(K-i)}{x}} dx \\ &\stackrel{(c)}{=} \lambda \sum_{i=0}^{K-1} \binom{K}{i} (-1)^{K-i} (K-i) \ln(K-i), \end{aligned} \quad (2.62)$$

where step (a) results from the binomial theorem. According to the Fubini theorem [48], we can exchange the order of the integral operation and summation giving (b). Step (c) can be derived based on the results in [58]. ■

We can see that t_{sel} is only related to the parameter λ and the number of potential relay nodes K . Fig. 2.15, which plots t_{sel}/λ as a function of K , verifies the accuracy of Theorem 2.5.

Clearly, the collision probability p_{coll} and the selection time t_{sel} are functions of the timer setting parameter λ . If λ is large, the collision probability can be reduced at the cost of a long selection time. By contrast, a small λ leads to a high collision probability and short selection time. If we can adaptively choose the value of λ , the spectral efficiency of fixed-rate TBRS is

$$\eta_{\text{TBRS}} = \frac{R_d(1 - p_{\text{out}})}{2} \max_{\lambda} \left[(1 - p_{\text{coll}}) \frac{(T - t_s - t_{\text{sel}})}{T} \right]. \quad (2.63)$$

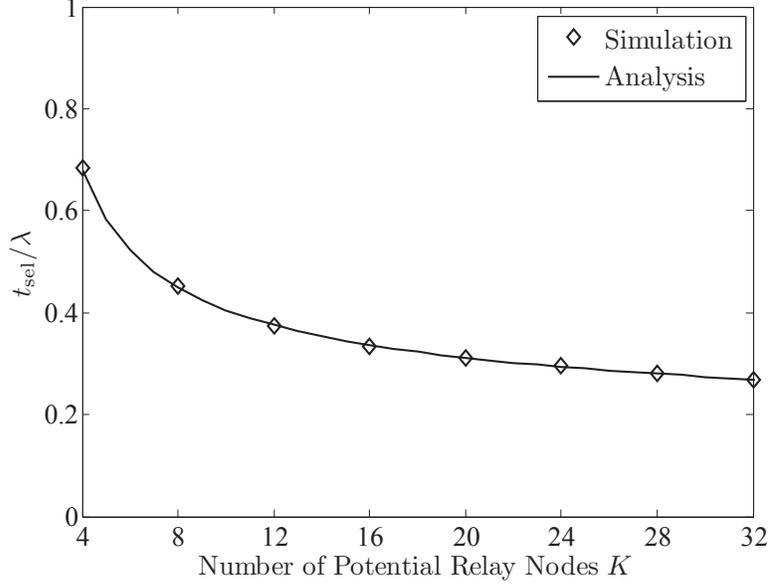


Figure 2.15: t_{sel}/λ versus the number of potential relay nodes K .

Optimal Timer

Instead of considering the inverse timer, which is a straightforward approach and only achieves suboptimal performance, the optimal metric-to-timer mapping should be investigated. Mathematically, the objective is to find an optimal mapping in the space of all monotone non-increasing functions that maximizes η_{TBRS} . Note that both the collision probability p_{coll} and the expected time overhead T_o depend on the mapping function. In general, finding the optimal mapping function is an intractable functional optimization problem.

In [54], it has been shown that an optimal mapping that minimizes p_{coll} within a maximum allowable selection time T_{max} maps the CDF of the channel power gain into several discrete timer values. That motivates us to decompose our goal into two subproblems: 1) minimizing p_{coll} for a fixed T_{max} and 2) choosing the appropriate value of T_{max} to maximize η_{TBRS} . Although the decomposition results in suboptimal solutions, we show that the performance of the proposed design is much better than existing designs.

For a fixed T_{max} , as the number of relays K goes to infinity, an asymptotic result

for p_{coll} is provided in [54]

$$p_{\text{coll}} = \underbrace{\exp(-[1 - \exp(\cdots - [1 - \exp(-1)]) \cdots])}_{K_T \text{ levels}} \quad (2.64)$$

where $K_T = \lfloor T_{\text{max}}/t_g \rfloor$ is the number of distinct timer values, and $\lfloor \cdot \rfloor$ is the rounding operator which chooses the largest integer which is not greater than the operand. Here, we choose an empirical approximation of (2.64) for analytical simplicity

$$p_{\text{coll}} \approx \exp\left(-\frac{c_1}{K_T}\right), \quad (2.65)$$

where $c_1 \approx 1.78$ can be obtained through numerical evaluations. Through simulation results in Section 2.4.3, we will show that these asymptotic and empirical results are tight approximations even when the number of relays K is small.

Another empirical observation is that the average selection time T_{sel} does not change as the number of relays K changes. In fact, we can approximate the average selection time as a linear function of T_{max} , i.e.,

$$T_{\text{sel}} \approx c_2 T_{\text{max}}, \quad (2.66)$$

where $c_2 \approx 1/3$. Note that the overhead time T_o also needs to include the time for transmitting additional pilot symbols so that the destination can accurately estimate the channel from the selected relay. In [57], it has been shown that the optimum length of the pilot symbols which maximizes the capacity is equal to the number of antennas. Therefore, we assume that the destination can obtain perfect knowledge of the i th node's channel power gain after it sends a single pilot symbol. In other words, the minimum time consumption for the channel estimation overhead is the symbol duration t_g .

Combining (2.65) and (2.66), we obtain the normalized net throughput

$$\eta_{\text{TBRS}} \approx \left[1 - \exp\left(-\frac{c_1}{\alpha\tau}\right)\right] \left(1 - \frac{1}{\tau} - c_2\alpha\right), \quad (2.67)$$

where $\alpha = T_{\text{max}}/T_c \in [0, 1]$ is the fraction of the maximum allowable selection time over the block length, and $\tau = T_c/t_g$ is the ratio between the block length and the symbol duration.

Now ρ can be optimized over α , and the optimal values of T_{\max} for any given T_c can be determined. According to the Karush-Kuhn-Tucker conditions, the value of α maximizing (2.67), α^* , satisfies

$$\exp\left(\frac{c_1}{\tau\alpha^*}\right) - \frac{c_1}{\tau\alpha^*} - \left(\frac{c_1}{\tau\alpha^*}\right)^2 \frac{1-\tau}{c_1c_2} = 1. \quad (2.68)$$

Although (2.68) does not have a closed-form analytical solution, we can easily obtain α^* via numerical methods.

The solution of (2.68) does not depend on the number of relays as we are using an asymptotic approximation ($K \rightarrow \infty$). Note that c_1 and c_2 are constants which can be evaluated based on the distribution of the channel power gain, and τ is also a constant which characterizes the time varying property of the wireless channel. So, once we have an accurate channel model, the optimal α can be uniquely determined.

2.4.3 Simulation Results

In this section, simulation results are presented to justify the empirical and analytical results. A comparison between centralized and decentralized schemes is also provided.

The plots in Fig. 2.16 validate our proposed empirical approximations (2.65) and (2.66). In the first two plots, the probability of successful selection $p_{\text{suc}} = 1 - p_{\text{coll}}$ is plotted as a function of K (the number of relays³) and τ (the ratio between total transmission time T_c and symbol length t_g). In the last plot, the expected selection time T_{sel} is plotted as a function of T_{\max}/t_g for different values of K . We observe that our approximations are very tight for almost all ranges of K and τ .

³ Here, we assume K is a fixed parameter. In practice, the number of active relays should be a random variable. Since our main focus is the comparison between the centralized and decentralized schemes, we assume that, for each realization, the number of active relays is the same for both schemes. Under this assumption, the randomness of the decoded set does not alter the comparison results.

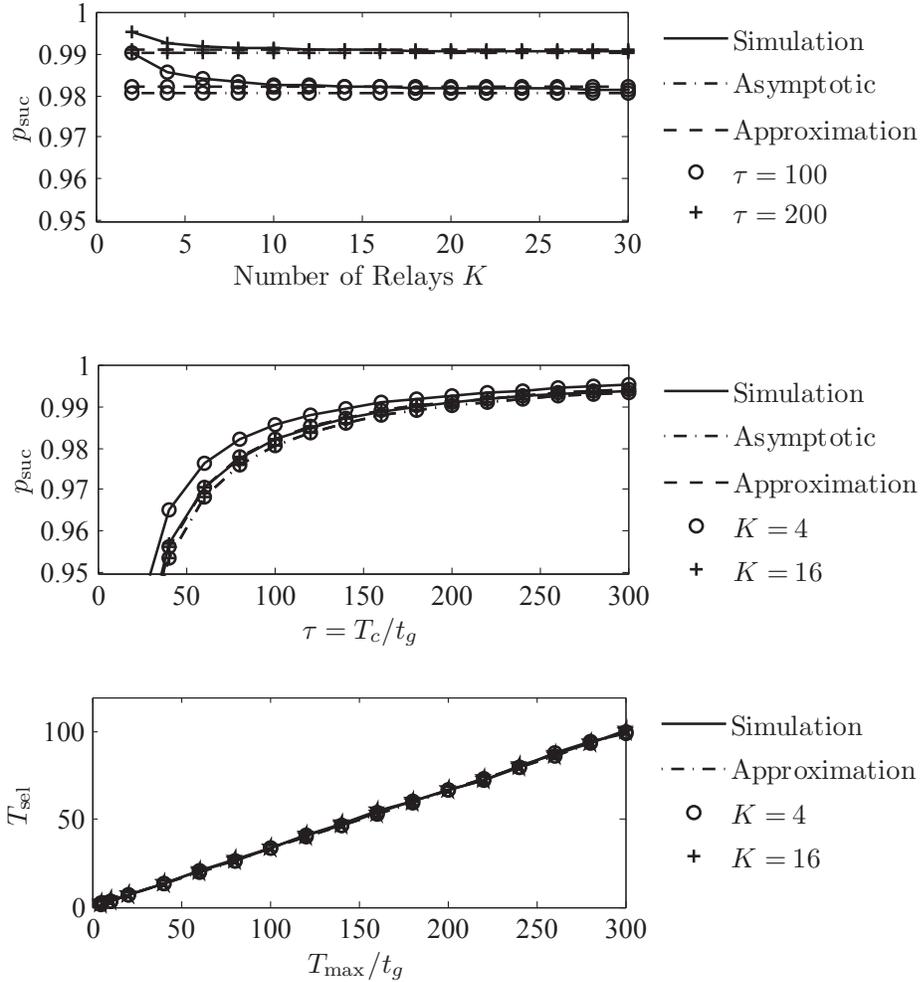


Figure 2.16: Validation of empirical approximations (2.65) and (2.66).

In Figs. 2.17 and 2.18, we present the normalized spectral efficiencies⁴ for different selection schemes: the proposed timer which maximizes the spectral efficiency (SE) defined by (2.67), the timer discussed in [54] which minimizes the collision probability, the inverse timer which maximizes (2.63), and the centralized feedback-based selection scheme. For the timer minimizing the collision probability, the maximum allowable selection time is set to be T_c . Fig. 2.17 illustrates the relationship between the normalized spectral efficiency and the number of relays K , where $\tau = 40$. In Fig. 2.18, the normalized spectral efficiency is plotted as a function of τ , which varies between 0 and 300. The number of relays K is 16, and the feedback transmission rate R_{fb} is set to be 6.5 Mbps over a 20 MHz band.

We observe that our proposed timer design achieves much better performance than the timer in [54] and the inverse timer, as expected. Note that the timer in [54] does not jointly consider the collision probability and the selection overhead, which explains why it does not perform well with respect to ρ . Another observation is that, when τ and T_c are small (i.e., the channel changes rapidly), the performance of the centralized scheme is very bad since the pilot symbols and feedback symbols occupy almost the entire transmission time. In this case, timer-based schemes are more appealing due to their low overhead requirements. By contrast, the centralized scheme outperforms all other schemes when τ and T_c are sufficiently large because of the unavoidable collisions in timer-based schemes. The impact of overhead on the net throughput becomes negligible as τ increases, and the performance for timer-based schemes is limited by the collision probability.

Although increasing K achieves a higher diversity order, it also incurs more overhead for channel estimation and relay selection (see Fig. 2.17). For any selection scheme, the spectral efficiency eventually degrades as K increases due to the increasing amount of overhead. We observe that the performance of the centralized scheme

⁴ Here, we normalized the spectral efficiency η by the desired transmission rate R_d . Since fixed-rate transmissions are assumed, the normalization does not change our results.

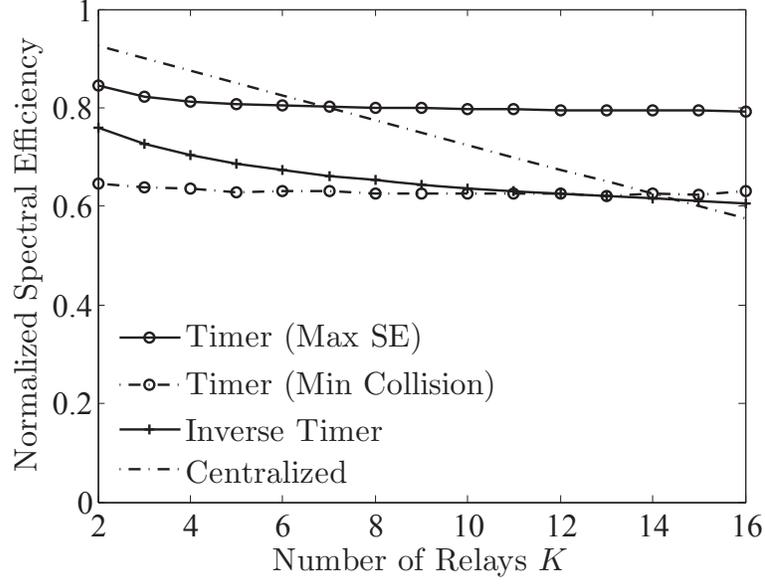


Figure 2.17: Normalized spectral efficiency as a function of the number of relays K .

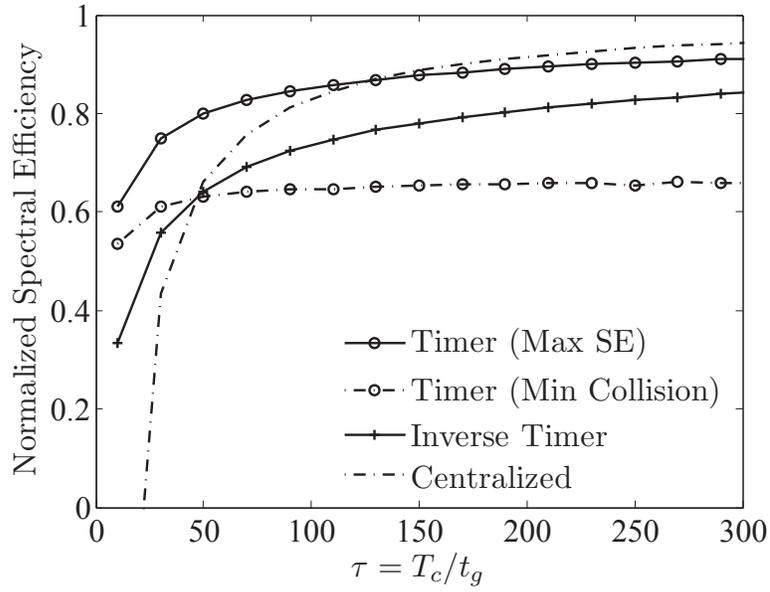


Figure 2.18: Normalized spectral efficiency as a function of $\tau = T_c/t_g$.

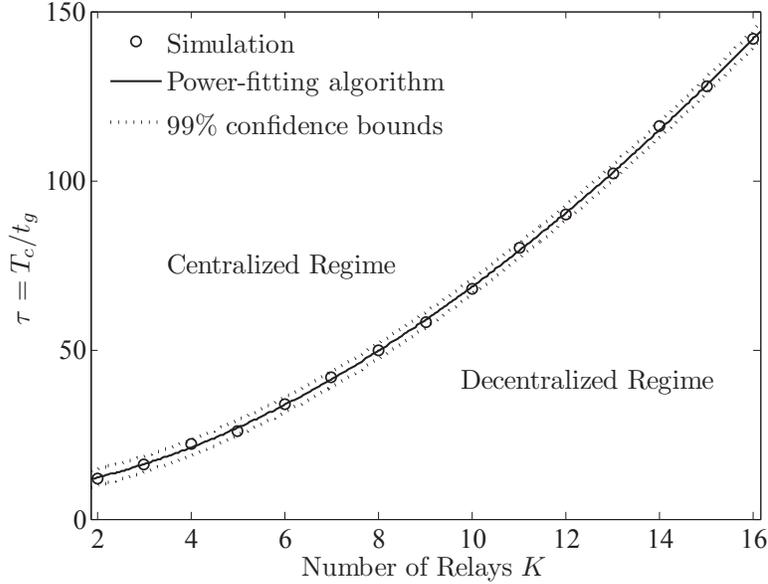


Figure 2.19: Operating regimes for centralized and decentralized relay selection schemes for different values of K and τ .

decreases linearly since the amount of channel estimation overhead is a linear function of K . On the other hand, timer-based schemes, which require less overhead than the centralized scheme, can support more relays, without losing much in performance.

The results in Figs. 2.17 and 2.18 depend on system parameters such as K and τ . In order to demonstrate the joint impact of these parameters, we plot the operating regimes in Fig. 2.19 to illustrate when we should use centralized or decentralized schemes. By using a power-fitting approximation [51], we provide criterion for determining which scheme should be applied: choose the centralized scheme when $\tau > c_1 K^{c_2} + c_3$. The parameters $c_i, i = 1, 2, 3$ can be estimated by numerical methods. Using this approximation, we can easily obtain a practical rule of thumb to efficiently perform relay selection.

The approximation is suitable since the differences between the estimated values and the true values are small, which is also observed from the 99% confidence bound in Fig. 2.19. In particular, if we choose $c_1 \approx 1.221$, $c_2 \approx 1.694$, and $c_3 \approx 8.31$, the root-mean-square estimation error in Fig. 2.19 is around 0.6869, which is negligible compared with the value of τ which is on the order of 10 or 100.

2.5 Summary

In this chapter, we studied an optimal cooperative STBC transmission strategy assuming no inter-relay communications or central control. A general optimization problem and a specialized case were formulated to minimize the outage probability subject to this constraint. We proved that when only Rayleigh fading is present, M -group decentralized STBC is the optimal transmission strategy. The optimal strategy for more realistic environments was also proposed; however, it is impractical due to the violation of the overhead constraint. Simulation results showed that M -group can achieve a near-optimal performance without incurring additional overhead.

We also used an information-theoretic approach to quantify the tradeoff between overhead and performance loss for a simplified cooperative system. A generalized optimization problem based on rate distortion theory was formulated to characterize the overhead-performance tradeoff. We discussed the asymptotic properties and an approximation of the rate distortion function, which were verified by simulation results.

Performance analyses for centralized and decentralized relay-selection schemes are also presented in this chapter. For different relay selection approaches, we quantified the required overhead and the optimum spectral efficiency. Closed-form analyses have been provided based on empirical approximations. A comparison between centralized and decentralized selection schemes is also provided. We show the superiority of centralized scheme when the impact of overhead is insignificant. Conversely, decentralized schemes are preferred if the amount of overhead has a significant effect on the performance.

There are several potential extensions of this work. For example, it might be possible to improve the performance of the STBC and M -group schemes by incurring a small and acceptable amount of overhead. Although we have proved that M -group is optimal under the zero-overhead constraint, the optimal schemes for more general cases have not been studied. Also, one could extend our comparisons between the general centralized and decentralized schemes to more realistic scenarios. For example, we can analyze other cooperative techniques such as beamforming and space-time coding.

Another example is to include the randomness of the first stage into our analysis. In addition, the comparisons in terms of ergodic capacity for variable-rate transmission might be further studied. Providing more accurate models for quantifying the required overhead is another challenging issue.

Chapter 3

MULTI-USER POISSON NETWORK WITH COOPERATION

As discussed in Chapter 2, it has been shown that the quality of a single wireless link can be significantly enhanced by means of cooperative communication. In a wireless network of moderate or large size, however, the situation becomes more complicated because there may exist many users which share the same time and frequency resources. Since cooperative communications usually involves multiple-node transmissions, it may generate additional spatial interference to other concurrent data transmissions in the network. The aggregate interference power can easily be strong and cause negative impact on the performance of the network. In other words, although each cooperative transmission has the potential to improve the reception quality of its destination, the overall performance taking into account the increased interference level remains unclear. Therefore, it can be expected that the impact of cooperation on the performance of a network is complicated. However, most of the existing literature focuses only on the local beneficial effect of cooperative communications and neglects its possible drawbacks.

Some efforts have been devoted to evaluating the performance of large-scale wireless networks using cooperative communications [59]. Due to the distributed and dynamic nature of the network, complex higher layer protocols are usually taken into account, which makes the analytical study extremely difficult. In general, sophisticated computer simulations are required in this line of work. One possible approach is to approximate dense networks with a continuum of nodes where the density of the network goes to infinity [60–62]. The continuum approximation, which has been verified to be accurate, can provide insights for investigating the performance of dense wireless networks; however, the observations and results obtained from this approach are usually

not applicable for sparse networks. On the other hand, emerging stochastic geometry tools provide an alternative way of thinking. Various aspects of non-cooperative wireless networks have been investigated following this line of inquiry [63–65]. In [66–68], cooperative communication systems with one source-destination pair are investigated using stochastic geometry. Most of these works, however, assume that the network is interference-free, which is clearly an idealized version of the problem.

For interference-limited cooperative networks, optimal centralized algorithms and suboptimal distributed algorithms have been proposed to achieve excellent performance for different scenarios in [69–72]. In particular, [69] studies spectrum management and relay selection in cooperative networks. In [70, 71], the results in [69] are extended, and the problem of jointly encoding rate control, power allocation, relay selection, and subcarrier assignment in a cooperative multimedia network is solved. In [72], distributed algorithms for cross-layer design in a multihop interference-limited cooperative system are proposed. Compared to [69–72], which focus on performance optimization for given cooperative networks, our objective is to derive a criterion for whether to use cooperation or not in a large-scale random network. The tools of stochastic geometry are applied to investigate a network with spatial interference.

Specifically, in this chapter, the outage performance of a random wireless network for both non-cooperative and cooperative strategies is analyzed; and, based on the outage, the tradeoff between cooperative diversity and the additional interference is studied. The criterion for whether cooperation among potential relay nodes should be used or not is derived; and we show that a lower outage probability is achieved by the non-cooperative strategy if and only if the extra interference caused by the cooperation (which depends on the intensity of the interfering sources, the source-destination distance, and the outage threshold) is larger than a threshold determined by the path loss exponent. If the network is crowded and there are a lot of interferers for any existing transmission, the extra interference due to cooperation degrades the performance. Asymptotic properties are also provided to validate this insight. As the network becomes more dense, the success probability of a non-cooperative strategy

decreases more slowly than the cooperative one. In contrast, the cooperative strategy outperforms the non-cooperative one if the number of interferers is relatively small. If the impact of the additional interference is negligible, the performance gain of cooperation can be fully achieved and the cooperative strategy is more attractive. The asymptotic analysis shows that the gap between the outage probabilities of the non-cooperative and cooperative strategies becomes continuously larger as the sparsity of the network increases.

3.1 System Model

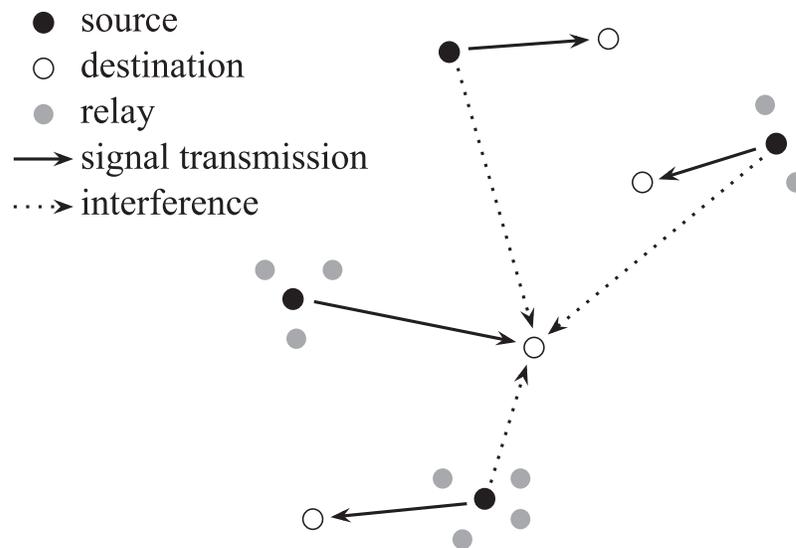


Figure 3.1: A large-scale wireless network with cooperative relays.

We consider a wireless network where a large number of nodes are spread over an infinite area. The entire network is synchronized in time and frequency. We also assume that each node makes transmission decisions independently, i.e., a slotted ALOHA protocol [73] is employed at the MAC layer. The active nodes are specified as sources, and their locations are typically dynamic due to their indiscriminate placement and the uncoordinated nature of the MAC protocol.

3.1.1 Non-Cooperative Strategy

We first present a non-cooperative strategy where no cooperation among nodes exists. We will use this strategy as a baseline for comparison. Such a strategy has been investigated in the literature [73–77]. To model the dynamic nature of the network, the locations of the sources are assumed to be spatially random. Without any *a priori* knowledge of the distribution pattern of the sources, we assume that the sources form a homogeneous Poisson process Π with intensity λ on the plane [78]. The parameter λ characterizes how densely the sources are distributed and depends not only on the density of the nodes but also the statistics of the incoming data traffic.

To measure the strength of a received signal, both large-scale path loss and small-scale fading effects will be taken into account. Specifically, we use a power-law path loss and a Rayleigh fading model. The channels are assumed to be quasi-static, i.e., the channel state information does not change over one time slot. Suppose that the distance from a given source to its corresponding destination is d . Then, the instantaneous signal power contributed by a source at distance d can be expressed as $P_t g d^{-\alpha}$ where P_t is the transmission power, g is an exponentially distributed random variable capturing the effects of Rayleigh fading and α is the path loss exponent¹.

It is highly likely that multiple sources may transmit concurrently, due to the uncoordinated nature of the ALOHA protocol. Therefore, the quality of reception at a destination will be severely affected by the spatial interference generated by other concurrent transmitting sources, or, other interferers. Without loss of generality, we suppose all sources use the same transmit power P_t . Let h_i denote the channel from the i th interferer to a destination at location x , the received interference power at the destination can be formulated as

$$P_I(x) = \sum_{x_i \in \Pi} P_t |h_i|^2 = \sum_{x_i \in \Pi} P_t g_i \|x_i - x\|^{-\alpha}, \quad (3.1)$$

¹ Here, we adopt the same path loss model (2.21) as we discussed in Chapter 2. If no specification is given, we will assume that $PL_0 = 1$ and $d > d_0 = 1$ in the following part of this dissertation.

where g_i characterizes the fading effects for the channel between the i th interferer and the destination, the x_i 's are the locations of the interferers and $\|\cdot\|$ represents Euclidean distance. Since Π is a stationary process and covers an infinitely large area, the statistics of $P_I(x)$ are invariant to the actual location x . In what follows, we use a simplified notation P_I to denote the interference.

In this chapter, the outage performance of the network will be investigated. For the sake of simplicity, we treat interference as noise and assume that a transmission is in outage if the received signal-to-interference-plus-noise ratio (SINR) is less than a pre-specified threshold γ_{th} . In a non-cooperative network, the received SINR at the destination is given by

$$\gamma_{\text{non}} = \frac{P_t g d^{-\alpha}}{P_I + P_N}, \quad (3.2)$$

where P_N is the noise power which is assumed to be the same for all links in the network.

3.1.2 Cooperative Strategy

In order to investigate the overall impact of cooperation, the non-cooperative strategy needs to be extended. Specifically, we suppose that every source has a set of relays in its vicinity which can help the communication to its intended destination. See Fig. 3.1 for a pictorial description. As introduced in Chapter 2, a time-domain, two-hop, decode-and-forward protocol is utilized for cooperation.

The direct link between a source and its intended destination is assumed to be extremely weak due to deep fading or shadowing, as is typically the case. Such a restriction can be easily relaxed and the analytical approach presented in this chapter is applicable to those cases. M -group STBC, which can provide a minimum overhead implementation as shown in Chapter 2, is adopted as the cooperative strategy in this chapter. With M -group, the relays in the vicinity of the source will independently and randomly divide themselves into M groups; each group then emulates one antenna of an M -antenna system. By applying a pre-specified STBC, an M -order cooperative diversity gain can be obtained. The received signal power at the destination P_s is then

determined by the particular STBC scheme, the number of groups M , and the number of relays.

The interference model is more complicated than the non-cooperative case because not only the sources but also the relays could interfere with each other. In the first hop, the interference comes from other concurrent sources and affects the signal reception at the relays. In the second hop, by contrast, the interference is generated by the relays of the other concurrent sources. To mitigate the interference, the exact statistics of the quantities and locations of both the sources and the relays would be indispensable. Such an approach would undoubtedly complicate or even hinder the derivation. Therefore, instead of using more complicated mathematical tools to model the system [79], we make the following assumptions and assume a simple, yet effective, model to facilitate our analysis.

1. In order to simplify the analysis, we assume that each vicinity is close to the corresponding source, and the geographical difference between the various relays in the same vicinity is negligible when viewed by a remote observer. Consequently, we can assert that the signals from the relays of a common source are indistinguishable to any destinations except the one associated with this source. Note that this assumption does not imply that the first hop transmission is perfect. Although this is an idealized simplification, our simulation results will show that this assumption still provides a tight approximation for some realistic scenarios. In practice, relay-assisted uplink transmission in a cellular network is a representative scenario which is similar to this model. The distance between the relay and destination is usually much larger than the distance between the source and relays [80].

2. The number of relays K for a given source is assumed to follow a probability distribution $\Pr(K = k)$, $k \in \mathcal{K} \subseteq \{0\} \cup \mathbb{Z}^+$, which may be a uniform distribution (\mathcal{K} is a finite set) or even a Poisson distribution, to name a few. Similarly, the number of relays K_v in every interfering vicinity is assumed to independently follow a common probability distribution $\Pr(K_v = k_v)$, $k_v \in \mathcal{K}_v \subseteq \{0\} \cup \mathbb{Z}^+$. Note that by introducing the randomness of K and K_v , the effect of imperfect first hop transmission is taken

into consideration.

These assumptions may be relaxed but at the cost of prohibitively complicating the derivation, and without providing additional insights. Similar to the non-cooperative case, we let all the sources and the relays use the same transmit power P_t , and assume that the distance between any source and its destination is d . The reader might be concerned that this assumption leads to an unfair comparison since cooperative transmission has a higher power budget than non-cooperative transmission. However, if we assume that the total power consumption is the same for both cooperative and non-cooperative scenarios, the total number of decoded nodes needs to be known at each node. To collect this information, a central controller, which incurs additional complexity and overhead, is required. In this chapter, we are interested in investigating an efficient distributed scheme with low overhead. Therefore, we assume that all nodes have the same power. Similarly, power control among relays is not considered here due to the excessive amount of overhead required.

Consider a given source-destination pair. The received interference power at location x in the first hop is

$$P_I^{(1)}(x) = \sum_{x_i \in \Pi} P_t g_i \|x_i - x\|^{-\alpha} \quad (3.3)$$

which has the same statistics as $P_I(x)$ given in (3.1). In the second hop, the received interference power is given by

$$P_I^{(2)}(x) = \sum_{x_i \in \Pi} k_{v,i} P_t g_i \|x_i - x\|^{-\alpha}, \quad (3.4)$$

where $k_{v,i}$ denotes the number of relays in the i th interfering vicinity, which can be a constant or a random number. Note that (3.3) and (3.4) use the same coordination set Π since we assume that the relay nodes are close to the corresponding source. In other words, a vicinity can be thought of as a single node with transmit power $k_{v,i} P_t$ from the perspective of the signal measurement.

Note that $P_I^{(1)}(x)$ and $P_I^{(2)}(x)$ can be denoted by $P_I^{(1)}$ and $P_I^{(2)}$, respectively, since their statistics are invariant to x . Without loss of generality, we also can arbitrarily

set up a coordinate system and place the destination at the origin. According to Slivnyak's Theorem of Poisson processes, a node can be added into Π without changing its statistics [81]. Consequently, we can always put a source at a particular position so that its intended destination is at the origin no matter where and how the coordinate system is set up. The results obtained remain general because the considered system models are homogeneous and stationary. Thus, we let $x = (0, 0)$ and rewrite (3.3) and (3.4) as

$$P_I^{(1)} = \sum_{x_i \in \Pi} P_t g_i \|x_i\|^{-\alpha}, \quad (3.5)$$

$$P_I^{(2)} = \sum_{x_i \in \Pi} k_{v,i} P_t g_i \|x_i\|^{-\alpha}. \quad (3.6)$$

3.2 Performance Analysis

In this section, we first discuss the outage probability for both non-cooperative and cooperative strategies, and then compare their outage performances for two different scenarios: a fixed number of relays and a random number of relays.

3.2.1 Non-Cooperative Strategy

The baseline strategy is considered, i.e., no cooperation is utilized. The outage probability for this strategy is defined by

$$p_{\text{out}}^{\text{non}} = \Pr(\gamma_{\text{non}} < \gamma_{\text{th}}), \quad (3.7)$$

where γ_{non} is given in (3.2).

A closed-form expression for $p_{\text{out}}^{\text{non}}$ has been derived in [74, 75], and is given by

$$p_{\text{out}}^{\text{non}} = 1 - \exp\left(-\frac{\gamma_{\text{th}} P_N}{P_t d^{-\alpha}} - A_\alpha \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}}\right), \quad (3.8)$$

where

$$A_\alpha = \frac{2\pi^2}{\alpha \sin\left(\frac{2\pi}{\alpha}\right)}. \quad (3.9)$$

The exponent has two parts: one representing the effects of noise, and the other interference. The first part $\frac{\gamma_{\text{th}}}{P_t d^{-\alpha} / P_N}$ measures the margin between the threshold and the

average received SNR. The second part $A_\alpha \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}}$ depends on the interference level. If we reformat the expression as

$$A_\alpha \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}} = A_\alpha \left(\frac{\gamma_{\text{th}}}{\frac{P_t d^{-\alpha}}{\lambda^{\frac{\alpha}{2}} P_t}} \right)^{\frac{2}{\alpha}}, \quad (3.10)$$

it can be seen that this component describes the margin between the threshold and the SIR.

3.2.2 Cooperative Strategy

The received SINR at the destination is given by

$$\gamma_{\text{co}} = \frac{P_s}{P_I^{(2)} + P_N}, \quad (3.11)$$

and the outage probability is

$$p_{\text{out}}^{\text{co}} = \Pr(\gamma_{\text{co}} < \gamma_{\text{th}}). \quad (3.12)$$

Note that the received signal power P_s is determined by the number of relays K in the current vicinity. On the other hand, the interference power $P_I^{(2)}$ is determined by the number of interferers and the number of relays K_v in each interfering vicinity. Both K and K_v are assumed to follow some predefined discrete probability distribution. Thus, (3.12) can be rewritten as

$$p_{\text{out}}^{\text{co}} = \sum_{k \in \mathcal{K}} \Pr(K = k) p(k), \quad (3.13)$$

where $p(k) = \mathbb{E}[\Pr(\gamma_{\text{co}} < \gamma_{\text{th}} | K = k)]$ is the conditional outage probability given the number of relays k in the current vicinity. The expectation $\mathbb{E}[\cdot]$ is taken with respect to K_v and $x_i \in \Pi$, i.e., the randomness of the interference has been averaged in $p(k)$. In the following theorems, we first derive the conditional outage probability $p(k)$ for different values of k .

When $k = 0$, meaning that the vicinity is empty, there is no signal transmission at all and the received SINR is 0, which implies that the conditional outage probability

is 1. Apart from this corner circumstance, the relays can help the transmission, and a performance gain can be obtained.

When $k = 1$, we cannot achieve any diversity gain, and the received signal power at the destination P_s is exactly the same as the signal power in the non-cooperative case. The received SINR is then given by

$$\gamma_{\text{co}} = \frac{P_t g d^{-\alpha}}{P_I^{(2)} + P_N}. \quad (3.14)$$

The following theorem provides a closed-form expression for $p(k)$ when $k = 1$.

Theorem 3.1. If $k = 1$, the conditional outage probability $p(k)$ is

$$p(1) = 1 - \exp\left(-\frac{\gamma_{\text{th}} P_N}{P_t d^{-\alpha}} - A_\alpha \lambda d^2 \mathbb{E}\left[(\gamma_{\text{th}} K_v)^{\frac{2}{\alpha}}\right]\right), \quad (3.15)$$

with the expectation being taken with respect to K_v .

Proof. According to (3.14), we have

$$\begin{aligned} & \Pr(\gamma < \gamma_{\text{th}}) \\ &= \int_{P_N}^{\infty} \Pr(P_t |h_i|^2 < z \gamma_{\text{th}}) d \Pr(P_I^{(2)} + P_N \leq z) \\ &= 1 - e^{-\frac{\gamma_{\text{th}}}{P_t d^{-\alpha}} P_N} \int_0^{\infty} e^{-\frac{\gamma_{\text{th}}}{P_t d^{-\alpha}} z'} d \Pr(P_I^{(2)} \leq z') \\ &= 1 - e^{-\frac{\gamma_{\text{th}}}{P_t d^{-\alpha}} P_N} \mathbb{E}\left[e^{-\frac{\gamma_{\text{th}}}{P_t d^{-\alpha}} P_I^{(2)}}\right]. \end{aligned} \quad (3.16)$$

Let S be the infinitely large plane on which the Poisson process $\Pi = \{X_j\}$ is defined.

Denoting $K_{v,j} g_j$ by M_j , i.e., we can rewrite (3.6) as

$$P_I^{(2)} = \sum_{x_j \in \Pi} M_j P_t \|x_j\|^{-\alpha}. \quad (3.17)$$

By the Marking Theorem of Poisson processes [78], the set of (M_j, X_j) is a Poisson process Π^* defined on (\mathbb{R}, S) with intensity

$$\lambda^*(m, x) = \lambda f_{M|X}(m|x) = \lambda f_M(m), \quad (3.18)$$

because $M_j = K_{v,j} g_j$ is independent with X_j . The function $f(\cdot)$ denotes the PDF of a random variable. Then, $P_I^{(2)}$ can be regarded as the sum of functions $\mathcal{F}(m, x) =$

$P_t m \|x\|^{-\alpha}$ over Π^* . By Campbell's Theorem [78], and letting $v = -\frac{\gamma_{\text{th}}}{P_t d^{-\alpha}}$, with $F(k_v)$ denoting the probability mass function $\Pr(K_v = k_v)$, we have

$$\begin{aligned}
\mathbb{E} \left[e^{-\frac{\gamma_{\text{th}}}{P_t d^{-\alpha}} P_I^{(2)}} \right] &= \exp \left(- \int_S \int_0^\infty (1 - e^{v\mathcal{F}(m,x)}) \lambda^*(m,x) dm dx \right) \\
&= \exp \left(-\lambda \int_S (1 - \mathbb{E}(e^{v\mathcal{F}(M,x)})) dx \right) \\
&= \exp \left(-\lambda \int_S \left(1 - \sum_{k_v=0}^\infty \frac{1}{1 - vk_v P_t \|x\|^{-\alpha}} \right) F(k_v) dx \right) \\
&= \exp \left(-\lambda \int_S \sum_{k_v=0}^\infty \frac{vk_v P_t \|x\|^{-\alpha}}{vk_v P_t \|x\|^{-\alpha} - 1} F(k_v) dx \right) \\
&= \exp \left(-\lambda 2\pi \sum_{k_v=0}^\infty \left(\int_0^\infty \frac{vk_v P_t r^{-\alpha}}{vk_v P_t r^{-\alpha} - 1} r dr \right) F(k_v) \right) \\
&\stackrel{(a)}{=} \exp \left(-\lambda 2\pi \sum_{k_v=0}^\infty \left(\frac{(-vk_v P_t)^{\frac{2}{\alpha}}}{\alpha} \int_0^\infty \frac{t^{\frac{2}{\alpha}-1}}{(1+t)^{\frac{2}{\alpha}+(1-\frac{2}{\alpha})}} dt \right) F(k_v) \right) \\
&\stackrel{(b)}{=} \exp \left(-\lambda 2\pi \Gamma(\frac{2}{\alpha}) \Gamma(1 - \frac{2}{\alpha}) \frac{(-v P_t)^{\frac{2}{\alpha}}}{\alpha} \sum_{k_v=0}^\infty k_v^{\frac{2}{\alpha}} F(k_v) \right) \\
&\stackrel{(c)}{=} \exp \left(-A_\alpha \lambda d^2 \mathbb{E} \left[(\gamma_{\text{th}} K_v)^{\frac{2}{\alpha}} \right] \right). \tag{3.19}
\end{aligned}$$

Note that $t = -\frac{r^\alpha}{vk_v P_t}$ in step (a). The Beta function

$$\text{B}(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)} = \int_0^\infty \frac{t^{x-1}}{(1+t)^{x+y}} dt \tag{3.20}$$

and the fact that $\Gamma(1) = 1$ are utilized in step (b), and $\Gamma(\frac{2}{\alpha})\Gamma(1 - \frac{2}{\alpha}) = \pi / \sin(\frac{2\pi}{\alpha})$ is used in step (c). Theorem 1 can be proved by substituting (3.19) into (3.16). \blacksquare

When $k \geq 2$, additional diversity gain can be obtained through cooperative techniques. In the following analysis, we will focus on the M -group STBC scheme and assume that $M = 2$. Although our results can certainly be extended to other STBC schemes, the derivation will be more complicated, without providing additional insights.

In this case, the received signal power P_s is the sum of the power contributed by each group, and the received SINR is

$$\gamma_{\text{co}} = \frac{|\sum_{i \in \mathcal{G}_1} \sqrt{P_t} h_i|^2 + |\sum_{i \in \mathcal{G}_2} \sqrt{P_t} h_i|^2}{P_I^{(2)} + P_N}, \quad (3.21)$$

where \mathcal{G}_1 and \mathcal{G}_2 denote two groups of relays. In Chapter 2, it has been shown that the optimal outage performance is achieved when the groups of relays are uniformly divided. In this chapter, we will assume that the optimal uniform grouping is achieved, i.e., \mathcal{G}_1 and \mathcal{G}_2 have the same size $k/2$. The following theorem provides a closed-form expression for the conditional outage probability when $k \geq 2$.

Theorem 3.2. If $k \geq 2$, the conditional outage probability $p(k)$ is

$$\begin{aligned} p(k) &= 1 - \left(1 + \frac{2\gamma_{\text{th}} P_N}{k P_t d^{-\alpha}} + \frac{2}{\alpha} A_\alpha \lambda d^2 \mathbb{E} \left[\left(\frac{2\gamma_{\text{th}} K_v}{k} \right)^{\frac{2}{\alpha}} \right] \right) \\ &\quad \times \exp \left(-\frac{2\gamma_{\text{th}} P_N}{k P_t d^{-\alpha}} - A_\alpha \lambda d^2 \mathbb{E} \left[\left(\frac{2\gamma_{\text{th}} K_v}{k} \right)^{\frac{2}{\alpha}} \right] \right). \end{aligned} \quad (3.22)$$

The expectations are again taken with respect to K_v .

Proof. We use $P_s^{(1)} = |\sum_{i \in \mathcal{G}_1} \sqrt{P_t} h_i|^2$ and $P_s^{(2)} = |\sum_{i \in \mathcal{G}_2} \sqrt{P_t} h_i|^2$ to denote the received signal power contributed by \mathcal{G}_1 and \mathcal{G}_2 , respectively. Let $\theta = -\frac{2\gamma_{\text{th}}}{k P_t d^{-\alpha}}$. According to (3.21), we have

$$\begin{aligned} \Pr(\gamma < \gamma_{\text{th}}) &= \Pr \left(\frac{P_s^{(1)} + P_s^{(2)}}{P_I^{(2)} + P_N} < \gamma_{\text{th}} \right) \\ &= \int_{P_N}^{\infty} \Pr(P_s^{(1)} + P_s^{(2)} < z\gamma_{\text{th}}) d\Pr(P_I^{(2)} + P_N \leq z) \\ &= 1 - \int_{P_N}^{\infty} (e^{\theta z} - e^{\theta z} \theta z) d\Pr(P_I^{(2)} + P_N \leq z) \\ &= 1 - e^{\theta P_N} (1 - \theta P_N) \int_0^{\infty} e^{\theta z'} d\Pr(P_I^{(2)} \leq z') + \theta e^{\theta P_N} \int_0^{\infty} z' e^{\theta z'} d\Pr(P_I^{(2)} \leq z') \\ &= 1 - e^{\theta P_N} (1 - \theta P_N) \mathbb{E}[e^{\theta P_I^{(2)}}] + \theta e^{\theta P_N} \mathbb{E}[P_I^{(2)} e^{\theta P_I^{(2)}}], \end{aligned} \quad (3.23)$$

where the expectations in the last step are taken with respect to $P_I^{(2)}$. We now need to calculate the two expectations $\mathbb{E}[e^{\theta P_I^{(2)}}]$ and $\mathbb{E}[P_I^{(2)} e^{\theta P_I^{(2)}}]$.

The first expectation can be derived in the same way as (3.19). That is,

$$\mathbb{E}\left[e^{\theta P_I^{(2)}}\right] = \exp\left(-A_\alpha \lambda d^2 \mathbb{E}\left[\left(\frac{2\gamma_{\text{th}} K_v}{k}\right)^{\frac{2}{\alpha}}\right]\right). \quad (3.24)$$

For the second expectation, we notice that Campbell's Theorem provides a closed-form expression for $\mathbb{E}[e^{\theta P_I^{(2)}}]$; however, we are interested in $\mathbb{E}[P_I^{(2)} e^{\theta P_I^{(2)}}]$. Similar to the derivation of Campbell's Theorem in [78], we consider a Poisson process Π defined on S with mean measure μ . Let \mathcal{F} be a real-valued function defined on S , which can only take a finite number of non-zero values $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_k$. The set $\Upsilon_j = \{x : \mathcal{F}(x) = \mathcal{F}_j, x \in S\}$ is measurable with $m_j = \mu(\Upsilon_j)$. Different Υ_j 's are disjoint and the number of nodes of Π falling in each Υ_j , denoted by N_j , independently follows a Poisson distribution with mean m_j . Define the sum of \mathcal{F} over Π as $\Sigma = \sum_{X \in \Pi} \mathcal{F}(X) = \sum_{j=1}^k \mathcal{F}_j N_j$. Then,

$$\begin{aligned} \mathbb{E}[\Sigma e^{\theta \Sigma}] &= \mathbb{E}\left[\sum_{i=1}^k \mathcal{F}_i N_i e^{\theta \sum_{j=1}^k \mathcal{F}_j N_j}\right] \\ &= \sum_{i=1}^k \mathbb{E}[\mathcal{F}_i N_i e^{\theta \mathcal{F}_i N_i} e^{\theta \sum_{j \neq i} \mathcal{F}_j N_j}] \\ &= \sum_{i=1}^k \mathbb{E}[\mathcal{F}_i N_i e^{\theta \mathcal{F}_i N_i}] \mathbb{E}[e^{\theta \sum_{j \neq i} \mathcal{F}_j N_j}] \\ &\stackrel{(d)}{=} \sum_{i=1}^k \mathcal{F}_i m_i e^{\theta \mathcal{F}_i} e^{-m_i(1-e^{\theta \mathcal{F}_i})} \prod_{j \neq i} e^{-m_j(1-e^{\theta \mathcal{F}_j})} \\ &= \sum_{i=1}^k \mathcal{F}_i m_i e^{\theta \mathcal{F}_i} \cdot \prod_{j=1}^k e^{-m_j(1-e^{\theta \mathcal{F}_j})} \\ &= \sum_{i=1}^k \int_{\Upsilon_i} \mathcal{F}(x) e^{\theta \mathcal{F}(x)} \mu(dx) \cdot e^{-\sum_{j=1}^k \int_{\Upsilon_j} (1-e^{\theta \mathcal{F}(x)}) \mu(dx)} \\ &= \int_S \mathcal{F}(x) e^{\theta \mathcal{F}(x)} \lambda(x) dx \cdot e^{-\int_S (1-e^{\theta \mathcal{F}(x)}) \lambda(x) dx}. \end{aligned} \quad (3.25)$$

In step (d), the following facts are used: suppose Y is a Poisson random variable with mean μ ; then, for any constant z , we have

$$\begin{aligned} \mathbb{E}[z^Y] &= e^{-\mu(1-z)} \\ \mathbb{E}[Y z^Y] &= \mu z e^{-\mu(1-z)} \end{aligned} \quad (3.26)$$

We could also show that (3.25) holds for more general forms of \mathcal{F} .

Using the same definitions of $\mathcal{F}(m, x)$ and $\lambda^*(m, x)$ as in Theorem 3.1, we can obtain a closed-form expression for the second expectation as

$$\begin{aligned}
\mathbb{E}\left[P_I^{(2)} e^{\theta P_I^{(2)}}\right] &= \int_S \int_0^\infty \mathcal{F}(m, x) e^{\theta \mathcal{F}(m, x)} \lambda^*(m, x) dm dx \\
&\quad \times \exp\left(-\int_S \int_0^\infty (1 - e^{\theta \mathcal{F}(m, x)}) \lambda^*(m, x) dm dx\right) \\
&= \lambda \int_S \sum_{k_v=0}^\infty \frac{k_v P_t \|x\|^{-\alpha}}{(\theta k_v P_t \|x\|^{-\alpha} - 1)^2} F(k_v) dx \\
&\quad \times \exp\left(-\int_S \int_0^\infty (1 - e^{\theta \mathcal{F}(m, x)}) \lambda^*(m, x) dm dx\right) \\
&= \lambda 2\pi \sum_{k_v=0}^\infty \left(\int_0^\infty \frac{k_v P_t r^{-\alpha}}{(\theta k_v P_t r^{-\alpha} - 1)^2} r dr\right) F(k_v) \\
&\quad \times \exp\left(-\int_S \int_0^\infty (1 - e^{\theta \mathcal{F}(m, x)}) \lambda^*(m, x) dm dx\right) \\
&= \lambda 2\pi \sum_{k_v=0}^\infty \left(\frac{-(-\theta k_v P_t)^{\frac{2}{\alpha}}}{\theta \alpha} \int_0^\infty \frac{t^{\frac{2}{\alpha}}}{(1+t)^{1+\frac{2}{\alpha}+1-\frac{2}{\alpha}}} dt\right) \\
&\quad \times F(k_v) \exp\left(-\int_S \int_0^\infty (1 - e^{\theta \mathcal{F}(m, x)}) \lambda^*(m, x) dm dx\right) \\
&= \lambda 2\pi \Gamma\left(1 + \frac{2}{\alpha}\right) \Gamma\left(1 - \frac{2}{\alpha}\right) \frac{-(-\theta P_t)^{\frac{2}{\alpha}}}{\theta \alpha} \mathbb{E}\left[K_v^{\frac{2}{\alpha}}\right] \exp\left(-\lambda A_\alpha (-\theta P_t)^{\frac{2}{\alpha}} \mathbb{E}\left[K_v^{\frac{2}{\alpha}}\right]\right) \\
&= \frac{2}{\alpha} A_\alpha \lambda d^{2-\alpha} \left(\frac{2\gamma_{\text{th}}}{k P_t}\right)^{-1} \mathbb{E}\left[\left(\frac{2\gamma_{\text{th}} K_v}{k}\right)^{\frac{2}{\alpha}}\right] \exp\left(-A_\alpha \lambda d^2 \mathbb{E}\left[\left(\frac{2\gamma_{\text{th}} K_v}{k}\right)^{\frac{2}{\alpha}}\right]\right). \quad (3.27)
\end{aligned}$$

Note that in deriving (3.27) the facts that $\Gamma(2) = 1$ and $\Gamma\left(1 + \frac{2}{\alpha}\right) \Gamma\left(1 - \frac{2}{\alpha}\right) = \frac{2\pi}{\alpha} / \sin\left(\frac{2\pi}{\alpha}\right)$ are utilized. Theorem 3.2 can be proved by substituting (3.19) and (3.27) into (3.23). ■

The closed-form expression for the outage probability $p_{\text{out}}^{\text{co}}$ can be obtained by combining (3.13), (3.15) and (3.22). If $\mathcal{K} = \mathcal{K}_v = \{k\}$ and $\Pr(K = k) = \Pr(K_v = k) =$

1, for instance, we can easily derive that

$$p_{\text{out}}^{\text{co}} = \begin{cases} 1 & k = 0 \\ 1 - \exp\left(-\frac{\gamma_{\text{th}} P_N}{P_t d^{-\alpha}} - A_\alpha \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}}\right) & k = 1 \\ 1 - \left(1 + \frac{2\gamma_{\text{th}} P_N}{k P_t d^{-\alpha}} + \frac{2}{\alpha} A_\alpha \lambda d^2 (2\gamma_{\text{th}})^{\frac{2}{\alpha}}\right) \exp\left(-\frac{2\gamma_{\text{th}} P_N}{k P_t d^{-\alpha}} - A_\alpha \lambda d^2 (2\gamma_{\text{th}})^{\frac{2}{\alpha}}\right) & k \geq 2 \end{cases} \quad (3.28)$$

Remark 3.1. If we assume that $\frac{P_N}{P_t d^{-\alpha}} \ll A_\alpha \lambda d^2$, i.e., the noise power is negligible in comparison with the interference power, the terms $\frac{\gamma_{\text{th}} P_N}{P_t d^{-\alpha}}$ in (3.8), (3.15) and (3.22) can be neglected. In such an interference-limited regime, since $\frac{\gamma_{\text{th}} P_N}{P_t d^{-\alpha}} \ll A_\alpha \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}}$, the outage probability for the non-cooperative strategy (3.8) can be simplified to

$$p_{\text{out}}^{\text{non}} \approx 1 - \exp\left(-A_\alpha \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}}\right). \quad (3.29)$$

For the cooperative strategy, we first define an auxiliary function

$$\Lambda(k) = A_\alpha \lambda d^2 \mathbb{E} \left[\left(\frac{2\gamma_{\text{th}} K_v}{k} \right)^{\frac{2}{\alpha}} \right], \quad k \geq 2 \quad (3.30)$$

with the expectation being taken with respect to K_v . In the interference-limited regime, the conditional probability can be approximately written as

$$p(k) \approx \begin{cases} 1 & k = 0 \\ 1 - \exp(-\Lambda(2)) & k = 1 \\ 1 - \exp(-\Lambda(k)) \left(1 + \frac{2}{\alpha} \Lambda(k)\right) & k \geq 2 \end{cases} \quad (3.31)$$

according to (3.15) and (3.22). As we can see from (3.31), when $k \geq 2$, there is an extra term $(1 + \frac{2}{\alpha} \Lambda(k))$ in the expression for $p(k)$ compared to the case when $k = 1$. Intuitively speaking, this additional factor comes from the benefits offered by cooperation.

If K and K_v are both discrete Poisson random variables with mean κ , for instance, by combining (3.13) and (3.31), the outage probability in the interference-limited regime can be obtained as

$$p_{\text{out}}^{\text{co}} \approx \sum_{i=2}^{\infty} \frac{\kappa^i}{i!} \exp(-\kappa) \left(1 - \exp(-\Lambda(i)) \left(1 + \frac{2}{\alpha} \Lambda(i)\right) \right)$$

$$+ \kappa \exp(-\kappa)(1 - \exp(-\Lambda(2))) + \exp(-\kappa). \quad (3.32)$$

If $\mathcal{K} = \mathcal{K}_v = \{k\}$ and $\Pr(K = k) = \Pr(K_v = k) = 1$, i.e., the randomness of the number of relays is neglected, the outage probability in the interference-limited regime can be obtained from (3.28) as

$$p_{\text{out}}^{\text{co}} \approx \begin{cases} 1 & k = 0 \\ 1 - \exp(-A_\alpha \beta) & k = 1 \\ 1 - \left(1 + \frac{2}{\alpha} 2^{\frac{2}{\alpha}} A_\alpha \beta\right) \exp\left(-2^{\frac{2}{\alpha}} A_\alpha \beta\right) & k \geq 2 \end{cases} \quad (3.33)$$

where $\beta = \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}}$.

3.2.3 To Cooperate or Not to Cooperate

In Sections 3.2.1 and 3.2.2, closed-form expressions for the outage probability in both the non-cooperative and cooperative cases have been derived as functions of key system parameters, such as the intensity of interfering sources λ , the source-destination distance d , the outage threshold γ_{th} , and the transmit power P_t . In this subsection, we will compare these two cases and see how the comparison results vary with the system parameters. Then, for any given system environment, we can determine a criterion for using cooperation.

Fixed K and K_v

We will first investigate a simplified scenario where the number of relays is fixed, i.e., $K = K_v = k$. This models the scenarios where every source employs the same number of relays in a deterministic way. Without loss of generality, we only consider $k \geq 2$. Intuitively, in the noise-limited scenario, cooperation definitely helps the performance by providing diversity gain and enhancing the received signal power. Therefore, we should always use cooperative strategies in this case. The following theorem presents the criterion for determining whether cooperation is beneficial in the interference-limited regime.

Theorem 3.3. In the interference-limited regime, we have

$$\begin{cases} p_{\text{out}}^{\text{co}} < p_{\text{out}}^{\text{non}} & \text{if } \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}} < \beta^* \\ p_{\text{out}}^{\text{co}} > p_{\text{out}}^{\text{non}} & \text{if } \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}} > \beta^* \end{cases} \quad (3.34)$$

where

$$\beta^* = \frac{1}{A_\alpha B_\alpha} (-C_\alpha - \mathcal{W}(-C_\alpha \exp(-C_\alpha))), \quad (3.35)$$

A_α is defined in (3.9), $B_\alpha = 2^{\frac{2}{\alpha}} - 1$, $C_\alpha = \frac{\alpha B_\alpha}{2(B_\alpha + 1)}$, and $\mathcal{W}(\cdot)$ is the principle branch of the Lambert \mathcal{W} function [82].

Proof. According to the interference-limited approximations (3.29) and (3.33), we can easily observe that the outage probabilities are increasing functions of $\beta = \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}}$. Hence, we need to solve the following equation in the interference-limited regime

$$p_{\text{out}}^{\text{non}}(\beta) - p_{\text{out}}^{\text{co}}(\beta) = 0, \quad (3.36)$$

which can be rewritten as

$$1 + \frac{1}{C_\alpha} y - \exp(y) = 0, \quad (3.37)$$

where $y = A_\alpha B_\alpha \beta$, $B_\alpha = 2^{\frac{2}{\alpha}} - 1$ and $C_\alpha = \frac{\alpha B_\alpha}{2(B_\alpha + 1)}$. According to (3.37), we have

$$\begin{aligned} & (-C_\alpha - y) \exp(-C_\alpha - y) \\ &= -C_\alpha \exp(-C_\alpha) \left(1 + \frac{1}{C_\alpha} y\right) \exp(-y) \\ &= -C_\alpha \exp(-C_\alpha), \end{aligned} \quad (3.38)$$

which indicates that

$$y = -C_\alpha - \mathcal{W}(-C_\alpha \exp(-C_\alpha)), \quad (3.39)$$

where $\mathcal{W}(\cdot)$ is the principle branch of the Lambert \mathcal{W} function [82]. We can then obtain the closed-form expression for β^* which satisfies (3.36). \blacksquare

We can see that β^* solely depends on the path loss exponent α . In practice, once α is determined by the channel model, β^* can then be viewed as a predefined threshold for system designers to choose whether cooperation should be used.

As we mentioned in Section 3.2.1, the term $\lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}}$ describes the margin between the received SIR and the outage threshold γ_{th} . Eq. (3.34) shows that if the spatial interference is not that strong, cooperation among the relays does help the transmission. One possible scenario satisfying this requirement is that potential sources in the network are relatively sparse compared with the distance between the sources and their destinations. It is easy to show from (3.8) and (3.28) that

$$\begin{aligned} \lim_{d^\alpha \rightarrow \infty} \frac{\ln(1 - p_{\text{out}}^{\text{non}})}{d^\alpha} &= -\frac{\gamma_{\text{th}} P_N}{P_t}, \\ \lim_{d^\alpha \rightarrow \infty} \frac{\ln(1 - p_{\text{out}}^{\text{co}})}{d^\alpha} &= -\frac{2\gamma_{\text{th}} P_N}{k P_t}. \end{aligned} \quad (3.40)$$

Eq. (3.40) implies that as the source-destination distance d increases, the success probability of the cooperative strategy decreases more slowly. In other words, cooperation among relays provides a larger transmission range for a given outage probability.

On the other hand, if the interference level is significant compared with the received signal, the cooperative strategy would provide a poorer outage performance. This happens if the concurrent sources are densely distributed and the interference dominates the performance. In this case, the extra interference caused by cooperation degrades the performance, and non-cooperative strategies are more appealing. It can be further obtained from (3.8) and (3.28) that, as the intensity of the interfering sources λ goes to infinity,

$$\begin{aligned} \lim_{\lambda \rightarrow \infty} \frac{\ln(1 - p_{\text{out}}^{\text{non}})}{\lambda} &= -A_\alpha d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}}, \\ \lim_{\lambda \rightarrow \infty} \frac{\ln(1 - p_{\text{out}}^{\text{co}})}{\lambda} &= -A_\alpha d^2 (2\gamma_{\text{th}})^{\frac{2}{\alpha}}. \end{aligned} \quad (3.41)$$

This shows that the success probabilities of both the non-cooperative and cooperative strategies decrease exponentially as the intensity of the interfering sources λ increases. The success probability of the cooperative strategy, however, decreases more rapidly.

Random K and K_v

We now move on to the scenarios where a random number of relays are employed by every source. Specifically, we assume that both K and K_v follow a common

probability distribution. Note that our derivations can also be extended to the cases where K and K_v have different distributions. Similar to the case with a fixed number of relays, the threshold for cooperation β^* can be numerically calculated by solving $P_{\text{out}}^{\text{co}}(\beta) - P_{\text{out}}^{\text{non}}(\beta) = 0$, where $\beta = \lambda d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}}$. Since it is difficult to obtain a closed-form expression for β^* , in the following analysis, we focus on the asymptotic comparison when the intensity of the interfering sources λ or the source-destination distance d goes to infinity. Numerical and simulation results for β^* when K and K_v follow a Poisson distribution will be presented in Section 3.3.2.

Specifically, as the source-destination distance d increases, from (3.13), (3.15) and (3.22), we obtain

$$\begin{aligned} & \lim_{d \rightarrow \infty} \frac{\ln(1 - p_{\text{out}}^{\text{co}})}{d^\alpha} \\ &= -\frac{\gamma_{\text{th}} P_N}{P_t} \Pr(K = 1) - \frac{\gamma_{\text{th}} P_N}{P_t} \sum_{k=2}^{\infty} \frac{2}{k} \Pr(K = k) \\ &> -\frac{\gamma_{\text{th}} P_N}{P_t} (1 - \Pr(K = 0)). \end{aligned} \quad (3.42)$$

Eq. (3.42) shows that, with common distributions for K and K_v , $1 - p_{\text{out}}^{\text{co}}$ decreases more slowly than $1 - p_{\text{out}}^{\text{non}}$ as the source-destination distance d increases, which is consistent with (3.40).

On the other hand, the outage probability given by (3.13) is the sum of a sequence of functions, which can be proved to be uniformly convergent. Therefore, by interchanging the order of the limit and sum, we have

$$\begin{aligned} & \lim_{\lambda \rightarrow \infty} \frac{\ln(1 - p_{\text{out}}^{\text{co}})}{\lambda} \\ &= -A_\alpha d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}} \mathbb{E} \left[K_v^{\frac{2}{\alpha}} \right] \Pr(K = 1) - A_\alpha d^2 (2\gamma_{\text{th}})^{\frac{2}{\alpha}} \mathbb{E} \left[K_v^{\frac{2}{\alpha}} \right] \sum_{k=2}^{\infty} k^{-\frac{2}{\alpha}} \Pr(K = k) \\ &< -A_\alpha d^2 \gamma_{\text{th}}^{\frac{2}{\alpha}} \mathbb{E} \left[K_v^{\frac{2}{\alpha}} \right] \sum_{k=1}^{\infty} k^{-\frac{2}{\alpha}} \Pr(K = k). \end{aligned} \quad (3.43)$$

Since

$$\mathbb{E} \left[K_v^{\frac{2}{\alpha}} \right] \sum_{k=1}^{\infty} k^{-\frac{2}{\alpha}} \Pr(K = k)$$

$$\begin{aligned}
&= \left(\sum_{k_v=1}^{\infty} k_v^{\frac{2}{\alpha}} \Pr(K_v = k_v) \right) \cdot \left(\sum_{k=1}^{\infty} k^{-\frac{2}{\alpha}} \Pr(K = k) \right) \\
&> \left(\sum_{k=1}^{\infty} \sqrt{k^{\frac{2}{\alpha}} \Pr(K = k)} \sqrt{k^{-\frac{2}{\alpha}} \Pr(K = k)} \right)^2 \\
&= 1 - \Pr(K = 0), \tag{3.44}
\end{aligned}$$

using the Cauchy-Schwarz inequality, we can conclude from (3.43) and (3.44) that the success probability for the cooperative strategy $1 - p_{\text{out}}^{\text{co}}$ decreases faster than that for the non-cooperative strategy $1 - p_{\text{out}}^{\text{non}}$ as the intensity of interfering sources λ increases, which is consistent with (3.41) where K and K_v are fixed.

3.3 Simulation Results

In this section, simulation results are presented to verify the theoretical results. Insights into the performance and strategies of networks are also drawn. In order to approximate an infinitely large network of sources, a two-dimensional homogeneous Poisson process is used. All the interferers are randomly located in a large two-dimensional circle S with radius 10^5 . The reference destination is located at the origin, and the corresponding source is located at $(d, 0)$. Since the distance d in our simulation setting is much less than 10^5 meters, the approximation error due to finite S is negligible. The path loss exponent α is assumed to be 4, and the noise power is 1. The SINR threshold γ_{th} is set to be 0 dB. The 2-group Alamouti STBC is adopted as the cooperative strategy. All simulation results are averaged over 10^3 network location configurations and 10^4 channel realizations for each network configuration.

3.3.1 Validation of Analysis

In our analysis, we assume that the size of each vicinity is small, and all relay nodes are placed to overlap their associated sources. In the simulation, we relax this assumption and assume that each source has a circular vicinity with radius r . All the relays for a given source are uniformly distributed in its vicinity. By comparing the

simulation and analytical results, we show that the performance difference is negligible for practical systems.

Specifically, Fig. 3.2 shows the outage probability as a function of the average received signal power $P_t d^{-\alpha}$ for both non-cooperative and cooperative strategies. In Fig. 3.2, the distance between the source and destination, d , is assumed to be 50 meters, and the intensity of the interfering sources λ is 10^{-6} . We also assume that K_v is a fixed number which equals K . The solid line represents the analytical results obtained from (3.8) and (3.28), and the dotted and dashed lines represent the simulation results for different values of the size of vicinity r . Comparisons for different numbers of K and K_v are also shown in Fig. 3.2.

Similarly, in Fig. 3.3, we assume that K and K_v are Poisson distributed random variables with the same mean value κ . The outage probabilities are plotted for different values of κ and r . We compare the analytical results from (3.32) and the simulation results.

We observe that the analytical results outperform the simulation results which include the impact of randomness in the locations of relays. As the size of the vicinity r decreases, the relays move closer to the sources, and the simulation results eventually converge to our analysis, as expected. Another observation is that the outage probabilities converge to a constant as the transmit power P_t goes to infinity, as shown in (3.29), (3.32) and (3.33).

3.3.2 To Cooperate or Not to Cooperate

In this subsection, we evaluate the outage performance for both non-cooperative and cooperative strategies. Two different scenarios are considered: a fixed number of relays and a random number of relays. Both analytical and simulation results are provided. In the following figures, we will use the lines to represent analytical results and the markers to denote the simulation results. In the simulation, we assume that the size of the vicinity $r = d/10$, and all nodes use the same transmit power P_t so that

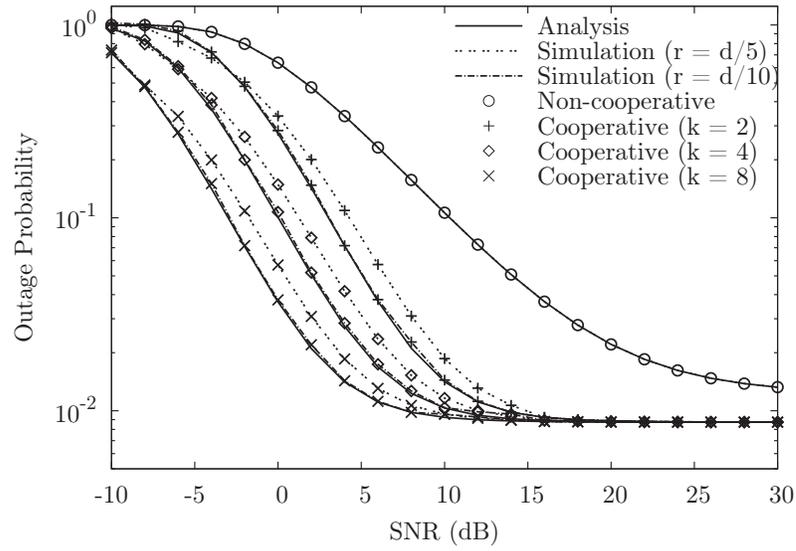


Figure 3.2: Outage probability versus the average received signal power ($d = 50$ m, $\alpha = 4$, $\lambda = 10^{-6}$, $K = K_v = k$).

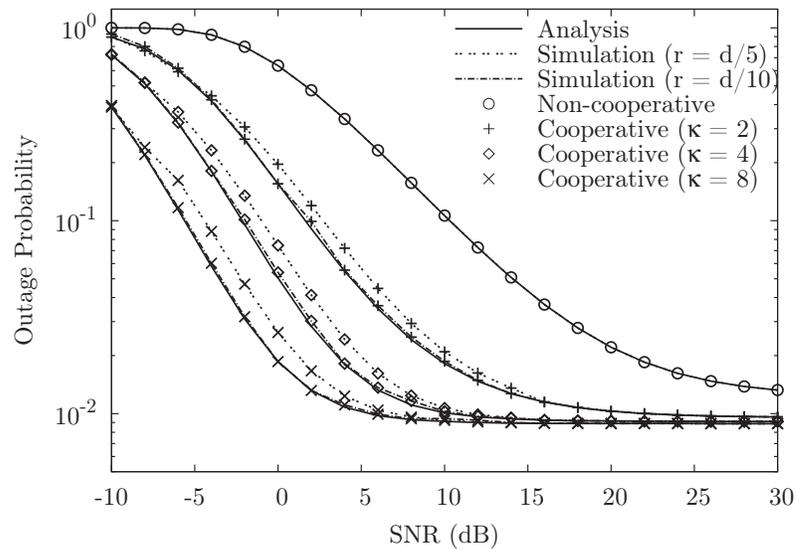


Figure 3.3: Outage probability versus the average received signal power. K and K_v follow a Poisson distribution with mean κ ($d = 50$ m, $\alpha = 4$, $\lambda = 10^{-6}$).

the average received signal power $P_t d^{-\alpha}$ (or, equivalently, the SNR) is 20 dB if $d = 50$ meters.

Fixed K and K_v

We first consider the scenario where K and K_v are assumed to be fixed and equal to 8. In Fig. 3.4, the outage probabilities of both non-cooperative and cooperative strategies are plotted as a function of λ , the intensity of the interfering sources. It can be clearly seen that the outage probability is always an increasing function of λ . According to Theorem 3, we can calculate that the threshold for cooperation β^* is around 0.4837 when $\alpha = 4$. Therefore, we know that cooperation is beneficial if and only if $\lambda < 0.4837/d^2 \approx 1.9348 \times 10^{-4}$. The simulation results in Fig. 3.4 verify that if λ less than about 1.94×10^{-4} , the cooperative strategies have smaller outage probability and achieve better performance²; otherwise, the non-cooperative strategy is superior but with only a small penalty.

In Figs. 3.5 and 3.6, the asymptotic behavior of the success probability is examined. In Fig. 3.5, the intensity of the interfering sources λ is set to 10^{-5} . We observe that the curves of success probabilities have an exponentially decreasing trend as the source-destination distance d becomes sufficiently large. The non-cooperative strategy has a faster decline, as shown in (3.40).

In Fig. 3.6, the source-destination distance is set to 50 meters. We can see that the success probabilities of both the non-cooperative and cooperative strategies exponentially decrease as the intensity of the interfering sources λ increases, which has been shown in (3.41). The slope of the non-cooperative strategy is smaller, which indicates that the cooperative strategy is more sensitive to the interference caused by the concurrent transmissions.

² Note that $\lambda \approx 1.9348 \times 10^{-4}$ implies that on average, there is an interferer which is around 78.4771 meters away from the source.

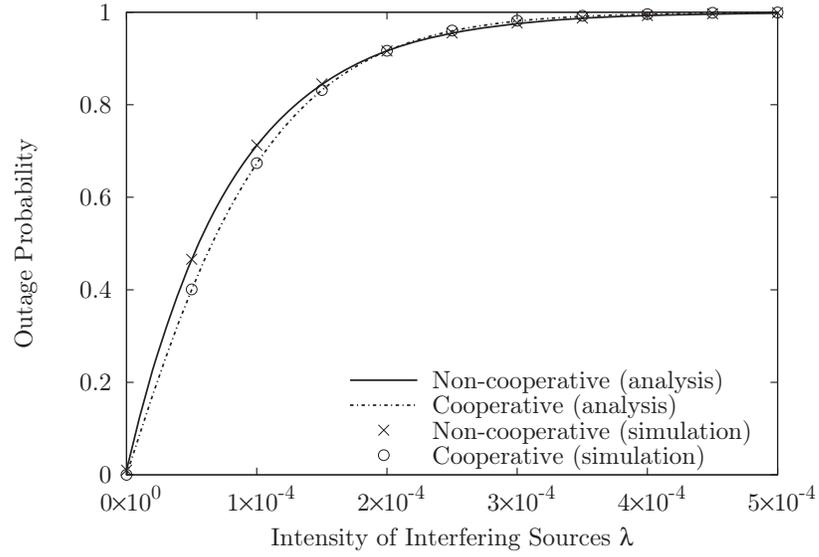


Figure 3.4: Outage probability versus the intensity of interfering sources λ ($d = 50$ m, $\alpha = 4$, $K = K_v = 8$).

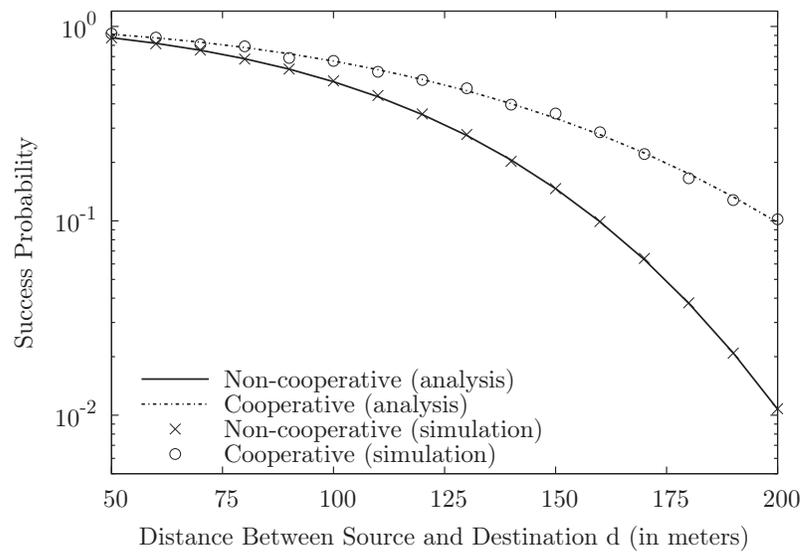


Figure 3.5: Asymptotic behavior of the success probability as the source-destination distance d increases ($\alpha = 4$, $\lambda = 10^{-5}$, $K = K_v = 8$).

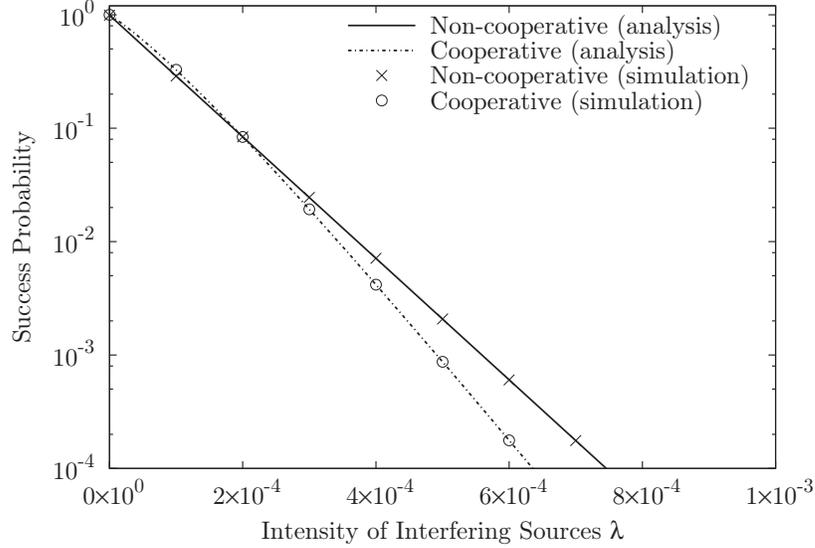


Figure 3.6: Asymptotic behavior of the success probability as the intensity of interfering sources λ increases ($d = 50$ m, $\alpha = 4$, $K = K_v = 8$).

Random K and K_v

Now we assume that K and K_v are Poisson distributed random variables with mean 16. In Fig. 3.7, the outage probabilities of both the non-cooperative and cooperative strategies are plotted as a function of λ . In this case, the threshold for cooperation β^* can be numerically obtained as $\beta^* \approx 0.5445$. Therefore, cooperation is beneficial if and only if $\lambda < 0.5445/d^2 \approx 2.178 \times 10^{-4}$, which is verified by the simulation results.

By comparing Figs. 3.4 and 3.7, it can be observed that the thresholds for cooperation are close to each other. Fig. 3.8 presents the numerical and simulation results for β^* when K and K_v follow a Poisson distribution with mean $\kappa = 16$. For the sake of comparison, the analytical and simulation results for the case with a fixed number of relays are also plotted in Fig. 3.8. The results show that β^* , in the case with fixed K and K_v , serves as a tight lower bound for the case with random K and K_v .

Figs. 3.9 and 3.10 illustrate the asymptotic behavior of the success probabilities. Similar to the case with a fixed number of relays, we observe from Fig. 3.9 that, with the same success probability, the cooperative strategy provides a longer transmission range

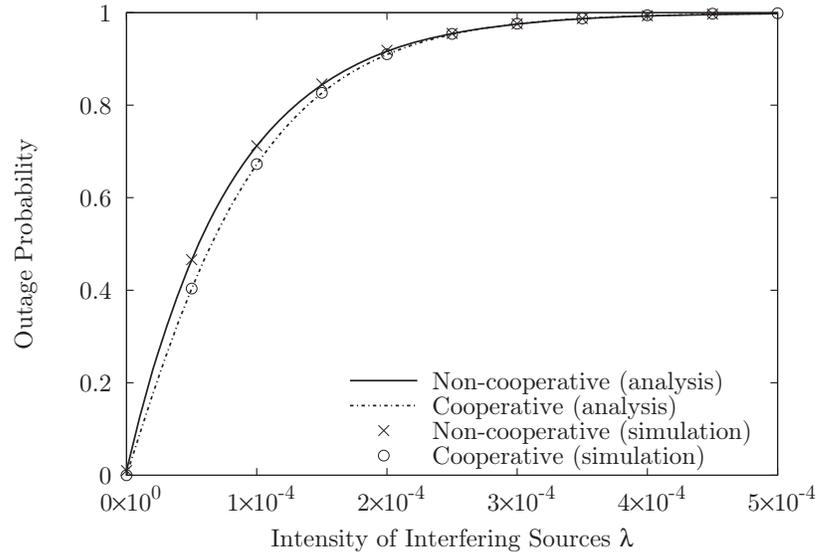


Figure 3.7: Outage probability versus the intensity of the interfering sources λ . K and K_v follow a Poisson distribution with mean 16 ($d = 50$ m, $\alpha = 4$).

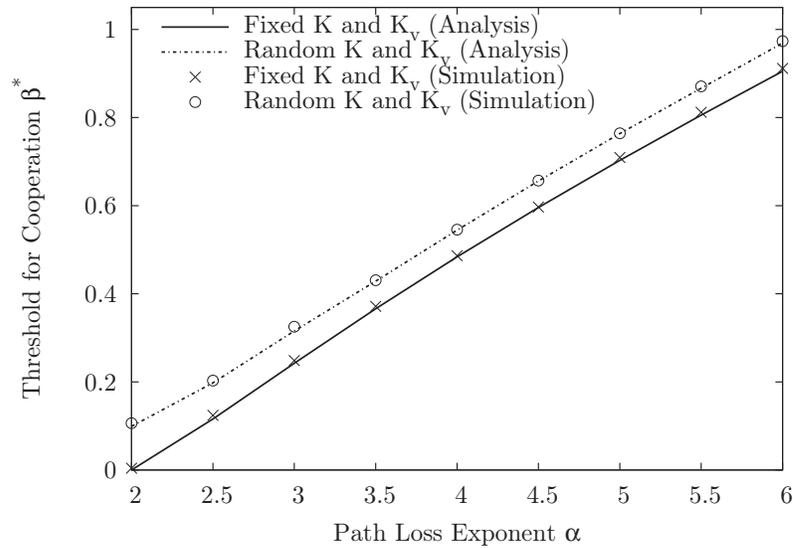


Figure 3.8: Threshold for cooperation β^* versus the path loss exponent α . In the case with a random number of relays, K and K_v follow a Poisson distribution with mean 16.

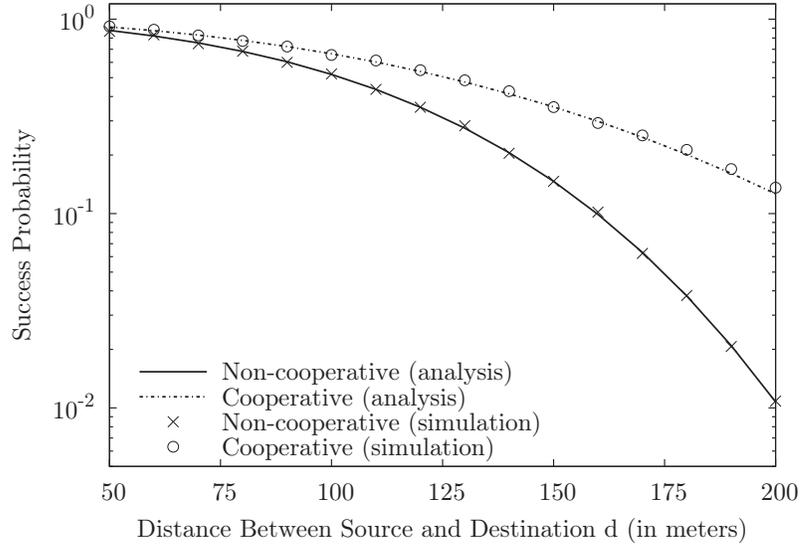


Figure 3.9: Asymptotic behavior of the success probability as the source-destination distance d increases. $\alpha = 4$, $\lambda = 10^{-5}$, K and K_v follow a Poisson distribution with mean 16.

d than the non-cooperative strategy, as expected. As the intensity of the interfering sources λ increases, the benefit of cooperation is eventually eliminated by the additional interference, as illustrated by the results in Fig. 3.10.

3.4 Summary

In this chapter, wireless networks with multiple concurrent transmissions and spatial interference were investigated. The closed-form outage performance for cooperative networks with an M -group STBC scheme was derived. By comparing the performance of non-cooperative and cooperative strategies, we derived a criterion which determines which strategy should be applied for a given network configuration. Our criterion shows that cooperation is beneficial if the margin between the SIR and the outage threshold γ_{th} is larger than a value that is a function of the path loss exponent α . This demonstrates the superiority of cooperative communication when the transmitting sources are sparsely distributed in the network. If the number of simultaneously transmitting sources in a network is small, using nearby relays could achieve a significant cooperative diversity benefit without creating too much interference to other

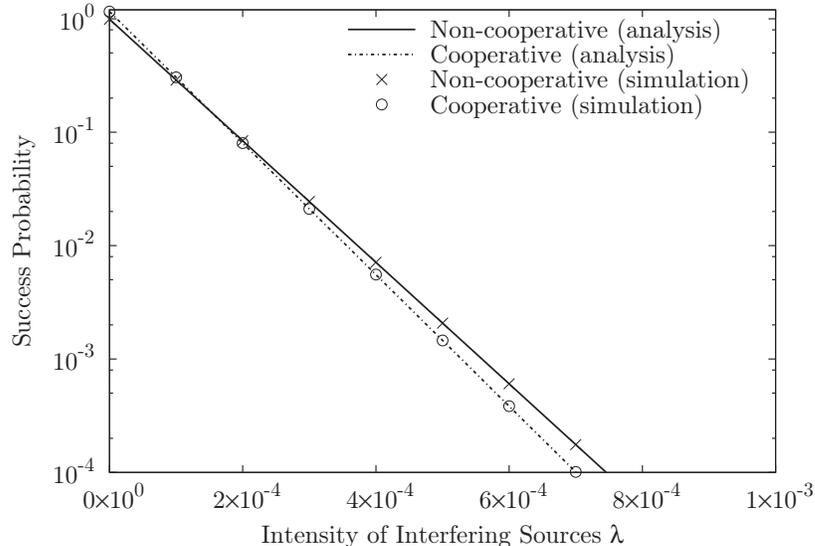


Figure 3.10: Asymptotic behavior of the success probability as the intensity of the interfering sources λ increases. $d = 50$ m, $\alpha = 4$, K and K_v follow a Poisson distribution with mean 16.

concurrent transmissions. Conversely, the non-cooperative strategy is preferred if the networks are dense and the interference dominates. The asymptotic properties of the outage probability as the intensity of the interfering sources λ or the source-destination distance d increases are studied. Simulation results, which verified our analyses, were also provided.

There are several possible extensions of this work. Other types of cooperative strategies, such as relay selection and beamforming, could be investigated. Although all these cooperative strategies achieve full diversity gain, their impacts on interference are very different. On the other hand, rate-adaptive transmission might also be considered. Since outage probability is only suitable for fixed-rate transmission, other corresponding performance measures such as outage capacity or ergodic sum rate, can be investigated. In addition, more practical networks should be studied. For example, the current analytical results are based on the assumption that the source-to-relay distance is relatively small compared to the source-to-destination distance. The impact of relay placement could be studied. Also, the performance analysis of multihop

wireless networks, which can be merged with routing protocols, would be a challenging task. Another direction could be a comparison of cooperative and non-cooperative schemes in a heterogeneous Poisson network. Finally, investigating the extra overhead incurred by cooperation and the overhead-performance tradeoff in a Poisson network is an interesting and valuable area for future study.

Chapter 4

MULTI-HOP LINEAR WIRELESS NETWORKS

Wireless ad hoc networks, such as Wireless Sensor Networks and MANET, have been extensively studied [83]. Relaying techniques are commonly applied for accomplishing the transmission in ad hoc networks since, in practical scenarios, the direct link between the source and destination could be very weak due to the possibly severe signal attenuation from path loss and shadow fading.

Multi-hop wireless networks use two or more wireless hops to convey information from a source to a destination. In other words, there are one or more intermediate nodes along the path that receive and forward information via wireless links. In previous chapters, we focused on special cases of multi-hop wireless networks where the number of hops is two. In this chapter, we will focus on general multi-hop networks.

4.1 Multi-hop Linear Network with Randomly Located Nodes

The number of hops is a very important metric for multi-hop communication networks. In general, any efficient routing protocol for a wired network [84] should have a small number of hops since more hops will lead to higher latency and lower reliability. The impact of hop counts on the end-to-end performance for wireless multi-hop networks, however, is not clear [85, 86]. In particular, it is not obvious whether it is advantageous to communicate over a large number of short hops, or over a small number of longer hops, or something in between.

A linear network is a commonly used, simplified, system model to investigate the impact of the number of hops. In a wireless network, the transmission process can be viewed as a linear network once the route is established by a given routing protocol. The problem of finding the optimum hop count for a linear network has been recently

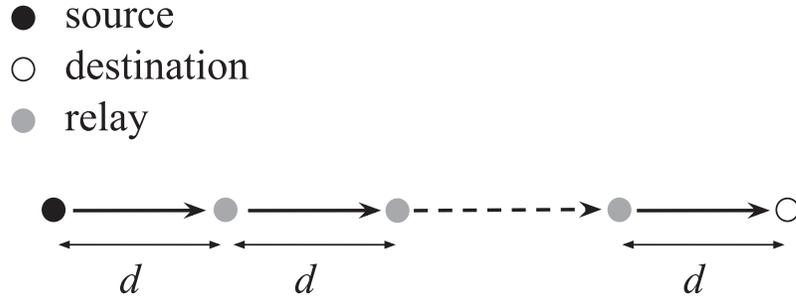


Figure 4.1: Linear network model with equidistant relay nodes.

studied in [86, 87]. In [86], a closed-form expression for the optimum hop count is derived for the simplest linear network without fading and interference. In [88], end-to-end capacity bounds over a linear network with fading are obtained, and a general equation for the number of hops which minimizes the outage probability is derived. In an AWGN channel, it is shown in [89] that equally spaced relays are optimum in a linear network. A simplified approach to determine the optimum number of hops is also provided in [89]. A new metric called the random access transport capacity is introduced in [90] to describe the performance of multi-hop wireless networks; the number of hops that maximizes this new performance measure is also provided.

To the best of our knowledge, most existing work considers a linear network model with equidistant nodes, as illustrated in Fig. 4.1. However, there are two notable practical issues with this model:

1. The hop link distance should be a random variable.
2. The source-destination distance varies with the actual route traversed.

In order to obtain a more reasonable linear network model, the randomness of the hop distance should be taken into account. In this section, we assume that the relay nodes are uniformly distributed over a short interval instead of assuming that all the nodes are equidistant from each other. In other words, only the nodes in a small area will be selected as relays to construct the linear multi-hop network. We also show that this novel model is a better approximation to reality than the traditional linear network model. By using some approximations, we obtain a closed-form expression for

the number of hops which can optimize the performance bound.

4.1.1 System Model

We propose a novel linear network model, called the *random shift model*, to characterize the randomness of the hop distance. The system under consideration consists of a source node and a destination node, and $N - 1$ intermediate relay nodes r_i , $i = 1, 2, \dots, N - 1$, randomly located on a straight line between the source and the destination. In other words, we consider a one-dimensional N -hop network. All the nodes in the network are assumed to have a single-antenna and be capable of only half-duplex decode-and-forward relaying. In particular, we consider a classic N -hop time-division decode-and-forward protocol, where each relay node r_i hears and fully decodes the data signal transmitted from r_{i-1} and forwards its re-encoded version to r_{i+1} .

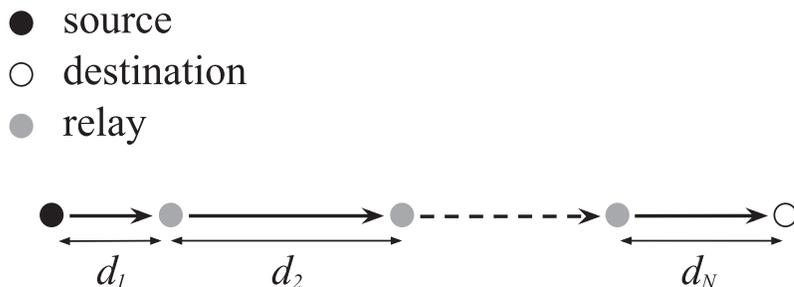


Figure 4.2: Random shift model: linear network with randomly located relay nodes.

In the random shift model, the source-to-destination distance is d_s . Notice that d_s is the actual distance traveled through the route, instead of the Euclidean distance between the source and destination. We assume that the locations of the intermediate relay nodes are independent random variables. Let x_i denote the position of the i -th relay on the line and assume the PDF of x_i is

$$f(x_i) = \begin{cases} \frac{1}{d\delta}, & d(i - \frac{\delta}{2}) \leq x_i \leq d(i + \frac{\delta}{2}) \\ 0, & \text{elsewhere} \end{cases} \quad (4.1)$$

where $d = \frac{d_s}{N}$, and δ is the *shift range*, which characterizes the randomness or uncertainty of the node position. Define the distance $d_i = x_i - x_{i-1}$, $i = 2, \dots, N-1$, with $d_1 = x_1$ and $d_N = d_s - x_{N-1}$. Then the CDF of d_i can be obtained as follows:

For $i = 1$ or N ,

$$F(d_i) = \begin{cases} 0, & d_i \leq d(1 - \frac{\delta}{2}) \\ \frac{1}{d\delta} (d_i - d(1 - \frac{\delta}{2})), & d(1 - \frac{\delta}{2}) \leq d_i \leq d(1 + \frac{\delta}{2}) \\ 1, & d_i \geq d(1 + \frac{\delta}{2}) \end{cases} \quad (4.2)$$

For $i = 2, \dots, N-1$,

$$F(d_i) = \begin{cases} 0, & d_i \leq d(1 - \delta) \\ \frac{1}{2} (\frac{1}{d\delta})^2 [d_i - d(1 - \delta)]^2, & d(1 - \delta) \leq d_i \leq d \\ \frac{1}{2} (\frac{1}{d\delta})^2 [d(1 + \delta) - d_i]^2, & d \leq d_i \leq d(1 + \delta) \\ 1, & d_i \geq d(1 + \delta) \end{cases} \quad (4.3)$$

We assume that the source node and all the relay nodes are supplied with finite transmit power P_t over the same frequency bandwidth B . Perfect time and frequency synchronization among all nodes in the system is also assumed. The signal received by node r_{i+1} can be expressed as

$$y_{i+1} = \sqrt{P_t d_i^{-\alpha}} s_i + n_i, \quad (4.4)$$

where s_i is the signal transmitted by node r_i , n_i is white Gaussian noise with zero mean and variance $\frac{N_0}{2}$ per dimension, and N_0 is the noise power spectral density. In (4.4), α is the path loss exponent (typically between 2 and 4). The SNR at the receiver of the i -th relay is then

$$\gamma_i = \frac{P_t}{P_N} d_i^{-\alpha}, \quad (4.5)$$

where $P_N = N_0 B$ is the noise power.

Our objective is to choose the number of hops to maximize the spectral efficiency, or the bandwidth-normalized achievable data rate η (in bits per second per hertz, bps/Hz). For a single hop, the spectral efficiency for a bandlimited AWGN channel is

$$\eta_{\text{sh}} = \log \left(1 + \frac{P_t}{P_N} d_s^{-\alpha} \right) \text{ bps/Hz}. \quad (4.6)$$

Divide the end-to-end transmission into N hops. Assume only one node is transmitting at any point in time, which implies that there is no interference at any receiver. Notice that, on each of the hops, nodes must transmit the same amount of information in $1/N$ -th of the channel uses available in the single hop case, i.e., the required per-hop spectral efficiency should be N times that of the single-hop case. For example, the spectral efficiency for a linear network with equidistant nodes can be expressed as

$$\eta_{\text{eq}} = \frac{1}{N} \log \left(1 + \frac{P_t}{P_N} \left(\frac{d_s}{N} \right)^{-\alpha} \right) \text{ bps/Hz.} \quad (4.7)$$

When the nodes are randomly placed, the system performance will be determined by the worst hop among all N hops. Since fading is not taken into account in our proposed model, the hop which has the worst performance is equivalent to the hop which has the largest distance d_i . The performance measure used in this case will be the spectral efficiency averaged over the randomly chosen distances

$$\begin{aligned} \eta &= \mathbb{E}_{d_1, \dots, d_N} \left[\frac{1}{N} \min_{i=1, \dots, N} \log \left(1 + \frac{P_t}{P_N} d_i^{-\alpha} \right) \right] \\ &= \mathbb{E}_{d_{\max}} \left[\frac{1}{N} \log \left(1 + \frac{P_t}{P_N} d_{\max}^{-\alpha} \right) \right] \text{ bps/Hz,} \end{aligned} \quad (4.8)$$

where $d_{\max} = \max_{i=1, \dots, N} d_i$, and $\mathbb{E}[\cdot]$ is the expectation operator.

4.1.2 Analysis

Here, we investigate the hop counts for a linear network with randomly located relay nodes. The objective is to find the number of hops N that maximizes the end-to-end spectral efficiency η . Using techniques similar to those in [91], the optimum N^* for a linear network with equidistant nodes is

$$N^* = \operatorname{argmax} \eta_{\text{eq}} \approx \left[\left(\frac{\omega_\alpha}{\gamma} \right)^{1/\alpha} \right]_+ \quad (4.9)$$

where

$$\omega_\alpha = \frac{-\alpha}{\mathcal{W}(-\alpha e^{-\alpha})} - 1 \quad (4.10)$$

is a constant which only depends on the path loss exponent α , $\gamma = \frac{P}{N_0 B} d_s^{-\alpha}$ is the receive SNR for the single-hop case, $[\cdot]_+$ is the operator which rounds the operand to the nearest positive integer, and $\mathcal{W}(\cdot)$ is the principal branch of the Lambert \mathcal{W} function [82].

In order to analyze the optimum number of hops for the random shift model, we need to obtain a closed-form expression for η , which requires knowledge of the CDF of d_{\max} . For example, when $N = 3$, we could derive the CDF of the maximum distance d_{\max} as

$$F(d_{\max}) = \begin{cases} 0, & d_{\max} < d \\ \frac{9}{2} \left(\frac{1}{d\delta}\right)^2 (d_{\max} - d)^2, & d \leq d_{\max} \leq d \left(1 + \frac{\delta}{4}\right) \\ 1 - \left(\frac{1}{d\delta}\right)^2 [d \left(1 + \frac{\delta}{2}\right) - d_{\max}] [d_{\max} - d \left(1 - \frac{3\delta}{2}\right)] \\ \quad - \frac{1}{2} \left(\frac{1}{d\delta}\right)^2 [d(1 + \delta) - d_{\max}]^2, & d \left(1 + \frac{\delta}{4}\right) \leq d_{\max} \leq d \left(1 + \frac{\delta}{2}\right) \\ 1 - \frac{1}{2} \left(\frac{1}{d\delta}\right)^2 [d(1 + \delta) - d_{\max}]^2, & d \left(1 + \frac{\delta}{2}\right) \leq d_{\max} \leq d(1 + \delta) \\ 1, & d_{\max} > d(1 + \delta) \end{cases} \quad (4.11)$$

The CDF of d_{\max} will get more complicated as N increases. In order to provide analytical and tractable solutions, several approximation techniques will be applied in the analysis which follows. As the function $\log(1 + x^{-\alpha})$ is a convex function of x , we can apply Jensen's inequality

$$\mathbb{E} \left[\log \left(1 + \frac{P_t}{P_N} d_{\max}^{-\alpha} \right) \right] \geq \log \left(1 + \frac{P_t}{P_N} (\mathbb{E}[d_{\max}^{-\alpha}]) \right) \quad (4.12)$$

Moreover, it is easy to show that

$$\begin{aligned} \frac{d_s}{N} \left(1 - \frac{\delta}{2} \right) &\leq d_1, d_N \leq \frac{d_s}{N} \left(1 + \frac{\delta}{2} \right) \\ \frac{d_s}{N} (1 - \delta) &\leq d_i \leq \frac{d_s}{N} (1 + \delta), i = 2, \dots, N - 1 \\ \frac{d_s}{N} &\leq d_{\max} \leq \frac{d_s}{N} (1 + \delta) \end{aligned} \quad (4.13)$$

Hence, lower and upper bounds for η can be obtained as

$$\eta \geq \frac{1}{N} \log \left(1 + \frac{P_t}{P_N} d_s^{-\alpha} N^\alpha (1 + \delta)^{-\alpha} \right) \quad (4.14)$$

$$\eta \leq \frac{1}{N} \log \left(1 + \frac{P_t}{P_N} d_s^{-\alpha} N^\alpha \right) = \eta_{\text{eq}} \quad (4.15)$$

Eq. (4.24) illustrates that the randomness of the locations degrades the system performance, i.e., equidistant placement of relays is optimal for a multi-hop linear network with AWGN channels, which has been shown in [89]. The number of hops that maximizes the upper bound is provided in (4.9). We can also obtain a closed-form expression for the number of hops that maximizes the lower bound in (4.14) as

$$N' \approx \left[\left(\frac{\omega_\alpha}{\gamma'} \right)^{1/\alpha} \right]_+, \quad (4.16)$$

where ω_α is given by (4.10), and

$$\gamma' = \frac{P_t}{P_N} d_s^{-\alpha} (1 + \delta)^{-\alpha}. \quad (4.17)$$

Although the N which maximizes the lower bound is not the exact solution, it follows the same trends and provides several insights. According to (4.16), N' only depends on the path loss exponent α , the transmit power P_t , the distance d_s , and the shift range δ . In general, when the received SNR for the single-hop case $\gamma = \frac{P_t}{P_N} d_s^{-\alpha}$ increases, the optimum number of hops will decrease for both the equidistant model and the random shift model. Intuitively, when the direct link between the source and destination is good enough, adding relay nodes will degrade the system performance since a higher per-hop spectral efficiency needs to be achieved in a time-division multi-hop linear network.

In order to accurately approximate the solution for N^* , we adopt another empirical approximation for $\mathbb{E}[d_{\max}]$,

$$\mathbb{E}[d_{\max}] \approx \frac{d_s}{N} \left(1 + \frac{\delta}{2} \frac{N-1}{N} \right), \quad (4.18)$$

We will validate this approximation via simulation. According to (4.18), we can easily show that the optimum number of hops $N^* = \operatorname{argmax} \eta_{\text{eq}}$ satisfies

$$(N^* - 1) \left[N^* - \left(1 + \frac{\delta}{2} \right) \left(\frac{\omega_\alpha^*}{\gamma} \right)^{1/\alpha} \right] = 0, \quad (4.19)$$

where

$$\omega_\alpha^* = \frac{-\nu}{\mathcal{W}(-\nu e^{-\nu})} - 1, \quad \nu = \alpha \left(1 + \frac{\delta}{(\delta + 2)N - \delta} \right). \quad (4.20)$$

Note that ω_α^* and ν are nonlinear and irregular functions of N , which indicates that we cannot get a closed-form expression for N^* from (4.19). Fortunately, this equation can be easily solved numerically. Moreover, when $\delta = 0$, (4.16) and (4.19) are equivalent to (4.9).

4.1.3 Simulation Results

To justify our random shift model, we consider a practical wireless communication system which consists of a source-destination pair, and several potential relay nodes which form a homogeneous two-dimensional Poisson point process Π of intensity λ [92]. All nodes are located in a circle of radius \mathcal{R}_c (centered at the source node). The destination node is located on the boundary of the circle.

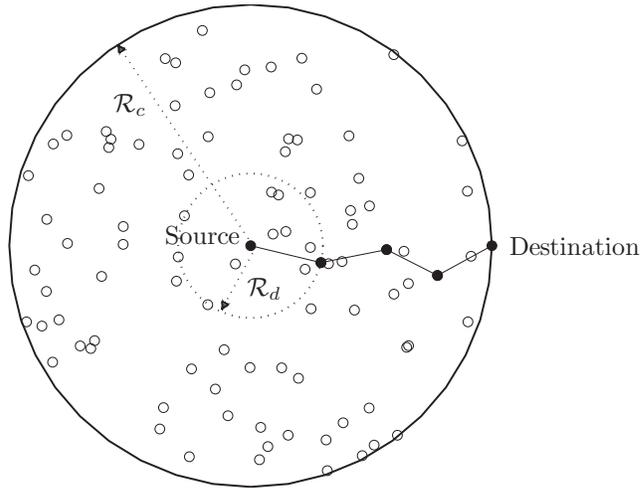


Figure 4.3: Illustration of a practical multi-hop wireless network

The route will be established based on the given \mathcal{R}_d as shown in Fig. 4.3. For example, we can use geographic routing algorithms such as greedy forwarding [93]. In the greedy forwarding protocol, each relay node forwards the message to its neighbors,

then the node that minimizes the distance to the destination will be chosen as the relay in each step.

The CDF of the maximized distance d_{\max} for greedy forwarding routing and the random shift model is plotted in Fig. 4.4. In order to fairly compare these two models, we constrain the number of hops to $N = 4$. In other words, we discard all the possibilities that there are more or less than 4 hops in a realistic network. It can be seen that by appropriately choosing d_s and δ in our model, the difference between the random shift model and a real system is negligible. Obviously, obtaining proper parameters for our model is a difficult task; this should be a topic of future work.

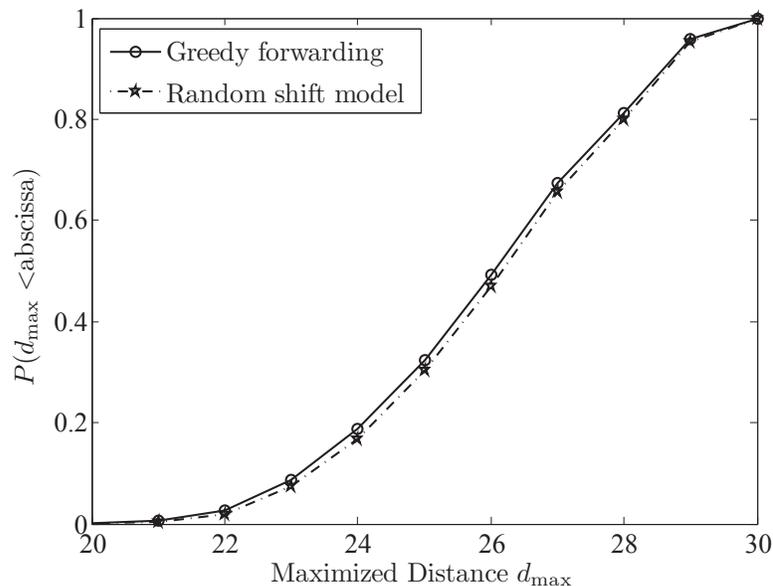


Figure 4.4: CDF of the maximized distance d_{\max} for a realistic system and the random shift model. ($\mathcal{R}_c = d_s = 100$ m, $\mathcal{R}_d = 30$ m, $N = 4$, $\delta = 0.2$)

We also evaluate the optimum number of hops for the random shift model. A linear network with randomly located relay nodes is considered. The end-to-end straight line distance $d_s = 100$ meters, and the i -th relay is uniformly distributed in $[\frac{d_s}{N}(1 - \delta), \frac{d_s}{N}(1 + \delta)]$. We consider the path loss model $\ell(d) = d^{-\alpha}$ with the path loss exponent α set to 4. The source and the relays use the same transmit power P_t , and the noise power P_N is assumed to be 1 for convenience.

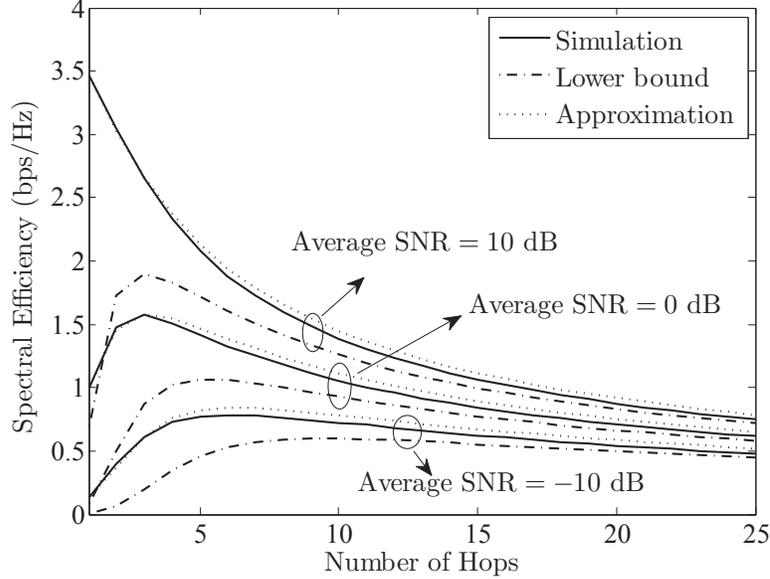


Figure 4.5: Lower bound and an approximation for the spectral efficiency using the random shift model (10^6 trials, $d_s = 100$ m, $\alpha = 4$, $\delta = 1$).

Fig. 4.5 illustrates the spectral efficiency for the random shift model; the lower bound is provided by (4.14) and the approximation by (4.18). The average SNR γ means the average receive SNR in the *single-hop* case, i.e., $\gamma = P_t d_s^{-\alpha}$. We consider three cases of average SNR γ , corresponding to three different power levels. According to Fig. 4.5, the approximation works well. We note that our approximation is not sensitive to the parameters, such as δ and d_s , although Fig. 4.5 only shows the special case when $\delta = 1$ and $d_s = 100$ m. It can also be observed that the lower bound, which follows the same trend of performance, cannot provide a suitable approximation for the optimum number of hops. In particular, the number of hops that maximizes the lower bound is much larger than the actual optimum number of hops.

Fig. 4.6 provides results for different values of average received SNR γ and shift range δ . When $\delta = 0$, the nodes are equidistant. In this case, the performance using the random shift model is exactly the same as for a conventional linear network model. As expected, we can also see that the overall performance decreases as the uncertainty increases (δ increases). Another observation is that the optimum number of hops changes as δ changes from 0 to 1; this indicates that results for the equidistant linear

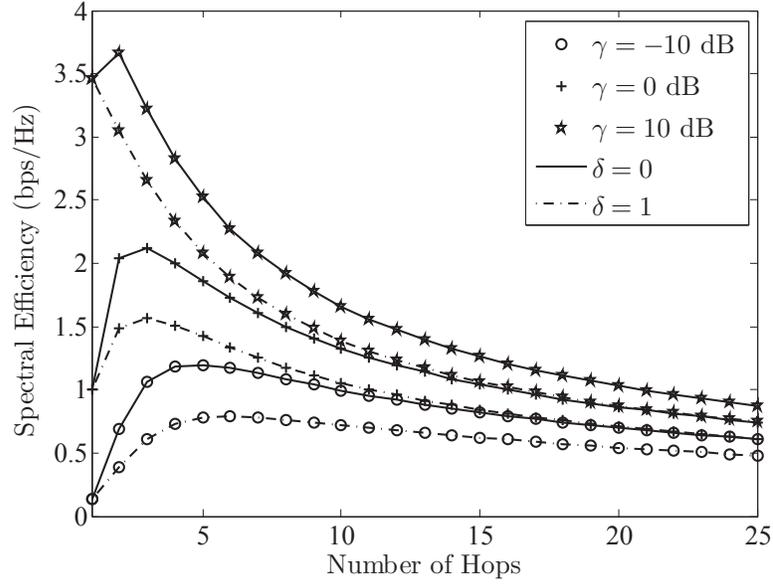


Figure 4.6: Spectral efficiency as a function of the number of hops for different values of average received SNR and shift range. (10^6 trials, $d_s = 100$ m, $\alpha = 4$).

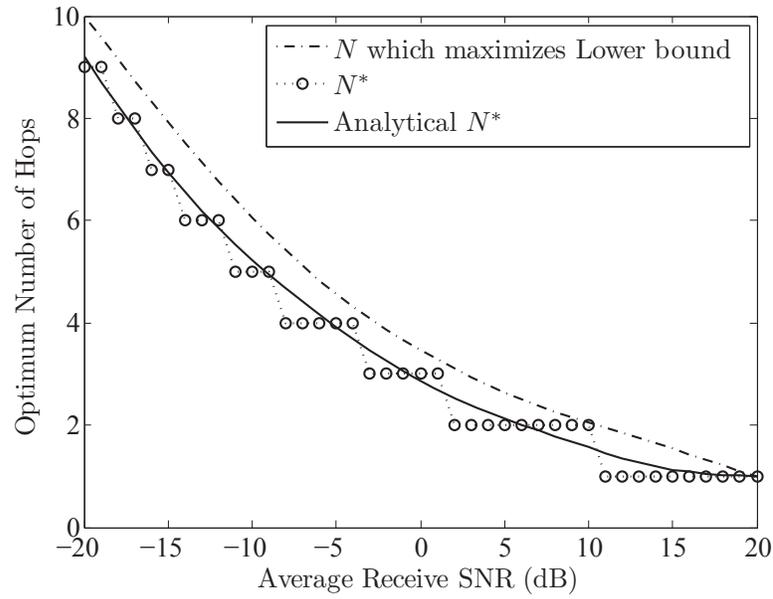


Figure 4.7: Optimum number of hops N^* (10^6 trials, $d_s = 100$ m, $\alpha = 4$, $\delta = 0.25$).

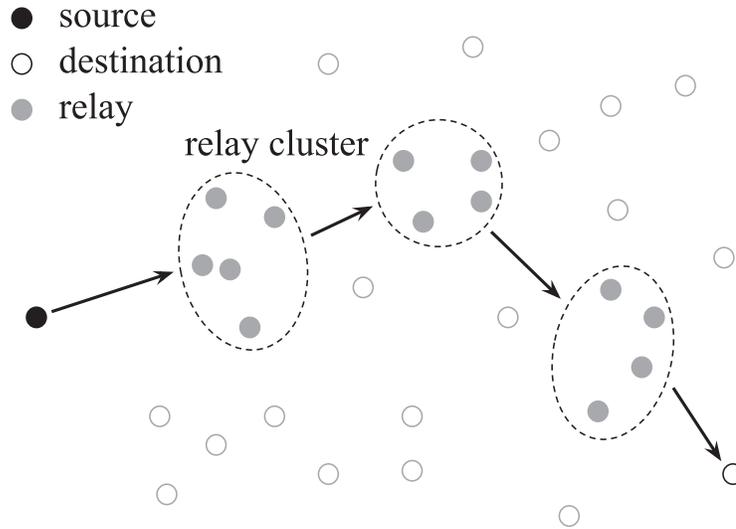


Figure 4.8: Multi-hop wireless ad hoc network. The route contains a source-destination pair and several relay clusters. The nodes in each cluster are selected from the potential relay nodes (small circles) according to a given routing protocol.

network model might not apply to practical topologies. By using numerical methods, we can also solve (4.19). The comparison between the analytical and simulation results for N^* are given in Fig. 4.7. By rounding the analytical result to the nearest positive integer, we can easily obtain the optimum number of hops.

4.2 Relay Deployment in Multi-Hop Linear Network with Cooperation

The relay deployment problem, which aims to optimally position the relays, is a key design issue that helps provide better network performance (network connectivity, lifetime, etc.) [94–97]. Relay deployment in cooperative networks has been extensively studied. For example, in [97], an approach for applying cooperative techniques and relay deployment to maximize the network lifetime, has been proposed. In [98], relay selection strategies have been designed to achieve full diversity gain for a multi-hop network.

Fig. 4.8 illustrates a multi-hop ad hoc network, which can be simplified to a linear network model as illustrated in Fig. 4.9. In this section, we focus on a multi-hop linear network with cooperative relays, and we strive to answer the following questions:

Where should the relay clusters be located? How many relays should be in each cluster? First, we derive the optimum relay cluster locations which minimize the end-to-end outage probability. Then, we consider the required cooperation overhead by using the overhead-performance tradeoff analysis in Chapter 2. A larger number of relays could lead to worse performance because of the extra overhead costs in implementing cooperation. The optimum number of relays, which maximizes the throughput, is then discussed.

4.2.1 System Model

We consider a generalized N -hop linear network model with cooperative relays. The system under consideration consists of a source node and a destination node, and $N - 1$ intermediate relay clusters which are located between the source and destination. The number of nodes in the j th relay cluster is denoted as K_j , $j = 1, 2, \dots, N - 1$, which implies that there are $\prod_{j=1}^{N-1} K_j$ distinct end-to-end paths in the network. Each path can be represented by a set containing the indices of the relays in all the clusters. The source-to-destination distance is assumed to be d_s .

We assume that the inter-cluster distance is much larger than the intra-cluster distance. Since it has been shown in [89] that equally spaced relays are optimum in a linear network, we also assume that the relay clusters are equidistant (i.e., the inter-cluster distance is d_s/N). We follow most of the assumptions in Section 4.1, including the channel model, time division system design, and perfect CSI at the receiver.

We consider a selective decode-and-forward, fixed-rate, relaying strategy (i.e., at each hop, only one relay node is selected to forward the signal at a constant transmission rate). Two selection schemes are investigated in this paper: optimal selection (select the “best” path from all end-to-end paths) and ad hoc selection (select the “best” path at each hop) [98]. Suppose that the transmission path (selected by some specific criteria) is represented by $\{r_1, r_2, \dots, r_{N-1}, r_N\}$, where r_j is the selected relay in the

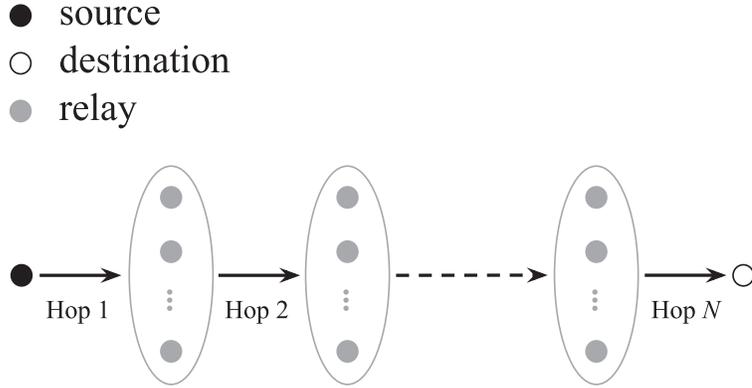


Figure 4.9: Linear network model with cooperative relays.

j th cluster, and r_N denotes the destination node. Then, the SNR at r_j is

$$\gamma_j = \frac{P_t}{P_N} \left(\frac{d_s}{N} \right)^{-\alpha} |h_{r_{j-1}r_j}|^2, \quad (4.21)$$

where $h_{r_{j-1}r_j}$ is the channel coefficient between r_{j-1} and r_j .

The end-to-end spectral efficiency can be expressed as

$$\eta = \frac{1}{N} \left[\log \left(1 + \min_{j=1,2,\dots,N} \gamma_j \right) \right], \quad (4.22)$$

and the outage probability is

$$\begin{aligned} p_{\text{out}} &= \Pr \left\{ \frac{1}{N} \left[\log \left(1 + \min_{j=1,2,\dots,N} \gamma_j \right) \right] < \eta \right\} \\ &= \Pr \left\{ \min_{j=1,2,\dots,N} \gamma_j < 2^{N\eta} - 1 \right\} \\ &= \Pr \left\{ \min_{j=1,2,\dots,N} |h_{r_{j-1}r_j}|^2 < \frac{2^{N\eta} - 1}{N^\alpha \gamma_d} \right\} \\ &= \Pr \left\{ g_{\min} < \frac{2^{N\eta} - 1}{N^\alpha \gamma_d} \right\}, \end{aligned} \quad (4.23)$$

where $\gamma_d = \frac{P_t}{P_N} d_s^{-\alpha}$ is the received SNR for the direct link, and $g_{\min} = \min |h_{r_{j-1}r_j}|^2$ is the channel gain for the worst hop in the selected transmission path. Obviously, the end-to-end outage is determined by the bottleneck hop.

We notice that the channel gains $|h_{r_{j-1}r_j}|^2$, $j = 1, 2, \dots, N$ are identically distributed exponential random variables, but they are not independent; the channel at the j th hop depends on which nodes are selected in the previous relay clusters.

4.2.2 Optimum Relay Deployment Strategy

In this section, we discuss the optimum number of hops for a multi-hop linear network with cooperative relays, and then determine the best relay placement. To simplify the analysis, we first assume that $K_1 = K_2 = \dots = K_{N-1} = K$. However, we also extend our analysis to more practical scenarios (for example, K_j is a random variable and not necessarily equal for all j). Then, we investigate the optimum number of relays per cluster, which can balance the system performance and the required overhead.

For the optimal relay selection strategy, the path which maximizes the channel gain for the bottleneck link, g_{\min} , will be chosen. Although there are $\prod_{j=1}^{N-1} K_j = d_s^{N-1}$ distinct end-to-end paths, some of these paths might share the same bottleneck link. Let \mathcal{S} denote the set that contains the bottleneck links of all possible paths, and ϑ is the number of distinct elements in \mathcal{S} . In other words, ϑ represents the degrees of freedom that can be utilized for diversity gain.

It has been shown in [98] that the end-to-end outage probability for optimal selection can be upper bounded by

$$p_{\text{out}} < \left(1 - \exp\left(-\frac{(2^{N\eta} - 1)}{N^\alpha \gamma_d}\right) \right)^\vartheta, \quad (4.24)$$

According to Lemma 1 in [98], $\vartheta \geq K$. So

$$p_{\text{out}} < \left(1 - \exp\left(-\frac{(2^{N\eta} - 1)}{N^\alpha \gamma_d}\right) \right)^K = p_{\text{out}}^*. \quad (4.25)$$

It is easy to show that the optimum N^* that minimizes the upper bound p_{out}^* satisfies

$$\alpha + N^* \eta 2^{N^* \eta} \ln 2 - \alpha 2^{N^* \eta} = 0. \quad (4.26)$$

Using techniques similar to those in [91], we have

$$N^* = \operatorname{argmin} p_{\text{out}}^* = \left[\frac{1}{\eta \ln 2} (\alpha + \mathcal{W}(-\alpha e^{-\alpha})) \right]_+. \quad (4.27)$$

In [98], an approximation for p_{out} is provided as

$$p_{\text{out}} \approx 2 \left(1 - \exp \left(-\frac{(2^{N\eta} - 1)}{N^\alpha \gamma_d} \right) \right)^K - \left(1 - \exp \left(-\frac{(2^{N\eta} - 1)}{N^\alpha \gamma_d} \right) \right)^{2K} + o \left(\left(-\frac{(2^{N\eta} - 1)}{N^\alpha \gamma_d} \right)^K \right), \quad (4.28)$$

where the last term is negligible in the high-SNR regime. We can prove that the number of hops N^* in (4.27) minimizes the approximation as well. On the other hand, for a fixed-rate scheme, the end-to-end throughput can be defined as $\eta(1 - p_{\text{out}})$. Therefore, N^* in (4.27) also maximizes the throughput.

Remark 4.1. Eq. (4.27) is exactly the same as the closed-form expression for the optimum number of hops for a linear network in an AWGN channel [87]. In [87], the power consumption that guarantees a given transmission rate is minimized. Note that there is an inherent power constraint in our model; our problem, which maximizes the rate by using a specific power, can be stated as a dual problem of the optimization problem in [87]. The duality gap is zero since both problems are convex and linearity constraint qualification conditions [99] are satisfied. This also explains why the optimum number of hops only depends on the rate and the path loss exponent.

Remark 4.2. According to (4.27), the optimum number of hops does not depend on the number of relays in each cluster, i.e., *diversity does not affect the optimum number of hops when the relay clusters are equidistant*. Obviously, the end-to-end outage performance can be significantly improved when we have diversity gain; however, the number of hops minimizing the outage probability remains the same. The intuitive explanation is that, after we select the path through all relay clusters, we form another linear network which only has one relay per hop. The diversity benefit helps increase the received SNR per hop, however, the structure of the linear network and the derivation of the optimum number of hops do not change.

Remark 4.3. Eq. (4.27) can be rewritten as

$$N\eta = \frac{1}{\ln 2} (\alpha + \mathcal{W}(-\alpha e^{-\alpha})). \quad (4.29)$$

The right-hand-side (RHS) of (4.29) is a constant which only depends on the path loss exponent α . For example, when $\alpha = 4$, we have $N\eta \approx 5.66$. In [100], the rate which maximizes the transport capacity is also given by the RHS of (4.29). The result in [100] can be considered as a special case of the work considered here: for single-hop transmission ($N = 1$) with a power constraint, maximizing the transport capacity is equivalent to minimizing the outage.

Note that these results only hold for a fixed-rate relaying scheme. If rate-adaptive techniques are also taken into account or the ergodic capacity is chosen as the performance measure, the optimum number of hops will be different.

If we assume the K_j 's are not necessarily all equal, the upper bound in (4.24) can still be used. The following theorem helps to determine the optimum number of hops in this scenario.

Theorem 4.1. \mathcal{S} includes at least $\min\{K_1, K_2, \dots, K_{N-1}\}$ distinct links, i.e., $\vartheta = |\mathcal{S}| \geq K_{\min} = \min\{K_1, K_2, \dots, K_{N-1}\}$.

Proof. Without loss of generality, we assume that $K_1 = K_{\min}$. In this case, a link in the first hop can be shared by at most $\prod_{j=2}^{N-1} K_j$ paths, which implies that at least K_1 links are required to cover all possible paths, that is, $\vartheta \geq K_{\min}$. ■

According to Theorem 4.1, we can rewrite (4.24) as

$$p_{\text{out}} < \left(1 - \exp\left(-\frac{(2^{N\eta} - 1)}{N^\alpha \gamma_d}\right)\right)^{K_{\min}} = p_{\text{out}}^*. \quad (4.30)$$

Obviously, N^* in (4.27) can also minimize p_{out}^* . In a practical wireless ad hoc network, the number of relays in each cluster should be a random variable. We can assume that all the nodes in an ad hoc network form a homogenous Poisson point process of intensity λ [79]. Then, the probability of finding k nodes in a bounded space \mathcal{A} is given by a discrete Poisson distribution

$$\mathcal{A}(k) = \Pr\{k \text{ nodes in } \mathcal{A}\} = e^{-\lambda \mathcal{M}(\mathcal{A})} \frac{(\lambda \mathcal{M}(\mathcal{A}))^k}{k!}, \quad (4.31)$$

where $\mathcal{M}(\mathcal{A})$ is a standard Lebesgue measure (area, volume, etc.) of \mathcal{A} . We can state that $\kappa = \lambda\mathcal{M}(\mathcal{A})$ is the average number of decoded nodes in the given relay cluster \mathcal{A} . Therefore, we assume that K_1, K_2, \dots, K_{N-1} are i.i.d. Poisson random variables with parameter κ , and

$$\Pr\{K_{\min} = k\} = \left(1 - \frac{\Gamma(k, \kappa)}{\Gamma(k)}\right)^{N-1} - \left(1 - \frac{\Gamma(k+1, \kappa)}{\Gamma(k+1)}\right)^{N-1}, \quad (4.32)$$

where $\Gamma(k, \kappa)$ is the incomplete Gamma function defined in (2.34).

Combining (4.30) and (4.32), we can obtain an upper bound on the average outage probability with a random number of potential relays

$$p_{\text{out}} < \sum_{k=0}^{\infty} \Pr\{K_{\min} = k\} \left(1 - \exp\left(-\frac{(2^{N\eta} - 1)}{N^{\alpha\gamma_d}}\right)\right)^k, \quad (4.33)$$

and show that N^* in (4.27) also minimizes the upper bound in (4.33). The proof is similar to the proof of (4.9) and is omitted here.

Since the diversity does not affect the optimum number of hops even if we consider the randomness of the decoded sets, we can easily extend the results for linear networks with a single relay per hop [87, 89, 91, 100] to our scenarios. In the following, we assume $K_1 = K_2 = \dots = K_{N-1} = K$.

In an ad hoc selection scheme, the relay selection is performed in a per-hop manner and performance is suboptimal. A high-SNR approximation is provided in [98]

$$p_{\text{out}} \approx (N - 2 + 2^K) \left(\frac{2^{N\eta} - 1}{N^{\alpha\gamma_d}}\right)^K, \quad (4.34)$$

We can show that the optimum N must satisfy

$$N(\ln N)^2 + (c_1 + c_2)N \ln N + c_2 \ln N - \alpha c_1 N + \alpha c_1 = 0, \quad (4.35)$$

where $c_1 = (2^K - 2)\eta \ln 2$, $c_2 = \alpha - \frac{1}{K}$. Although a closed-form expression for the optimum N cannot be obtained since (4.35) is a nonlinear transcendental equation, we can solve it by numerical methods. Through simulation results in Section 4.2.3, we will show that the optimum number of hops from (4.35) is very close to the value that satisfies (4.27).

Suppose we have a source-destination pair with desired spectral efficiency η . Then, we can easily obtain the optimum N^* based on (4.27) and then equally place $N^* - 1$ relay clusters between the source and the destination. Note that N^* does not depend on the number of relays K , which indicates that we can separate the relay deployment problem into two parts: (1) deciding the locations of the relay clusters and (2) determining the number of relays in each cluster. We have already addressed the first part by determining the optimum number of hops. Now we will investigate the optimum number of relays which achieves the optimal overhead-performance tradeoff.

Obviously, the outage probability decreases as the number of relays K increases because we have more diversity gain. The outage capacity, $\eta_{\text{out}} = \max \eta(1 - p_{\text{out}}(\eta))$ (bps/Hz), is thus a monotonically increasing function of K . However, in a realistic system, the receivers require knowledge of the CSI so that the signal can be successfully decoded. This can be facilitated by sending training symbols. Also, the receivers need to feedback some information, such as the CSI or the index of the best path, to implement the relay selection strategies. Intuitively, when we have more relays, we will incur more overhead for the training and selection tasks. The training and feedback overhead, which is also a monotonically increasing function of K , should also be considered. In that case, more relays does not necessarily lead to better performance.

According to [57], the smallest number of training symbols for a multiple-antenna system is equal to the number of transmit antennas. This result can be directly applied to our scenarios. For the optimal selection scheme, we have $K^2(N - 2) + 2K$ links in the network, which requires at least $K^2(N - 2) + 2K$ training symbols to guarantee meaningful channel estimation. For the ad hoc selection scheme, we need at least K training symbols per hop, and the overall number of training symbols is NK .

The feedback overhead for relay selection has been studied in Chapter 2. In general, $\log K$ feedback bits are required to implement perfect selection among K possible links. Intuitively, $\log K$ is the entropy (uncertainty) of the index for the best link from all K links. For the ad hoc selection scheme, we require $(N - 1)\log K$ feedback bits to choose the path. The analysis for the optimal selection scheme is more

complicated. Here, we use a simplified analysis to characterize the feedback overhead bits for the optimal selection scheme: since we have to select the best path from all K^{N-1} paths, it is reasonable to use at least $\log K^{N-1} = (N-1)\log K$ bits to denote the index of the selected path.

We note that the optimal and ad hoc selection schemes require the same amount of feedback overhead. However, optimal selection costs much more in training overhead than the ad hoc selection scheme. Note also that the analysis here only provides lower bounds for the required overhead.

Suppose that the feedback signals for updating the selected path are sent periodically with period T (which is usually chosen as 10% of the channel coherence time), and the training symbol duration is T_s , i.e., there are T/T_s symbols per block. Then, the spectral efficiency for optimal selection is

$$\eta_{\text{se,opt}} = \left(1 - \frac{K^2(N-2) + 2K}{T/T_s}\right) \eta_{\text{out}} - \frac{(N-1)\log K}{BT} \text{ bps/Hz}, \quad (4.36)$$

and the spectral efficiency for ad hoc selection is

$$\eta_{\text{se,ad hoc}} = \left(1 - \frac{NK}{T/T_s}\right) \eta_{\text{out}} - \frac{(N-1)\log K}{BT} \text{ bps/Hz}. \quad (4.37)$$

The optimal K which maximizes (4.36) or (4.37) can be obtained numerically.

4.2.3 Simulation Results

Assume we have a source-destination pair at a distance $d_s = 1$ km. The objective is to place several relays between the source and destination, such that the end-to-end outage performance is optimized. All nodes in the network, including the source and the destination, are supplied with transmit power $P_t = 20$ dBm over the frequency bandwidth $B = 10$ MHz. The path loss exponent α is assumed to be 4, and the noise spectral density $N_0 = -174$ dBm/Hz. When no relay is employed, the average received SNR at the destination is $\gamma_d = 4$ dB. Note that the general conclusions, which can be observed from the following simulation results, do not depend upon the specific values of these parameters.

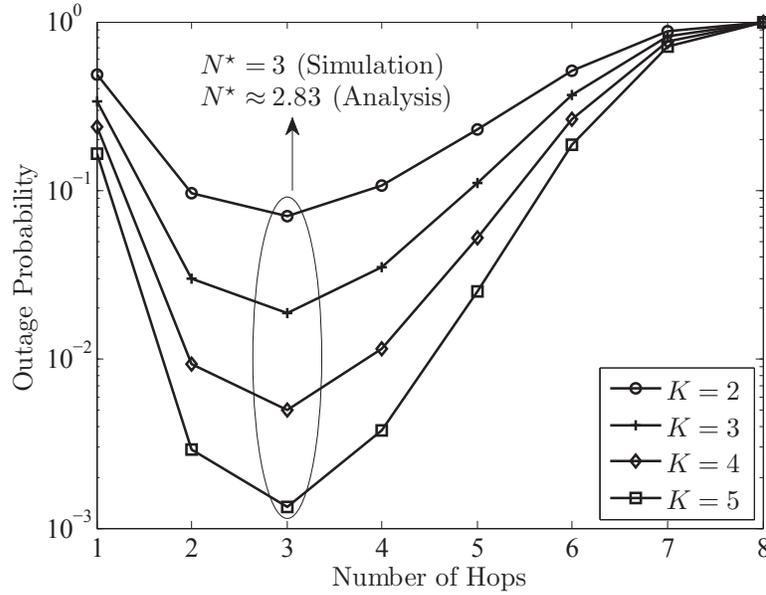


Figure 4.10: Optimum number of hops for different values of K (optimal selection), $\eta = 2$ bps/Hz, $\gamma_d = 4$ dB.

Where Should The Relay Clusters Be Located?

Since equidistant relay clusters have been shown to be optimal, once we know the number of hops which can minimize the outage probability, the optimal relay deployment strategy can be determined. In Figs. 4.10 and 4.11, the outage probability is plotted as a function of the number of hops for the optimal and ad hoc selection schemes. The desired spectral efficiency η is assumed to be 2 bps/Hz. As expected, we can observe from Fig. 4.10 that the optimum number of hops N^* does not depend on the number of relays K for the optimal selection scheme. The simulation results also verify our analytical results (4.27). According to Fig. 4.11, the optimum number of hops for ad hoc selection is almost the same as that given by (4.27). This indicates that we can use the analysis for the optimal selection scheme to obtain approximate results for the ad hoc selection scheme.

Fig. 4.12 provides results for different values of the desired spectral efficiency η . The number of relays K is chosen to be 4 in Fig. 4.12. By rounding the analytical result to the nearest positive integer, we can easily obtain the optimum number of hops

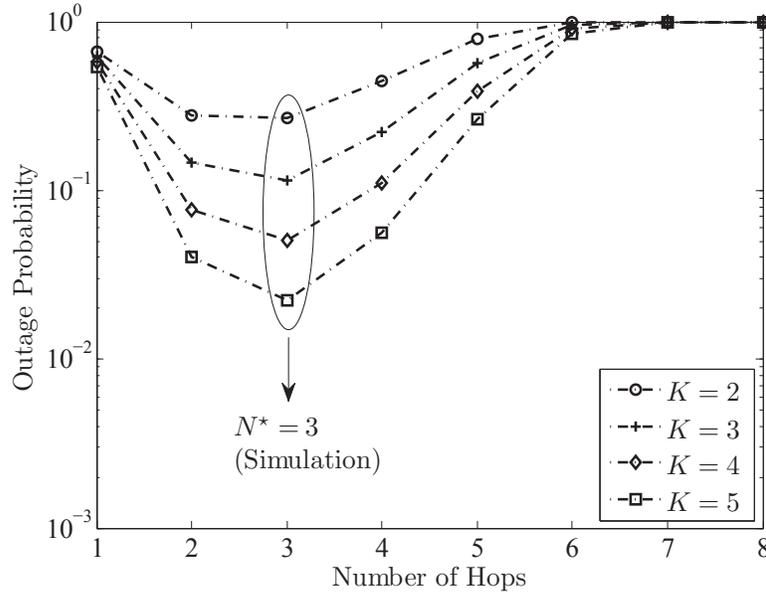


Figure 4.11: Optimum number of hops for different values of K (ad hoc selection), $\eta = 2$ bps/Hz, $\gamma_d = 4$ dB.

and determine where to place the relay clusters. For example, if η is 1.5 bps/Hz, the optimum number of hops is 4 and the optimum per-hop distance is 250 meters.

How Many Relays Should Be In Each Cluster?

Now we determine the optimum number of relays by investigating the required training and feedback overhead. Consider a wireless system with moderate mobility such that the coherence time is 10 msec [53]. The feedback signals for updating the selected path are sent every 1 msec. The training symbol duration is assumed to be $T_s = 1 \mu\text{sec}$, i.e., there are 1000 symbols to be transmitted in each block. Intuitively, the impact of overhead becomes negligible with large coherence time and small training duration. The simulation results also verify this intuition; however, here we only present the results for specific parameters to show the importance of overhead.

Figs. 4.13 and 4.14 illustrate the tradeoff between the throughput (bps/Hz) and the number of relays per cluster for the optimal and ad hoc selection schemes. One observation is that when the amount of overhead is small, we can always obtain a gain by adding more relays. However, the diversity gain is eventually canceled by the

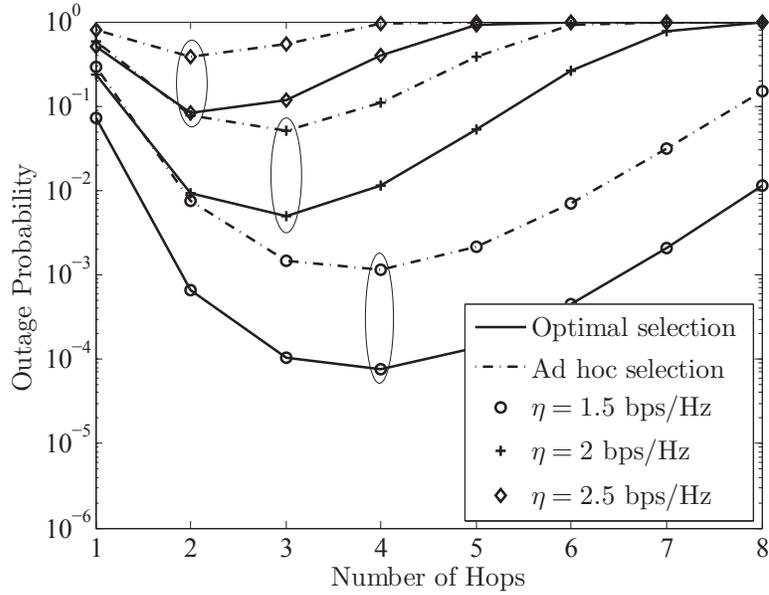


Figure 4.12: Optimum number of hops for different η (optimal and ad hoc selection schemes). $K = 4$, $\alpha = 4$, $\gamma_d = 4$ dB.

excessive amount of overhead when the number of relays increases. In Fig. 4.13, the required overhead occupies all the transmission resources when N and K are large, and no meaningful data can be transmitted through the multi-hop network. Another observation is that, although ad hoc selection is sub-optimal in outage, sometimes it provides higher throughput than optimal selection which requires a significant amount of overhead.

4.3 Summary

In this section, we investigated the optimum number of hops for a linear network with randomly located nodes. By proposing a novel model that permits randomness in the node locations, we analyzed the number of hops that maximizes the lower bound and an approximation for the spectral efficiency. Although a closed-form expression for the optimum number of hops is difficult to obtain, bounds and approximations were provided. Simulation results showed that the random shift model can characterize a realistic system.

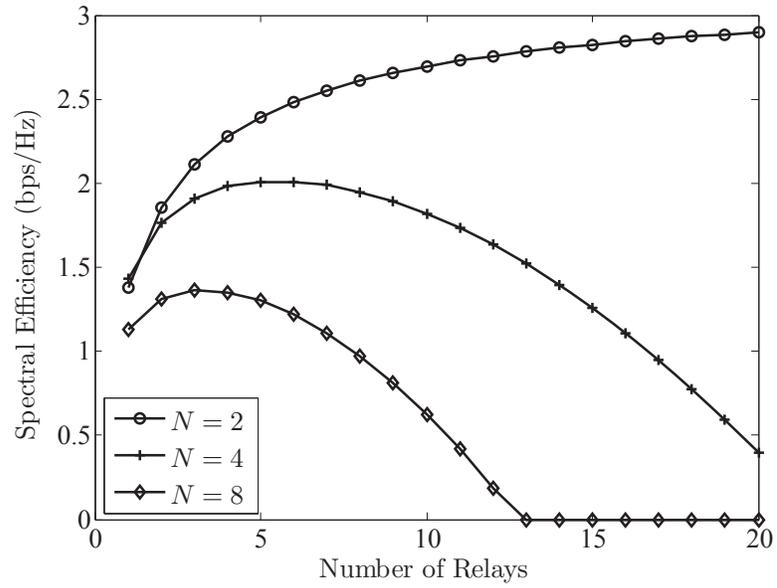


Figure 4.13: Optimum number of relays for different values of N (optimal selection), $\gamma_d = 4$ dB.

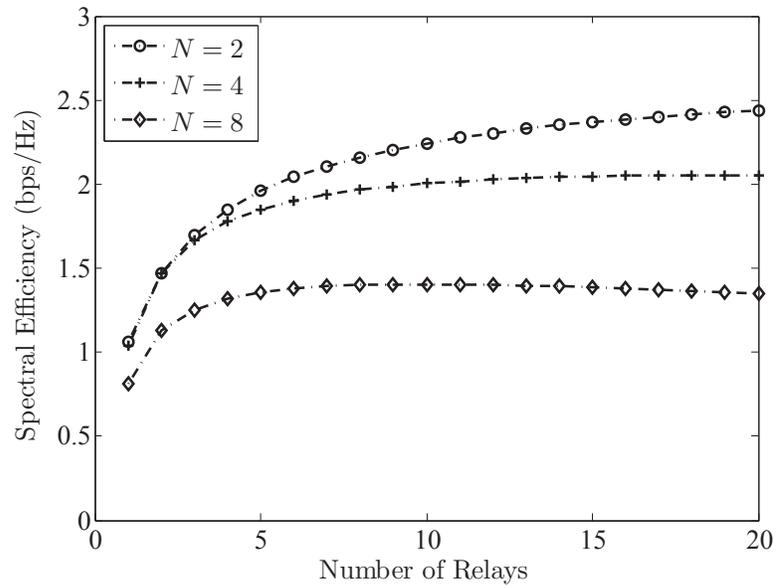


Figure 4.14: Optimum number of relays for different values of N (ad hoc selection), $\gamma_d = 4$ dB.

For a multi-hop linear network with cooperative relays, we investigated the optimal relay deployment strategy. Two different techniques were considered: optimal selection and ad hoc selection. First, we derived a closed-form expression for the number of hops that minimizes the end-to-end outage probability. We proved that the diversity gain does not affect the optimum number of hops, which means that most existing results for a linear network can also be applied to our scenarios. We also provided lower bounds on the required training and selection overhead for cooperation, and then determined the number of relays that maximizes the throughput.

Possible future directions include quantifying the required overhead more precisely so that the optimum relay deployment can be better understood. Instead of assuming perfect channel estimation and relay selection, we could extend the work to determining the optimal strategy with imperfect information. Retrieving a robust design which guarantees worst-case performance is also a fruitful area for research.

Chapter 5

MULTI-USER DOWNLINK NETWORKS

5.1 Introduction and System Model

MU-MIMO, which has been included in the newest versions of wireless standards such as 3GPP Long Term Evolution Advanced (LTE-A) [3] and IEEE 802.11ac [4], is a promising technology providing impressive performance. Compared with a conventional single-user (SU) system, MU-MIMO enables the AP or BS to communicate with multiple users simultaneously over the same spectrum, and, thereby, improves the sum rate [101]. Since the user terminal usually has a very limited number of antennas in commodity systems, MU-MIMO can better utilize the available resources compared with SU systems.

The performance of MU-MIMO, however, relies heavily on the accuracy of the CSI; perfect CSI at the AP/BS is required so that the interference among users can be canceled. In real communication systems, perfect CSI at the transmitter is difficult to obtain. It is more likely that only imperfect CSI, such as quantized observations of the channel, is available at the AP/BS. It has been shown that quantized CSI severely degrades the performance of MU-MIMO systems [101–103].

In this chapter, we focus on a downlink MU-MIMO system with L users, as illustrated in Fig. 5.1. Note that this model is applicable to either cellular systems or wireless local area networks, and we will use AP to represent the transmitter from this point forward without loss of generality. We assume that the number of transmit antennas N_t is equal to the aggregated number of receive antennas, i.e., $N_t = LN_r$, and N_r data streams are sent from the AP to each user. We also assume a block fading channel model, i.e., the channel remains constant over a given time period.

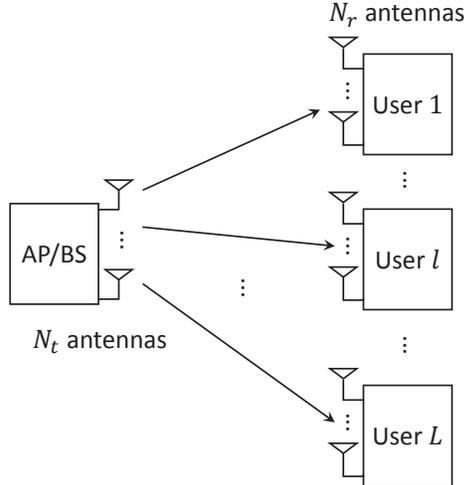


Figure 5.1: The multi-user MIMO broadcast channel with L users. The AP/BS is equipped with N_t antennas and each user has N_r antennas. We assume $N_t = LN_r$.

Let \mathbf{x}_l be an $N_r \times 1$ vector of data symbols for the l th user, with the power constraints $\mathbb{E}[\mathbf{x}_l \mathbf{x}_l^H] = \mathbf{P}_l$ and $\sum_{l=1}^L \text{tr}(\mathbf{P}_l) \leq P_t$, where P_t is the total transmit power. Here we assume that the total power is uniformly distributed among all data streams, i.e., $\mathbf{P}_l = \frac{P_t}{N_t} \mathbf{I}_{N_r}, \forall l = 1, 2, \dots, L$. Our results can be easily extended to scenarios with non-uniform power allocation.

At the AP, \mathbf{x}_l is beamformed by a $N_t \times N_r$ matrix \mathbf{V}_l and sent through N_t antennas. Then, the received signal at the l th user is

$$\mathbf{y}_l = \mathbf{H}_l^H \mathbf{V}_l \mathbf{x}_l + \sum_{j=1, j \neq l}^L \mathbf{H}_l^H \mathbf{V}_j \mathbf{x}_j + \mathbf{n}_l, \quad (5.1)$$

where \mathbf{H}_l is the $N_t \times N_r$ channel matrix and \mathbf{n}_l is the $N_r \times 1$ noise vector at the l th user. We assume that the channel experiences Rayleigh fading across time and space, i.e., the entries of \mathbf{H}_l and \mathbf{n}_l are modelled by i.i.d. complex Gaussian random variables with zero mean and unit variance.

The transmit beamforming matrix \mathbf{V}_l could be designed for canceling interference among different users. Block diagonalization (BD) is a widely used technique because of its analytical simplicity and asymptotic optimality at high SNR [102]. With perfect CSI at the AP, BD decomposes a downlink MU-MIMO channel into multiple

parallel, independent, point-to-point MIMO links; thus each user can receive its own signals with no interference. In particular, \mathbf{V}_j is chosen to satisfy $\mathbf{H}_l^H \mathbf{V}_j = 0, \forall j \neq l$ so that the interference from other users is eliminated completely. Note that we usually assume that $\mathbf{V}_j^H \mathbf{V}_j = \mathbf{I}_{N_r}$ because of the power constraint. A closed-form expression for \mathbf{V}_j can be derived based on the singular value decomposition (SVD) of the aggregated channel matrix of the other users [104].

To cancel interference through beamforming, the AP is required to collect global CSI which can only be obtained using a CSI feedback frame from each user. In practical systems, the channel matrix is quantized before it is sent to the AP due to the limited rate of the feedback channel. In particular, the channel from the AP to the l th user \mathbf{H}_l is represented by its quantized version $\hat{\mathbf{H}}_l$. Since \mathbf{H}_l is not available, the AP uses $\hat{\mathbf{H}}_l$ to obtain the transmit beamforming matrix \mathbf{V}_j such that $\hat{\mathbf{H}}_l^H \mathbf{V}_j = 0, \forall j \neq l$. In this case, the interference has not been canceled yet, and $\sum_{j=1, j \neq l}^L \mathbf{H}_l^H \mathbf{V}_j \mathbf{x}_j$ represents the residual interference caused by quantized CSI.

Sum rate, R_{sum} , is a widely-used metric to measure the performance of a communication system. The maximum rate at the l th user is represented by the Shannon capacity. Thus, the instantaneous sum rate is given by [40]

$$R_{\text{sum}} = \sum_{l=1}^L \log \det (\mathbf{I}_{N_r} + \boldsymbol{\Psi}_l^{-1} \mathbf{H}_l^H \mathbf{V}_l \mathbf{P}_l \mathbf{V}_l^H \mathbf{H}_l), \quad (5.2)$$

where

$$\boldsymbol{\Psi}_l = \mathbf{I}_{N_r} + \sum_{j=1, j \neq l}^L \mathbf{H}_l^H \mathbf{V}_j \mathbf{P}_j \mathbf{V}_j^H \mathbf{H}_l \quad (5.3)$$

is the interference-plus-noise covariance matrix at the l th user.

However, the sum rate performance ignores the effect of feedback overhead. In practical systems, the CSI feedback frames occupy the time or spectrum resources for data transmission and thereby degrade the actual data rate. Intuitively, in order to get a higher sum rate, the receiver needs to feedback more bits of CSI, which contribute to the overhead. On the other hand, a larger amount of overhead which happens frequently causes the data transmission to be less efficient.

The impact of the overhead on the performance highly depends on the system configuration. For example, if the channel never changes, the receiver only needs to feedback the CSI once. The time or/and frequency resources occupied by feedback packets are negligible compared with the those allocated for data transmission. In this case, the impact of overhead is insignificant and the minimization of feedback bits is unnecessary. However, for a fast-fading channel, the receiver is required to feedback CSI very often and the large extra cost due to feedback severely degrades the beamforming gain achieved by accurate CSI feedback. In order to characterize the implicit overhead-performance tradeoff, we introduce a metric, net capacity,

$$R_{\text{net}} = R_{\text{sum}} (1 - \theta), \quad (5.4)$$

where $\theta = T_f/\tau$ is the overhead ratio in time, $T_f = LB/R_{\text{fb}}$ represents the transmission time spent on feedback packets, B is the number of feedback bits for each user, and R_{fb} is the transmission rate for the feedback channel. We assume that CSI measurements are made at time intervals of length τ , where τ is large compared to the symbol length. We also choose τ based on the autocorrelation function $\rho = J_0(2\pi f_d \tau)$, where f_d is the Doppler frequency determined by the radio frequency wavelength, λ , the speed of the mobile user v (i.e., $f_d = v/\lambda$), and $J_0(\cdot)$ is the zeroth-order Bessel function of the first kind. The ergodic sum rate can be approximated as a function of the number of feedback bits; accordingly, the approximate average net capacity can be derived.

Note that we ignore the preamble and header in the CSI feedback frame, and only the quantization bits are considered as overhead in our analysis. In a practical multicarrier system, each user needs to feed back the quantized channels for all subcarriers; this implies that quantization bits dominate the overall feedback bits.

In this chapter, we first compare the performance, overhead, and complexity of different quantization techniques. Based on our analysis and simulation results, for a specific criterion and system configuration, we can provide the guidelines for determining the preferred technique. Then, we investigate the overhead-performance tradeoff by applying the rate distortion approach. The amount of overhead required

for interference cancellation can be quantified by minimizing the mutual information between the correct and incorrect channel information, where the distortion constraint guarantees that the performance loss is less than a given threshold.

5.2 Quantization Techniques for Downlink MU-MIMO

Different quantization techniques have been proposed and applied to practical downlink MU-MIMO systems. For example, scalar quantization (SQ) is an inefficient but easy-to-implement approach and has been adopted in IEEE 802.11 standards. In SQ, the scaled real and imaginary entries in the channel matrices are quantized, respectively. Thus, a large number of feedback bits is required for SQ, especially when the number of antennas is large. In WLAN environments, the channel state usually remains constant for several transmission blocks; So, SQ is favored due to its low complexity.

Vector quantization (VQ), in which the original information is represented by a vector codeword, is another approach which has been supported by 3GPP LTE for single-user MIMO with two or four transmit antennas [3]. Although VQ significantly reduces the number of feedback bits, it incurs high computational complexity and large storage requirements, especially for MU-MIMO systems [103, 105–107]. In [102], it has been shown that, for the medium SNR regime of a MU-MIMO downlink system consisting of a four-antenna AP and two two-antenna users, a 20-bit codebook is required to maintain a performance gap of no more than 3 dB with respect to the case with perfect CSI. This indicates that the AP has to generate and store a codebook containing 2^{20} codewords, and then select the appropriate codeword from such a huge pool. Since the AP repeats the quantization process whenever the channel condition changes, the excessive amount of extra computations makes VQ impractical.

Several approaches have been introduced to reduce the complexity of VQ. In [105], instead of using a large number of bits to quantize the entire channel matrix, partitioned VQ (P-VQ) first divides the channel into several small blocks and then implements VQ for each small block. The drawback of P-VQ is that it only improves

performance when the number of feedback bits is not large. In [106] and [107], compressive sensing is applied to reduce the dimension of the CSI matrix and thereby simplify the quantization process. The main focus of [106] and [107] is to compress the CSI matrix, but the quantization complexity and the quantization error after compression are ignored. Utilizing multiple feedback signals is considered in [108–110] to get an accurate CSI reconstruction at the AP. The feedback signals are jointly reconstructed by the AP according to the principle of compressive sensing so that the complexity in each single time slot can be reduced. In [111] and [112], the time and spatial correlations in the MIMO channel are utilized to reduce the quantization complexity.

In this chapter, we propose a new quantization method, sparse coding quantization (SCQ). Inspired by the paradigm of sparse coding [113], SCQ uses a sparse representation of the codebook to better approximate the original message. Compared with conventional VQ, SCQ uses a codebook with a much smaller size by exploiting a linear combination of several codewords to characterize the channel. Though extra feedback bits are needed to specify the sparse representation, the complexity of SCQ is significantly reduced without much performance loss. The theoretical approximation of the sum rate achieved by SCQ is derived, and verified to be tight through simulation results. We show that, in order to guarantee the same performance, SCQ requires much lower search complexity and memory requirements than VQ.

Since the accuracy and performance of different quantization approaches vary significantly, the relevance of each quantization approach changes across applications. In this chapter, we also evaluate different quantization techniques, SQ, VQ and SCQ, in terms of complexity and net capacity (5.4). By investigating the total number of feedback bits required to maintain a constant rate loss, we study the computational complexity and memory requirements for each quantization method. The analysis and simulation provide guidelines for determining the preferred technique, for a specific criterion and system configuration.

5.2.1 Quantization Methods

In this section, we study three quantization methods, namely scalar quantization (SQ), vector quantization (VQ), and sparse coding quantization (SCQ). We approximate and compare the performance, the required number of feedback bits, and the complexity for these quantization schemes. In the following, the subscripts (user index) in the channel and beamforming matrices are omitted because the quantization is a general process which can be applied for all users.

Scalar Quantization (SQ)

Since the AP only needs the spatial direction of the channel to eliminate the interference, the channel matrix \mathbf{H} is normalized and then quantized using SQ at each user. The real and imaginary parts of each complex element h_{ij} located in the i th row and j th column of in \mathbf{H} are quantized to B_{SQ} bits, respectively. Note that one bit is reserved for the sign of each of the real and imaginary parts. The quantized version of h_{ij} is

$$\hat{h}_{ij} = \frac{1}{2^{B_{\text{SQ}}-1}} \left\lfloor \frac{h_{ij}}{m} (2^{B_{\text{SQ}}-1} - 1) \right\rfloor, \quad (5.5)$$

where m is a scaling ratio which guarantees the real/imaginary element in the normalized channel matrix is always less than or equal to one. In particular, m can be chosen as the maximum value among all real and imaginary elements of the channel matrix \mathbf{H} ,

$$m = \max \left\{ \max_{i,j} \{\Re(h_{ij})\}, \max_{i,j} \{\Im(h_{ij})\} \right\}, \quad (5.6)$$

where $\Re(\cdot)$ and $\Im(\cdot)$ represent the real and imaginary parts of a complex element, respectively. Therefore, the number of feedback bits needed at each user is $B = 2N_t N_r B_{\text{SQ}}$, and the total amount of feedback overhead is $LB = 2N_t^2 B_{\text{SQ}}$.

A closed-form expression for the relationship between the rate loss and the number of feedback bits is usually intractable. A feasible approach is to approximate the quantization error as a random variable with a given distribution, for example, uniformly distributed in $[-2^{-B_{\text{SQ}}+1}, 2^{-B_{\text{SQ}}+1}]$. For the sake of analytical simplicity, we assume the quantization error is a Gaussian random variable with zero mean and

variance $\sigma_{\text{SQ}}^2 = \frac{1}{12}2^{-2B_{\text{SQ}}+2}$ [114]. It has been show that the Gaussian assumption provides an upper bound to characterize the quantization error. Since the sum rate is a function of the variance, by applying a similar approach as in Theorem 2 in [102], an approximation for the number of feedback bits required to maintain a performance gap of no more than 3 dB with respect to the case with perfect CSI is derived to be

$$B \approx 2N_t N_r \left(\frac{P_{\text{dB}}}{3} - \frac{1}{2} \log \frac{N_r}{12} \left(2^{\frac{N_t}{N_r}} - 1 \right) \right), \quad (5.7)$$

where $P_{\text{dB}} = 10 \log_{10} P$ is the normalized transmit power in units of dB.

The overall computation time for SQ is $O(N_t^2)$, which is independent of B_{SQ} . Furthermore, SQ does not require additional storage. Thus, increasing the resolution of the SQ quantizer does not affect the complexity. However, a large number of feedback bits might be required, especially when the number of antennas is large.

Vector Quantization (VQ)

In a limited feedback system using VQ, $\hat{\mathbf{H}}_l$ is chosen from a codebook according to

$$\hat{\mathbf{H}}_l = \operatorname{argmax}_{\mathbf{C} \in \mathcal{C}} d^2(\mathbf{H}_l, \mathbf{C}), \quad \mathbf{C} \in \mathcal{C} \quad (5.8)$$

where \mathcal{C} is a quantization codebook of size 2^B , i.e., $(\mathbf{C}_1, \mathbf{C}_2, \dots, \mathbf{C}_{2^B})$, which is known at both the AP and the users. B is the number of feedback bits per user. How to design an optimal or near-optimal codebook \mathcal{C} has been investigated in [115, 116]. Without loss of generality, we focus on the performance obtained by a random codebook, since it provides asymptotically optimal performance with closed-form analytical results [102]. That is, each codeword $\mathbf{C} \in \mathcal{C}$ is a $N_t \times N_r$ unitary matrix, and is independently and uniformly chosen from a unit sphere defined in an $N_t \times N_r$ -dimensional complex space [115, 117]. The chordal distance [118]

$$d^2(\mathbf{H}_l, \mathbf{C}) = N_r - \operatorname{tr}(\tilde{\mathbf{H}}_l^H \mathbf{C} \mathbf{C}^H \tilde{\mathbf{H}}_l) \quad (5.9)$$

is chosen as the metric, where $\tilde{\mathbf{H}}_l$ is an orthonormal basis for the subspace spanned by the columns of \mathbf{H}_l .

Each user chooses an appropriate codeword and sends back the index of the codeword. Thus, the number of feedback bits per user is $B = B_{\text{VQ}}$. The relation between the quantization error of VQ and B_{VQ} has been extensively investigated. In general, the quantization error can be modeled as an additive Gaussian noise with zero mean and variance [118]

$$\begin{aligned} Q &= \mathbb{E} \left[\min_{\mathbf{C} \in \mathcal{C}} d^2(\mathbf{H}_l, \mathbf{C}) \right] \\ &\leq \frac{1}{T} \Gamma \left(\frac{1}{T} \right) (C_{N_t N_r})^{-\frac{1}{T}} 2^{-\frac{B}{T}} = \bar{Q}, \end{aligned} \quad (5.10)$$

where $T = N_r(N_t - N_r)$, $C_{N_t N_r} = \frac{1}{T!} \prod_{i=1}^{N_r} \frac{(N_t - i)!}{(N_r - i)!}$, and $\Gamma(\cdot)$ represents the Gamma function.

An approximation for the required number of bits for VQ to maintain the 3-dB performance gap is provided in [102]:

$$B \approx T \left(\frac{P_{\text{dB}}}{3} - \log N_r \right) - \log C_{N_t N_r}. \quad (5.11)$$

It has been shown that VQ requires significantly fewer feedback bits than SQ. However, the main drawback of VQ is the computational complexity and storage requirement, especially when the codebook is large. The overall time complexity of searching codewords and calculating the chordal distance is $O(L2^{B_{\text{VQ}}})$. The storage requirement, on the other hand, is $O(N_t^2 2^{B_{\text{VQ}}})$. Using the simulation model in [103] as an example, 20-bit VQ is required to achieve acceptable performance. This indicates that each user must generate and store a codebook containing 2^{20} codewords and select the appropriate codeword from this large codebook.

Sparse Coding Quantization (SCQ)

For a given codebook \mathbf{C} and a given vector \mathbf{y} , the sparse coding problem [113] considers how to find a vector \mathbf{z} such that

$$\mathbf{y} = \mathbf{C}\mathbf{z}. \quad (5.12)$$

Here, the number of non-zero elements in \mathbf{z} is K , which is small compared with the dimension of \mathbf{z} . Accordingly, \mathbf{z} is called a K -sparse representation of \mathbf{y} using codebook

C. In general, there is no solution for (5.12). Thus, we solve the following optimization problem [113] instead,

$$\begin{aligned} \min_{\mathbf{z}} \quad & d^2(\mathbf{y}, \mathbf{Cz}) \\ \text{s.t.} \quad & \|\mathbf{z}\|_{\ell_0} \leq K \end{aligned} \tag{5.13}$$

where $d^2(\cdot, \cdot)$ could be any distance metric, for example, chordal distance in (5.9). The ℓ_0 -norm $\|\mathbf{z}\|_{\ell_0}$ denotes the number of non-zero elements in \mathbf{z} , and $\|\mathbf{z}\|_{\ell_0} \leq K$ is the sparsity constraint. Several efficient algorithms have been proposed to solve (5.13), such as orthogonal matching pursuit (OMP) [119].

The main factor causing the impracticality of VQ is that the search complexity and storage requirements increase exponentially as the number of feedback bits increases. Here, we propose a way to use multiple codewords instead of a single codeword to describe \mathbf{H} , resulting in a reduction in the size of the codebook.

The algorithm includes two steps: (i) vectorize the channel matrix and (ii) choose the sparse representation of the vectorized channel by applying OMP. With SCQ, we first vectorize an $N_t \times N_r$ channel matrix \mathbf{H}_l into an $N_t N_r \times 1$ vector \mathbf{h}_l . Similarly, the codebook \mathcal{C} is transformed into an $N_t N_r \times 2^\beta$ matrix \mathbf{C} which contains 2^β codewords. Here the codebook size is β which is less than or equal to the total number of feedback bits B . Then, the quantized channel $\hat{\mathbf{h}}_l = \mathbf{Cz}$ is chosen such that

$$\begin{aligned} \min_{\mathbf{z}} \quad & d^2(\mathbf{h}_l, \hat{\mathbf{h}}_l) = d^2(\mathbf{h}_l, \mathbf{Cz}) \\ \text{s.t.} \quad & \|\mathbf{z}\|_{\ell_0} \leq K \end{aligned} \tag{5.14}$$

The l th user feeds back K codeword indices and K non-zero coefficients in \mathbf{z} to the AP. A critical issue is how to feed back these K coefficients. Here, we apply VQ to quantize a $K \times 1$ vector that contains these K coefficients through a random codebook with 2^μ codewords. Based on the received K indices and K non-zero coefficients, the AP calculates the quantized vectorized channel and converts it into an $N_t \times N_r$ matrix $\hat{\mathbf{H}}_l$. Since all operations are linear, the transformations between vectors and matrices do not affect the quantization error.

	VQ	SCQ
Number of feedback bits	B	$K\beta + \mu$
Search complexity	$O(2^B)$	$O(K2^\beta + 2^\mu)$
Storage requirement	$O(N_t N_r 2^B)$	$O(N_t N_r 2^\beta + K2^\mu)$

Table 5.1: Comparisons between VQ and SCQ in terms of the number of feedback bits, search complexity, and storage requirements.

The total number of feedback bits for SCQ is $K\beta + \mu$, using β bits for each of K indices and μ bits for the $K \times 1$ coefficient vector. Since the OMP algorithm takes K iterations [119] and the entire codebook is searched in each iteration, the time complexity of SCQ is $O(K2^\beta + 2^\mu)$. Furthermore, SCQ only needs to store 2^β codewords ($N_t \times N_r$ matrices) for quantizing the channel and 2^μ codewords ($K \times 1$ vectors) for representing the coefficients. The comparisons between VQ and SCQ are presented in Table 5.1.

There are three key parameters for SCQ: K (the number of linear coefficients), β (the number of bits for each codeword index), and μ (the number of bits for representing the K coefficients). Thus, (K, β, μ) -SCQ is defined to precisely declare a given quantization scheme. VQ is equivalent to a special case of SCQ that finds the one-sparse representation of the actual channel \mathbf{H} , i.e., $(1, B, 0)$ -SCQ. In this case, $\mu = 0$ because only one codeword is fed back, and the user does not need to send any coefficients. The theorem below provides a tight approximation to the sum rate achieved by SCQ.

Theorem 5.1. For a (K, β, μ) -SCQ scheme, the average sum rate can be approximated as

$$R_{\text{sum}} \approx N_t \left[(L-1) \log \left(\frac{PN_r \sigma^2 c_1 + N_t}{PN_r \sigma^2 c_2 + N_t} \right) + \log \left(1 + \frac{PN_r}{N_t} c_1 \right) + \log \frac{c_2}{c_1} + (c_1 - c_2) \log(e) \right], \quad (5.15)$$

where

$$\sigma^2 = \frac{\Gamma\left(\frac{1}{T}\right)}{KT} (C_{N_t N_r})^{-\frac{1}{T}} 2^{-\frac{\beta}{T}} + \frac{\Gamma\left(\frac{1}{K-1}\right)}{K-1} 2^{-\frac{\mu}{T}}, \quad (5.16)$$

$T = N_r(N_t - N_r)$, $C_{N_t N_r} = \frac{1}{T!} \prod_{i=1}^{N_r} \frac{(N_t-i)!}{(N_r-i)!}$; c_1 and c_2 are the positive solutions to

$$\begin{aligned} c_1 + N_r P c_1 N_r P c_1 + N_t + (L-1) \frac{N_r P \sigma^2 c_1}{N_r P \sigma^2 c_1 + N_t} &= 1, \\ c_2 + (L-1) \frac{N_r P \sigma^2 c_2}{N_r P \sigma^2 c_2 + N_t} &= 1. \end{aligned} \quad (5.17)$$

Proof. With a (K, β, μ) -SCQ scheme, the channel matrix \mathbf{H}_l can be approximated by a sparse representation

$$\mathbf{H}_l^{(K)} = \sum_{i=1}^K z_i \mathbf{C}_i, \quad (5.18)$$

where z_i is the non-zero element in vector \mathbf{z} , and \mathbf{C}_i is the corresponding codeword. Since the OMP algorithm is a greedy algorithm [119], according to the performance bound in [120], the estimation error is bounded by

$$\begin{aligned} d^2(\mathbf{H}_l, \mathbf{H}_l^{(K)}) &\leq \frac{1}{K} d^2(\mathbf{H}_l, \mathbf{H}_l^{(1)}) \\ &\stackrel{(a)}{\leq} \frac{\Gamma\left(\frac{1}{T}\right)}{KT} (C_{N_t N_r})^{-\frac{1}{T}} 2^{-\frac{\beta}{T}} = \sigma_\beta^2, \end{aligned} \quad (5.19)$$

where $T = N_r(N_t - N_r)$, $C_{N_t N_r} = \frac{1}{T!} \prod_{i=1}^{N_r} \frac{(N_t-i)!}{(N_r-i)!}$. Step (a) comes from the facts that VQ is equivalent to SCQ with $K = 1$ and the corresponding quantization error is bounded in (5.10). According to [103], we provide a new expression for the channel \mathbf{H}_l

$$\mathbf{H}_l^{(K)} = \mathbf{H}_l + \mathbf{E}_\beta, \quad (5.20)$$

where the entries in \mathbf{E}_β are i.i.d. complex Gaussian random variables with zero mean and variance σ_β^2 .

Similarly, the coefficients $z_i, i = 1, 2, \dots, K$ are quantized by using a codebook with 2^μ codewords. Let $\bar{\mathbf{z}} = [z_1, z_2, \dots, z_K]$, the quantized version $\hat{\mathbf{z}} = [\hat{z}_1, \hat{z}_2, \dots, \hat{z}_K]$ can be approximated by $\hat{\mathbf{z}} \approx \bar{\mathbf{z}} + \mathbf{e}_z$ [103], where the entries e_i in \mathbf{e} are i.i.d. complex Gaussian random variables with zero mean and variance

$$\sigma_\mu^2 = \frac{\Gamma\left(\frac{1}{K-1}\right)}{K-1} 2^{-\frac{\mu}{K-1}}, \quad (5.21)$$

where (5.21) is obtained by choosing $N_t = K$ and $N_r = 1$ in (5.10). Then, the quantized channel can be written as

$$\widehat{\mathbf{H}}_l = \sum_{i=1}^K \widehat{z}_i \mathbf{C}_i = \mathbf{H}_l^{(K)} + \underbrace{\sum_{i=1}^K e_i \mathbf{C}_i}_{\mathbf{E}_\mu} = \mathbf{H}_l^{(K)} + \underbrace{\mathbf{E}_\beta + \mathbf{E}_\mu}_{\mathbf{E}}. \quad (5.22)$$

Since \mathbf{C}_i is orthonormal, the multiplications by \mathbf{C}_i do not change the distribution of \mathbf{E}_μ . Therefore, the entries in $\mathbf{E} = \mathbf{E}_\beta + \mathbf{E}_\mu$ are i.i.d. complex Gaussian random variables with zero mean and variance $\sigma_\beta^2 + \sigma_\mu^2$. By using derivations similar to the proof in [103], we show that the average rate achieved by (K, β, μ) -SCQ is approximated by (5.15). ■

Theorem 5.1 provides the performance of SCQ by applying the results for the sparse coding algorithm in [119, 120] and the approach in [103]. The result in [103] is a specific example of (5.15) with $K = 1$.

Remark 5.1. Consider the simple case where $N_r = 1$. Comparing Theorem 1 with the analytical results for VQ in [102, 103], we can obtain a sufficient condition for SCQ to achieve the same performance as conventional VQ with a B -bit codebook. That is,

$$\beta = B - (N_t - 1) \log K, \quad \mu = \frac{K - 1}{N_t - 1} B. \quad (5.23)$$

Note that (5.23) is a *sufficient* condition, i.e., the required values of β and μ could be smaller in practice, as will be shown in the simulation results.

From (5.23), we observe that it is usually not worth increasing K . As K increases, β decreases logarithmically and μ increases linearly, indicating that the complexity increases eventually. For example, if $K \geq N_t$, we have $\mu \geq B$, which implies that SCQ requires higher complexity than VQ in this case. In most scenarios, setting $K = 2$ achieves the benefit of sparse representation without incurring much complexity. Another observation is that when N_t is large, SCQ is preferable because it requires much less complexity than VQ.

Remark 5.2. Based on (5.11), the required codebook sizes for SCQ to maintain the 3-dB performance gap can be approximated as

$$\begin{aligned}\beta &\approx T \left(\frac{P_{\text{dB}}}{3} - \log KN_r \right) - \log C_{N_t N_r}, \\ \mu &\approx \frac{K-1}{T} \left[\left(\frac{P_{\text{dB}}}{3} - \log N_r \right) - \log C_{N_t N_r} \right].\end{aligned}\tag{5.24}$$

In most scenarios, by setting $K = 2$, the benefit of SCQ can be achieved without incurring much overhead. Accordingly, the required number of feedback bits is

$$B \approx \frac{2T^2 + 1}{T} \left(\frac{P_{\text{dB}}}{3} - \log N_r \right) - 2T - \frac{1+T}{T} \log C_{N_t N_r}.\tag{5.25}$$

Remark 5.3. Based on Theorem 5.1, we can obtain the approximated net capacity of SCQ

$$C_{\text{net}} = R_{\text{sum}} \left(1 - \frac{K\beta + \mu}{R_{\text{fb}}\tau} \right).\tag{5.26}$$

In general, the net capacity of SCQ is lower than that of VQ as SCQ requires more feedback bits to achieve the same performance. For example, suppose $N_r = 1$ and the amount of feedback overhead for a VQ scheme is B bits. To achieve the same sum rate performance, the amount of overhead for SCQ should be $\left[K + \frac{K-1}{N_t-1} - \frac{K \log K}{B} (N_t - 1) \right] B$ bits according to (5.23).

5.2.2 Simulation Results

In this section, simulation results are presented to verify the theoretical results, compare the performance of different quantization techniques, and show the impact of the number of feedback bits. In the simulation, we generate 10^4 independent realizations of the quantization codebooks for each user. For each possible codebook, simulation results are averaged over 10^4 channel realizations. For all scenarios, uniform power allocation is assumed, and the feedback rate R_{fb} is assumed to be 6.5 Mbps, which is the lowest data rate in the 802.11ac standard [4].

In Fig. 5.2, the simulation results (solid curves) and analytical results (dash-dotted curves) are compared in terms of sum rate. The first case presents a two-user

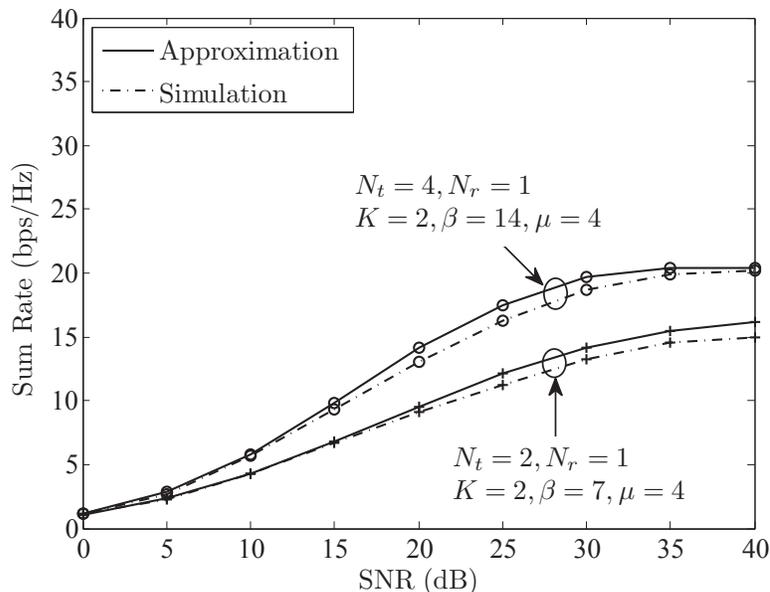


Figure 5.2: Analytical approximation and simulation results of the sum rate achieved by SCQ.

downlink MIMO network where the AP is equipped with two antennas and each user has a single antenna. In this case, SCQ applies a linear combination of two 7-bit codewords and a codebook of size 4 bits to represent the channel matrix and these two coefficients, respectively. The approximation given in Theorem 5.1 is close to the simulation result of the (2, 7, 4)-SCQ scheme. Similar insights can be observed for the second case which considers an AP with four antennas simultaneously serving four single-antenna users. For the (2, 14, 4)-SCQ scheme, the relative ratio of approximation error with respect to the simulation result is less than 4% for the entire SNR regime of interest. Based on these two different system configurations, we can see that the approximation is tight. This scheme and analysis can also be applied in a straightforward way to systems with a large number of antennas.

In Fig. 5.3, we illustrate the sum rate versus SNR for VQ and SCQ in the case where four users are simultaneously served by the AP. An appropriate codebook size is chosen for each quantization method such that the percent rate loss compared to the sum rate achieved with perfect CSI is 1% for 20-dB SNR. Thus, VQ uses a 26-bit

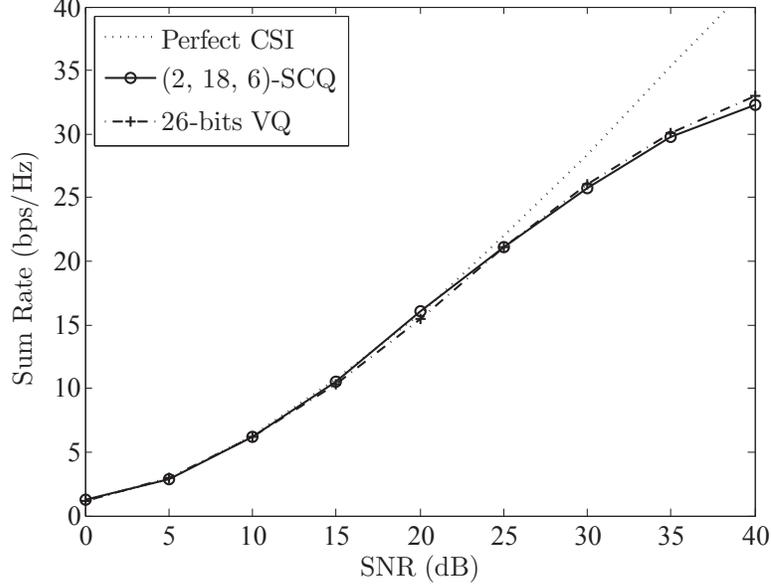


Figure 5.3: Sum rate comparison between VQ and SCQ to achieve 1% rate loss with respect to the performance with perfect CSI at 20 dB SNR. $N_t = 4$, $N_r = 1$ and $L = 4$.

codeword, while SCQ employs a linear combination of two 18-bit codewords for the sparse representation of the channel matrix and a 6-bit codebook for the quantization of two coefficients. Compared with VQ, SCQ reduces the overall computation complexity from $O(2^{26})$ to $O(2^{19})$, as the complexity of obtaining 6-bit quantization is negligible. In addition, for both quantization methods, as SNR increases, the gap between the performance with perfect CSI and that with quantized CSI becomes larger and larger; thus, more bits would be needed to limit rate loss to 1% for SNR values above 20 dB.

The impact of the number of feedback bits depends on the channel block length τ and how often each user sends the feedback frames. It motivates us to investigate the net capacity which reveals the overhead-performance tradeoff. We assume that the carrier frequency f_c is 5 GHz, τ is determined by the ρ -coherence time, i.e., $\rho = J_0(2\pi f_d \tau)$, and the transmission rate R_{fb} on the feedback channel is 6.5 Mbps. Note that ρ and f_d are parameters which describe the system environments, including channel quality and user mobility.

In Fig. 5.4, the required feedback overhead (in bits) per user for each quantization technique is illustrated; the sum rate performance is 3 dB less than that with

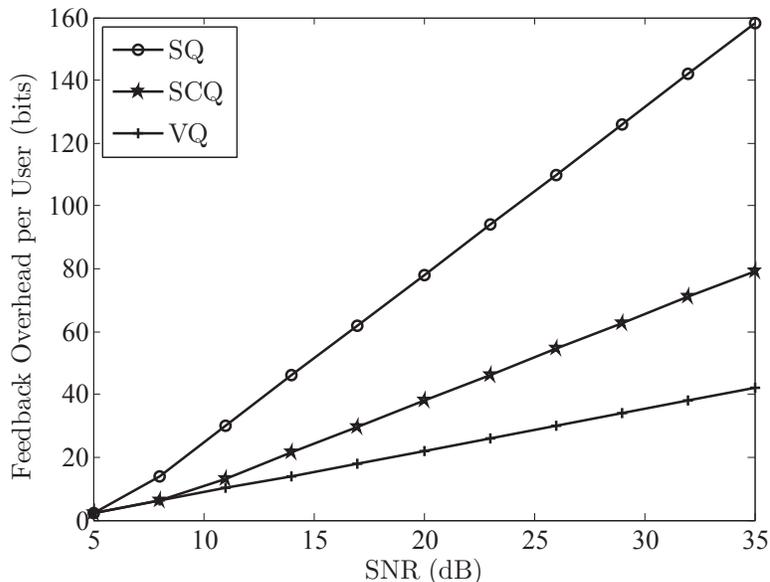


Figure 5.4: Feedback overhead for the 3-dB performance gap versus SNR. $N_t = 4$, $N_r = 2$ and $L = 2$.

perfect CSI at the AP. Here, the quantization error is assumed to be Gaussian, so the results are upper bounds on the required number of feedback bits. We can see that, as SNR increases, more feedback bits are needed so that stronger interference can be reduced to a specific level. Furthermore, SCQ incurs more overhead than VQ but requires a smaller number of feedback bits than SQ. Another observation from Fig. 5.4 is that VQ requires a 29-bit codebook for each user when SNR is 24 dB, implying that the computational complexity of VQ is $O(2^{29})$. Meanwhile, the complexity of SQ, which is independent of the number of feedback bits, remains constant ($O(1)$). According to (5.24), the complexity of SCQ is $O(2^{26})$, which indicates that SCQ provides an option for balancing complexity and overhead. Though the given example is for the scenario of high-rate feedback, our analysis is applicable to the low-rate feedback regime as well. From Fig. 5.4, we can see that only a small number of feedback bits is required for any quantization method at low SNR; in that range the analytical and simulation results reveal that SCQ also provides a complexity benefit compared to VQ.

In Fig. 5.5, the net capacity achieved by these three quantization methods is plotted as a function of the overhead ratio $\theta = B/R_{fb}\tau$, where ρ is the autocorrelation

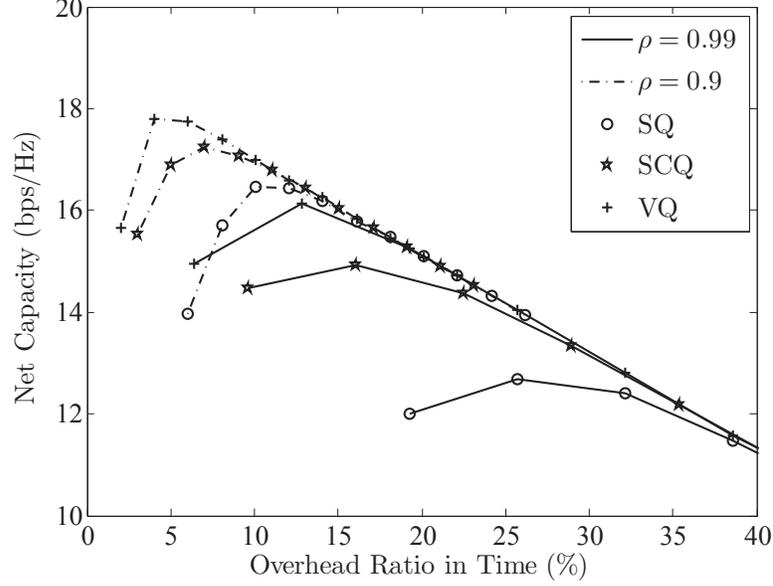


Figure 5.5: Net capacity versus the overhead time ratio for different values of ρ . The carrier frequency $f_c = 5$ GHz, velocity $v = 25$ m/s, and doppler frequency $f_d = 500$ Hz. $N_t = 4$, $N_r = 2$, $L = 2$ and SNR = 20 dB.

of the random path gain for time separation τ . For given R_{fb} and τ (determined by ρ), the overhead ratio θ is determined by the total number of feedback bits, which varies from 16 to 128. We consider a high mobility environment: the velocity v is assumed to be 25 m/s, i.e., around 90 km/h. The average received SNR is assumed to be 20 dB. The results indicate that, as θ increases, the net capacity first increases because the CSI becomes more accurate. However, the capacity eventually decreases because of the excessive amount of overhead. Although VQ outperforms SCQ in terms of net capacity, the performance gap is acceptable. In addition, SQ gives the worst performance as expected. When θ is sufficiently large, where the net capacity is overwhelmed by the overhead required, the performance of each technique reduces to the same level.

By choosing an appropriate number of feedback bits, the net capacity can be maximized for given system configurations. Fig. 5.6 illustrates the relation between the optimal net capacity and the velocity of the mobile users v . We can see that, as the mobility increases, the optimal net capacity decreases for any of these quantization techniques because the channel changes more rapidly. Another observation from Figs.

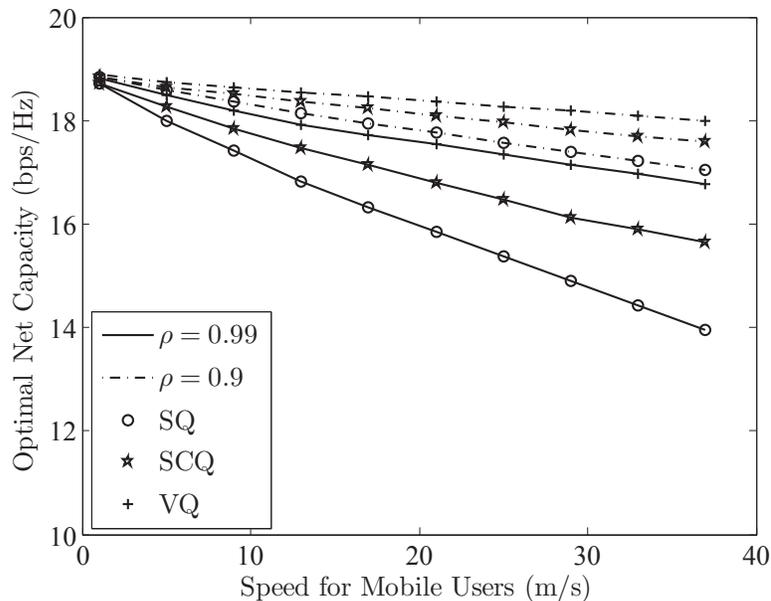


Figure 5.6: Net capacity versus the velocity of the user v . $N_t = 4$, $N_r = 2$, $L = 2$ and $\text{SNR} = 20$ dB.

5.5 and 5.6 is that the impact of overhead becomes less important as ρ decreases, indicating a preference for SQ due to its simplicity.

5.3 Overhead-Performance Tradeoff for Downlink MU-MIMO Systems

The design and analysis of limited feedback techniques for downlink MU-MIMO have been studied extensively over the past few years. Much of the work analyzes the system performance by utilizing and extending the well known performance bound for vector quantization. In particular, for MIMO systems with finite-rate feedback, the quantization of eigenchannel vectors is related to quantization on the Grassmann manifold [115, 118, 121, 122].

In this section, by using rate distortion theory, we quantify and analyze the optimal tradeoff between the sum rate performance and feedback overhead. Compared with the conventional rate-distortion tradeoff where the distortion is defined as the minimum mean square error (MSE) between the actual channel and the quantized channel achievable for a given quantization codebook size, we adopt a different approach to characterize the amount of feedback overhead as a function of allowable rate loss due

	Classic Rate Distortion Results	Our Approach
Information Source	Actual channel	Rate with perfect CSI
Destination	Quantized channel	Rate with quantized CSI
Distortion	MSE	Rate loss
Mutual Information	Codebook size	Amount of overhead

Table 5.2: Classic rate distortion results versus our approach

to the CSI quantization. A similar framework has been applied in Chapter 2 for quantifying the overhead-performance tradeoff in cooperative networks. Table 5.2 compares the original terms used in the classic results and the terms used in our approach. Although the classic rate distortion tradeoff for quantization on the Grassmann manifold has been well-studied, to the best of our knowledge, there is no existing work focusing on the applications of rate distortion theory for the overhead-performance tradeoff in downlink MU-MIMO systems.

We follow the same system model and consider the simplest scenario ($N_t = L = 2, N_r = 1$) as a first step. In other words, we consider a system with a two-antenna BS and two single-antenna users. The BD scheme has been applied to cancel the inter-user interference. The beamforming vectors at the BS are determined based on noiseless and zero-delay feedback of B bits per user. Instead of discussing how to design a VQ codebook of 2^B codewords, we try to answer the fundamental question: how many feedback bits are required for a given rate loss tolerance?

The rate distortion formulation can be written as

$$\begin{aligned}
R(D) = & \min_{f(\hat{\mathbf{h}}_i), f(\hat{\mathbf{h}}_i|\mathbf{h}_i)} \mathcal{I}(\mathbf{h}_1, \mathbf{h}_2; \hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2) \\
\text{s.t.} & \quad \mathbb{E} \left[\log_2 \left(1 + \frac{P_t}{2} |\mathbf{h}_1^H \hat{\mathbf{v}}_2|^2 \right) + \log_2 \left(1 + \frac{P_t}{2} |\mathbf{h}_2^H \hat{\mathbf{v}}_1|^2 \right) \right] \leq D
\end{aligned} \tag{5.27}$$

where $\mathcal{I}(\mathbf{h}_1, \mathbf{h}_2; \hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2)$ denotes the mutual information between the actual CSI and the available CSI at the BS, and $f(\hat{\mathbf{h}}_i)$ and $f(\hat{\mathbf{h}}_i|\mathbf{h}_i)$ denote the probability distributions to be optimized. In particular, we search among all potential distributions which satisfy the capacity loss constraint. It is difficult to obtain the optimal distributions directly, so we will make several assumptions to simplify the problem.

First, we assume that the channel magnitudes, which can be viewed as side information for rate distortion, are available at the BS. In practical systems, this could be done by measuring and feeding back the channel quality index (CQI) at the receivers. Mathematically, we normalize all the channel vectors \mathbf{h}_i :

$$\mathbf{h}_i = \|\mathbf{h}_i\| \bar{\mathbf{h}}_i = \sqrt{a_i} \bar{\mathbf{h}}_i$$

where a_i is the channel power gain which is perfectly known¹ at the BS, $\bar{\mathbf{h}}_i$ is the normalized channel vector (or, the channel direction). It has been shown in [46] that a_i has a chi-squared distribution, and the channel direction $\bar{\mathbf{h}}_i$ is isotropically distributed over the 4-dimensional unit sphere. Since the essential part for interference cancellation is the channel direction, we also assume that the receivers only quantize the direction $\bar{\mathbf{h}}_i$ to $\hat{\mathbf{h}}_i$. The rate distortion formulation in (5.27) can then be rewritten as

$$\begin{aligned} R(D) = \min \quad & \mathcal{I}(\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2; \hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2 | a_1, a_2) \\ \text{s.t.} \quad & \mathbb{E} \left[\log_2 \left(1 + \frac{P_t}{2} a_1 |\bar{\mathbf{h}}_1^H \hat{\mathbf{v}}_2|^2 \right) + \log_2 \left(1 + \frac{P_t}{2} a_2 |\bar{\mathbf{h}}_2^H \hat{\mathbf{v}}_1|^2 \right) \right] \leq D \end{aligned} \quad (5.28)$$

Note that the quantized channels $\hat{\mathbf{h}}_i$ are also located on the 2-dimensional complex unit sphere. The distributions of $\hat{\mathbf{h}}_i$ depend on the specified designs for the quantizer.

We also observe that the rate loss only depends on the magnitude of the two inner products, $|\bar{\mathbf{h}}_1^H \hat{\mathbf{v}}_2|$ and $|\bar{\mathbf{h}}_2^H \hat{\mathbf{v}}_1|$, which represent the residual interference caused by imperfect CSI. We define two scalar random variables $s_1 = |\bar{\mathbf{h}}_1^H \hat{\mathbf{v}}_2|$ and $s_2 = |\bar{\mathbf{h}}_2^H \hat{\mathbf{v}}_1|$. Then, the rate loss constraint in (5.27) can be rewritten as

$$\mathbb{E} \left[\log_2 \left(1 + \frac{P_t}{2} a_1 s_1^2 \right) + \log_2 \left(1 + \frac{P_t}{2} a_2 s_2^2 \right) \right] \leq D \quad (5.29)$$

Note that BD is assumed; then we have

$$\hat{\mathbf{v}}_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \hat{\mathbf{h}}_2, \quad \hat{\mathbf{v}}_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \hat{\mathbf{h}}_1, \quad (5.30)$$

¹ In practice, a_i is also quantized. For example, in LTE-A Release 11 [3], a_i is represented as CQI and is then quantized by a 4-bit SQ. Since we focus on the interference-limited scenarios where the channel directions dominate the overall performance, the impact of quantized CQI is ignored.

which implies that s_i only depends on $\bar{\mathbf{h}}_i$

$$s_i = \left| \bar{\mathbf{h}}_i^H \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \hat{\mathbf{h}}_i \right| \quad (5.31)$$

When $\hat{\mathbf{h}}_i = \bar{\mathbf{h}}_i$, $s_i = 0$, i.e., perfect CSI at the BS can be exploited to completely eliminate the interference, and therefore achieve zero distortion.

Since h_1 and h_2 are independent, by using an orthogonal coordinate transformation (one-to-one mapping similar to the transformation from Cartesian to polar coordinates) from $\bar{\mathbf{h}}_i$ to a pair (s_i, \mathbf{t}_i) , where \mathbf{t}_i denotes the remaining coordinates, we can easily show that

$$\mathcal{I}(\bar{\mathbf{h}}_1, \bar{\mathbf{h}}_2; \hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2 | a_1, a_2) = \mathcal{I}(\bar{\mathbf{h}}_1; \hat{\mathbf{h}}_1 | a_1) + \mathcal{I}(\bar{\mathbf{h}}_2; \hat{\mathbf{h}}_2 | a_2) \quad (5.32)$$

Since $\bar{\mathbf{h}}_i$ has four dimensions (two complex numbers), and s_i is a scalar, \mathbf{t}_i has three dimensions. Similar to Theorem 1 in [46], we can rewrite the mutual information as

$$\begin{aligned} \mathcal{I}(\bar{\mathbf{h}}_i; \hat{\mathbf{h}}_i | a_i) &\geq D_{\text{KL}}(f(s_i | a_i) \| f(s_i) | f(a_i)) \\ &= \int_0^\infty \int_0^1 f(a_i) f(s_i | a_i) \log_2 \left(\frac{f(s_i | a_i)}{f(s_i)} \right) ds_i da_i \end{aligned} \quad (5.33)$$

where D_{KL} denotes the relative entropy between the two probability distributions, and $f(\cdot)$ represents the PDF. Note that s_i has to be located in $[0, 1]$, since $\|\hat{\mathbf{h}}_i\|$ and $\|\bar{\mathbf{h}}_i\|$ are normalized vectors. In [46], it is also shown that equality can be achieved if 1) $\hat{\mathbf{h}}_i$ is isotropic over the unit sphere (which is true for well-designed codebooks including a random codebook), and 2) the distribution of s_i conditioned on $\hat{\mathbf{h}}_i$ does not depend on $\hat{\mathbf{h}}_i$. In other words, s_i contains all the important information and can fully describe the channel information. Here, we use $D_{\text{KL}}(f(s_i | a_i) \| f(s_i) | f(a_i))$ as a lower bound to the mutual information.

Choosing the relative entropy as the objective function can significantly simplify the problem. First, we only need to find the optimum conditional probability $f(s_i | a_i)$, since $f(s_i)$ and $f(a_i)$ can be easily determined once the channel distribution is given. In the Rayleigh fading case, we have [121]

$$f(a_i) = a_i e^{-a_i}, a_i > 0$$

and

$$f(s_i) = 2s_i, 0 \leq s_i \leq 1$$

Instead of solving a complicated high-dimension optimization problem, we only need to solve a variational inequality problem with two scalar random variables. Now, the rate distortion function becomes

$$\begin{aligned} \min_{f(s_1|a_1), f(s_2|a_2)} & \sum_{i=1}^2 \int_0^\infty \int_0^1 a_i e^{-a_i} f(s_i|a_i) (\log_2 f(s_i|a_i) - \log_2 2s_i) ds_i da_i \\ \text{s.t.} & \sum_{i=1}^2 \int_0^\infty \int_0^1 a_i e^{-a_i} f(s_i|a_i) \log_2 \left(1 + \frac{P}{2} a_i s_i^2 \right) ds_i da_i \leq D \\ & \int_0^1 f(s_i|a_i) ds_i = 1, \forall i, a_i > 0 \end{aligned}$$

Then, we decompose the problem into two parts. First, for fixed channel power gain a_i , we solve the following optimization problem to obtain $R(D_1(a_1), D_2(a_2))$

$$\begin{aligned} R(D_1(a_1), D_2(a_2)) &= \min_{f(s_1|a_1), f(s_2|a_2)} \sum_{i=1}^2 \int_0^1 f(s_i|a_i) (\log_2 f(s_i|a_i) - \log_2 2s_i) ds_i \\ & \text{s.t.} \int_0^1 f(s_1|a_1) \log_2 \left(1 + \frac{P}{2} a_1 s_1^2 \right) ds_1 \leq D_1(a_1) \\ & \int_0^1 f(s_2|a_2) \log_2 \left(1 + \frac{P}{2} a_2 s_2^2 \right) ds_2 \leq D_2(a_2) \\ & \int_0^1 f(s_i|a_i) ds_i = 1, \forall i, a_i > 0 \end{aligned} \quad (5.34)$$

Given the rate distortion function for any given a_i , the second part is to optimize the rate loss functions $D_i(a_i)$

$$\begin{aligned} \min_{D_1(a_1), D_2(a_2)} & \sum_{i=1}^2 \int_0^\infty a_i e^{-a_i} R(D_1(a_1), D_2(a_2)) da_i \\ \text{s.t.} & \sum_{i=1}^2 \int_0^\infty a_i e^{-a_i} D_i(a_i) da_i = D \end{aligned} \quad (5.35)$$

It is easy to see that the rate distortion problem is equivalent to (5.34) and (5.35).

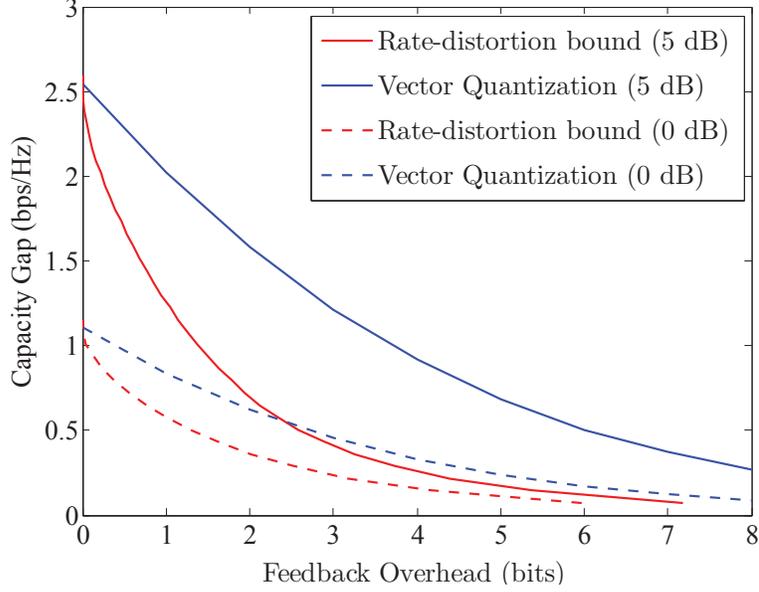


Figure 5.7: Rate distortion bounds versus vector quantization for different values of transmit power. $N_t = 2, N_r = 1, L = 2$. The channel magnitudes are assumed to be fixed as 1 for both users.

Then, by using optimal control theory such as Pontryagin's minimum principle and Mangasarian's sufficient condition², we can obtain the optimal distribution $f(s_i|a_i)$ for (5.34)

$$f(s_i|a_i) = 2s_i e^{-1-p_i} \left(1 + \frac{P}{2} a_i s_i^2\right)^{-q_i},$$

where p_i and q_i satisfy

$$p_i = \ln \left(\int_0^1 2e^{-1} s_i \left(1 + \frac{P}{2} a_i s_i^2\right)^{-q_i} ds_i \right)$$

$$\int_0^1 s_i \left(1 + \frac{P}{2} a_i s_i^2\right)^{q_i} \left[\log_2 \left(1 + \frac{P}{2} a_i s_i^2\right) - D_i(a_i) \right] ds_i = 0$$

An intuitive solution of (5.35) is to assume that $D_1(a_1) = D_2(a_2) = \frac{D}{2}$ for any given a_i , i.e., different users have the same amount of performance loss. In Fig. 5.7, we plot the derived rate distortion bounds and the simulation results for vector quantization techniques. The channel magnitudes are assumed to be fixed and equal

² These are extensions of the KKT conditions from general optimization to functional optimization.

to 1 for both users. We can observe that as SNR increases, the required feedback overhead for achieving the same level of performance significantly increases. Another observation is the zero overhead scenario. Intuitively, when no overhead is allowed, the precoding vectors are pure noise. We can easily derive the closed-form expression for the performance loss.

5.4 Summary

In this chapter, we first proposed a new method, SCQ, to reduce the computational complexity of conventional VQ with negligible performance loss, by utilizing a linear combination of multiple codewords. A closed-form expression for the sum rate achieved by SCQ has been established and verified by simulations. It has also been shown that SCQ is preferable compared with VQ due to the simplicity of implementation, especially for large-scale MIMO systems. We also evaluated and compared three quantization methods in terms of the required number of feedback bits, the net capacity, and the complexity. The results illustrate the tradeoff among these schemes and show that SCQ is an option for balancing performance and complexity. SQ is more preferable when the overhead effect is negligible, for example, when the channel condition changes very slowly.

Besides considering specific quantization techniques, we also discussed the fundamental tradeoff between the feedback overhead and the performance of MU-MIMO systems. By solving the functional optimization problems, we presented a closed-form rate distortion function which can characterize the minimum amount of feedback bits required for effective interference cancellation. Numerical results show that the gap between the commonly-used VQ and the rate distortion bound is non-negligible. How to achieve the optimal rate distortion bound is left for future work.

Chapter 6

HETEROGENEOUS NETWORKS

Heterogeneous Networks (HetNets) are a promising avenue for providing the performance and capacity leap needed to meet the ever growing demands from mobile wireless users. By using a mix of conventional macro cells and small cells (including micro cells, femto cells, relay stations, and WiFi access points), HetNets effectively reduce the distance between the transmitter and the receiver, and, thereby, increase the area spectral efficiency [123].

Without a breakthrough in battery technology, reducing the power consumption at the mobile users is imperative and has great practical interest [124–126]. As shown in [126], in addition to reducing the traffic over macrocells, HetNets can save a significant amount of energy for the mobile terminals. Specifically, mobile users can choose to communicate with the base station of a closer small cell with better channel quality, rather than with the base station of a macrocell. On the other hand, delivering traffic over WiFi networks can be more power-efficient for mobile devices [127]. This motivates our investigation of power efficient resource allocation algorithms that take advantage of HetNets to minimize the power consumption of mobile users.

Reliable broadcast/multicast services in cellular networks have become a growing interest, largely due to the rapid increase of the demand for multimedia data [128]. Evolved Multimedia Broadcast/Multicast Service (eMBMS)[129] has been proposed in LTE-A, focusing on optimizing the support for broadcast/multicast services. In this chapter, we consider the scenario of multicast services (for example, software updates and popular news feeds) in HetNets.

6.1 Energy-Efficient User Pairing

The issue of cell association (assigning users to different cells) is one of the key challenges in resource allocation for HetNets [130]. For example, in [130], user-cell pairing and power control are jointly optimized with the objective of maximizing the system throughput. The authors in [131] take this a step further and jointly consider backhaul capacity, a transmit power limitation, and user demands when determining the user-cell pair, transmit power allocation and channel frequency assignment. These studies mainly focus on physical-layer resource management and rely on physical-layer information such as link quality or interference level. Other works use application-specific information to assist the physical resource allocation. In [132], “context information” (such as whether the data is multimedia data) is observed from the user’s domain and the information is used to schedule the transmission in the network. Another example [133] uses information about the hand-off mechanisms for different radio access technologies and QoS requirements to determine whether a user should be handed off to a different network.

In addition to efficient cell association, peer-to-peer (P2P) streaming among the mobile terminals has been demonstrated to be feasible and is shown to be important in recent research. With the increasing density of mobile users, exploiting the resources of neighboring users becomes imperative. For example, in [134], the authors show that P2P cooperation can be used to provide near-live TV streaming in a resource constrained mobile environment. In [135], a cross-layer P2P-based solution is proposed to distribute live video streaming over a mobile ad hoc network. While most existing work uses P2P networks to reduce the Internet access requirement [136] and to improve the streaming quality [137], less attention has been paid to the power consumption of the mobile users.

In [138], energy-efficient user pairing is considered. The main focus in this paper is to reduce the power consumption and hence prolong the battery life for the mobile users. In addition, MIMO techniques and cooperative communications are employed to improve the energy efficiency. Note that cooperative communications has already been

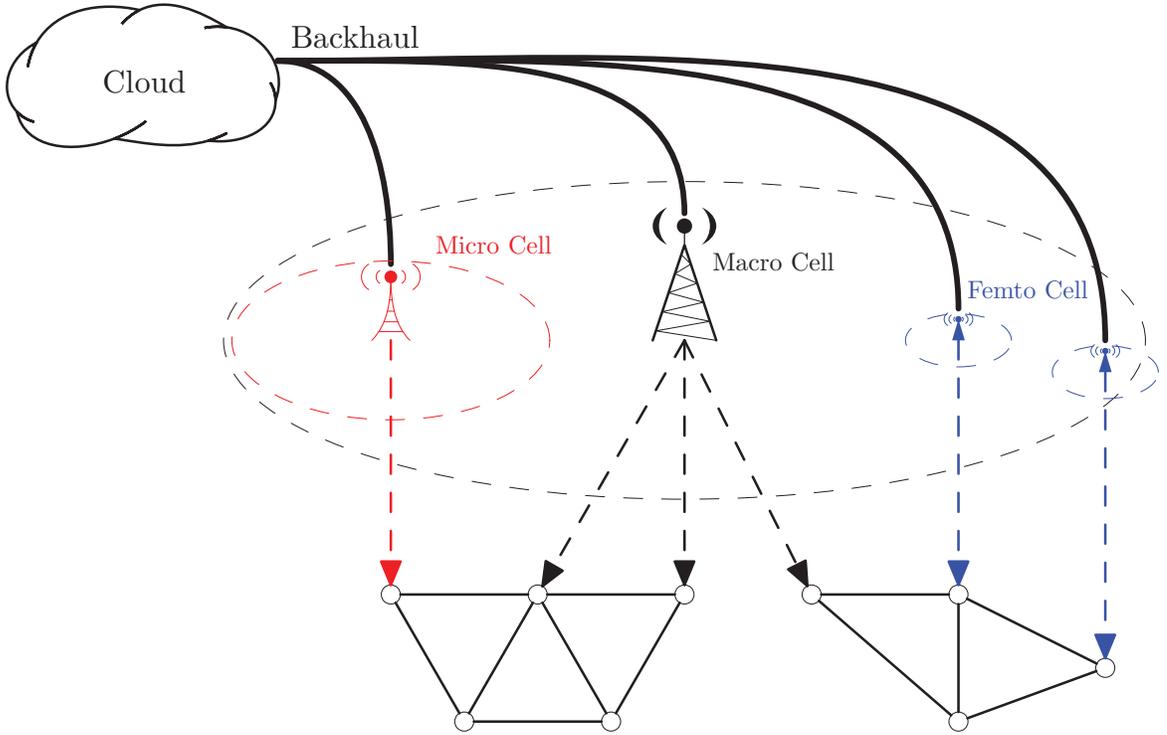


Figure 6.1: An example of user pairing scenario with cooperative communications and P2P cooperation.

incorporated into cellular standards (for example, relay technology and CoMP [139]). The impact of employing P2P cooperation among mobile users has also been included in [138], showing that the applications of cooperation and P2P streaming significantly reduce the power consumption at mobile terminals.

Let \mathcal{N} be the set of mobile users (MUs) and \mathcal{S} the set of base stations (BSs). We denote user pairing by an *indicator matrix*, \mathcal{U} , where $\mathcal{U}_{ij} = 1$, $i \in \mathcal{S}$, $j \in \mathcal{N}$, if and only if MU j is associated with BS i . Fig. 6.1 depicts one user pairing scenario with cooperative communications and P2P cooperations. Note that the users are located in different clusters; in each cluster, only a subset of users need to retrieve data from the BSs. Other users can utilize P2P links, and thereby reduce power consumption. Also, the additional antennas equipped at the BSs can also be utilized for enhancing link quality and improving energy efficiency. Here, we determine the optimal user pairing scheme which can minimize the maximum power consumption among all mobiles users.

In order to focus on the design of user pairing schemes, perfect backhaul with high capacity and low latency is assumed to support the cooperation among distributed BSs. We also assume orthogonal channel assignment to mitigate interference. In particular, the set of orthogonal subcarriers is equally divided among the active transmissions. We assume that the clustering of MUs is based on proximity since P2P streaming only helps if the peers are close to each other. As in [138], our study on user pairing focuses on associating the P2P networks with the BSs in a power efficient manner, while satisfying the QoS constraints.

If the number of BS antennas is sufficient, cooperative beamforming [40], which realizes maximal ratio combining at the receiver, will be applied to increase the data rate and thus improve the energy efficiency. In this case, the data rate achieved along the cellular link(s) is

$$r_i^C = \frac{B^C}{N^C} \log_2 \left\{ 1 + \frac{\sum_{v \in \mathcal{S}} P_v \mathcal{U}_{vi} \mathcal{H}_{vi}^C}{N_0^C \frac{B^C}{N^C}} \right\} \quad (6.1)$$

where B^C is the entire bandwidth available in the cellular network, N^C is the number of MUs that are associated with BSs in the cellular network, i.e., the number of columns in \mathcal{U} that are not all 0's, P_v is the transmit power at BS v , \mathcal{H}_{vi}^C is the channel power gain from BS v to MU i , and N_0^C is the power spectral density of the noise.

The P2P network can be represented by an adjacency matrix, \mathcal{M} , where $\mathcal{M}_{ij} = 1, i, j \in \mathcal{N}$, if and only if there is a reliable P2P connection from MU i to MU j . Clearly, \mathcal{M} also provides the clustering of the MUs. For the sake of simplicity, we assume that perfect scheduling is performed to eliminate intra-cluster interference, and uniform resource allocation is applied. Therefore, the data rate achieved at MU j along the P2P link from MU i is

$$r_{ij}^P = \frac{B^P}{N^j} \log_2 \left\{ 1 + \frac{P_i \mathcal{M}_{ij} \mathcal{H}_{ij}^P}{N_0^P \frac{B^P}{N^j}} \right\} \quad (6.2)$$

where B^P is the bandwidth available for each cluster in the P2P network, N^j is the number of active P2P links in the cluster that MU j belongs to (calculated from \mathcal{M}),

P_i is the transmit power at MU i , \mathcal{H}_{ij}^P is the channel power gain for the link from MU i to MU j , and N_0^P is the power spectral density of the noise. Note that the entire spectrum for the P2P network can be reused in every cluster, which means that the inter-cluster interference is assumed negligible.

To receive the desired data with an acceptable quality, for any MU i , a minimum throughput S is required for a period T . Ideally, we need to consider both the data flow scheduling and the cell association so that the optimum power consumption¹ can be achieved. However, jointly optimizing the amount of data received via the cellular links and the P2P links is complicated and intractable. In fact, determining the optimal data flow for the P2P networks is NP-Hard [140].

Without loss of generality, as in [138], we simplify the P2P network to a multi-hop, chain-like network with a single cluster head. In this case, for each cluster, one node is selected as the cluster head to receive the “entire” data via the cellular link; then, a multi-hop chain-like network is formed, with the cluster head as the source, to relay the information from the BS(s) to the P2P network. Hence, for MU i acting as a cluster head, the throughput via the cellular link is S , and the cluster head needs to forward the data to the next hop, so the outgoing P2P traffic has throughput S . The power consumed by a cluster head is

$$p_i = \frac{P_r^C}{r_i^C} S + \frac{P_t^P}{r_{i(i+1)}^P} S, \quad \forall i \in \mathcal{N}^C \quad (6.3)$$

where \mathcal{N}^C is the set of cluster heads, and $i + 1$ denotes the next hop. For any other MU j in the cluster, there is no data coming directly from the BSs, and the incoming and outgoing P2P traffic has exactly the same throughput S because of the multi-hop structure. Obviously, for the leaf nodes in the cluster, that is, the destination nodes in the multi-hop chain-like network, there are no outgoing P2P transmissions. So,

$$p_j = \frac{P_r^P}{r_{(j-1)j}^P} S + \frac{P_t^P}{r_{j(j+1)}^P} S, \quad \forall j \in \mathcal{N} - \mathcal{N}^C \quad (6.4)$$

¹ We focus on the power consumed by the cellular and the WiFi (assuming it is used for P2P communications) radio interfaces.

where $j - 1$ denotes the previous hop of MU j .

To minimize the maximum power consumption among the MUs P_{\max} , we formulate the following problem:

$$\begin{aligned}
 & \min_{\{\mathcal{U}\}} P_{\max} & (6.5) \\
 & \text{s.t.} \quad (a) \text{ MU power:} \\
 & \quad \quad p_i \leq P_{\max}, & \forall i \in \mathcal{N} \\
 & \quad \quad (b) \text{ Single cluster head:} \\
 & \quad \quad \sum_{i \in \mathcal{C}} \text{sgn} \left(\sum_{v \in \mathcal{S}} \mathcal{U}_{vi} \right) = 1, & \forall \mathcal{C}
 \end{aligned}$$

Constraint (b) means that there is exactly one cluster head for any cluster \mathcal{C} , and $\text{sgn}(\cdot)$ is the sign function.

6.2 Robust User Pairing

In order to solve the proposed optimization problem (6.5), we assume we have complete and perfect information for the link rates, which are determined by the instantaneous CSI. However, the non-deterministic nature of wireless communications makes this assumption unrealistic. Typically, CSI is estimated at the receivers by sending pilot symbols. The transmitter usually obtains CSI via a feedback channel or from past received signals, exploiting the channel reciprocity in time-division systems. In practice, estimation error, feedback delay, and other types of uncertainty are inevitable. In the conventional approach, the presence of uncertainty in CSI, on which the objective function crucially depends, is neglected. Consequently, the solutions for (6.5) may be the best decision according to the imperfect (known) CSI but not the best choice according to the real CSI. The solutions for the ideal problems can easily become infeasible or provide suboptimal performance.

To circumvent these problems, we introduce here a robust approach that incorporates the presence of uncertainty into the optimization problem (6.5). To abstract the physical layer details and simplify the analysis, we summarize the impact of all the

uncertainties into the channel power gain. In this section, we consider two different scenarios: (1) some statistics are known about the channel uncertainty and (2) the channel uncertainty lies in some bounded region and no other statistical information is available.

First, we assume the channel power gain is known within some mean-square error. This is a well-known method to model the imperfect CSI. In particular, the channel power gain can be modeled as a known part with a probabilistic additive component as $\mathcal{H} = \widehat{\mathcal{H}} + \mathcal{E}$, where $\mathbb{E}[\mathcal{H}] = \widehat{\mathcal{H}}$ and $\mathbb{E}[\mathcal{E}] = 0$. Here $\mathbb{E}[\cdot]$ denotes the expectation operation. Intuitively, we view $\widehat{\mathcal{H}}$ as a corrupted measurement and \mathcal{E} as a zero-mean uncertainty with variance σ^2 . Note that if the random component is assumed to be Gaussian, the uncertainty is unbounded. This means that errors caused by this uncertainty might be much larger than the noise variance, although the probability is small. We might also assume that the noise follows other bounded distributions, such as a uniform distribution (for example, more typical when modeling quantization errors). The data rate achieved along each wireless link \widehat{r} is determined by $\widehat{\mathcal{H}}$, and the robust user pairing problem can then be formulated as

$$\begin{aligned} \min_{\{\mathcal{U}\}} \quad & P_{\max} & (6.6) \\ \text{s.t.} \quad & (a) \Pr\{\widehat{p}_i \leq P_{\max}\} \geq p_{\text{th}}, \quad \forall i \in \mathcal{N} \\ & (b) \text{Single cluster head constraint in (6.5)}, \end{aligned}$$

where p_{th} is a predefined probability threshold, and

$$\widehat{p}_i = \begin{cases} \frac{P_r^C}{\widehat{r}_i^C} S + \frac{P_t^P}{\widehat{r}_{i(i+1)}^P} S & i \in \mathcal{N}^c \\ \frac{P_r^P}{\widehat{r}_{(j-1)j}^P} S + \frac{P_t^C}{\widehat{r}_{j(j+1)}^C} S & i \in \mathcal{N} - \mathcal{N}^c \end{cases} \quad (6.7)$$

In contrast to the original non-robust version (6.5), the robust version, which includes the effect of uncertainty, only achieves suboptimal performance but is a more realistic approach to managing the presence of uncertainty in the channel. A more realistic assumption is that the statistical properties of the uncertainty are not known to the system designer. In this case, we assume the channel uncertainty belongs to

some bounded *uncertainty set* or *region* [141], and this is the only knowledge available for solving the optimization problems. In other words, we know that the errors due to the uncertainty cannot exceed specified thresholds.

Taking this error model into account, we formulate the optimization problem as

$$\begin{aligned} \min_{\{\mathcal{U}\}} \quad & P_{\max} & (6.8) \\ \text{s.t.} \quad & (a) \ \widehat{p}_i \leq P_{\max}, \quad \forall i \in \mathcal{N}, \forall \mathcal{H} \in \mathcal{R}_{\mathcal{U}} \\ & (b) \ \text{Single cluster head constraint in (6.5),} \end{aligned}$$

where \widehat{p}_i is given in (6.7), and $\mathcal{R}_{\mathcal{U}}$ is the uncertainty set. The shape and the size of the uncertainty set $\mathcal{R}_{\mathcal{U}}$ depends on the physical phenomena that produce it. For simplicity, we assume that the uncertainty set is convex; however, we can also deal with a non-convex uncertainty set by taking its convex hull. In this chapter, we focus on the following two types of uncertainty set:

Polyhedron uncertainty set: For quantization in the channel estimation, $\mathcal{R}_{\mathcal{U}}$ is a polyhedron around the estimated channel. In particular, the channel power gain \mathcal{H} satisfies some linear inequalities:

$$|h_{ij} - \widehat{h}_{ij}| \leq e_p, \forall i, j \quad (6.9)$$

where h_{ij} is the i th row and j th column of \mathcal{H} , and e_p is the size of the polyhedron.

Ellipsoid uncertainty set: Assume that the uncertainty is bounded by an ellipsoid. For example, $\mathcal{R}_{\mathcal{U}}$ might be represented as

$$\sum_{i,j} |h_{ij} - \widehat{h}_{ij}|^2 \leq e_c, \quad (6.10)$$

where e_c is a predetermined error bound based on the system configuration.

For a solution to be feasible, the constraints have to be satisfied for all possible \mathcal{H} in the uncertainty set. The original optimal solution which is based on a fixed \mathcal{H} might not be feasible for the robust problem (6.8).

Both (6.6) and (6.8) are non-convex minimax optimization problems with integer constraints. In general, they are combinatorially hard to solve in their current forms.

One way to solve such problems is by transforming the non-convex problems into their convex equivalents. Powerful numerical algorithms, such as sub-gradient methods and interior-point methods, can then be applied to search the solution efficiently.

In [138], a scheme called *Ordered Best-K w/ Coop* is proposed for solving (6.5), where K is the number of clusters. In this approach, we first select the cluster head for the smallest cluster. The node that has the highest data rate gets selected as the cluster head and also gets paired with the BS that provides this highest data rate. Then, we apply cooperative communications to increase the data rate and hence reduce the power consumption at the bottleneck node, i.e., the node with the largest p_i . It has been shown that combining user pairing and cooperation can achieve almost the optimal performance that is obtained via an exhaustive search. Therefore, due to its significant performance gain and low complexity, *Ordered Best-K w/ Coop* is a promising technique to minimize the maximum power consumption at the MUs. Here, we show that *Ordered Best-K w/ Coop*, which is not designed for solving (6.6) and (6.8), is also robust to uncertainty in the channel measurements.

We consider a small system, where there are (a) two base stations (one macro BS with three antennas and one micro BS with a single antenna), and (b) four mobile users comprising two clusters, one with three nodes and another with a single node. Frequency-flat Rayleigh channels are assumed in all simulations to capture the effect of fading and path loss. The results are averaged over 10^4 independent channel realizations.

A rich body of literature has been dedicated to measuring the power consumption of cellular and WiFi interfaces for mobile users. Although a variety of power consumption models have been proposed and studied, one general conclusion is that, in spite of comparable power consumption, WiFi is much more power efficient in sending/receiving the same amount of data because of its higher data rates. The experimental measurements from [142–144] are used in the following simulations (shown in Table 6.1).

First, we assume that the estimated channel power gain is corrupted by additive

Power consumption of cellular reception P_r^C	700 mW
Power consumption of WiFi reception P_r^P	800 mW
Power consumption of WiFi transmission P_t^P	1100 mW
Streaming rate S	1 Mbps
Average cellular data rate r^C	1-5 Mbps
Average WiFi data rate r^P	20 Mbps

Table 6.1: Simulation Parameters

Gaussian noise with zero mean and variance σ^2 . In Fig. 6.2, through simulations, we show the impact of this uncertainty on the performance of the algorithms that are designed assuming perfect information. With increasing noise variance (i.e., the amount of uncertainty), the performance of both schemes degrade. Note that the approach that is optimal with perfect information becomes worse than *Ordered Best-K w/ Coop* when the noise power exceeds a certain level. This means that *Ordered Best-K w/ Coop* is more robust to uncertainty in the channel measurements, at a cost of suboptimal performance when the noise is negligible. According to Fig. 6.2, we can also see that the maximum power consumption decreases as the cellular data rate r^C increases, as expected. In general, most of the energy consumed is due to the low cellular data rate, and the user which acts as the cluster head will consume more power than the other users. When r^C increases from 1 Mbps to 5 Mbps, the maximum power consumption is reduced by 25%.

Figs. 6.3 and 6.4 show the maximum power consumption versus the size of the uncertainty set. Fig. 6.3 assumes a polyhedron uncertainty set defined by (6.9), while Fig. 6.4 assumes an ellipsoid uncertainty set defined by (6.10). The estimated channel is assumed to be uniformly distributed over the uncertainty set. We observe that the impact of the uncertainty is less significant compared with Fig. 6.2, as expected. If the uncertainty is assumed to be Gaussian, the uncertainty is unbounded and it is possible that the errors might be much larger than the noise variance, although the probability is small. It has been shown that a Gaussian random variable has the largest entropy (uncertainty) amongst all random variables of equal variance.

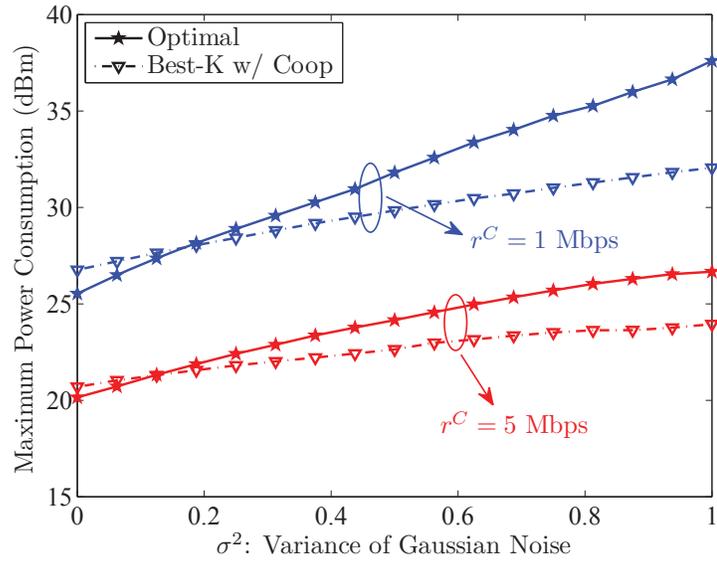


Figure 6.2: Performance of user pairing with Gaussian uncertainty.

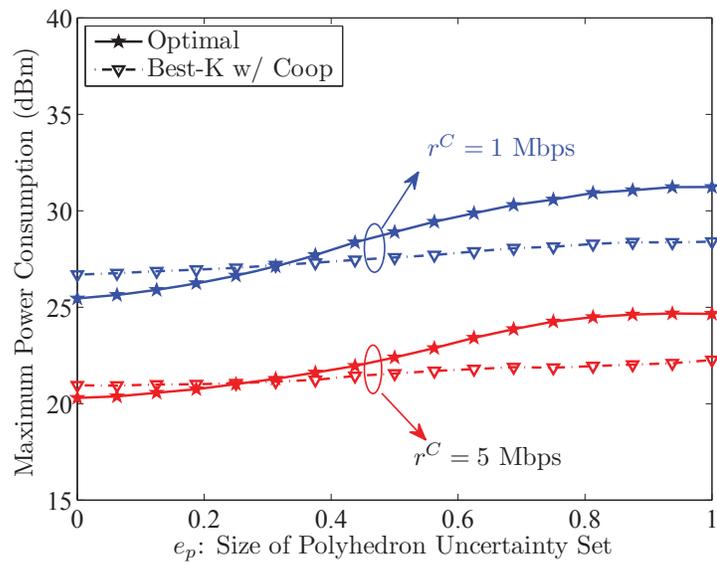


Figure 6.3: Performance of user pairing with a polyhedron uncertainty set.

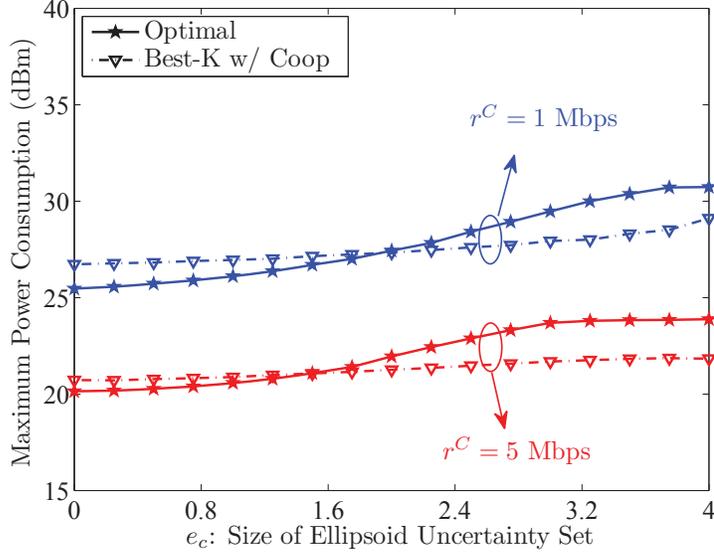


Figure 6.4: Performance of user pairing with an ellipsoid uncertainty set.

6.3 Summary

In this chapter, we proposed a user pairing algorithm with cooperation and P2P streaming that is robust under imperfect channel knowledge. We discussed both stochastic and worst-case robust optimization frameworks for the user pairing problems, and evaluated the robustness of the proposed algorithm for different types of uncertainty models. The results illustrate that the proposed robust algorithm can provide a guaranteed level of performance even if the measurements of the channel power gain are corrupted.

Investigating the analytical solutions of the robust optimization problems (6.6) and (6.8) is critical for future work. Instead of assuming a simple P2P network model, analyzing realistic P2P network models in practical HetNets is another challenging topic. In addition to the uncertainty of channel characteristics, uncertainties caused by other factors also need to be taken into account. For example, the number of MUs in the network and the network topology are also varying.

Chapter 7

CONTRIBUTIONS AND FUTURE WORK

This chapter summarizes the main contributions of the dissertation and discusses potential future research. Section 7.1 summarizes our work and gives some concluding remarks. Section 7.2 discusses a variety of directions for future research relating to our viewpoint on wireless networks.

7.1 Contributions

The insatiable demand from users of mobile wireless networks is driving wireless technologies towards their limits. As we introduced in Chapter 1, different directions for improving network performance have to be investigated. In this dissertation, we provided theoretical frameworks and practical insights on overhead, uncertainty and interference.

Practical Performance Analyses with Considerations of Overhead

The impact of overhead on practical network performance is usually neglected. However, it is unfair to compare different schemes if we ignore the overhead part. In this dissertation, we analyzed overhead-aware designs for different systems, including cooperative networks and downlink MU-MIMO. In particular, in Section 2.2, we showed the optimality of a low-overhead design for a single-user, two-hop, cooperative network. If central control and inter-node communications are not allowed, the M -group scheme discussed in 2.2 is the optimal distributed STBC strategy. In Section 2.4 and Chapter 4, we presented detailed performance analyses for both centralized and decentralized relay selection schemes in cooperative networks. The entire time frame, including the channel estimation, relay selection, and data transmission, has been included in

our analysis. Novel comparisons between centralized and decentralized schemes are provided, showing that the centralized scheme might outperform the decentralized scheme if the amount of overhead is insignificant. In Section 5.2, the sum rate and net capacity performance of different quantization techniques have been studied. By taking the feedback overhead into account, we obtained criterion for determining which quantization techniques should be implemented in practice.

Rate Distortion Approach for Overhead-Performance Tradeoff

Rate distortion theory provides a useful analysis tool for investigating the relationship between the amount of overhead and network performance. Though the original purpose of rate distortion theory is to find a lower bound on lossy source coding problems, the concepts of distortion and mutual information can be extended to network performance measures and the quantity of overhead, respectively. In this dissertation, by exploiting this information-theoretic framework, we demonstrated the overhead-performance tradeoff for both cooperative networks and MU-MIMO systems. In Section 2.3, we quantified the required feedback bits for effective relay selection. Asymptotic properties and approximations are also discussed and verified by the simulation results. In Section 5.3, the required feedback bits for characterizing CSI and canceling inter-user interference is represented as a function of tolerable rate loss. The proposed theoretical bounds can help in designing practical systems where the impact of overhead cannot be neglected.

Robust Designs under Channel Uncertainty

Control, coordination, and optimization across the entire network are typically required for achieving the performance limits. While such system-wide operations can be performed with complete and perfect information for all the required parameters, the non-deterministic nature of wireless communications and networks makes such assumptions unrealistic. In practice, centralized control is more likely to obtain incomplete or erroneous information (due to the inherent delay between any centralized controller

and individual radio units, for instance) and will have to make imperfect decisions. Optimizing the network performance without consideration of uncertainty typically yields significant performance degradation. In Chapter 6, we considered robust user pairing with cooperative communication and peer-to-peer streaming. Specifically, we showed that a solution that is optimal with perfect information will become worse than a heuristic solution when the noise exceeds a threshold; this demonstrates that our proposed solution is more robust to uncertainty in the channel measurements.

Performance Analyses for Interference-Limited Networks

Due to the broadcast nature of wireless medium, it is inevitable that multiple users will communicate over a shared wireless channel. If the power of each simultaneous transmission is increased, the signal and interference power will increase proportionally while the noise power will usually remain constant. Thus, at some point thermal noise becomes approximately negligible and any further increases in transmission power provides essentially no benefit. The design and analysis of such interference-limited configurations, even in their simplest forms, are different and challenging problems. In Chapter 3, we investigated a large-scale Poisson network with multiple ad hoc users. The outage performance for both non-cooperative and cooperative schemes was studied. We not only provided closed-form performance bounds and approximations, but also derived the criterion for determining which scheme should be applied in practice. In Section 5.2, an infrastructure network with multiple concurrent transmissions was investigated. The impact of residual interference, due to the finite rate feedback channel, was included in our analysis.

7.2 Future Works

The work we presented here constitutes only a small portion of the open problems. Future research will look beyond the simplified problems considered in this dissertation to more practical and complicate systems. In particular, the following topics are of interest:

Overhead-Performance Tradeoff for Multi-Hop Networks

In Chapters 2 and 5, we investigated the overhead-performance tradeoff for two-hop cooperative networks and single-hop downlink channels. Many applications, including sensor networks, mobile ad hoc networks, and military tactical networks, require reliable and efficient multi-hop transmissions due to energy and distance limitations. Extending our existing work to multi-hop networks is an interesting and challenging task. In Chapter 4, we investigated the overhead-aware multi-hop network by applying the results in Chapter 2. However, the fundamental tradeoff for multi-hop networks is still unsolved.

Overhead Models

In this dissertation, the overhead was typically quantified in terms of time. The models adopted in this dissertation usually assumed that the system is static in a limited period of time. The time occupied by the “non-data” parts, including channel estimation, feedback, selection, and so on, was considered as overhead. However, there are other resources such as frequency and energy consumption which should also be included in the overhead model. The analytical results in this dissertation might differ with the results for other, more complicated, overhead models.

Systematic Robust Designs

In Chapter 6, we provided a heuristic approach and showed that this approach is actually robust to channel uncertainty. However, the proposed robust optimization problems have not been analytically solved. The analytical solutions will help in designing practical and systematic algorithms which guarantee reasonable performance even if the uncertainties are involved. Unlike the heuristic approaches which cannot be generalized, the systematic robust designs will be applicable to different scenarios and are of great practical interest.

Joint Optimization of Overhead, Uncertainty, and Interference

Most of the work presented in this dissertation focuses on a particular aspect of real systems. For example, we ignored channel uncertainty when we investigated the overhead-performance tradeoff and interference-limited networks. Similarly, the impact of overhead on the performance was not included when we discussed robust designs. However, all of these performance-limiting factors are related. In general, a non-robust interference-limited system requires very accurate information to guarantee the performance, and thereby incurs a significant amount of overhead. If we can jointly optimize the entire system with respect to overhead, uncertainty, and interference, the actual performance might be boosted.

BIBLIOGRAPHY

- [1] “Global mobile statistics: mobile subscribers, handset market share, and mobile operators,” [online] <http://mobithinking.com/mobile-marketing-tools/latest-mobile-stats/a>.
- [2] “The mobile economy,” GSMA Intelligence, Tech. Rep., 2014, [online] http://www.gsmamobileeconomy.com/GSMA_ME_Report_2014_R_NewCover.pdf.
- [3] “Evolved universal terrestrial radio access (E-UTRA); physical layer procedures (release 11),” *3GPP TS 36.213*, 2012.
- [4] “Part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications,” *IEEE Standard 802.11ac/Draft 4.0*, 2012.
- [5] L. J. Cimini, “Analysis and simulation of a digital mobile channel using orthogonal frequency division multiplexing,” *IEEE Transactions on Communications*, vol. 33, no. 7, pp. 665–675, 1985.
- [6] S. Aubert, J. Tournois, and F. Nouvel, “On the implementation of MIMO-OFDM schemes using perturbation of the QR decomposition: Application to 3GPP LTE-a systems,” in *Proceedings of International Conference on Acoustics, Speech and Signal Processing*, May 2011.
- [7] Q. Li, G. Li, W. Lee, M. Lee, D. Mazzaresse, B. Clerckx, and Z. Li, “MIMO techniques in WiMAX and LTE: a feature overview,” *IEEE Communications Magazine*, vol. 48, no. 5, pp. 86–92, 2010.
- [8] E. Lahetkangas, K. Pajukoski, E. Tiirola, J. Hamalainen, and Z. Zheng, “On the performance of LTE-Advanced MIMO: How to set and reach beyond 4G targets,” in *Proceedings of European Wireless Conference*, April 2012.
- [9] N. Prasad, H. Zhang, M. Jiang, G. Yue, and S. Rangarajan, “Resource allocation in 4G MIMO cellular uplink,” in *Proceedings of Global Telecommunications Conference*, December 2011.
- [10] R. G. Gallager, “Low-density parity-check codes,” *IEEE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962.

- [11] W. Tranter, D. Taylor, R. Ziemer, N. Maxemchuk, and J. Mark, “Near optimum error correcting coding and decoding: Turbocodes,” in *The Best of the Best: Fifty Years of Communications and Networking Research*. Wiley-IEEE Press, 2007, pp. 45–55.
- [12] L. J. Cimini, C.-C. Shen, and L. Zhang, “The cost of using cooperation in a wireless network,” in *Proceedings of Asilomar Conference on Signals, Systems and Computers*, November 2009.
- [13] N. Marchenko and C. Bettstetter, “Impact of relay selection overhead in cooperative diversity protocols,” in *Proceedings of Vehicular Technology Conference Fall*, September 2011.
- [14] Y. Xiao and L. J. Cimini, “Impact of overhead on spectral efficiency of cooperative relaying,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 5, pp. 2228–2239, 2013.
- [15] B. P. Dunn, “Overhead in communication systems as the cost of constraints,” Ph.D. dissertation, University of Notre Dame, 2010.
- [16] A. Ben-Tal and A. Nemirovski, “Robust convex optimization,” *Mathematics of Operational Research*, vol. 23, no. 4, pp. 769–805, 1998.
- [17] ———, “Robust solutions of uncertain linear programs,” *Operational Research Letters*, vol. 25, no. 1, pp. 1–13, 1999.
- [18] A. Sendonaris, E. Erkip, and B. Aazhang, “User cooperation diversity - Part I: System description,” *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1927–1938, 2003.
- [19] ———, “User cooperation diversity - Part II: Implementation aspects and performance analysis,” *IEEE Transactions on Communications*, vol. 51, no. 11, pp. 1939–1948, 2003.
- [20] J. N. Laneman, D. N. C. Tse, and G. W. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.
- [21] H. Wan, J.-F. Diouris, and G. Andrieux, “Effect of synchronization errors on alamouti coding in wireless sensor networks,” in *Proceedings of Wireless Communications and Networking Conference*, April 2009.
- [22] T. Hossain, S. Kandeepan, and D. B. Smith, “Decode-and-forward cooperative communications: Performance analysis with power constraints in the presence of timing errors,” in *Mobile Multimedia Communications*. Springer Berlin Heidelberg, 2012, vol. 77, pp. 463–475.

- [23] Q. Li and D. Rus, "Global clock synchronization in sensor networks," in *Proceedings of International Conference on Computer Communications*, March 2004.
- [24] T. M. Schmidl and D. C. Cox, "Robust frequency and timing synchronization for OFDM," *IEEE Transactions on Communications*, vol. 45, no. 12, pp. 1613–1621, 1997.
- [25] F. Tufvesson, O. Edfors, and M. Faulkner, "Time and frequency synchronization for OFDM using PN-sequence preambles," in *Proceedings of Vehicular Technology Conference-Fall*, September 1999.
- [26] S. Srinivasa and S. Jafar, "The optimality of transmit beamforming: A unified view," *IEEE Transactions on Information Theory*, vol. 53, no. 4, pp. 1558–1564, 2007.
- [27] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1451–1458, 1998.
- [28] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Transactions on Information Theory*, vol. 45, no. 5, pp. 1456–1467, 1999.
- [29] G. Scutari and S. Barbarossa, "Distributed space-time coding for regenerative relay networks," *IEEE Transactions on Wireless Communications*, vol. 4, no. 5, pp. 2387–2399, 2005.
- [30] J. Luo, R. S. Blum, L. J. Cimini, L. J. Greenstein, and A. M. Haimovich, "Link-failure probabilities for practical cooperative relay networks," in *Proceedings of Vehicular Technology Conference Spring*, June 2005.
- [31] Y. Jing and B. Hassibi, "Distributed space-time coding in wireless relay networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3524–3536, 2006.
- [32] B. S. Mergen and A. Scaglione, "Randomized space-time coding for distributed cooperative communication," *IEEE Transactions on Signal Processing*, vol. 55, no. 10, pp. 5003–5017, 2007.
- [33] L. Zhang and L. J. Cimini, "Efficient power allocation for decentralized distributed space-time block coding," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1102–1106, 2009.
- [34] S. Savazzi and U. Spagnolini, "Distributed orthogonal space-time coding: Design and outage analysis for randomized cooperation," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4546–4557, 2007.

- [35] Y. Jing and H. Jafarkhani, "Single and multiple relay selection schemes and their achievable diversity orders," *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp. 1414–1423, 2009.
- [36] A. Bletsas, A. Khisti, D. P. Reed, and A. Lippman, "A simple cooperative diversity method based on network path selection," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 659–672, 2006.
- [37] A. Bletsas, H. Shin, and M. Z. Win, "Cooperative communications with outage-optimal opportunistic relaying," *IEEE Transactions on Wireless Communications*, vol. 6, no. 9, pp. 3450–3460, 2007.
- [38] R. Yim, N. B. Mehta, and A. F. Molisch, "Fast multiple access selection through variable power transmissions," *IEEE Transactions on Wireless Communications*, vol. 8, no. 4, pp. 1962–1973, 2009.
- [39] V. Shah, N. B. Mehta, and R. Yim, "The relay selection and transmission trade-off in cooperative communication systems," *IEEE Transactions on Wireless Communications*, vol. 9, no. 8, pp. 2505–2515, 2010.
- [40] A. Goldsmith, *Wireless Communications*. Cambridge University Press, 2005.
- [41] B. Hayes, "The easiest hard problem," *American Scientist*, vol. 90, no. 2, pp. 113–117, 2002.
- [42] R. G. Gallager, "Basic limits on protocol information in data communication networks," *IEEE Transactions on Information Theory*, vol. 22, no. 4, pp. 385–398, 1976.
- [43] J. Hong and V. O. K. Li, "Impact of information on network performance - an information-theoretic perspective," in *Proceedings of Global Communications Conference*, November-December 2009.
- [44] D. Wang and A. A. Abouzeid, "On the cost of knowledge of mobility in dynamic networks: An information-theoretic approach," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 995–1006, 2012.
- [45] R. K. Ganti and M. Haenggi, "Throughput versus routing overhead in large ad hoc networks," in *Proceedings of Wireless Internet Conference*, November 2008.
- [46] M. Xu, D. Guo, and M. L. Honig, "Multicarrier beamforming with limited feedback: A rate distortion approach," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 916–927, 2013.
- [47] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.

- [48] H. L. Royden, *Real Analysis*. Prentice Hall, 1988, vol. 4.
- [49] S. Verdú, “Spectral efficiency in the wideband regime,” *IEEE Transactions on Information Theory*, vol. 48, no. 6, pp. 1319–1343, 2002.
- [50] P. T. Boggs and J. W. Tolle, “Sequential quadratic programming,” *Acta Numerica*, vol. 4, no. 1, pp. 1–51, 1996.
- [51] G. A. F. Seber and A. J. Lee, *Linear Regression Analysis*. John Wiley & Sons, 2003.
- [52] B. Gui, L. Dai, and L. J. Cimini, “Selective relaying in cooperative OFDM systems: Two-hop random network,” in *Proceedings of Wireless Communications and Networking Conference*, March-April 2008.
- [53] J. G. Andrews, S. Shakkottai, R. W. Heath, N. Jindal, M. Haenggi, R. Berry, D. Guo, M. Neely, S. Weber, S. Jafar, and A. Yener, “Rethinking information theory for mobile ad hoc networks,” *IEEE Communications Magazine*, vol. 46, no. 12, pp. 94–101, 2008.
- [54] V. Shah, N. B. Mehta, and R. Yim, “Optimal timer based selection schemes,” *IEEE Transactions on Communications*, vol. 58, no. 6, pp. 1814–1823, 2010.
- [55] R. Talak and N. B. Mehta, “Optimal timer-based best node selection for wireless systems with unknown number of nodes,” *IEEE Transactions on Communications*, vol. 61, no. 11, pp. 4475–4485, 2013.
- [56] V. Shah, N. B. Mehta, and R. Yim, “Splitting algorithms for fast relay selection: generalizations, analysis, and a unified view,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 4, pp. 1525–1535, 2010.
- [57] B. Hassibi and B. M. Hochwald, “How much training is needed in multiple-antenna wireless links?” *IEEE Transactions on Information Theory*, vol. 49, no. 4, pp. 951–963, 2003.
- [58] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed. New York: Dover, 1964.
- [59] J. Yackoski, L. Zhang, C.-C. Shen, L. J. Cimini, and B. Gui, “Networking with cooperative communications: Holistic design and realistic evaluation,” *IEEE Communications Magazine*, vol. 47, no. 8, pp. 113–119, 2009.
- [60] B. S. Mergen and A. Scaglione, “A continuum approach to dense wireless networks with cooperation,” in *Proceedings of International Conference on Computer Communications*, March 2005.

- [61] —, “On the power efficiency of cooperative broadcast in dense wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 497–507, 2007.
- [62] A. Kailas, “On the performance of alternating concurrent cooperative transmissions in the high path-loss attenuation regime,” *Network Protocols and Algorithms*, vol. 4, no. 2, pp. 68–81, 2012.
- [63] Z. Sheng, D. Goeckel, K. K. Leung, and Z. Ding, “A stochastic geometry approach to transmission capacity in wireless cooperative networks,” in *Proceedings of International Symposium on Personal, Indoor, and Mobile Radio Communications*, September 2009.
- [64] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, “Stochastic geometry and random graphs for the analysis and design of wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 27, pp. 1029–1046, 2009.
- [65] J. G. Andrews, R. K. Ganti, M. Haenggi, N. Jindal, and S. Weber, “A primer on spatial modeling and analysis in wireless networks,” *IEEE Communications Magazine*, vol. 48, no. 11, pp. 156–163, 2010.
- [66] A. Altieri, L. R. Vega, C. G. Galarza, and P. Piantanida, “Cooperative strategies for interference-limited wireless networks,” in *Proceedings of International Symposium on Information Theory*, July-August 2011.
- [67] H. Wang, S. Ma, and T.-S. Ng, “On performance of cooperative communication systems with spatial random relays,” *IEEE Transactions on Communications*, vol. 59, no. 4, pp. 1190–1199, 2011.
- [68] H. Wang, S. Ma, T.-S. Ng, and H. V. Poor, “A general analytical approach for opportunistic cooperative systems with spatial random relays,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 12, pp. 4122–4129, 2011.
- [69] Z. Guan, T. Melodia, D. Yuan, and D. A. Pados, “Distributed spectrum management and relay selection in interference-limited cooperative wireless networks,” in *Proceedings of International Conference on Mobile Computing and Networking*, September 2011.
- [70] Z. Guan, L. Ding, T. Melodia, and D. Yuan, “On the effect of cooperative relaying on the performance of video streaming applications in cognitive radio networks,” in *Proceedings of International Conference on Communications*, June 2011.
- [71] Z. Guan, T. Melodia, and D. Yuan, “Jointly optimal rate control and relay selection for cooperative wireless video streaming,” *IEEE/ACM Transactions on Networking*, vol. 21, no. 4, pp. 1173–1186, 2013.

- [72] L. Ding, T. Melodia, S. Batalama, and J. Matyjas, “Distributed routing, relay selection, and spectrum allocation in cognitive and cooperative ad hoc networks,” in *Proceedings of International Conference on Sensing, Communications, and Networking*, June 2010.
- [73] N. Abramson, “The ALOHA system: another alternative for computer communications,” in *Proceedings of National Computer Conference, American Federation of Information Processing Societies*, November 1970.
- [74] F. Baccelli, B. Blaszczyszyn, and P. Mühlethaler, “An ALOHA protocol for multihop mobile wireless networks,” *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 421–436, 2006.
- [75] S. Weber, J. G. Andrews, and N. Jindal, “An overview of the transmission capacity of wireless networks,” *IEEE Transactions on Communications*, vol. 58, no. 12, pp. 3593–3604, 2010.
- [76] P. C. Pinto, A. Giorgetti, M. Z. Win, and M. Chiani, “A stochastic geometry approach to coexistence in heterogeneous wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1268–1282, 2009.
- [77] M. Z. Win, P. C. Pinto, and L. A. Shepp, “A mathematical theory of network interference and its applications,” *Proceedings of the IEEE*, vol. 97, no. 2, pp. 205–230, 2009.
- [78] J. F. C. Kingman, *Poisson Processes*. Oxford University Press, 1993.
- [79] M. Haenggi and R. K. Ganti, *Interference in Large Wireless Networks*. NOW Publishers Inc., 2009.
- [80] S. Khakurel, M. Mehta, and A. Karandikar, “Optimal relay placement for coverage extension in LTE-A cellular systems,” in *Proceedings of India National Conference on Communication*, February 2012.
- [81] D. Stoyan, W. S. Kendall, J. Mecke, and L. Ruschendorf, *Stochastic Geometry and Its Applications*, 2nd ed. John Wiley & Sons Ltd, 1995.
- [82] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, “On the Lambert W function,” *Advances in Computational Mathematics*, vol. 5, no. 1, pp. 329–359, 1996.
- [83] C. K. Toh, *Ad Hoc Mobile Wireless Networks: Protocols and Systems*. Prentice Hall, 2001.
- [84] D. Bertsekas and R. Gallager, *Data Networks*. Prentice Hall, 1992.
- [85] M. Haenggi and D. Puccinelli, “Routing in ad hoc networks: a case for long hops,” *IEEE Communications Magazine*, vol. 43, no. 10, pp. 93–101, 2005.

- [86] Y. Chen and J. G. Andrews, “An upper bound on multi-hop transmission capacity with dynamic routing selection,” in *Proceedings of International Symposium on Information Theory*, June 2010.
- [87] M. Sikora, J. N. Laneman, M. Haenggi, D. J. Costello, and T. E. Fuja, “Bandwidth- and power-efficient routing in linear wireless networks,” *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2624–2633, 2006.
- [88] O. Oyman and S. Sandhu, “Non-ergodic power-bandwidth tradeoff in linear multi-hop networks,” in *Proceedings of International Symposium on Information Theory*, July 2006.
- [89] A. Florea and H. Yanikomeroglu, “On the optimal number of hops in infrastructure-based fixed relay networks,” in *Proceedings of Global Telecommunications Conference*, December 2005.
- [90] J. G. Andrews, S. Weber, M. Kountouris, and M. Haenggi, “Random access transport capacity,” *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 2101–2111, 2010.
- [91] D. Chen, M. Haenggi, and J. N. Laneman, “Distributed spectrum-efficient routing algorithms in wireless networks,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 5297–5305, 2008.
- [92] F. Baccelli and B. Blaszczyszyn, *Stochastic Geometry and Wireless Networks*. NOW Publishers Inc., 2009.
- [93] I. Stojmenovic, “Position-based routing in ad hoc networks,” *IEEE Communications Magazine*, vol. 40, no. 7, pp. 128–134, 2002.
- [94] P. Pabst, B. H. Walke, D. C. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. D. Falconer, and G. P. Fettweis, “Relay-based deployment concepts for wireless and mobile broadband radio,” *IEEE Communications Magazine*, vol. 42, no. 9, pp. 80–89, 2004.
- [95] K. Xu, H. Hassanein, G. Takahara, and Q. Wang, “Relay node deployment strategies in heterogeneous wireless sensor networks: multiple-hop communication case,” in *Proceedings of Communications Society Conference on Sensor and Ad Hoc Communications and Networks*, September 2005.
- [96] A. Venkateswaran, V. Sarangan, T. F. La Porta, and R. Acharya, “A mobility-prediction-based relay deployment framework for conserving power in manets,” *IEEE Transactions on Mobile Computing*, vol. 8, no. 6, pp. 750–765, 2009.

- [97] T. Himsoon, W. P. Siriwongpairat, Z. Han, and K. J. R. Liu, “Lifetime maximization via cooperative nodes and relay deployment in wireless networks,” *IEEE Journal on Selected Areas in Communications*, vol. 25, no. 2, pp. 306–317, 2007.
- [98] B. Gui, L. Dai, and L. J. Cimini, “Routing strategies in multihop cooperative networks,” *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 843–855, 2009.
- [99] S. K. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [100] Y. Xiao and L. J. Cimini, “Maximal throughput routing in wireless linear networks,” in *Proceedings of Conference on Information Sciences and Systems*, March 2012.
- [101] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, “An overview of limited feedback in wireless communication systems,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1341–1365, 2008.
- [102] N. Ravindran and N. Jindal, “Limited feedback-based block diagonalization for the MIMO broadcast channel,” *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1473–1482, 2008.
- [103] Q. Wang, L. Cimini, L. Greenstein, and D. Chan, “Single-user and multi-user mode selection for MIMO broadcast channels with imperfect channel state information,” in *Proceedings of Conference on Information Sciences and Systems*, March 2012.
- [104] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath, and B. L. Evans, “Sum capacity of multiuser MIMO broadcast channels with block diagonalization,” *IEEE Transactions on Wireless Communications*, vol. 6, no. 6, pp. 2040–2045, 2007.
- [105] M. Kibria, H. Murata, S. Yoshida, K. Yamamoto, D. Umehara, S. Denno, and M. Morikura, “Partitioned vector quantization for MU-MIMO downlink broadcasting,” in *Proceedings of Vehicular Technology Conference-Fall*, September 2012.
- [106] H. Song, W. Seo, and D. Hong, “Compressive feedback based on sparse approximation for multiuser MIMO systems,” *IEEE Transactions on Vehicular Technology*, vol. 59, no. 2, pp. 1017–1023, 2010.
- [107] R. N. de Sá Netto and C. C. Cavalcante, “Quantization and noise impact over feedback reduction of MIMO systems using compressive sensing,” in *Proceedings of International Symposium on Wireless Communication Systems*, August 2012.

- [108] S. T. Qaseem and T. Y. Al-Naffouri, “Compressive sensing for reducing feedback in MIMO broadcast channels,” in *Proceedings of International Conference on Communications*, June 2010.
- [109] K. Kim, S. Jang, D. Kim, and J. Seo, “An efficient feedback scheme using compressive sensing for MIMO broadcast channel with random beamforming,” in *Proceedings of International Conference on Computer and Automation Engineering*, February 2010.
- [110] W. Lu, Y. Liu, and D. Wang, “Efficient feedback scheme based on compressed sensing in MIMO wireless networks,” *Computers and Electrical Engineering*, vol. 39, no. 6, pp. 1587–1600, 2012.
- [111] P.-H. Kuo, H. Kung, and P.-A. Ting, “Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays,” in *Proceedings of Wireless Communications and Networking Conference*, April 2012.
- [112] Y. Li and R. Song, “Novel scheme of CSI feedback compression for multi-user MIMO-OFDM system,” in *Proceedings of International Workshop on Information and Electronics Engineering*, March 2012.
- [113] A. Bruckstein, D. Donoho, and M. Elad, “From sparse solutions of systems of equations to sparse modeling of signals and images,” *Society for Industrial and Applied Mathematics Review*, vol. 51, no. 1, pp. 34–81, 2009.
- [114] R. M. Gray and D. L. Neuhoff, “Quantization,” *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2325–2383, 1998.
- [115] D. Love, R. Heath, and T. Strohmer, “Grassmannian beamforming for multiple-input multiple-output wireless systems,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2735–2747, 2003.
- [116] B. M. Hochwald, C. B. Peel, and A. L. Swindlehurst, “A vector-perturbation technique for near-capacity multiantenna multiuser communication – part II: Perturbation,” *IEEE Transactions on Communications*, vol. 53, no. 3, pp. 537–544, 2005.
- [117] L. Zheng and D. Tse, “Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel,” *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, 2002.
- [118] W. Dai, Y. Liu, and B. Rider, “Quantization bounds on Grassmann manifolds and applications in MIMO systems,” *IEEE Transactions on Information Theory*, vol. 54, no. 3, pp. 1108–1123, 2008.

- [119] J. C. Ye and S. Y. Lee, “Non-iterative exact inverse scattering using simultaneous orthogonal matching pursuit (S-OMP),” in *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, March-April 2008.
- [120] R. A. DeVore and V. N. Temlyakov, “Some remarks on greedy algorithms,” *Advances in Computational Mathematics*, vol. 5, no. 3, pp. 173–187, 1996.
- [121] K. K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang, “On beamforming with finite rate feedback in multiple-antenna systems,” *IEEE Transactions on Information Theory*, vol. 49, no. 10, pp. 2562–2579, 2003.
- [122] J. Zheng, E. R. Duni, and B. D. Rao, “Analysis of multiple-antenna systems with finite-rate feedback using high-resolution quantization theory,” *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1461–1476, 2007.
- [123] J. G. Andrews, “Seven ways that hetnets are a cellular paradigm shift,” *IEEE Communications Magazine*, vol. 51, no. 3, pp. 136–144, 2013.
- [124] C. S. Bontu and E. Illidge, “DRX mechanism for power saving in LTE,” *IEEE Communications Magazine*, vol. 47, no. 6, pp. 48–55, 2009.
- [125] F.-S. Chu, K.-C. Chen, and G. Fettweis, “Green resource allocation to minimize receiving energy in OFDMA cellular systems,” *IEEE Communications Letters*, vol. 16, no. 3, pp. 372–374, 2012.
- [126] K. Lee, J. Lee, Y. Yi, I. Rhee, and S. Chong, “Mobile data offloading: How much can WiFi deliver?” *IEEE/ACM Transactions on Networking*, vol. 21, no. 2, pp. 536–550, 2013.
- [127] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, “A close examination of performance and power characteristics of 4G LTE networks,” in *Proceedings of International Conference on Mobile Systems, Applications, and Services*, June 2012.
- [128] F. Gabin, M. Kampmann, T. Lohmar, and C. Priddle, “3GPP mobile multimedia streaming standards [standards in a nutshell],” *IEEE Signal Processing Magazine*, vol. 27, no. 6, pp. 134–138, 2010.
- [129] D. Lecompte and F. Gabin, “Evolved multimedia broadcast/multicast service (embms) in LTE-advanced: overview and rel-11 enhancements,” *IEEE Communications Magazine*, vol. 50, no. 11, pp. 68–74, 2012.
- [130] A. Ganti, T. Klein, and M. Haner, “Base station assignment and power control algorithms for data users in a wireless multiaccess framework,” *IEEE Transactions on Wireless Communications*, vol. 5, no. 9, pp. 2493–2503, 2006.

- [131] S. Hong, J. Mehlman, and D. O’Neill, “Sortingham: holistic resource optimization for heterogeneous cellular networks,” Stanford University, Tech. Rep., 2011.
- [132] M. Proebster, M. Kaschub, T. Werthmann, and S. Valentin, “Context-aware resource allocation for cellular wireless networks,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2012, no. 1, p. 216, 2012.
- [133] P. Bellavista, A. Corradi, and L. Foschini, “Context-aware handoff middleware for transparent service continuity in wireless networks,” *Pervasive and Mobile Computing*, vol. 3, no. 4, pp. 439–466, 2007.
- [134] M. Stiemerling and S. Kiesel, “A system for peer-to-peer video streaming in resource constrained mobile environments,” in *Proceedings of the ACM Workshop on User-provided Networking*, December 2009.
- [135] N. B. Niraula, K. Kanchanasut, and A. Laouiti, “Peer-to-peer live video streaming over mobile ad hoc network,” in *Proceedings of International Wireless Communications and Mobile Computing Conference*, June 2009.
- [136] X. Jin and Y.-K. Kwok, “Cloud assisted P2P media streaming for bandwidth constrained mobile subscribers,” in *Proceedings of International Conference on Parallel and Distributed Systems*, December 2010.
- [137] F. Qiu, J. Bai, Y. Cui, and Y. Xue, “Optimal rate allocation in peer-to-peer streaming over wireless networks,” in *Proceedings of International Conference on Collaboration Technologies and Systems*, May 2011.
- [138] Y. Xiao, Y. Guan, L. J. Cimini, and C.-C. Shen, “Power efficient user pairing for multicasting in heterogeneous wireless networks,” in *Proceedings of Military Communications Conference*, November 2013.
- [139] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, “Coordinated multipoint: Concepts, performance, and field trial results,” *IEEE Communications Magazine*, vol. 49, no. 2, pp. 102–111, 2011.
- [140] B. Zhang, S.-H. G. Chan, G. Cheung, and E. Y. Chang, “Localtree: An efficient algorithm for mobile peer-to-peer live streaming,” in *Proceedings of International Conference on Computer Communications*, June 2011.
- [141] D. Bertsimas and D. Brown, “Constructing uncertainty sets for robust linear optimization,” *Operational Research*, vol. 57, no. 6, pp. 1483–1495, 2009.
- [142] N. Balasubramanian, A. Balasubramanian, and A. Venkataramani, “Energy consumption in mobile phones: a measurement study and implications for network applications,” in *Proceedings of Internet Measurement Conference*, November 2009.

- [143] A. Carroll and G. Heiser, “An analysis of power consumption in a smartphone,” in *Proceedings of the USENIX Annual Technical Conference*, June 2010.
- [144] S. Alawnah and A. Sagahyroon, “Modeling smartphones power,” in *Proceedings of European Conference on Electrotechnics*, July 2013.