

Performance of Relaying: Infinite Blocklength Regime vs. Finite Blocklength Regime

Von der Fakultät für Elektrotechnik und Informationstechnik
der Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors
der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

M.Sc. Yulin Hu

Aus Anhui, China

Berichter: Universitätsprofessor Dr.-Ing. Anke Schmeink
Universitätsprofessor Dr.-Ing. James Gross

Tag der mündlichen Prüfung: 22. Dezember 2015

Diese Dissertation ist auf den Internetseiten der
Hochschulbibliothek online verfügbar.

Declaration

Some of the research presented in this thesis has been published in some journals and conference proceedings. These publications are listed after the abstract of the thesis.

All work presented in this thesis as original is so, to the best knowledge of the author. References and acknowledgments to other researchers have been given as appropriate.

Aachen, 8. Januar 2016

Yulin Hu

Acknowledgements

My most sincere thanks go to my advisors, Prof. James Gross and Prof. Anke Schmeink. James has been extremely patient and tremendously inspiring throughout different stages of my PhD study. Anke has offered me invaluable interactions and freedom when I was preparing this dissertation as well as other journal and conference publications. Their thoughtful advices have greatly helped me in building my long-term career path.

Special thanks go to my defense committee members, Prof. Gerd Ascheid and Prof. Dorit Merhof for spending the time, offering oral exam, giving me feedback, and attending my defense session.

Several colleagues of the former MNP group as well as ISEK and TI have made my life more energetic and diverse. I would like to thank Di Li, Oscar Punal, Christian Dombrowski, Donald Parruca, Marco Weyres, Tong Wang, Qinwei He, Grit Classen, Simon Goertzen, Christopher Schnelling, Christoph Schmitz, Michael Reyer, Ehsan Zandi and Omid Taghizadeh.

I would also like to thank my family: My parents and my sister for supporting me spiritually throughout writing this thesis and my life in general. And finally, infinite thanks to my lovely wife as well as my son.

Abstract

Motivated by the increasing demand for higher throughput, broader coverage and higher-level quality of service (QoS), a major effort is being made to study the relaying technology in wireless networks. However, most existing research on investigating the relaying performance over wireless fading channels typically rely on simplifying assumptions, such as perfect channel state information (CSI) and infinite blocklength, which may be inaccurate in realistic wireless systems.

In this dissertation, we contribute to developing more extensive and systematical investigation models of relaying. On one hand, both the physical-layer and the higher-layer performance models of relaying are developed while multiple performance metrics are taken into account, e.g., reliability, capacity and energy efficiency. On the other hand, various scenarios with different assumptions regarding CSI and blocklength (e.g., perfect or imperfect CSI and finite or infinite blocklengths) are considered. In general, the following fundamental questions are addressed based on our investigation models: Does relaying become more efficient or pay off less under these scenarios with more realistic assumptions (e.g., imperfect CSI or/and finite blocklengths)? What are performance differences if we do not consider or only consider one of these realistic assumptions? Is it essential to consider these realistic assumptions? What can we suggest for designing relaying systems especially for realistic applications?

We achieve the above ultimate aim and find answers for these fundamental questions by two steps. In the first step, we investigate above performance metrics of relaying networks in the infinite blocklength regime (where blocklengths are infinite) and in the finite blocklength regime (where blocklengths are finite), respectively. In particular, both the perfect CSI scenario and the imperfect CSI scenario are considered. In other words, by this step relaying performance models are studied under different assumptions regarding CSI and blocklength, respectively. In the second step, we simulatively compare these relaying performance models under different scenarios/assumptions. By these two steps, we finally observe some surprising performance advantages of relaying in the finite blocklength regime (in comparison to in the infinite blocklength regime) and conclude a set of important guidelines for the design of efficient relaying systems.

List of Publications

Publications included by Chapter 4 regarding relaying in the infinite blocklength regime:

- [J1] Y. Hu, J. Gross and A. Schmeink, "QoS-Constrained energy efficiency of cooperative ARQ in multiple DF relay systems", *IEEE Trans. Veh. Technol.*, to appear.
- [C1] Y. Hu, J. Gross, A. Schmeink and T. Wang, Maximizing energy efficiency for multiple DF relay system with QoS constraint", *Proc. IEEE ISWCS*, Ilmenau, Germany, Aug 2013.
- [C2] Y. Hu and J. Gross, On the outage probability and effective capacity of multiple decode-and-forward relay system,"*Proc. IFIP/IEEE Wireless Days*, Nov. 21-23, 2012, Dublin, Ireland.

Publications included by Chapter 5 and Chapter 6 regarding relaying in the finite blocklength regime:

- [J2] Y. Hu, J. Gross and A. Schmeink, "Blocklength-limited performance of relaying under quasi-static Rayleigh channels", *IEEE Trans. Wireless. Commun.*, under review (after major revision).
- [J3] Y. Hu, J. Gross and A. Schmeink, On the performance advantage of relaying under the finite blocklength regime", *IEEE Commun. Lett.*, vol. 19, no. 5, pp. 779-782, May 2015.
- [J4] Y. Hu, J. Gross and A. Schmeink, On the capacity of relaying with finite blocklength", *IEEE Trans. Veh. Technol.*, to appear.

Publications not included:

Publications by the author, which are not included in this thesis, are listed below.

- [C3] Y. Hu, J. Gross and A. Schmeink, Outage probability of a multi-relay cognitive network with an uncertain number of forwarding relays", *Proc. IEEE ISWCS*, Barcelona, Spain, Aug. 2014.
- [C4] Y. Hu, J. Gross and Z. Ding, The outage performance of realtime transmission in multiple asynchronous relays enhanced OFDM system", *Proc. IEEE ICNC*, San Diego, USA, Jan. 2013.
- [C5] Y. Hu and L. Qiu, A novel multiple relay selection strategy for LTE-Advanced relay systems", *Proc. IEEE VTC*, Budapest, May 2011
- [C6] M. Serror, Y. Hu, C. Dombrowski, K. Wehrle and J. Gross, Performance of cooperative ARQ with perfect CSI in wireless industrial networks", *Proc. IEEE WOWMOM 2016*, under review.
- [J5] Y. Hu, L. Qiu, Optimal antenna locations for distributed antennas in Cellular networks." *J. China Univ. Sci. and Tech.*, 2009, vol. 10, pp.1091-1096.

Inhaltsverzeichnis

1	Introduction	1
2	Background Issues	5
2.1	Relaying	5
2.2	Effective Capacity	6
2.2.1	Effective Bandwidth	7
2.2.2	Effective Capacity	9
2.3	Point-to-Point (P2P) Scenario: Infinite Blocklength Regime vs. Finite Blocklength Regime	13
2.4	Maximum Ratio Combining	16
2.5	Convex/Concave and Quasi-convex/Quasi-Concave Functions	18
2.5.1	Convex and Concave Functions	18
2.5.2	Quasi-Convex and Quasi-Concave Functions	19
3	Related Work and Scope of the Thesis	23
3.1	Relaying in the Infinite Blocklength Regime: State of the Art	23
3.2	Relaying in the Finite Blocklength Regime: State of the Art	26
3.3	Scope of the Thesis	27
4	Relaying Performance in the Infinite Blocklength Regime	31
4.1	System Model	32
4.1.1	General System Description	32
4.1.2	Channel, SNR and Energy Consumption Models	33
4.1.3	Multi-Relay C-ARQ Protocols	35
4.1.3.1	Proactive/Reactive Behavior	35
4.1.3.2	Static/Dynamic Relaying Behavior	35

4.2	Methodology	37
4.3	Reliability Issue in the Infinite Blocklength Regime – Outage Probability	39
4.3.1	Multi-DF-Relay Assisted Transmissions	39
4.3.2	Multi-DF-Relay Assisted Retransmissions	43
4.3.2.1	Simplified Case – C-ARQ without MRC	44
4.3.2.2	General Case – C-ARQ with MRC	45
4.4	Throughput and Effective Capacity	51
4.4.1	Physical-layer Throughput	51
4.4.2	Higher-layer Effective Capacity	55
4.5	Energy Efficiency	61
4.5.1	Energy Consumption	61
4.5.2	Physical-layer Energy Efficiency (PEE)	67
4.5.2.1	The Derivation of Physical-layer Energy Efficiency	67
4.5.2.2	The Performance Evaluation of PEE	67
4.5.2.3	The Performance Improvement of PEE	68
4.5.3	Higher-layer Energy Efficiency– Effective Energy Efficiency	76
4.5.3.1	The Definition of Effective Energy Efficiency (EEE)	76
4.5.3.2	Performance Evaluation regarding EEE	77
4.5.3.3	The Performance Improvement of EEE	79
4.6	Performance Evaluation of C-ARQ Protocols	80
4.6.1	Evaluation of the System Performance on Relay Numbers:	81
4.6.2	Performance Loss due to Relay Number Uncertainty	83
4.6.3	Optimality of Dynamic Protocols for Different Topologies	84
4.7	Summary and Discussion	86
5	Relaying Performance in the Finite Blocklength Regime	89
5.1	System Model	89
5.1.1	System General Description	89
5.1.2	Channel Model and CSI Assumptions	90
5.2	Methodology	93
5.3	Reliability Issue in the Finite Blocklength Regime – Error Probability	94
5.3.1	Perfect CSI Scenario with Static Channels	94
5.3.2	Perfect CSI Scenario with Quasi-Static Channels	97

5.3.3	Average CSI Scenario with Quasi-Static Channels	98
5.4	BL-Capacity (Physical-Layer Throughput) and Effective Capacity . . .	100
5.4.1	The Blocklength-Limited Capacity: BL-Capacity	100
5.4.1.1	Perfect CSI Scenario with Static Channels	101
5.4.1.2	Perfect CSI Scenario with Quasi-Static Channels . . .	104
5.4.1.3	Average CSI Scenario with Quasi-Static Channels . . .	104
5.4.2	Effective Capacity	108
5.4.2.1	Perfect CSI Scenario with Static Channels	109
5.4.2.2	Perfect CSI Scenario with Quasi-Static Channels . . .	109
5.4.2.3	Average CSI Scenario with Quasi-Static Channels . . .	110
5.5	Energy Efficiency	115
5.5.1	Energy Consumption	115
5.5.2	The Physical-Layer Energy Efficiency (PEE)	116
5.5.3	The Higher-Layer Energy Efficiency (EEE)	116
5.6	Summary and Discussion	118
6	Relaying Performance Comparison: Infinite Blocklength Regime vs. Finite Blocklength Regime	121
6.1	Performance Advantage of Relaying in a Single Relay System	121
6.1.1	Simulative Investigation	121
6.1.1.1	Performance Advantage of Relaying in the Finite Blocklength Regime: Relaying vs. Direct Transmission.	123
6.1.1.2	The Performance Advantage of Relaying under Perfect CSI Scenario vs. Average CSI Scenario.	125
6.1.1.3	The Performance Loss Due to a Finite Blocklength: Perfect CSI Scenario vs. Average CSI Scenario	126
6.1.2	A Theoretical Explanation of the Performance Advantage of Relaying	129
6.2	Performance Advantage of Relaying under a Multi-Relay Scenario . . .	136
6.2.1	Simulative Investigation	137
6.2.1.1	The Impact of Relay Number on the Performance Advantage of Relaying	138

6.2.1.2	Evaluating the Performance Advantage of Multi-Relay in the Finite Blocklength Regime	139
6.2.2	An Explanation of the Performance Advantage of Multi-Relay in the Finite Blocklength Regime	142
6.3	Summary	144
7	Conclusions and Future Work	145
7.1	Summary and Discussion	145
7.2	Novelty and Contributions	149
7.3	Future Research Directions	151
A	Proof of Proposition in Chapter 4	157
A.1	The Proof of Equation (4.6)	157
A.2	The Proof of Proposition 1	157
B	Proofs of Propositions in Chapter 5	159
B.1	The poof of Proposition 3	159
B.2	The poof of Proposition 4	160
B.3	The poof of Proposition 5	161
B.4	The poof of Proposition 6	162
B.5	The poof of Proposition 7	166
B.6	The poof of Proposition 8	166
	Glossary	169
	List of Symbols and Notations	171
	Literaturverzeichnis	175

Abbildungsverzeichnis

1.1	Scenarios considered in this thesis regarding different assumptions of CSI and blocklength.	2
2.1	A queuing model of the effective bandwidth analysis.	7
2.2	A queuing model [22] of the effective capacity analysis.	10
2.3	A function is quasi-convex but neither convex nor continuous.	20
3.1	A general summary of related work of relaying in the infinite blocklength regime from a perspective of research topic.	25
3.2	A specific summary (from a perspective of system assumption) of related work regarding higher-layer performance of relaying.	26
3.3	Methodology of this thesis.	28
3.4	The scope and the contribution of this thesis (from a scenario perspective).	29
3.5	Scopes of Chapters 4, 5 and 6 which involve all the research work in this thesis.	29
4.1	Example of the considered multiple relay system scenario.	33
4.2	The difference between proactive protocols and reactive protocols.	36
4.3	The difference between static protocols and dynamic protocols.	36
4.4	Methodology of developing investigation models of relaying in the infinite blocklength regime.	38
4.5	Outage probability performance under different topologies (i.e., different radii of relay-group area).	42
4.6	Validation of the distribution of τ of the C-ARQ protocol without MRC (packet size 70 bits).	48

4.7	Validation of distributions of τ of these C-ARQ protocols with MRC (packet size 85 bits).	49
4.8	The delay comparison between these protocols with and without MRC (packet size is set to 70 bits).	50
4.9	Outage performance of these three C-ARQ protocols (packet size 70 bits).	51
4.10	Physical-layer throughputs of the C-ARQ protocol without MRC.	54
4.11	Physical-layer throughputs of C-ARQ protocols with MRC	54
4.12	Physical-layer throughputs of C-ARQ protocols with different assumptions of MRC	55
4.13	Variance of the C-ARQ protocol without MRC.	57
4.14	MSDR of the C-ARQ protocol without MRC.	57
4.15	Variance of C-ARQ protocols with MRC.	58
4.16	MSDR of C-ARQ protocols with MRC.	59
4.17	Variance of service process increments under studied C-ARQ protocols with different MRC assumptions.	60
4.18	MSDR of the C-ARQ protocols under different MRC assumptions.	60
4.19	Energy consumption of the C-ARQ protocol without MRC.	65
4.20	Energy consumptions of C-ARQ protocols with MRC.	66
4.21	Energy consumptions of C-ARQ protocols under different MRC assumptions.	66
4.22	Physical-layer energy efficiencies of C-ARQ protocols with MRC.	68
4.23	Physical-layer energy efficiencies of C-ARQ protocols under different MRC assumptions.	69
4.24	The flow chart of the process of solving the tradeoff problem	73
4.25	The energy efficiency and the outage probability on tradeoff factor	74
4.26	The relationship between energy efficiency and outage probability	76
4.27	Higher-layer energy efficiencies of C-ARQ protocols with MRC.	78
4.28	Higher-layer energy efficiencies of C-ARQ protocols under different MRC assumptions.	79
4.29	The energy consumption, MSDR and EEE performance of proposed protocol with $\omega = 0.2$	80
4.30	MSDR and physical-layer throughput with dynamic number of relays deployed in the system. The QoS parameters are $\{d, P_d\} = \{20T_f, 10^{-3}\}$	81

4.31	EEE and PEE of the protocols with dynamic number of relays deployed in the system. The QoS parameters are $\{d, P_d\} = \{20T_f, 10^{-3}\}$	82
4.32	EEE Performance Loss due to relay number uncertainty.	83
4.33	The EEE comparison between the dynamic reactive protocol and the dynamic proactive protocol while varying the location of the destination.	84
4.34	The EEE comparison between proactive protocols and reactive protocols, where locations of the source and relays are fixed.	85
5.1	Example of the considered single relay system scenario.	90
5.2	Three scenarios considered in this chapter.	90
5.3	Methodology of developing investigation models of relaying in the finite blocklength regime.	94
5.4	The relationship between overall error probability and error probability of the bottleneck link.	102
5.5	BL-capacities under static channels with $\gamma_1 = -5$ dB, $\gamma_2 = \gamma_3 = 4$ dB and $\gamma_{\text{MRC}} = 4.5$ dB.	103
5.6	The performance comparison between relaying with average CSI and relaying with perfect CSI while varying blocklength.	106
5.7	The relationship between BL-capacity and coding rate under the average CSI scenario.	106
5.8	BL-capacity (in bit/ch.use) vs. weight factor under the average CSI scenario.	107
5.9	How the weight factor introduces tradeoff between error probability and coding rate and further influences the BL-capacity.	108
5.10	MSDR performance of relaying under static channels.	110
5.11	The relationship between MSDR and coding rate under both the perfect CSI scenario and the average CSI scenario.	112
5.12	The relationship between MSDR and coding rate under the average CSI scenario.	113
5.13	The relationship between MSDR and weight factor under the average CSI scenario.	113
5.14	The relaying performance versus blocklength under the average CSI scenario.	114

5.15	Energy efficiencies of relaying under perfect CSI scenario (with static channels).	117
5.16	energy efficiencies of relaying under quasi-static channels.	117
6.1	The performance comparison between relaying (with average CSI) and direct transmission (with average CSI) while varying the weight factor η . In the simulation, the blocklength at each hop of relaying is 500 symbols.	123
6.2	The performance comparison between relaying (with average CSI) and direct transmission (with average CSI) under dynamic blocklength ($\eta = 0.2$).	124
6.3	The performance comparison between relaying with average CSI and relaying with perfect CSI while varying the weight factor.	125
6.4	The performance comparison between relaying with average CSI and relaying with perfect CSI under dynamic blocklength ($\eta = 0.2$).	126
6.5	The performance comparison between relaying in the infinite blocklength regime and relaying in the finite blocklength regime (blocklength at each hop of relaying is 500 symbols).	127
6.6	The performance comparison between relaying in the infinite blocklength regime and relaying in the finite blocklength regime while varying the blocklength. In the simulation, $\eta = 0.2$	128
6.7	The performance advantage of relaying in the finite blocklength regime (from a perspective of topology).	133
6.8	Performance ratios of relaying and direct transmission while the blocklength of each phase of relaying and the blocklength of direct transmission are 100 and 200. We vary p_{tx} to obtain different SNR while setting $C_R = C_D$	134
6.9	Performance advantage ratio of relaying to direct transmission ($\varepsilon_R = \varepsilon_D = 10^{-3}$).	135
6.10	The comparison between relaying and direct transmission with finite blocklengths while the SNR of the direct link is 5 dB and $\varepsilon_R = \varepsilon_D = 10^{-3}$.	136
6.11	Shannon capacity vs. BL-capacity while varying the number of relays deployed in the system.	138

6.12	Shannon capacity vs. blocklength-limited MSDR while varying the number of relays deployed in the system.	139
6.13	Relaying performance under two-relay scenario.	140
6.14	The performance of two-relay transmission while varying the blocklength ($\eta = 2$).	141
7.1	Error models under different scenarios/assumptions	147

1 Introduction

The 5th Generation of wireless communication (5G) is more about providing services at the appropriate QoS, said Marcus Weldon, a chief technology officer of Alcatel-Lucent. In fact, future wireless networks are expected to provide certain QoS guarantees so that acceptable performance and quality levels can be met for various end-applications. In particular, latency and error probability are major concerns of the design of future wireless networks. First, in voice over IP (VoIP), interactive-video (e.g., videoconferencing) and streaming-video applications where the end-users are human beings, data packets are expected to satisfy certain delay and error probability constraint according to user experience. In addition, in some applications where the end-users are machines, e.g., machine-to-machine (M2M) networks, Internet of things and industrial control applications, requirements of latency and error probability become critical performance metrics. For instance, the synchronization of generators of smart grid generally have requirements of at most 5 *ms* delay and maximal error probability 10^{-5} . For such networks, performance evaluation and system design are also interesting and challenging.

To satisfy the requirements of future wireless networks, various technologies are proposed, studied and improved. Among them, relaying has been considered as one of key technologies in the standardization of future wireless communications, e.g., by 3rd Generation Partnership Project (3GPP) [1] and by 5G pioneers [2, 3]. It is also well known from literature [4, 5, 6, 7] as an efficient way to mitigate wireless fading by exploiting spatial diversity. However, most existing research on investigating the relaying performance over wireless fading channels typically rely on simplifying assumptions, such as perfect CSI, infinite blocklength and vanishing error probability. Although these assumptions are reasonable for most of the current wireless communication applications, they may be inaccurate in future wireless systems for applications of communicating between machines with critical QoS constraints.

In this thesis, we develop fundamental investigation and evaluation models for relaying networks. On one hand, we study the reliability, capacity and energy efficiency of relaying from both the physical-layer and the higher-layer perspectives. On the other hand, our investigation models cover all these application scenarios (with and without critical QoS constraints). As shown in Figure 1.1, we consider various scenarios with different assumptions regarding CSI and blocklength (e.g., perfect or imperfect CSI and finite or infinite blocklengths).

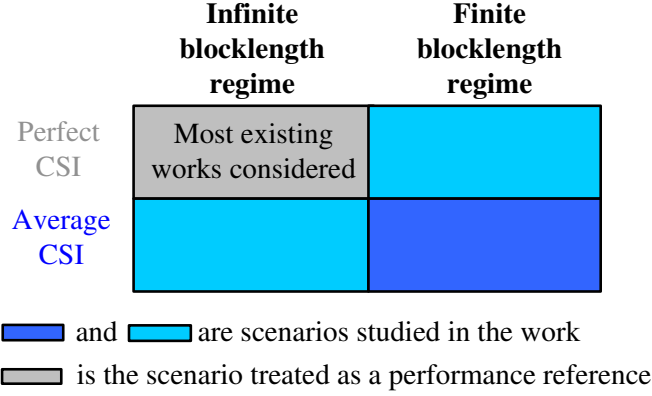


Abbildung 1.1: Scenarios considered in this thesis regarding different assumptions of CSI and blocklength.

The above ultimate aim of this thesis is addressed by several steps. Firstly, we respectively study the relaying performance in the infinite blocklength regime as well as the finite blocklength regime¹. In the infinite blocklength regime (the left half of Figure 1.1), we analyze the performance of relaying networks with only average CSI while relay-assisted retransmission protocols are considered to improve the transmission reliability. In addition, different from existing works in the literature, the (acknowledgment) feedback overhead of time and energy is taken into account. Moreover, we assume that (in the retransmission process of a packet) these fail-decoded signals of the initial transmission and previous retransmissions are combined with the new retransmitted signal

¹ We adopt the name finite blocklength regime" (which is used in the fundamental work [8] regarding the performance analysis of a transmission with finite blocklength) to represent the scenario with finite blocklengths. In addition, the well-known Shannon capacity theory indicates the achievable data rate of a transmission where blocklengths are assumed to be infinite and therefore noise can be perfectly averaged by these infinite blocklengths. In this work, we refer to the infinite blocklength regime as the scenario with infinite blocklengths.

before being decoded by the destination. Under the described system, we systematically investigate the physical-layer performance (e.g., outage probability, throughput, physical-layer energy efficiency) as well as the higher-layer performance (e.g., effective capacity, higher-layer energy efficiency) of relaying. In addition, by introducing two simple operations, we improve the physical-layer energy efficiency and the higher-layer energy efficiency, respectively. Then, we consider the finite blocklength regime which corresponds to the right half of Figure 1.1. Finite blocklength regime differs from the infinite blocklength regime in that a finite blocklength makes the error probability of the communication become no longer negligible. In particular, a fundamental work [8] shows that the performance loss due to a finite blocklength is considerable and becomes more significant when the blocklength is relatively short. Considering that two-hop relaying exploits spatial diversity and at the same time it halves the blocklength of the transmission (if equal time division is considered), the relaying performance in the finite blocklength regime becomes interesting. In this part of the work, we investigate this relaying performance with a finite blocklength while the focus is on the question if halving the blocklength makes relaying pay off less in the finite blocklength regime. The error probability, blocklength-limited capacity, energy efficiency as well as the higher-layer performance of relaying are derived under both the perfect CSI scenario and the average CSI scenario. Finally, we simulatively compare relaying behaviors in the infinite blocklength regime and in the finite blocklength regime while all the scenario in Figure 1.1 are addressed. We observe some surprising performance advantages of relaying in the finite blocklength regime in comparison to in the infinite blocklength regime.

By the above analytical investigations as well as simulative evaluations, we find significant performance differences among relaying networks under different assumptions regarding CSI and blocklength. Our key findings are as follows:

General findings:

- Relaying is an efficient way to improve the system performance under all the scenarios with different assumptions regarding CSI and blocklength.
- Deploying multiple relays boots the system performance significantly while overhead should be spent to acquire the exact number of relays present in the system.

Specific findings regarding blocklength and CSI:

- In the infinite blocklength regime (with infinite blocklengths), both the physical-layer energy efficiency and the higher-layer energy efficiency of relaying are quasi-concave in the packet size. In the finite blocklength regime, the physical-layer energy efficiency is increasing in the blocklength while the higher-layer energy efficiency is quasi-concave in the blocklength. Moreover, with a finite blocklength both these two energy efficiencies are concave or quasi-concave in the coding rate of the block.
- A finite blocklength introduces a performance loss in comparison to the Shannon capacity. However, relaying (which halves the blocklength) is able to mitigate this loss by setting the coding rate at each hop aggressively. Moreover, we found that this performance advantage of relaying (with a single relay) under the average CSI scenario is more significant than under the perfect CSI scenario. In other words, relaying (with a single relay) is more efficient in systems with short latency/blocklengths and average CSI.
- Surprisingly, under a multi-relay scenario relaying introduces a significantly high blocklength-limited performance (i.e., performance in the finite blocklength regime). In particular, with only average CSI relaying has similar performance in the infinite blocklength regime and in the finite blocklength regime. However, if perfect CSI is assumed to be available at the source, the blocklength-limited performance of relaying is possible and more likely to be higher than the Shannon capacity limit.

The rest of the thesis is structured as follows: After providing an overview of related background issues (Chapter 2), the state of the art regarding performance analysis of relaying is presented (Chapter 3). The scope of the thesis is discussed in detail in Section 3.3. Subsequently, analytical models for investigating relaying performance in the infinite blocklength regime as well as in the finite blocklength regime are presented in Chapter 4 and 5, respectively. Then, in Chapter 6 we compare the relaying performance/behaviors under these two regimes by simulations. Finally, in Chapter 7 we draw our conclusions and discuss issues for future work.

2 Background Issues

2.1 Relaying

Relaying is well-known as a fundamental technique to improve the performance of a wireless link. It deploys additional transceiver(s) such that the destination likely receives a much stronger signal forwarded by the relay(s) in comparison to direct transmission by the source. Hence the coverage of the network and the transmission reliability are improved [9, 10].

There are two main relaying duplex modes: Half-duplex relaying and full-duplex relaying. In half-duplex relaying [11], the relay can only either transmit or receive at the same time and at the same frequency band. Thus, a transmission of a half-duplex relaying network is divided into two frames. In the first frame, the source transmits to the relay and later on in the second frame the relay forwards the signal to the destination. Under the full-duplex relaying mode [12], the relay receives a signal from the source and transmits a signal to the destination at the same time and at the same frequency band. This results in self-interference, which is actually a major limitation of full-duplex relaying. In this thesis, half-duplex relaying mode is assumed.

In addition, there are also two major data processing principles that have been established for relaying [13]: In amplify-and-forward (AF) relaying protocols, the analog samples of the received signal are simply amplified and passed on. No decoding of the baseband samples is performed. As AF relays have low costs and short sizes, they are usually used as signal repeaters [14]. The advantage of AF relaying is that the destination definitely receives something from relay(s) while the drawback is that it cannot detect errors in the received signal. Contrarily, if the relay decodes the signal first and then forwards it, this is referred to as decode-and-forward (DF) relaying. As the relay decodes the signal, errors can be detected by the relay. The disadvantage of

DF relaying is that if the relay fails to decode the signal the destination will receive nothing from the relay. In this work, we consider the DF relaying principle.

According to the number of deployed relays, relaying systems also can be divided into single-relay systems and multi-relay systems. In comparison to a single-relay system, multiple relays in general boost the system performance in two ways: If only the best relay is selected for forwarding the data packet, a selective diversity is achieved [15, 16, 17]; If more than one relay are selected to assist the transmission, this introduces a cooperative diversity [18, 19, 20]. In this thesis, we consider both single-relay and multi-relay scenarios.

2.2 Effective Capacity

As we have discussed in the introduction, QoS is one of the major concerns in the design of future wireless communication networks. Considering that it is possible that different services have different levels of QoS requirements need to be satisfied, this is one of the great challenges during the systems design of future networks. On the one hand, it is difficult to provide deterministic QoS guarantees in wireless networks due to the random fading nature of wireless channels. On the other hand, it is usually infeasible to guarantee that the queue length at the transmitter buffer never exceeds a certain threshold.

To facilitate the design of QoS-support systems, a more practical approach is to consider not the deterministic QoS provisioning but the statistical one. The original study of applying statistical QoS guarantees in dynamic queuing analyses were in the context of effective bandwidth [21], where the key part of the derivation is based on the large deviation principle. Based on the effective bandwidth analysis, the famous effective capacity model is developed [22]. In the following, before providing the derivation of the effective capacity model we first introduce the effective bandwidth which is actually the beginning of the story of effective capacity.

2.2.1 Effective Bandwidth

The effective bandwidth [21] is defined as the minimum constant service rate required to guarantee a specified statistical QoS constraint θ which is called QoS exponent in the effective capacity model [22]. In particular, as shown in Figure 2.1, the effective

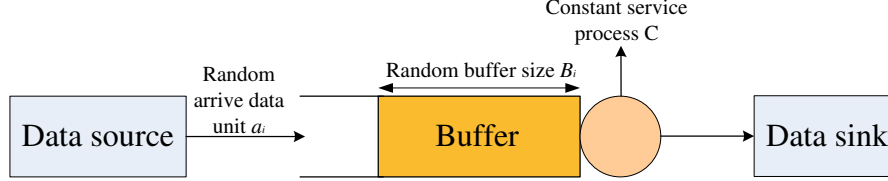


Abbildung 2.1: A queuing model of the effective bandwidth analysis.

bandwidth model considers a discrete-time system where a random increment of data units a_i arrive during slot i to an infinite buffer. If the buffer is served a first-come-first-served data stream with constant rate C , the random buffer size at time i is given by:

$$B_i = \max_{0 \leq k \leq i} (A_i - A_k - (i - k) \cdot C), \quad (2.1)$$

where A_i is the cumulative arrival function with $A_0 = 0$. In addition, A_i is assumed to have stationary increments.

The focus of effective bandwidth analysis is to study the distribution of B_i , especially when $i \rightarrow \infty$. However, based on the above model we notice that as i increases the arrival increments are not supersets of smaller values of i . Hence, it is difficult to analyze B_i for $i \rightarrow \infty$ based on Equation (2.1). Then, let us consider the following trick instead: Call the current frame (time slot) 0 (instead of i) and assume i time units ago the buffer to be empty the last time. Denote the time adjusted cumulative arrivals by A' . Then, we can reformulate the above equation into:

$$B'_i = \max_{0 \leq k \leq i} (A'_0 - A'_{-k} - k \cdot C). \quad (2.2)$$

Obviously, B_i and B'_i have the same distribution as the increments of A/A' are stationary.

2 Background Issues

Based on this reformulated equation, we fix a sample path A'_i and let $i \rightarrow \infty$. As we are considering an increasing set over which we build the maximum, hence B'_i is non-decreasing and converges to:

$$B^\infty = \max_{0 \leq s} (A'_0 - A'_{-s} - s \cdot C). \quad (2.3)$$

Hence, B_i also converges to this limit. Therefore, the considered queuing system is stable if B^∞ is finite with probability 1 which happens precisely when:

$$\lim_{i \rightarrow \infty} \frac{A_i}{i} = \lim_{i \rightarrow \infty} \frac{A'_0 - A'_{-i}}{i} = r \leq C. \quad (2.4)$$

In the long-time, fraction of the time that the buffer exceeds a certain threshold t is given by:

$$\Pr \{B^\infty > t\} = \Pr \left\{ \max_{0 \leq k} (A'_0 - A'_{-k} - k \cdot C) > t \right\}. \quad (2.5)$$

Define $S_{-k} = A'_0 - A'_{-k} - k \cdot C$ and consider $\Pr \{B'_i > t\}$. We have:

$$\begin{aligned} \Pr \{B'_i > t\} &= \Pr \left\{ \max_{0 \leq k \leq i} (S_{-k}) > t \right\} \\ &= \Pr \left\{ \bigcup_{0 \leq k \leq i} (S_{-k} > t) \right\} \\ &\leq \sum_{k=0}^i \Pr \{S_{-k} > t\}. \end{aligned} \quad (2.6)$$

Considering the Chernoff's Bound $\Pr \{x > 0\} \leq \mathbb{E} [e^{\theta x}]$ where $\mathbb{E} [e^{\theta x}]$ is the expected value of $e^{\theta x}$, we have:

$$\begin{aligned} \Pr \{B'_i > t\} &\leq \sum_{k=0}^i \Pr \{S_{-k} > t\} = \sum_{k=0}^i \Pr \{S_{-k} - t > 0\} \\ &\leq \sum_{k=0}^i \mathbb{E} [e^{\theta(S_{-k}-t)}] = e^{-\theta t} \sum_{k=0}^i \mathbb{E} [e^{\theta S_{-k}}]. \end{aligned} \quad (2.7)$$

Now assume that the log-moment generation function exists: $\Gamma(\theta) = \lim_{k \rightarrow \infty} \frac{1}{k} \ln \mathbb{E} [e^{\theta(A_k - A_0)}]$. Hence, for large k we have the approximation:

$$\mathbb{E} \left[e^{\theta(A'_0 - A'_{-k})} \right] = \mathbb{E} \left[e^{\theta(A_k - A_0)} \right] \approx e^{k\Gamma(\theta)}. \quad (2.8)$$

Therefore, for large k we have:

$$\begin{aligned} \mathbb{E} \left[e^{\theta S_{-k}} \right] &= \mathbb{E} \left[e^{\theta(A'_0 - A'_{-k} - Ck)} \right] \\ &= e^{-\theta Ck} \mathbb{E} \left[e^{\theta(A'_0 - A'_{-k})} \right] \approx e^{k[\Gamma(\theta) - \theta C]} \end{aligned} \quad (2.9)$$

If $\frac{\Gamma(\theta)}{\theta} < C$, then for large i we have:

$$\Pr \{B'_i > t\} \leq e^{-t\theta} \sum_{k=0}^i \mathbb{E} \left[e^{\theta S_{-k}} \right] = K \cdot e^{-t\theta}. \quad (2.10)$$

Note that the elements of the sum are geometrically decreasing with k , hence we have:

$$\Pr \{B^\infty > t\} \leq K \cdot e^{-t\theta} \quad (2.11)$$

$\frac{\Gamma(\theta)}{\theta}$ is so-called effective bandwidth of the source. As long as the server rate C is larger than this value, we can approximate the tail of the queue length distribution by an exponential function with QoS-exponent θ .

2.2.2 Effective Capacity

The effective capacity model is developed based on the effective bandwidth. In fact, it is strongly associated with the effective bandwidth model. The effective capacity characterizes the (maximum) arrival rate to the buffer due to random service rate/process. Recall that the effective bandwidth is the minimum constant service rate required to guarantee a specified statistical QoS constraint. They actually apply more or less the same set of mathematical techniques to model and measure the link-layer performance of the system from different angles.

The effective capacity is a link-layer capacity which relates the stochastic characterization of the service of the queuing system to the queue-length or delay constraints of

the flow. In the following we continue to provide a brief introduction to the derivation of effective capacity.

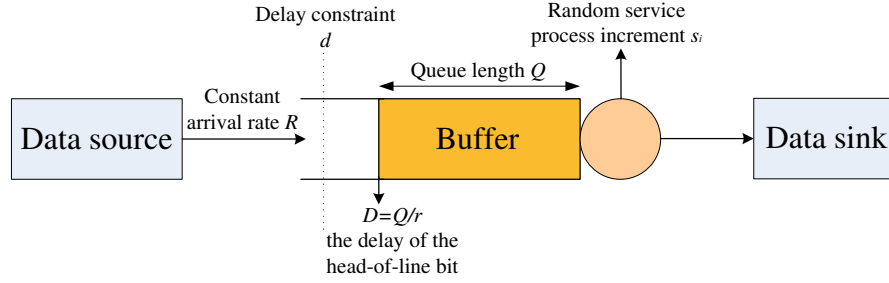


Abbildung 2.2: A queuing model [22] of the effective capacity analysis.

The starting point for the analysis is the queuing model as shown in Figure 2.2. The buffer is at the source/transmitter of a system while the system is expected to support a service with QoS constraints of delay and delay violation probability as $\{d, P_d\}$. At frame i , R bits arrives in the buffer from the data source. At the same time, there is also a amount of bits (due to the channel fading) that effectively leave the queue at frame i . We call these bits as the service process increment at frame i , and denote this increment by s_i . Then, the cumulative service process is $S_i = \sum_{n=0}^i s_n$. Therefore, the queue length is a random variable which is subject to both the arrival data rate and the service process. The random queue length at frame i is given by:

$$Q_i = \max_{0 \leq k \leq i} ((i - k) \cdot R - (S_i - S_k)). \quad (2.12)$$

Then, let us consider that the arrival and service process are stationary. Furthermore, assume that the queue is stable as the average service rate is larger than the average arrival rate. Hence, the random queue length Q_i at time i converges to the steady-state random queue length Q . We are interested in characterizing the long-term statistics $\Pr.\{Q\}$ of the queue length. Similar to the derivation of Equation (2.11), the framework of effective capacity also gives us an upper bound:

$$\Pr\{Q > x\} \approx K \cdot e^{-\theta^* x}, \quad (2.13)$$

where K is the probability that the queue is non-empty and recall that θ is the QoS exponent. It is shown in [22] that the above approximation is more accurate for an x with smaller a value. Moreover, for a constant rate source with r the exponent θ has to fulfill the constraint:

$$r < -\Lambda(-\theta^*)/\theta^*, \quad (2.14)$$

where $\Lambda(\theta)$ is the log-moment generating function of the cumulative service process S_i defined as (assuming the increments to be stationary as well):

$$\Lambda(\theta) = \lim_{i \rightarrow \infty} \frac{1}{i} \log \mathbb{E} \left[e^{\theta \cdot (S_i - S_0)} \right]. \quad (2.15)$$

The ratio $-\Lambda(-\theta)/\theta$ is referred to as effective (service) capacity, as the exponential decay of the queue length distribution in Equation (2.13) is only witnessed if the ratio $-\Lambda(-\theta)/\theta$ is bigger than the constant arrival rate r of the source for some θ^* .

So far, we have only considered the random queue length. Then, denote by D_i the random queuing delay of the head-of-line bit during frame i . If the constant arrival rate at the source is R , with a queue length of $Q = q$, a current delay of the head-of-line bit is given by $D = q/R$. This yields the following approximation for the steady-state delay distribution:

$$\Pr \{D > d\} \leq K \cdot e^{-\theta \cdot R \cdot d}. \quad (2.16)$$

A considerable challenge in determining the effective service capacity is the characterization of the log-moment generating function. If s_i is independent and identically distributed (i.i.d.), a convenient simplification is to obtain the log-moment generating function via the law of large numbers. Hence, the effective service capacity is given by [23]:

$$\begin{aligned} -\frac{\Lambda(-\theta)}{\theta} &= \lim_{i \rightarrow \infty} \frac{1}{i \cdot \theta} \log \mathbb{E} \left[e^{-\theta \cdot s_i} \right] \\ &= \mathbb{E} [s_i] - \frac{\theta}{2} \text{Var} [s_i]. \end{aligned} \quad (2.17)$$

Therefore, the queuing performance of the system is determined by the mean and the variance of the random increment of the service process given by s_i .

The above analysis allows to determine a bound on the maximum outage probability if the (constant) arrival rate is given. In contrast, we can also fix the delay and outage target and derive the maximum arrival rate R^* that can be supported by the random service process. Based on Equation (2.16), by upper bounding K by 1 we have the following relationship:

$$\frac{-\ln(P_d) + \ln(K)}{d} \geq r \cdot \theta \Leftrightarrow R^* \cdot \theta \approx \frac{-\ln(P_d)}{d}, \quad (2.18)$$

where the approximation results from the fact that upper bounding K by one can underestimate the maximum sustainable rate if especially the delay target is quite low.

On the other hand, based on Equation (2.14) and (2.17) we can obtain an upper limit on the QoS exponent given as $\theta_{\max}$:

$$\theta_{\max} = 2 \frac{\mathbb{E}[s_i] - R}{\text{Var}[s_i]}. \quad (2.19)$$

By substituting this expression in Equation (2.18), the maximum arrival rate at the source R^* that can be supported by the random service process for given QoS requirements $\{d, P_d\}$ is obtained:

$$R_{\text{MS}}|_{\{d, P_d\}} = R^* \approx \frac{\mathbb{E}[s_i]}{2} + \frac{1}{2} \sqrt{(\mathbb{E}[s_i])^2 + \frac{2 \ln(P_d)}{d} \cdot \text{Var}[s_i]}. \quad (2.20)$$

In this thesis, we call the maximum source rate as Maximum Sustainable Data Rate (MSDR) and refer to MSDR as a metric for the queuing performance. Based on (2.20), the major challenge for determining the MSDR is to obtain the mean and variance of the increment of the service process for relaying networks.

2.3 Point-to-Point (P2P) Scenario: Infinite Blocklength Regime vs. Finite Blocklength Regime

In electrical engineering and information theory, it is well-known that the Shannon capacity of a complex channel is given by (in bits per channel use) [24]:

$$C(\gamma) = \log_2(1 + \gamma), \quad (2.21)$$

where γ is the SNR of the channel. The Shannon's capacity theory indicates the achievable data rate (in bits per channel use) of a transmission with an arbitrarily small decoding error probability. Moreover, to achieve the arbitrarily small error probability, coding is assumed to be performed using a block with an infinite length. In other words, the Shannon capacity bound is achievable as long as the transmission has an infinite blocklength.

In practice, a system is only possible to have a block with a finite length. A short blocklength makes the error probability of the communication become no longer arbitrarily small. In other words, certain performance loss/gap (in comparison to the Shannon capacity) is introduced by a finite blocklength. This performance loss/gap was initially recognized by Shannon. It is shown in his work [24] that the first-order achievable rate over a P2P Gaussian channel with SNR γ is characterized by :

$$R(m, \varepsilon) \leq C(\gamma) + o(1), \quad (2.22)$$

where the little-o notation intuitively means that $o(1)$ is much smaller than 1, i.e., $o(x)/x \rightarrow 0$ as $x \rightarrow 1$. A key problem of the above equation is that the coding rate does not depend on the error probability and the blocklength. Hence, it does not help us to characterize the transmission behavior and the tradeoff among the coding rate, the error probability and blocklength/delay.

Recently, an attempt [8] to find the dependencies of the coding rate upon blocklength and error probability is to refine the above result and in particular to characterize the details of the $o(1)$ term. The key idea is introducing a factor called channel dispersion

which is defined at the beginning of reference [8] as:

$$V \triangleq \lim_{\varepsilon \rightarrow \infty} \lim_{m \rightarrow \infty} m \left(\frac{C(\gamma) - R(m, \varepsilon)}{Q^{-1}(\varepsilon)} \right), \quad (2.23)$$

where m is the blocklength. In addition, $Q^{-1}(\cdot)$ is the inverse Q-function, where the Q-function is given by:

$$Q(w) = \int_w^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt. \quad (2.24)$$

Based on the channel dispersion, the expression of the $o(1)$ term is improved. As a result, the second-order achievable rates over a P2P Gaussian channel is provided in [8]:

$$R(m, \varepsilon) \leq C(\gamma) - \sqrt{\frac{V}{m}} Q^{-1}(\varepsilon) + o\left(\frac{1}{\sqrt{m}}\right), \quad (2.25)$$

More recently, Equation (2.25) has been further refined to the third-order term:

$$R(m, \varepsilon) \leq C(\gamma) - \sqrt{\frac{V}{m}} Q^{-1}(\varepsilon) + \frac{1}{2} \frac{\log_2 m}{m} + O\left(\frac{1}{m}\right), \quad (2.26)$$

where big-O notation $O(\cdot)$ describes the limiting behavior of a function when the argument tends towards a particular value or infinity, usually in terms of simpler functions. And the improvement from Equation 2.25 with a term of little-o notation to Equation 2.26 with a term of big-O notation is facilitated by the work in [25].

A very important contribution of reference [8] is that it provides an accurate approximation for the maximal channel coding rate:

$$\begin{aligned} R(m, \varepsilon) &\approx C(\gamma) - \sqrt{\frac{V}{m}} Q^{-1}(\varepsilon) + O\left(\frac{\log_2 m}{m}\right), \\ &\approx C(\gamma) - \sqrt{\frac{V}{m}} Q^{-1}(\varepsilon) \end{aligned} \quad (2.27)$$

More importantly, reference [8] derives the channel dispersion for a real AWGN channel, which is given by:

$$V_{\text{real}} = \frac{\log^2 e}{2} \cdot \frac{\gamma(\gamma + 2)}{(\gamma + 1)^2}. \quad (2.28)$$

2.3 Point-to-Point (P2P) Scenario: Infinite Blocklength Regime vs. Finite Blocklength Regime

Moreover, it was reported in [8] that the above approximation of coding rate is accurate for AWGN channels for blocklengths and error probabilities of practical interest.

Under a quasi-static fading channel model, each channel state is assume to be static during a transmission period. Therefore, in each transmission period a quasi-static fading channel with fading coefficient h can be viewed as an AWGN channel with channel gain h^2 . Therefore, the above result with a AWGN channel has been reasonably extended to a complex quasi-static fading channel model in [26]: with an SNR γ the coding rate of a transmission period (in bits per channel use) is given by:

$$r = R(\gamma, \varepsilon, m) \approx C(\gamma) - \sqrt{\frac{V_{\text{comp}}}{m}} Q^{-1}(\varepsilon), \quad (2.29)$$

where the channel dispersion of a complex Gaussian channel is twice the one of a real Gaussian channel: $V_{\text{comp}} = 2V_{\text{real}} = \gamma \frac{\gamma+2}{(1+\gamma)^2} (\log_2 e)^2 = \left(1 - \frac{1}{(1+\gamma)^2}\right) (\log_2 e)^2$.

Then, for a single hop transmission of a transmission period of quasi-static fading channel, with blocklength m and coding rate r , the decoding (block) error probability at the receiver is given by:

$$\varepsilon = P(\gamma, r, m) \approx Q\left(\frac{C(\gamma) - r}{\sqrt{V_{\text{comp}}/m}}\right). \quad (2.30)$$

In this thesis, the relaying performance under both of these two regimes are considered, studied and compared. Moreover, we will use Equation (2.29) and Equation (2.30) to evaluate each single hop of relaying under static channels and quasi-static fading channels.

Comparing Equation (2.29) with Equation (2.21), we observe significant difference regarding the performance models in the infinite blocklength regime and the finite blocklength regime, which is the second part of the right hand side of Equation (2.29). The fundamental difference is actually the consideration of the impact of blocklength as well as decoding error probabilities in the latter case. From a perspective of application, Shannon capacity bounds are still tight for most of the current wireless communication systems. The reason is that these systems typically have significant blocklengths and are more or less robust to slight errors. However, there are also some applications with critical delay and error probability constraints, e.g., industrial control applicati-

ons, military applications and future M2M communications. Wireless communication systems with these applications are required to have significantly short blocklengths. Then, the finite blocklength model (which clearly describes the relationship among error probability, blocklength and capacity) is more appropriate to be applied in the study and design of such systems.

2.4 Maximum Ratio Combining

In wireless communications, maximum ratio combining (MRC) is a method of linear diversity combining. During the combining, each signal branch is multiplied by a weight factor that is proportional to the amplitude of this signal branch. As a result, branches with strong signal are further amplified, while weak signals are attenuated. It is known [27] that MRC is the optimum combining scheme in a noise-only system but is suboptimal in the presence of interference.

In the following we provide a brief introduction to the principle of MRC based on [27, 28]. Denote the original signal transmitted at the source by $\mathbf{x}$. Due to various reasons, e.g., multi-path and the fact that the signal is forwarded by multiple relays or the signal is retransmitted several times, the destination receives multiple signals from different channels/links/paths. The i th received signal is given by:

$$\mathbf{y}_i = h_i \mathbf{x} + \mathbf{n}_i, \quad (2.31)$$

where h_i is the channel and $\mathbf{n}_i$ is the noise with power σ^2 . Hence, the SNR of i th received signal is given by

$$\gamma_i = |\mathbf{x}|^2 \cdot \frac{|h_i|^2}{\sigma^2}. \quad (2.32)$$

Assume the destination in total receives L signals. By linear combining, the output of the combining is given by:

$$\begin{aligned} \mathbf{y} &= \sum_{i=1}^L w_i (h_i \mathbf{x} + \mathbf{n}_i) \\ &= \mathbf{x} \sum_{i=1}^L w_i h_i + \sum_{i=1}^L w_i \mathbf{n}_i, \end{aligned} \quad (2.33)$$

where w_i is the weight of the i th signal. Hence, the combined SNR is given by:

$$\gamma_{\text{comb}} = \frac{\text{E} \left[|\mathbf{x}|^2 \right] \left| \sum_{i=1}^L w_i h_i \right|^2}{\sigma^2 \sum_{i=1}^L |w_i|^2}, \quad (2.34)$$

where $\text{E} \left[|\mathbf{x}|^2 \right]$ is the average/expected value of $|\mathbf{x}|^2$. According to the Cauchy-Schwarz inequality, the following relationship holds:

$$\left| \sum_{i=1}^L w_i h_i \right|^2 \leq \sum_{i=1}^L |w_i|^2 \cdot \sum_{i=1}^L |h_i|^2. \quad (2.35)$$

In particular, the above relationship holds with equality if w_i or h_i is a multiple of the other. As a result, the combined SNR is maximized if we set each weight as a multiple of the corresponding signal amplitude. This is actually why we call this optimal combining as maximum ratio combining. Normally, we set $w_i = h_i$. Then, the combined power to noise ratio (SNR) of MRC is given by:

$$\begin{aligned} \gamma_{\text{MRC}} &= \frac{\text{E} \left[|\mathbf{x}|^2 \right] \left| \sum_{i=1}^L w_i h_i \right|^2}{\sigma^2 \sum_{i=1}^L |w_i|^2} \\ &= \frac{\text{E} \left[|\mathbf{x}|^2 \right]}{\sigma^2} \cdot \sum_{i=1}^L |h_i|^2 = \sum_{i=1}^L \gamma_i, \end{aligned} \quad (2.36)$$

which is actually the sum of SNRs of all combined signals.

It should be mentioned that the key idea of MRC, which is to boost the strong signal components and attenuate the weak (relatively noisy) components, also performs in the Rake receiver for detecting direct-sequence CDMA signals over a dispersive channel.

In this thesis, we apply MRC for combining signals forwarded by multiple relays as well as for combining the retransmitted signal with the previous fail-decoded ones.

2.5 Convex/Concave and Quasi-convex/Quasi-Concave Functions

Optimization problems are important in the study of communication systems, e.g., in information theory, wireless network power control, and maximizing network utilization. In particular, these problems can be significantly facilitated if the objective functions can be proved as convex/concave or quasi-convex/quasi-concave functions. In this section, we introduce the basic definitions and concepts for convex, concave and quasi-convex quasi-concave functions [29].

2.5.1 Convex and Concave Functions

A set $\mathcal{C} \subseteq \mathbb{R}^q, q \in \mathbb{N}$, is convex if for all $t_1, t_2 \in \mathcal{C}$ and $\lambda \in [0, 1]$:

$$\lambda t_1 + (1 - \lambda) t_2 \in \mathcal{C}. \quad (2.37)$$

A function f is convex if its domain $\mathcal{C}$ is convex and for all $t_1, t_2 \in \mathcal{C}$ and $\lambda \in [0, 1]$ the following inequality holds:

$$f(\lambda t_1 + (1 - \lambda) t_2) \leq \lambda f(t_1) + (1 - \lambda) f(t_2). \quad (2.38)$$

A function f is concave if $-f$ is convex. A concave function with domain $\mathcal{C}$ satisfies:

$$f(\lambda t_1 + (1 - \lambda) t_2) \geq \lambda f(t_1) + (1 - \lambda) f(t_2). \quad (2.39)$$

Characterization of Convex Functions

Consider a non-empty and open set $\mathcal{C} \subseteq \mathbb{R}^n$. the gradient of a function f is given by: $\nabla f(\mathbf{x}) = \left(\frac{\partial f(\mathbf{x})}{\partial x_1}, \frac{\partial f(\mathbf{x})}{\partial x_2}, \dots, \frac{\partial f(\mathbf{x})}{\partial x_n} \right)$. Moreover, the Hessian of f is given by $\nabla^2 f(x) \in \mathbb{C}^n$ $\nabla^2 f(x)_{ij} = \frac{\partial^2 f(x)}{\partial x_i \partial x_j}, i, j, \dots, n$ Then, we call a function $f: \mathcal{C} \rightarrow \mathbb{R}$ differentiable on $\mathcal{C}$ if $\nabla f(\mathbf{x})$ exists at each $x \in \mathcal{C}$. We call f twice differentiable on $\mathcal{C}$ if $\nabla^2 f(x) \in \mathbb{C}^n$ exists at each $x \in \mathcal{C}$. There are several important characteristics of convex functions:

1. First-order condition: Let $f: \mathcal{C} \rightarrow \mathbb{R}$ be differentiable. Then f is convex if and only if $f(t_1) \geq f(t_2) + \nabla f(t_2)^T (t_1 - t_2)$, for all $t_1, t_2 \in \mathcal{C}$.
2. Second-order condition: Let $f: \mathcal{C} \rightarrow \mathbb{R}$ be twice differentiable. Then f is convex if and only if $\nabla^2 f(t) \geq 0$.
3. Minimizing a convex function: Let $f: \mathcal{C} \rightarrow \mathbb{R}$ be convex and differentiable. Then, the following three assertions are equivalent:
 - t^* is a global minimum.
 - t^* is a local minimum.
 - t^* is a critical point, i.e., $\nabla f(t^*) = 0$.

Preservation of Convexity

In addition to the above first-order and second-order conditions, another important way to prove a function to be convex is to show that this function is obtained from simple convex functions by operations that preserve convexity. There are several operations that preserve convexity:

1. Nonnegative weighted sums: Let $f_1, \dots, f_m: \mathbb{R} \rightarrow \mathbb{R}$ be convex and $w_1, w_2, \dots, w_m$ are nonnegative. Then, the sum $g = w_1 f_1 + w_2 f_2 + \dots w_m f_m$ is convex.
2. Composition: Consider $g: \mathbb{R}^n \rightarrow \mathbb{R}$ and $h: \mathbb{R} \rightarrow \mathbb{R}$. Then the composition $f(x) = h(g(x))$ is convex, if one of the following assertions holds
 - g is convex and h is convex and increasing.
 - g is concave and h is convex and decreasing.
3. Maximum: Let $f_1, \dots, f_k: \mathcal{C} \rightarrow \mathbb{R}$ be convex. Then $f: \mathcal{C} \rightarrow \mathbb{R}$ defined by: $f(t) = \max\{f_1(t), \dots, f_k(t)\}$ is convex.

2.5.2 Quasi-Convex and Quasi-Concave Functions

A function f is quasi-convex if its domain $\mathcal{C}$ is convex and for all $t_1, t_2 \in \mathcal{C}$ and $\lambda \in [0, 1]$ the following inequality holds:

$$f(\lambda t_1 + (1 - \lambda) t_2) \leq \max\{f(t_1), f(t_2)\}. \quad (2.40)$$

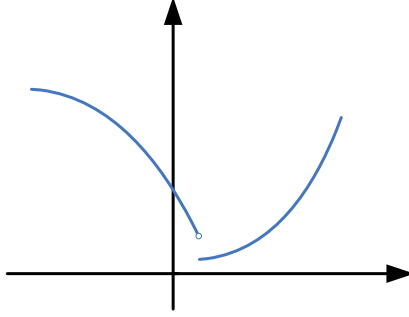


Abbildung 2.3: A function is quasi-convex but neither convex nor continuous.

Informally, saying $f(t)$ is quasi-convex in t means $f(t)$ is first decreasing and then increasing in t . A convex function is definitely a quasi-convex function. However, as shown in Figure 2.3 a quasi-convex function is not necessarily a convex function.

Similarly, a function f is quasi-concave if its domain $\mathcal{C}$ is convex and for all $t_1, t_2 \in \mathcal{C}$ and $\lambda \in [0, 1]$ the following inequality holds:

$$f(\lambda t_1 + (1 - \lambda) t_2) \geq \min \{f(t_1), f(t_2)\}. \quad (2.41)$$

Informally, saying $f(t)$ is quasi-convex in t means $f(t)$ is first increasing and then decreasing in t .

Characterization of Quasi-Convex Functions

Here we prove some commonly-used characteristics of quasi-convex/quasi-concave functions:

1. A convex function is quasi-convex.
2. A concave function is possible to be quasi-convex. For example, a monotonic concave function is quasi-convex. In fact, a monotonic function is both quasi-convex and quasi-concave.
3. A quasi-convex/quasi-concave function is not necessarily continuous.
4. If $f(t)$ $g(w)$ are positive convex decreasing functions, then $f(t)g(w)$ is quasi-convex in (t, w) .

Preservation of Quasi-Convexity

Similar to convex/concave functions, there are also several operations that preserve quasi-convexity. For example, the non-negative weighted maximum of quasi-convex functions is quasi-convex, i.e., $f = \max \{w_1 f_1, \dots, w_n f_n\}$ with w_i non-negative.

It should be mentioned that there are some operations preserving convexity but not preserving quasi-convexity:

1. The sum of quasi-convex functions defined on the same domain need not be quasi-convex: i.e., if $f(t)$, $g(t)$ are quasi-convex, then the sum function $h(t) = f(t) + g(t)$ need not be quasi-convex.
2. The sum of quasi-convex functions defined on different domains, i.e., if $f(t)$, $g(w)$ are quasi-convex, $h(t, w) = f(t) + g(w)$ need not be quasi-convex.

Both convex/concave and quasi-convex/concave functions play important roles in studies and applications in communication systems as it is easy to find the global optimal solution by Descent Methods. In this thesis, we will show the relationship between relaying performance and some factors based on the above concave and quasi-concave functions.

3 Related Work and Scope of the Thesis

The field of relaying has been under investigation for several decades so far. In the following, we present the related work of relaying regarding the infinite blocklength regime as well as the finite blocklength regime, respectively.

3.1 Relaying in the Infinite Blocklength Regime: State of the Art

In the infinite blocklength regime, the analysis and design of cooperative transmission protocols for wireless networks has attracted a lot of interest during the passed two decades. Of particular interest are two-hop channels where a relay terminal assists in the communication between a source terminal and a destination terminal. By deploying additional transceiver(s), the destination likely receives a much stronger signal forwarded by the relay(s) in comparison to the direct transmission by the source. This leads to an improvement in the coverage and the transmission reliability of the link [9, 10].

Based on these very mature relaying techniques, more recently cooperative communications has attracted a lot of research interest with respect to the infinite blocklength regime. In cooperative communications the fundamental assumption is that there are multiple potential relays. Hence, by spending some control overhead, the best positioned relay or a set of best positioned relays are identified to (jointly) forward the signal to the destination. This boosts the system performance as it exploits cooperative diversity [30] which is a form of spatial/multi-user diversity. Many different techniques have been proposed how to specifically exploit this cooperative diversity. If for example

the source has instantaneous CSI, [31] proposes to apply distributed space-time block coding schemes at the source and relays. The paper shows that this leads to a higher spectral efficiency in comparison to pure relaying. Under the same assumption of CSI at the source, [32, 33] let each relay transmit separately for a certain time fraction. By optimizing the transmission duration of each relay, the outage probability [32] and capacity [33] can then be improved in comparison to pure relaying without cooperation. In general, cooperative communications is well known to effectively increase the throughput and reliability in case that multiple relays are present [4, 34]. However, this fundamentally comes at the price of an increased complexity, energy consumption and hardware cost.

On top of the above mentioned cooperative communications, there is the possibility to further improve the reliability by combining the cooperation principle with automatic repeat request (ARQ) protocol mechanisms. These combined schemes are referred to as cooperative-ARQ (C-ARQ) protocols. A C-ARQ protocol lets relay(s) retransmit the data packet to the destination when the initial transmission fails instead of having the source retransmit the packet. Hence, the reliability of the system is improved due to cooperative diversity. Furthermore, if the destination can combine the received signals of the initial transmission and retransmissions, the reliability is additionally boosted by the combined channel qualities. It has been shown over the last few years [35, 36, 37, 38, 39, 40, 41, 42] that - from a pure physical-layer perspective - C-ARQ protocols have an outstanding throughput and reliability performance in comparison to pure relaying and cooperative communications.

Nevertheless, as with cooperative communications in general, C-ARQ protocols lead to a higher system complexity and energy requirement. Few works attempt to address this tradeoff. For instance, from a pure physical-layer perspective [36] shows that C-ARQ has higher energy efficiency than the non-cooperative ARQ schemes under certain channel conditions. In addition, [43, 44] show that the energy efficiency of C-ARQ has a strong dependence on the number of retransmissions. Apart from the additional energy consumption, C-ARQ protocols also lead to a more complex framing-structure which likely introduces additional stochastic delays to the information flow from the source to the destination. While studies exist which have been investigating the delay of C-ARQ protocols [45, 41, 46], they are typically limited to the physical-layer and

do not take higher-layer queuing effects into account. However, a joint delay analysis is particularly relevant for delay-sensitive higher-layer applications. It typically comes with a limited arrival rate (instead of the often studied full-buffer assumption) but also with constraints regarding the reception delay and the violation probability of this target delay (for instance, for voice flows a typical scenario is to keep the delay below 200 ms while the delay violation probability is around 5% depending on the used codec). In a word, C-ARQ protocols actually introduce a tripartite tradeoff among throughput, delay and energy consumption. However, to date this tradeoff is only understood from a physical-layer perspective and a more complete analysis of this tradeoff with respect to higher-layer performance is missing. In Figure 3.1, we in general summarize the related work from a perspective of research topic (without taking into account the difference between system scenarios/assumptions). In comparison to the rest, research issues regarding higher-layer performance of C-ARQ protocols are quite open.

	Pure relaying	Cooperative relaying	C-ARQ
Physical-layer performance	Well-studied	Well-studied	Well-studied
Higher-layer performance	Well-studied	Well-studied	Open

Well-studied
 Open

Abbildung 3.1: A general summary of related work of relaying in the infinite block-length regime from a perspective of research topic.

On the other hand, existing publications regarding the higher-layer performance of either C-ARQ or even pure relaying (without retransmission) are generally with following assumptions: (1) Channel knowledge is perfect [47, 48, 49, 50, 51, 52, 53]. (2) ARQ feedback overhead is not considered [47, 48, 49, 50, 51, 52, 53, 54]. (3) Do not make use of these fail-decoded signals in the initial transmission and previous retransmissions [48, 49, 50, 51, 52, 53, 54], even these signals are already received by the destination. However, theses assumptions are over-optimistic in a practical system. In fact, the picture in practice is quite the opposite. First, the CSI feedback is imperfect (e.g., only the average CSI is available at the transmitter). Second, ARQ feedback costs

time and energy. Third, it is more reasonable to make use of these fail-decoded signals, e.g., by applying MRC.

To the best of our knowledge, in the related papers few of them consider the above practical assumptions. We show the summary of related researches with regarding two of the assumptions (average CSI and applying MRC) in Figure 3.2. All these issues with these two assumptions are open from the higher-layer perspective. Furthermore, if consider all the three assumptions (including ARQ overhead), all the issues in Figure 3.2 are open, no matter from a physical-layer perspective or a higher-layer perspective.

	Perfect CSI	Average CSI
C-ARQ Without MRC		
C-ARQ With MRC		

Well-studied
 Open

Abbildung 3.2: A specific summary (from a perspective of system assumption) of related work regarding higher-layer performance of relaying.

Summarizing all this, the performance of C-ARQ systems is not well understood, both from the theoretical point of view (higher-layer capacity, energy efficiency and in particular the tripartite tradeoff among throughput, delay and energy consumption) and the practical point of view (considering more realistic scenarios).

3.2 Relaying in the Finite Blocklength Regime: State of the Art

As discussed in the previous section, in the infinite blocklength regime relaying [4, 5, 6] is well known as an efficient way to mitigate wireless fading by exploiting spatial diversity. Specifically, two-hop decode-and-forward (DF) relaying protocols significantly improve the capacity and quality of service [55, 56, 57, 58]. However, all the above studies of the advantages of relaying are under the ideal assumption of communicating arbitrarily reliably at Shannon's channel capacity.

For communication with arbitrarily small decoding error probability, coding is assumed to be performed using a block with an infinite length. If the codeword is restricted to a reasonable size, i.e., to a finite blocklength, the error probability of the communication becomes no longer negligible. Hence, in the finite blocklength regime it is essential to consider the error probability while investigating the communication performance. Taking the error probability into account, [8] identifies a tight bound of coding rate of a single-hop transmission system. The authors show that the performance loss due to a finite blocklength is considerable and becomes more significant when the blocklength is relatively short.

For a two-hop relaying system (with equal time division), the blocklength of direct transmission is twice as long as the blocklength in each hop of relaying. Hence, one could suspect that relaying pays off less in the finite blocklength regime compared to the infinite blocklength regime because of halving the blocklength. However, to the best of our knowledge, the performance of relaying in the finite blocklength regime has not been studied so far.

3.3 Scope of the Thesis

From Section 3.1 and 3.2, we learn that the relaying performance is not well-understood. On the one hand, most existing research on investigating the relaying performance over wireless fading channels typically rely on simplifying assumptions, such as perfect CSI and infinite blocklength. On the other hand, these existing researches mainly focuses on the physical-layer performance of relaying while a systematical higher-layer performance model of relaying is missing. In this thesis, the general scope is to develop extensive and systematical investigation models for relaying. Both the physical-layer and the higher-layer performance models of relaying are introduced while multiple performance metrics are taken into account, e.g., reliability, capacity and energy efficiency. Moreover, we study the relaying performance under various scenarios with different assumptions regarding CSI and blocklength (e.g., perfect or imperfect CSI and finite or infinite blocklengths). As shown in Figure 1.1 and discussed in Chapter 1, we first investigate the relaying performance in the infinite blocklength regime as well as in the finite blocklength regime. Afterwards, we look into the performance

difference in the two regimes regarding relaying. In the following, we discuss the scope of the thesis in detail.

Firstly, we show in Figure 3.3 a set of metrics we considered in our investigation and evaluation models: Reliability (e.g., outage probability and error probability), throughput and capacity as well as energy efficiency. Moreover, both physical-layer performance (throughput/blocklength-limited capacity and energy efficiency) and higher-layer performance (effective capacity and effective energy efficiency) of relaying networks are studied.

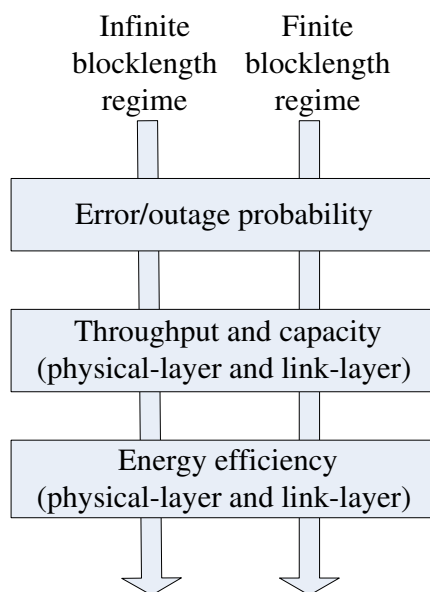


Abbildung 3.3: Methodology of this thesis.

From a scenario perspective, we study relaying systems under different topologies and CSI assumptions. As shown in Figure 3.4, these scenarios (differ in the topologies and CSI assumptions) are considered under both the infinite blocklength regime and the finite blocklength regime. In the figure, blocks with different colors are associated with our research methods which are either analytical and simulative investigations or pure simulative investigations. Our theoretical and simulative investigations (with blue blocks) are mainly regarding these open issues in the literature. These investigations regarding the infinite blocklength regime and the finite blocklength regime are presented in Chapter 4 and Chapter 5, respectively. Moreover, in Chapter 6 we finally

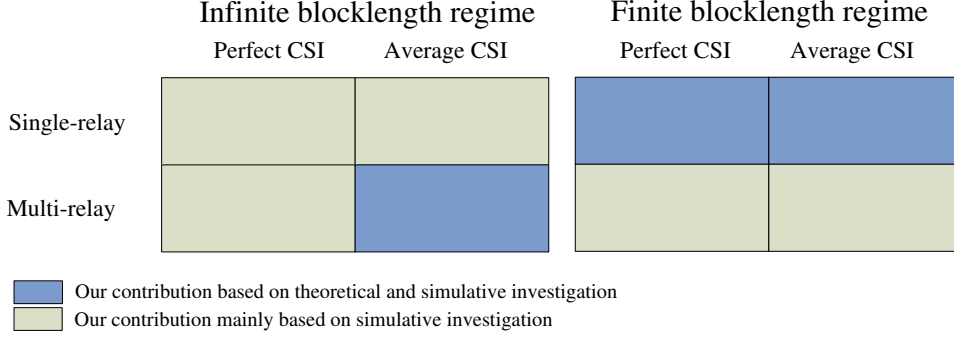


Abbildung 3.4: The scope and the contribution of this thesis (from a scenario perspective).

compare the relaying performance/behaviors in the infinite blocklength regime and in the finite blocklength regime. This comparison is mainly based on simulations under various scenarios.

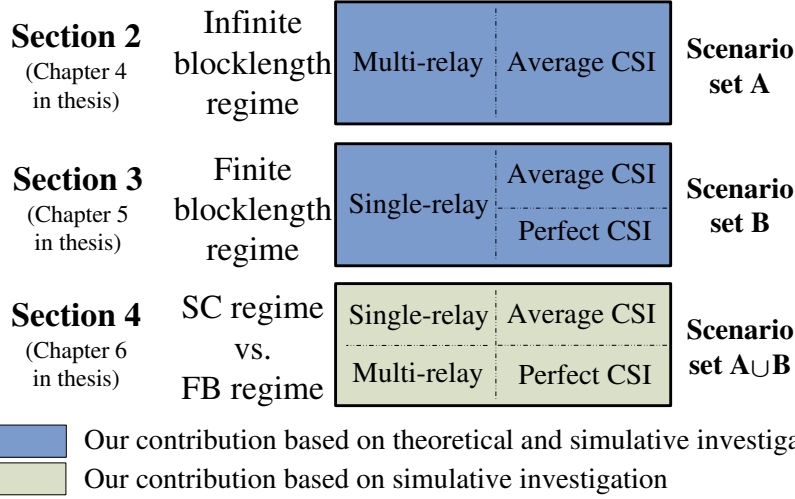


Abbildung 3.5: Scopes of Chapters 4, 5 and 6 which involve all the research work in this thesis.

Summarizing, Chapters 4, 5 and 6 are core chapters of this thesis. We show in Figure 3.5 scopes of these three chapters and the relationship among them. We analytically as well as simulatively study the relaying performance under corresponding specific regime and specific scenarios in Chapter 4 and Chapter 5, respectively. Then,

3 Related Work and Scope of the Thesis

in Chapter 6 we compare the relaying performance/behaviors in these two regimes by simulations while all the scenarios in Chapter 4 and Chapter 5 are considered.

4 Relaying Performance in the Infinite Blocklength Regime

In this chapter, we systematically investigate the relaying performance model in the infinite blocklength regime where blocklengths are assumed to be infinite. The relay-assisted retransmissions (C-ARQ) are considered in our investigation model to analyze the delay, throughput as well as the higher-layer performance. Different from existing research [47, 48, 49, 50, 51, 52, 53, 54], we consider a multi-relay scenario with following assumptions:

- The channel knowledge is not perfect. i.e., only the average CSI is available at the source.
- ARQ feedback overhead of time and energy is considered.
- In the retransmission process of a packet, fail-decoded signals of the initial transmission and previous retransmissions are combined with the new retransmitted one (by MRC) before the destination decoding it.

Under above assumptions, the performance of C-ARQ protocols is derived. Firstly, we investigate the outage probability and the transmission delay distribution. Secondly, the (physical-layer) throughput as well as the (higher-layer) effective capacity are derived.. Thirdly, we further evaluate the energy efficiency by considering the ratio of either throughput or effective capacity to energy consumption of the considered C-ARQ protocols. In the following, we refer to these ratios as physical-layer energy efficiency (PEE) and effective energy efficiency (EEE).

Based on these performance metrics, we study four C-ARQ protocols with MRC over all the received signals of the same packet. These protocols represent different choices with respect to the integration of the direct link into the protocol as well as

the dynamic choice of relays for forwarding. We mainly consider these protocols with MRC while at the same time a C-ARQ protocol without MRC is also presented as a performance reference.

The contents of this chapter are as follows. We first describe the system model in Section 4.1. Subsequently, in Section 4.2 we discuss the methodology of our systematical analysis which considers various metrics. Then, in Sections 4.3, 4.4 and 4.5, the outage probability, capacity and energy efficiency are theoretically investigated, respectively. At the same time, these analytical models are validated by simulations in their corresponding sections. In Section 4.6, we continue to evaluate these C-ARQ protocols to derive guidelines for the design of efficient C-ARQ networks. Finally, we conclude the work of this chapter in 4.7.

4.1 System Model

In this section, we first describe the system, and then present models of channel behavior and energy consumption. At last, investigated C-ARQ protocols are introduced.

4.1.1 General System Description

We consider a DF relaying scenario which contains a source S , a destination D and J DF relays $R_1, R_2, \dots, R_J$ as schematically shown in Fig. 5.1. Distances between relays are assumed to be much shorter than the distance either from the source to relays or from relays to the destination.

The entire system operates in a slotted fashion where time is divided into frames of length T_f . Depending on the exact operational rule of the protocol, these frames are either broadcasting frames, forwarding/relaying frames or feedback frames with acknowledgment (ACK) or negative-ACK (NACK). In a broadcasting frame, the source takes a fixed amount of data ρ out of its buffer and then broadcasts this packet to the destination and relays. Due to the random channel behavior, a random amount of relays is able to decode the broadcasted packet successfully and will forward the packet to the destination in the subsequent forwarding frame. Once the destination received

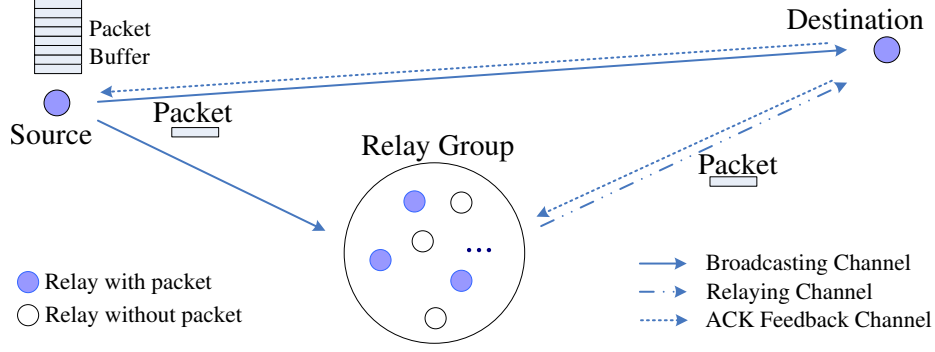


Abbildung 4.1: Example of the considered multiple relay system scenario.

the packet (either successfully or not) it gives feedback through an ACK/NACK frame. We assume this feedback to be error-free. Due to C-ARQ protocols, a packet is retransmitted by relays if the initial transmission fails. The destination is assumed to apply maximal ratio combining (MRC) during the retransmissions¹ (the combination of MRC and ARQ is also known as Hybrid-ARQ II [59]). This results in a significantly stronger signal at the destination.

At the source, a constant data flow originates with arrival rate r bits per frame. The transmission of this data is subject to QoS requirements $\{d, P_d\}$ (e.g., for audio or video applications) where d stands for a maximum tolerable delay and P_d denotes the delay violation probability. Data that cannot be transmitted immediately at the source is put into a first-in-first-out buffer. Note finally that relays do not queue data, they only store the currently transmitted data packet.

4.1.2 Channel, SNR and Energy Consumption Models

Channels are considered to experience Rayleigh blockfading where they are invariant during one frame but vary independently from one frame to the next. We denote the channel gains from the source to the destination, from the source to relay j and from relay j to the destination in frame i by $z_{S,D,i}$, $z_{S,j,i}$ and $z_{j,D,i}$ ($j = 1, 2, \dots, J$). The corresponding average channel gains are $\bar{z}_{S,D}$, $\bar{z}_{S,j}$ and $\bar{z}_{j,D}$. In addition, we denote p_{tx}

¹ Compared with the SNR of a relaying frame, the SNR of a direct link is significantly lower and can be treated as negligible. To simplify the derivation, we only consider combining the signals transmitted/retransmitted by relays.

as the transmit power at the source and each relay, and denote p_{fb} as the transmit power at the destination for feedback. The noise power is denoted by σ^2 . Also, we assume no interference to be present. Hence, the instantaneous signal-to-noise ratio (SNR) from the source to the destination in frame i is $\gamma_{\text{S,D},i} = p_{\text{tx}} z_{\text{S,D},i} / \sigma^2$. Similarly, the instantaneous SNRs in frame i from the source to relay j and from relay j to the destination are $\gamma_{\text{S,j},i} = p_{\text{tx}} z_{\text{S,j},i} / \sigma^2$ and $\gamma_{\text{j,D},i} = p_{\text{tx}} z_{\text{j,D},i} / \sigma^2$. As multiple relays forward the same packet in a forwarding frame, the destination obtains a joint instantaneous SNR as the sum of the instantaneous SNRs of links from these relays as $\gamma_{\xi_i, \text{D}, i} = \sum_{R_j \in \xi_i} \gamma_{\text{j,D}, i}$, where ξ_i is the set of relays which forwards/reforwards the packet during frame i .

Moreover, by applying MRC over the initially transmitted and retransmitted signals a cumulative joint SNR is obtained at the destination, which refers to the sum of the joint instantaneous SNRs in the initial transmission and retransmissions. This cumulative joint SNR is given by $\gamma_{\text{C-ARQ}} = \sum_{i \in \pi} \gamma_{\xi_i, \text{D}, i}$, where π is a set of indices of frames during which the packet is forwarded and re-forwarded by relays.

Recall that we assume only the average CSI to be available at the source and each relay. To account for this, we consider an error model for the transmitted data at relays and the destination. Given an SNR γ , at most $T_f B \log_2(1 + \gamma)$ bits can be conveyed correctly per frame, where B is the bandwidth. Hence, a currently transmitted/retransmitted packet of size ρ is successfully received if the instantaneous/cumulative SNR is above the threshold $\gamma^* = 2^{\rho/BT_f} - 1$. On the other hand, if the instantaneous/cumulative SNR is lower than the threshold, an outage occurs which leads to a retransmission by the corresponding C-ARQ protocol.

Regarding the energy consumption, we adopt the energy consumption model introduced in [60, 61]. In general, we consider two effects that contribute differently to the energy consumption. The first one is the energy consumption due to transmission. The other one is the basic energy consumption which is spent for signal processing, battery backup, cite cooling and so on. We assume that the basic energy consumption at the source, each relay and the destination are the same while the corresponding power is denoted by p_c .

4.1.3 Multi-Relay C-ARQ Protocols

C-ARQ is a link-level ARQ protocol that exploits relay(s) for retransmission. In particular, a multi-relay C-ARQ protocol requires multiple relays to retransmit a packet if the initial transmission or the previous retransmission fails. In this work, we study four multi-relay C-ARQ protocols which have different behaviors regarding the trade-off between throughput, delay and energy consumption. The protocols differ in the way they deal with the initial transmission, and in the way they compose the group of retransmission relays.

4.1.3.1 Proactive/Reactive Behavior

A proactive protocol activates relays to proactively forward the packet in the initial transmission as well as retransmissions. Hence, the initial transmission of a proactive protocol is a relay assisted transmission consisting of a broadcasting frame and a relaying frame. Contrarily, a reactive protocol uses a direct link transmission as the initial transmission and requires relays to forward the packet only if the direct link transmission fails. In other words, reactive protocols let the destination try to decode the packet by a one-frame direct link transmission. If the direct link transmission succeeds, as shown at the top right of Fig. 4.2 the transmission delay of this packet is only two frames while saving transmit power at the relays. However, if the direct link transmission fails, the reactive protocol introduces an additional delay and consumes additional energy.

4.1.3.2 Static/Dynamic Relaying Behavior

The second category of C-ARQ protocols concerns the number of relays participating during the retransmission frames. After the source has broadcasted a packet, not necessarily all the relays decode the packet successfully due to channel fading. Let us call these relays which decode the packet from the source correctly as active relay set and call the rest as passive relay set. Obviously, only active relays are able to participate in the first retransmission. As shown in Fig. 4.3 (left), static protocols only exploit these active relays for retransmission and let the passive relays idle. Hence, the number of active relays under a static protocol is fixed during the retransmission process of a

4 Relaying Performance in the Infinite Blocklength Regime

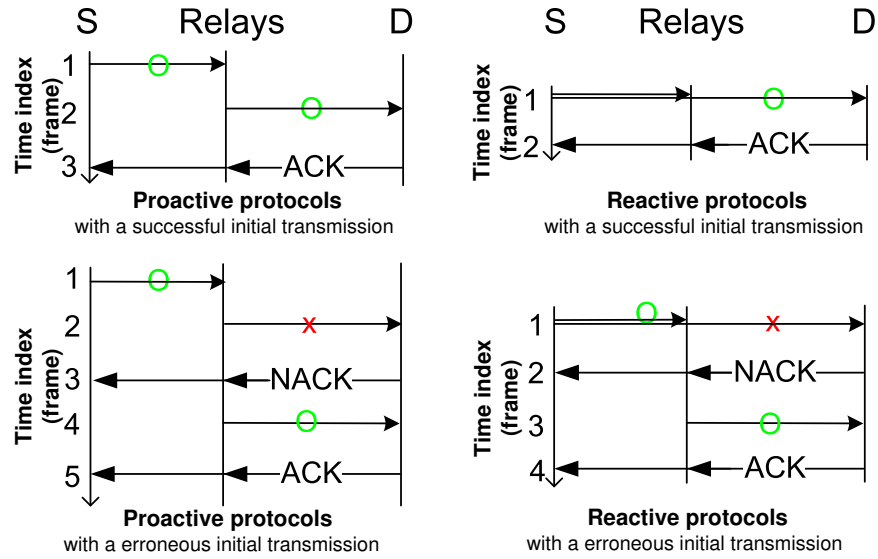


Abbildung 4.2: The difference between proactive protocols and reactive protocols.

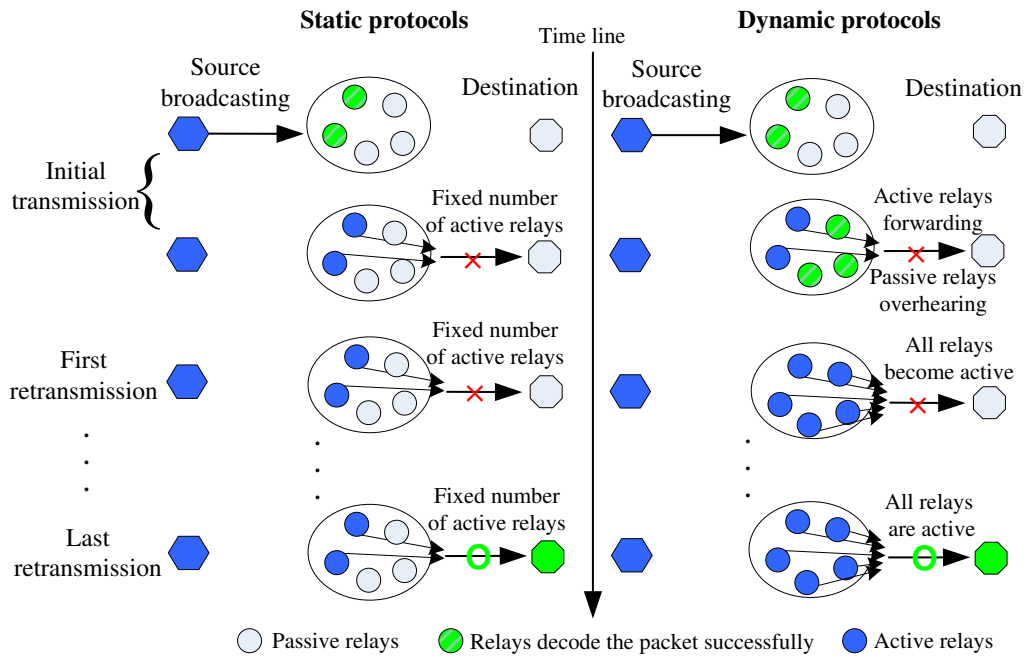


Abbildung 4.3: The difference between static protocols and dynamic protocols.

packet. On the other hand, dynamic protocols require all the passive relays to overhear the packet when the active relays forward the packet to the destination. Therefore,

more relays might join the set of the active relays later on in the retransmission rounds, as shown in Fig. 4.3 (right). Regarding the trade-off between throughput, delay and energy consumption, a dynamic protocol likely has a higher throughput and shorter latency while consuming more energy due to a higher number of (initially passive) relays participating in the retransmission process after a while.

In the rest of this chapter, we will investigate the performance of four C-ARQ protocols: dynamic proactive (DP) protocol, dynamic reactive (DR) protocol, static proactive (SP) protocol and static reactive (SR) protocol, which are actually combinations of these two presented protocol categories.

4.2 Methodology

For a complex channel with an SNR γ , based on the Shannon capacity theory at most $N \log_2(1 + \gamma)$ bits can be conveyed correctly per N channel use. If the transmitter have the perfect instantaneous CSI/SNR of the channel, it is able to choose a suitable packet size $\rho \leq N \log_2(1 + \gamma)$ to ensure the success of the transmission. However, in practice it is more likely that the CSI at the transmitter is not perfect, e.g., the transmitter only has the average CSI. Then, if the transmitter sends a packet with size ρ , to make the receiver decode the packet correctly the channel SNR is required to be higher than a threshold $\gamma^* = 2^{\rho/N} - 1$. Unfortunately, owing to the random channel fading, it is possible that the instantaneous channel SNR performs lower than the threshold. This results in an outage and the probability of this (instantaneous channel SNR is lower than the threshold) is called outage probability.

Note that the decoding error probability due to noise is assumed to be arbitrarily small under infinite blocklength regime. Hence, the reliability of a transmission in the infinite blocklength regime is totally subject to the outage probability. Besides the reliability, the throughput and higher-layer performance of the system are influenced by the outage probability, too. A high outage probability means a significant percentage of incorrect transmissions, which leads to a low (average) throughput. Moreover, the packet should be retransmitted after an outage. This introduces an additional delay to the transmitted packet (which influences the higher-layer performance) and consumes more energy. Summarizing, the outage probability is a very important metric in a relay-

assisted system with retransmissions, as it directly or indirectly affects the physical-layer performance as well as the higher-layer performance.

Under these considered C-ARQ protocols with DF relays, to make the destination decode a packet correctly a random number (non-negative integer) of retransmissions is required due to the random channel fading. This leads to a random transmission delay with length τ (frames). The distribution of τ is a very important link in the analysis in this chapter. Based on it, we can obtain the average energy consumption of successfully transmitting/retransmitting a packet. In addition, with a given packet size we can calculate the throughput and the effective capacity (by determining the mean and variance of the corresponding service process increments) based on the distribution of τ . Then, we have physical-layer and higher-layer energy efficiencies based on the throughput, the effective capacity as well as the energy consumption.

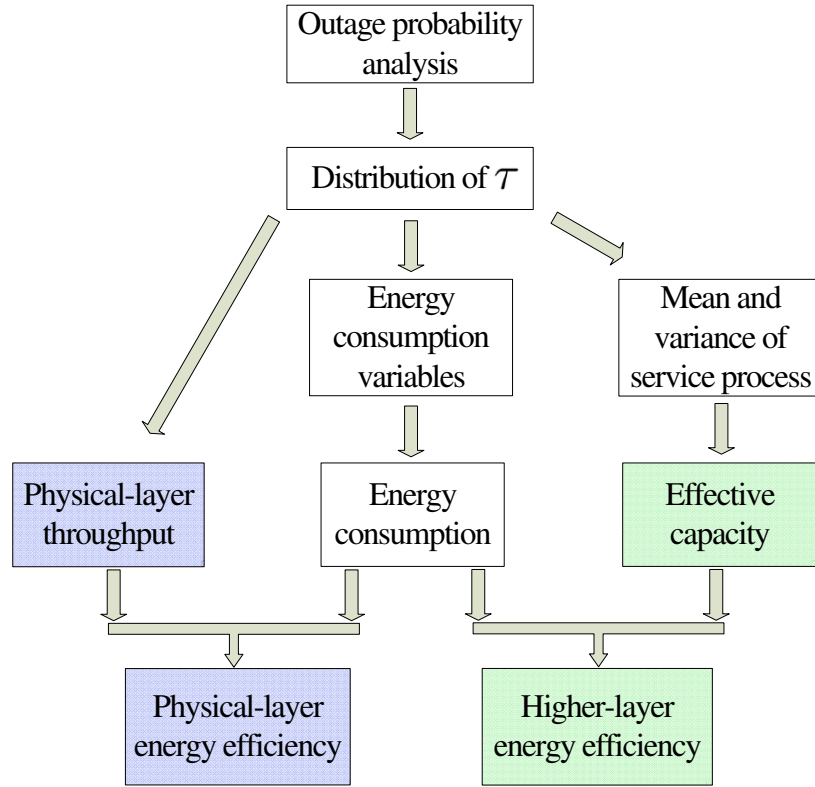


Abbildung 4.4: Methodology of developing investigation models of relaying in the infinite blocklength regime.

The above methodology is also summarized in Figure 4.4. Based on this methodology, in the following sections we first study the outage probability as well as the distribution of τ of C-ARQ protocols. Subsequently, making use of the distribution of τ , we investigate the throughput and effective capacity. At last, we analyze the energy consumption model based on which physical-layer and higher-layer energy efficiencies are further derived, evaluated and improved.

4.3 Reliability Issue in the Infinite Blocklength Regime – Outage Probability

In this section, we first study the outage probability of the multi-relay-assisted transmission (without retransmission mechanism). Then, based on this the outage probability of multi-relay-assisted retransmissions as well as the distribution of the transmission delay τ are investigated.

4.3.1 Multi-DF-Relay Assisted Transmissions

As briefly discussed in Section 4.2, (in the infinite blocklength regime) for a single-hop link if the transmitter sends a packet with size ρ , to make the receiver decode the packet correctly the channel SNR is required to be higher than a threshold $\gamma^* = 2^{\rho/N} - 1$. For a Raleigh-fading channel, the probability that the instantaneous channel SNR is lower than the threshold is given by:

$$P = 1 - \exp\left(-\gamma^* \sigma^2 / \bar{z} p_{\text{tx}}\right), \quad (4.1)$$

where $\bar{z}$, p_{tx} and σ^2 are the average channel gain, the transmit power and the noise power.

Now let us consider our multi-relay system as shown in Figure 5.1. We denote the average channel gain of the link from source S to relay j by $\bar{z}_{\text{S},j}$. Therefore, the outage probability of the link is given by:

$$P_{\text{S},j} = 1 - \exp\left(-\gamma^* \sigma^2 / \bar{z}_{\text{S},j} p_{\text{tx}}\right). \quad (4.2)$$

4 Relaying Performance in the Infinite Blocklength Regime

Due to random channel fading, it is not necessarily that all the relays decode the packet successfully (i.e., no outage happens). Denote the active relay set as Ω . This set consists of all the relays which decode the packet successfully from the source. Hence, the probability mass function (PMF) of Ω is given by:

$$P(\Omega) = \prod_{j \notin \Omega} P_{S,j} \prod_{j \in \Omega} (1 - P_{S,j}) . \quad (4.3)$$

From this we immediately obtain the outage probability of the relaying frame with forwarding relay set as Ω :

$$P_{\Omega,D} = 1 - \exp \left(-\gamma^* \sigma^2 \left/ \sum_{j \in \Omega} \bar{z}_{j,D} p_{\text{tx}} \right. \right) . \quad (4.4)$$

However, this fairly general case is hard to analyze further since the Ω is unpredictable.

Recall that we assumed the distance among relays to be significantly small compared to the distance from each relay to the source or to the destination. We therefore propose to consider a simplification in the topology to make all the relays have same value of pathloss in broadcasting frame. Likewise, we also simplify the model to assume pathlosses of links from different relays to the destination to be also the same. This topology simplification is very good, as it makes the outage probabilities of all the links from the source to different relays be the same, i.e., $P_{S,j} = P_1, \forall j$. This results in that the number of active relays becomes Binomial-distributed (recall that we assume the channel fading of links from the source to different relays to be i.i.d.). We denoted by n the number of active relays. In other words, in total J relays deployed in the system, n relays decode the packet from the source successfully. If n equals 0, this means that all relays fail in decoding and the outage probability equals 1. Otherwise, n is a positive integer. Then, the probability of this is given based on the characteristic of the binomial distribution by:

$$P_B(n, J, P_1) = \binom{n}{J} (1 - P_1)^n P_1^{J-n} , \quad (4.5)$$

where P_1 is the outage probability of a source-relay link.

Let us move on to the subsequent relaying phase. Due to similar distances between the relays and destination, the sum of the fading signals from all the relays in the active relay set becomes a Gamma distributed random variable.

$$\gamma_D = \frac{p_{tx}}{\sigma^2} \sum_{j \in \Omega} z_{j,d} \sim \Gamma\left(n, \frac{p_{tx} \bar{z}_{R,D}}{\sigma^2}\right), \quad (4.6)$$

Beweis. See Appendix A.1 □

Based on this characteristic of the joint SNR at the destination, the outage probability of a relaying phase (with n active/forwarding relays) can be obtained based on the Gamma cumulative distribution function (CDF), and is given by:

$$P_2 = \Pr\{\gamma_D \leq \gamma^*\} = F(\gamma^*; n, \beta), \quad (4.7)$$

where β is the average of the signal strength and given by:

$$\beta = p_{tx} \bar{z}_{R,D} / \sigma^2. \quad (4.8)$$

According to [62], the CDF of the Gamma distribution P_2 is further given by:

$$P_2 = \begin{cases} 1 - \sum_{j=0}^{n-1} \frac{1}{j!} \left(\frac{\gamma^*}{\beta}\right)^j e^{-\frac{\gamma^*}{\beta}}; & n > 0 \\ 1; & n = 0 \end{cases}. \quad (4.9)$$

Therefore, according to the law of total probability, the mathematical expectation of outage probability of the overall multi-relay assisted two-hop transmission (from the source to the destination) is given by:

$$\begin{aligned} P^{\text{out}} &= \sum_{n=0}^J F(\gamma^*; n, \beta) P_B(n, J, P_1) \\ &= P_1^J + \sum_{n=1}^J \sum_{j=0}^{n-1} \frac{1}{j!} \left(\frac{\gamma^*}{\beta}\right)^j e^{-\frac{\gamma^*}{\beta}} P_B(n, J, P_1). \end{aligned} \quad (4.10)$$

Hence, we have derived a closed-form expression of outage probability for the studied two-hop transmission assisted by multi-DF-relay by introducing a topology approximation.

Validation

By simulation, we validate above derivations as well as evaluate the error of the proposed topology approximation. In the simulation, we randomly deploy 9 relays in a circular area with radius R , and set distances from the source to the center of relay group and from the relay group center to the destination to be the same as 100m. In addition, the frame length is set to $T_f = 10ms$ while the transmit power at the source, the transmit power at each relay and the noise power are set to 25dBm, 20dBm and -95dBm, respectively. At last, we utilize the well-known COST [63] model for calculating the path-loss while the center frequency is set to 2 GHz.

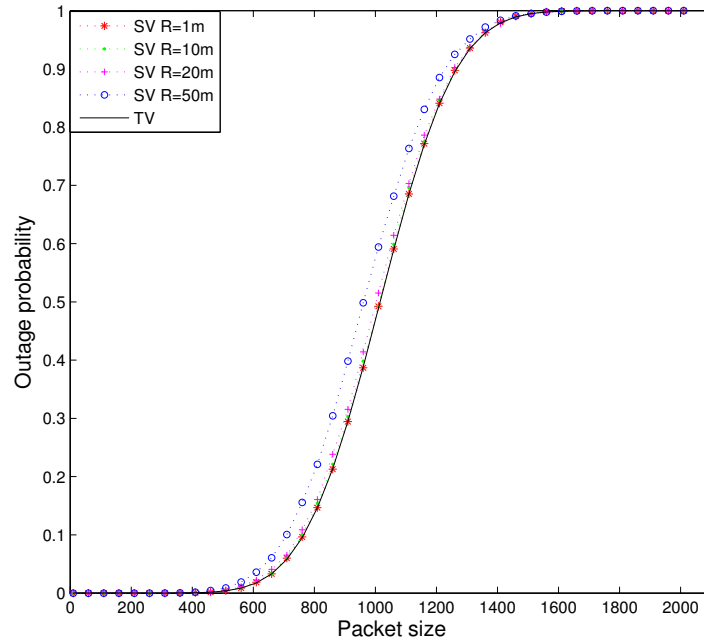


Abbildung 4.5: Outage probability performance under different topologies (i.e., different radii of relay-group area).

We show the performance gap between theoretical value (based on the approximation) and simulation values (based on real locations of relays) in Figure 4.5, where

we consider different radii of the circular area for randomly deploying relays. Firstly, we observe that the more separate relays are the more significant the performance mismatch (approximation error) is. Secondly, the figure also shows that the performance mismatch is tiny even when $R = 20\text{m}$ which means the maximum distance among relays could be $2/5$ of the distance from the source to relays or from relays to the destination. Therefore, under a multi-relay scenario where the distances between relays are much shorter than the distance either from the source to relays or from relays to the destination, the error of the proposed topology approximation is negligible.

4.3.2 Multi-DF-Relay Assisted Retransmissions

As introduced in Section 4.1.3, the relay-assisted retransmission (C-ARQ) protocols not only improve the system reliability but also introduce additional delays by retransmitting packets. In fact, spending more time for retransmissions results in a higher reliability and a longer delay. In other words, the reliability is strongly associated with the transmission delay. In the following, we study the outage probability performance together with the transmission delay of each considered C-ARQ protocol. We mainly investigate four C-ARQ protocols with MRC while another protocol without MRC is considered as a comparison scheme.

For the comparison scheme which does not apply MRC in the retransmission process, the received SNR of a retransmission is totally subject to the corresponding instantaneous channel quality. As the channel states in different frames are i.i.d., the SNRs of the initial transmission and a retransmission are also i.i.d.. Therefore, a retransmission has the same (expected) outage probability as the initial transmission. It results in that the number (either 1 or 0) of successfully decoded packet per single transmission/retransmission (from the source to the destination) becomes Bernoulli distributed. This simplifies all the performance analysis related to the outage probability. Therefore, in this work we call a retransmission protocol without MRC by **Simplified Case** (of the retransmission protocols).

Although discarding fail-decoded signals facilitates the theoretical analysis, it is more beneficial and reasonable in practice to make use of these already received but fail-decoded signals. We call by **General Case** if a retransmission protocol applies MRC

over the new-retransmitted signal and previous fail-decoded signals (for decoding). This combination of MRC and ARQ is also known as Hybrid-ARQ II [59]. This results in a significantly high (combined) SNR at the destination in the retransmission process, which refers to the sum of the instantaneous SNRs in the initial transmission and retransmissions. A C-ARQ protocol with MRC lets SNRs of the initial transmission and retransmissions no longer be i.i.d.. As a result, the outage probability analysis of these C-ARQ protocols become complicated.

In the following, we discuss the outage probabilities under these two cases respectively.

4.3.2.1 Simplified Case – C-ARQ without MRC

Recall that under Rayleigh block fading channels SNRs of the initial transmission and retransmissions are i.i.d.. Therefore, a retransmission has the same outage probability as the initial transmission. Considering a packet with transmission delay τ frames ($\tau > 0$), if each transmission/retransmission consists of M frames (e.g., $M = 3$ for a two-hop relaying transmission together with a feedback frame), obviously $\frac{\tau}{M}$ is an integer. The probability mass function (PMF) of τ , i.e., the probability of the event that the destination successfully decodes the packet after τ frames transmission and retransmissions, is given by:

$$P(\tau) = \left(P^{\text{out}}\right)^{\frac{\tau}{M}-1} \left(1 - P^{\text{out}}\right). \quad (4.11)$$

After transmitting/retransmitting τ frames, it is possible the destination still can not decode the packet successfully. The probability of this is defined as the outage probability of a C-ARQ protocol with transmission delay τ . Hence, the outage probability of the simplified C-ARQ protocol (without MRC) is given by:

$$P_{\text{simplified C-ARQ}}^{\text{out}}(\tau) = 1 - \sum_{i=0}^{\frac{\tau}{M}-1} \left(P^{\text{out}}\right)^i \left(1 - P^{\text{out}}\right). \quad (4.12)$$

4.3.2.2 General Case – C-ARQ with MRC

In the following, we study the outage performance of several C-ARQ protocols with MRC. In Section 4.1.3, we have introduced four C-ARQ protocols with MRC which are dynamic proactive (DP), dynamic reactive (DR), static proactive (SP) and static reactive (SR). Recall that these C-ARQ protocols with MRC differ in the way they deal with the initial transmission, and in the way they compose the group of retransmission relays. Hence, they have different transmission delay distributions and different outage performances. In the following, we first study the delay distributions of these protocols respectively, then derive the outage probability based on the delay distributions.

Different from the C-ARQ without MRC, these four C-ARQ protocols have more flexible and dynamic retransmissions strategies, e.g., the packet can be transmitted by source and retransmitted by relays and number of retransmitting relays is dynamic. In general, under these four C-ARQ protocols the number of retransmission frames of a packet consists of re-broadcasting frames and re-forwarding frames. The source re-broadcasts the packet if no relay decodes the packet correctly while a re-forwarding frame occurs if the previous forwarding/re-forwarding failed. If we denoted by k_1 and k_2 the number of broadcasting frames (including the initial broadcasting and re-broadcastings) and the number of forwarding frames (including the initial forwarding and re-forwardings), the random variable τ (the number of initial transmission(s) and retransmission frames) actually totally depends on k_1 and k_2 . Taking into account the frame cost for feedback, expressions of τ under proactive and reactive protocols are given by:

$$\tau_{k_1, k_2} = \begin{cases} 2k_1 + 2k_2 - 1, & \text{proactive protocols;} \\ 2k_1 + 2k_2, & \text{reactive protocols.} \end{cases} \quad (4.13)$$

Assuming the packet to be successfully decoded at some relay after k_1 broadcasting attempts, the probability of this equals $P_1^{J(k_1-1)} \cdot P_B(P_1)$. Similarly, if the destination finally decodes the packet correctly based on combining signals transmitted/retransmitted k_2 times from n active relays, the probability of this under a static proactive protocol equals $F(\gamma^*; n(k_2 - 1), \beta) - F(\gamma^*; nk_2, \beta)$.

For a dynamic protocol, the computation of a successful transmission during the relaying phase is actually more complex, as one needs to consider the dynamic number

of relays during the retransmission attempts. If the initial transmission of a dynamic proactive protocol succeeds (which means $k_2 = 1$) and at the same time the number of active relays is n , then the probability of this equals $F(\gamma^*; 0, \beta) - F(\gamma^*; n, \beta)$. Otherwise, the destination finally decodes the packet correctly based on combining the signals transmitted once by n active relays and retransmitted $k_2 - 1$ ($k_2 > 1$) times from all the J relays. This probability is equal to $F(\gamma^*; J(k_2 - 2) + n, \beta) - F(\gamma^*; J(k_2 - 1) + n, \beta)$. Therefore, the PMF of τ under the static proactive protocol and the dynamic proactive protocol are given by (4.14) and (4.15).

$$\begin{aligned} & P\{\tau_{\text{SP}} = 2k_1 + 2k_2 - 1\} \\ &= \sum_{n=1}^J \left\{ P_1^{J(k_1-1)} \cdot P_B(n; J, P_1) \cdot [F(\gamma^*; n(k_2 - 1), \beta) - F(\gamma^*; nk_2, \beta)] \right\}. \end{aligned} \quad (4.14)$$

$$\begin{aligned} & P\{\tau_{\text{DP}} = 2k_1 + 2k_2 - 1\} \\ &= \sum_{n=1}^J \left\{ P_1^{J(k_1-1)} \cdot P_B(n; J, P_1) \cdot [F(\gamma^*; \max\{J(k_2 - 2) + n, 0\}, \beta) - F(\gamma^*; J(k_2 - 1) + n, \beta)] \right\}. \end{aligned} \quad (4.15)$$

Similarly, the PMF of τ under the static reactive protocol and the dynamic reactive protocol are given by:

$$\begin{aligned} & P\{\tau_{\text{SR}} = 2(k_1 + k_2)\} \\ &= \begin{cases} 1 - P_0, & k_1 = 1; \\ P_0 \sum_{n=1}^J \left\{ P_1^{J(k_1-1)} \cdot P_B(n; J, P_1) \cdot [F(\gamma^*; n(k_2 - 1), \beta) - F(\gamma^*; nk_2, \beta)] \right\}, & k_1 > 1, \end{cases} \end{aligned} \quad (4.16)$$

$$\begin{aligned} & P\{\tau_{\text{DR}} = 2(k_1 + k_2)\} \\ &= \begin{cases} 1 - P_0, & k_1 = 1; \\ P_0 \sum_{n=1}^J \left\{ P_1^{J(k_1-1)} \cdot P_B(n; J, P_1) \cdot [F(\gamma^*; J(k_2 - 2) + n, \beta) - F(\gamma^*; J(k_2 - 1) + n, \beta)] \right\}, & k_1 > 1, \end{cases} \end{aligned} \quad (4.17)$$

where P_0 is the outage probability of the source-destination direct link transmission. According to (4.1), P_0 is given by $P_0 = 1 - \exp(-\gamma^* \sigma^2 / 2 \bar{z}_{S,D} p_{tx})$, where $z_{S,D}$ is the average channel gain of the direct link.

Based on the distributions of τ as shown in (4.14), (4.15), (4.16) and (4.17), we can immediately obtain the residual outage probability of a packet after τ transmission and retransmissions frames (for these studied four C-ARQ protocols with MRC):

$$P_{\text{C-ARQ}}^{\text{out}}(\tau) = 1 - \sum_{k_1=1}^{\lceil \tau/2 \rceil} \sum_{k_2=\lceil \tau/2 \rceil - k_1}^{\lceil \tau/2 \rceil} P[\tau_{k_1, k_2}], \quad (4.18)$$

where $\lceil \cdot \rceil$ is the rounding function (rounding to the next highest integer).

Validation

By simulations, we validate our derivations in Section 4.3.2. In the simulation, we randomly deployed 5 relays in a circle with radius 20 m. The S-R and R-D distances are both set to 200 m. In addition, we set $p_{tx} = 20$ dBm, $p_{fb} = 15$ dBm, $p_c = 17$ dBm and noise power $\sigma^2 = -95$ dBm, respectively.

We first validate the delay distribution of the C-ARQ protocol without MRC in Figure 4.6. From the figure, we observe a very good match of simulation results (S) in comparison to our theoretical models (T). In addition, the figure shows that the probability of τ is strictly decreasing in τ , which also matches with (4.11).

Delay distributions of these C-ARQ protocols with MRC are validated in Figure 4.7. Firstly, we observe again a good match between simulation results with our theoretical models. Secondly, different from the C-ARQ protocol without MRC, the probability of τ under a C-ARQ protocol with MRC first increases and then decreases as τ increases, i.e., the probability of τ are concentrated at the low delay regime. This difference is actually due to MRC in the retransmission process. Applying MRC significantly boosts the received SNR at the destination, this more likely makes a packet be decoded correctly based on a very small number of retransmissions. As shown in Figure 4.7, for a proactive protocol the probability of $\tau = 3$ (frames) is about 0.01. In other words, the outage probability of the initial transmission is 99%. At the same time the outage probability of the initial transmission of a reactive protocol (recall that the initial

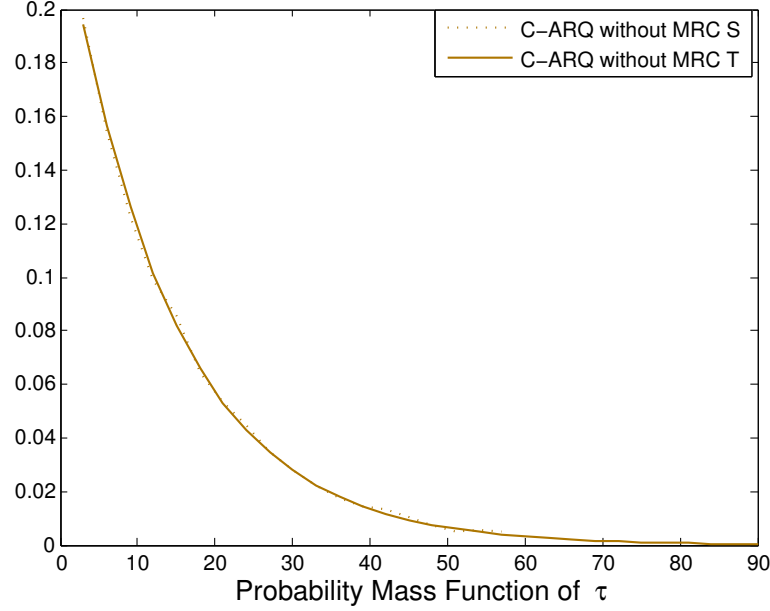


Abbildung 4.6: Validation of the distribution of τ of the C-ARQ protocol without MRC (packet size 70 bits).

transmission of a reactive protocol is a S-D direct link transmission) is almost 100%. Both 99% and 100% are very terrible values for the outage probability. Fortunately, the system has retransmission mechanism. By several times of retransmitting, a packet is very likely to be decoded successfully at the destination. For example, under dynamic proactive protocol 37% of packets can be successfully decoded at the destination just relying on the first round retransmitting and 38% of them by the second round retransmitting. Therefore, although introducing additional delay, relay-assisted retransmission protocols (C-ARQ) significantly promote the transmission reliability. Thirdly, among four C-ARQ protocols with MRC, we observe that the probability density function of random variable τ is more concentrated for dynamic protocols than for static protocols. This indicates that dynamic protocols are more time-efficient than static protocols. Finally, proactive protocols outperform reactive protocols by having low expected transmission delay. Moreover, in comparison to the corresponding proactive protocol each reactive protocol has a similar curve but also 2 frames additional delay. Recall that the initial transmission of a reactive protocol is a S-D direct link transmission. As the direct link is too weak to transmit the data packet, a reactive protocol just wastes

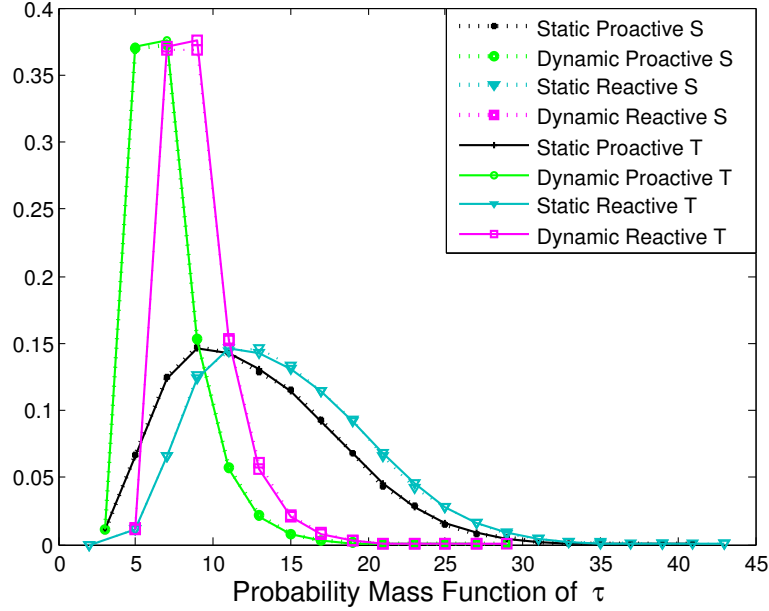


Abbildung 4.7: Validation of distributions of τ of these C-ARQ protocols with MRC (packet size 85 bits).

two frames (a direct link transmission frame and a NACK frame) for this initial transmission. That is the reason why the expected transmission delay of a reactive protocol is higher than the corresponding proactive protocol (about 2 frames).

We further compare delay distributions of these C-ARQ protocols under different MRC consumptions in Figure 4.8. To make the comparison fair, we only consider comparing two proactive protocols with MRC to the protocol without MRC as all these three protocols let relays assist the initial transmission as well as retransmissions. In particular, the only difference between the protocol without MRC and the static proactive protocol is if apply MRC or not. By the comparison among the above three protocols, we intent to observe the importance of applying MRC.

Figure 4.8 shows that these two protocols with MRC have better delay performance than the protocol without MRC, i.e., the former have higher probability to have short transmission delay τ . This actually illustrates the advantage and the importance of applying MRC from a perspective of delay.

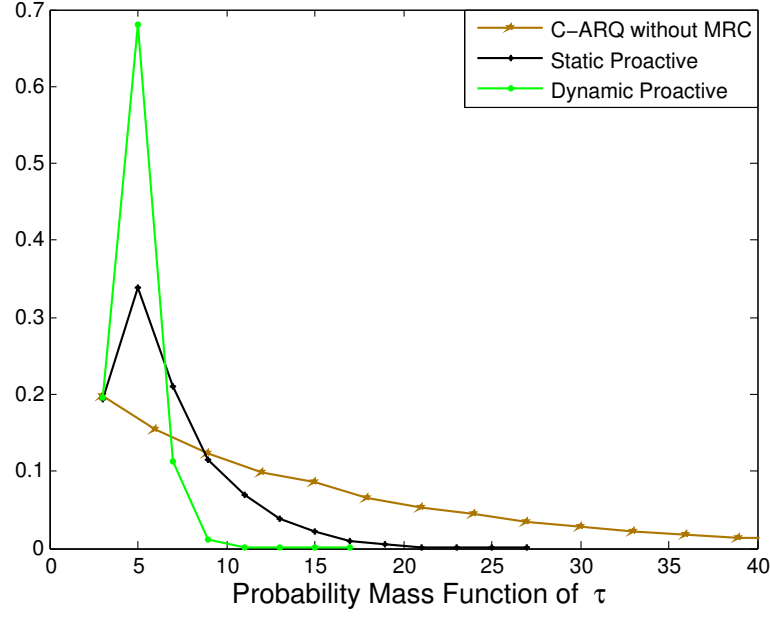


Abbildung 4.8: The delay comparison between these protocols with and without MRC (packet size is set to 70 bits).

The corresponding outage probability comparison is shown in Figure 4.9. It shows a similar result as Figure 4.8 that by applying MRC the reliability performance is significantly boosted.

Summary

In Section 4.3, we first derived the outage probability of a relay-assisted two-hop transmission which is actually the initial transmission or a retransmission of a C-ARQ protocol. We then studied the outage probability as well as the distribution of the transmission delay τ for each considered C-ARQ protocol. Two cases (with different assumption of MRC) of C-ARQ protocols were considered. In addition, by simulation we validated our derivations.

In fact, the distribution of τ plays a key role in the analysis of this chapter. This also can be observed from Figure 4.4. In the next section, we make use of the distribution of τ to investigate the physical-layer throughput and higher-layer capacity of each C-ARQ protocol.

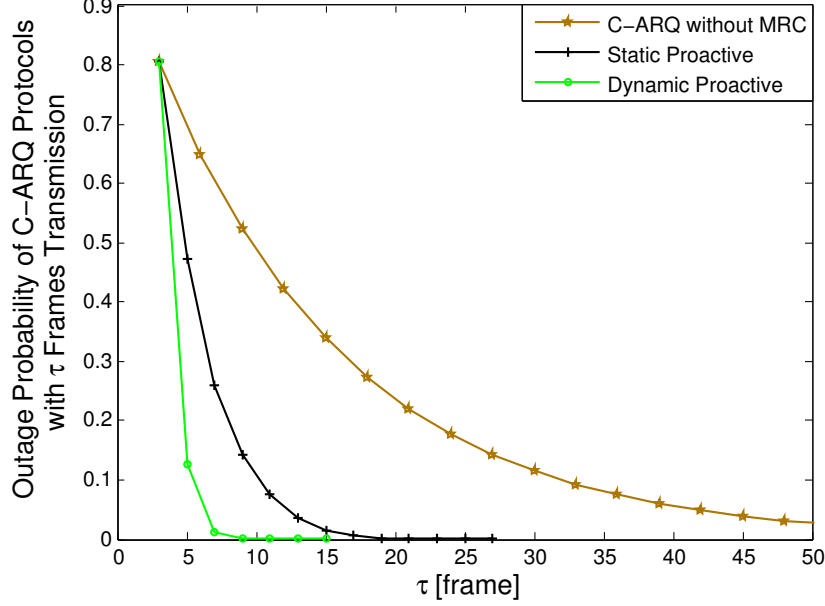


Abbildung 4.9: Outage performance of these three C-ARQ protocols (packet size 70 bits).

4.4 Throughput and Effective Capacity

In this section, we investigate the physical-layer throughput and higher-layer effective capacity of a multi-relay-assisted network with retransmission protocols.

4.4.1 Physical-layer Throughput

If a packet with size ρ is transmitted successfully at the cost of τ frames transmission and retransmissions, the average of throughput of system during these τ frames is given by ρ/τ per frame. Based on this, we introduce the service process model to the system. As introduced in Section 2.2, a service process increment s_i denotes the amount of bits that effectively leave the queue at frame i . Then, we adopt a service process model [64]² where the incremental service process equals $s_i = \rho/\tau, s_{i+1} = \rho/\tau, \dots, s_{i+\tau} = \rho/\tau$.

² In fact, the real service process increment s_i takes the form $s_i = 0, s_{i+1} = 0, \dots, s_{i+\tau} = \rho$ over the time span of the τ frames until successful transmission (as the packet is taken out of the source queue only if it was successfully transmitted). The advantage of this simplified model [64] is that we only need to determine the statistics of τ in order to obtain the mean and variance of the service

Therefore, the throughput of C-ARQ protocols become the means of service process increments of the protocols. In the following, we study the throughput of C-ARQ protocols by determining their means of the service process increments.

Simplified Case – C-ARQ without MRC

Under the simplified retransmission scenario, the outage probability of a transmission and a retransmission are the same. Therefore, the service process increment s_i is binomial distributed. With packet size ρ and the overall outage probability (of a two-hop transmission or retransmission) P^{out} , the expected/average throughput (per frame) is given by:

$$C_{\text{ave}} = \mathbb{E}[s_i] = \rho(1 - P^{\text{out}})/M. \quad (4.19)$$

Recall that M is the number of frames for each transmission or retransmission. In the studied system, we have $M = 3$ as each transmission or retransmission consists of two frames (two-hop) for relaying and one for feedback.

General Case – C-ARQ with MRC

For these C-ARQ protocols with MRC, throughputs of these C-ARQ protocols also can be obtained by calculating their means of service process increments. However, the service process models of these protocols are relatively more complicated than the simplified case. This is due to the fact that under each protocol with MRC outage probabilities of a transmission and of a retransmission are different. Therefore, in order to obtain the mean of service process increments of a protocol with MRC we first derive the PMF of the service process increment s_i , which is actually subject to the PMF of τ .

It should be clarified that the PMF of τ and the PMF of the service process increment s_i are different. Recall that if we take randomly one packet from all the packets that have been transmitted the probability that this packet has a transmission delay τ is $\Pr[\tau_{k_1, k_2}]$. This is the PMF of τ . Based on our service process model, the corresponding

process increments. This comes at the price of being an approximation, however, we validate in the following that this assumption is not impacting the system performance significantly.

service process increment of this packet has value ρ/τ and lasts τ frames. In other words, when the service process increment shows up with a certain value ρ/τ , they show up in a group with size τ . This group behavior needs to be taken into account. Therefore, the PMF of s_i can be obtained based on the PMF of τ by scaling the probability for each possible value of τ . Hence, the PMF of s_i is given by:

$$\Pr(s_i = \frac{\rho}{\tau_{k_1, k_2}}) = \frac{\Pr[\tau_{k_1, k_2}] \cdot \tau_{k_1, k_2}}{\sum_{k_1, k_2 \in \mathbb{N}} \tau_{k_1, k_2} \Pr[\tau_{k_1, k_2}]}, \quad (4.20)$$

where the numerator is the expected value of τ_{k_1, k_2} given by $E[\tau_{k_1, k_2}]$. Then, the mean of service process increments, which is also the physical-layer throughput (per frame) of a C-ARQ protocol with MRC, is given by:

$$\begin{aligned} C_{\text{ave}} = E[s_i] &= \sum_{k_1, k_2 \in \mathbb{N}} \frac{\rho}{\tau_{k_1, k_2}} \cdot \Pr(s_i = \frac{\rho}{\tau_{k_1, k_2}}) \\ &= \sum_{k_1, k_2 \in \mathbb{N}} \frac{\rho}{\tau_{k_1, k_2}} \cdot \frac{\Pr[\tau_{k_1, k_2}] \cdot \tau_{k_1, k_2}}{E[\tau_{k_1, k_2}]} = \frac{1}{T_f} \cdot \frac{\rho}{E[\tau_{k_1, k_2}]}. \end{aligned} \quad (4.21)$$

Validation

Here we provide simulation results obtained from the model parameterization which is the same as the simulation in Section 4.3. Figure 4.10 and Figure 4.11 show physical-layer throughputs of the C-ARQ protocol without MRC and these C-ARQ protocols with MRC, respectively. First of all, we observe again a perfect match between our analytical models and simulations. In addition, Figure 4.11 shows that proactive protocols have higher throughputs than proactive protocols and that dynamic protocols are more stable for dynamic packet size in comparison to static protocols. Moreover, we find that the throughput of each protocol can be maximized by choosing an appropriate packet size. This is an important guideline for the system design.

Then, we compare the protocol C-ARQ without MRC and these two proactive protocols (with MRC) in Figure 4.12. We observe that C-ARQ protocols with MRC outperform the one without MRC and at the same time are more stable for dynamic packet size.

4 Relaying Performance in the Infinite Blocklength Regime

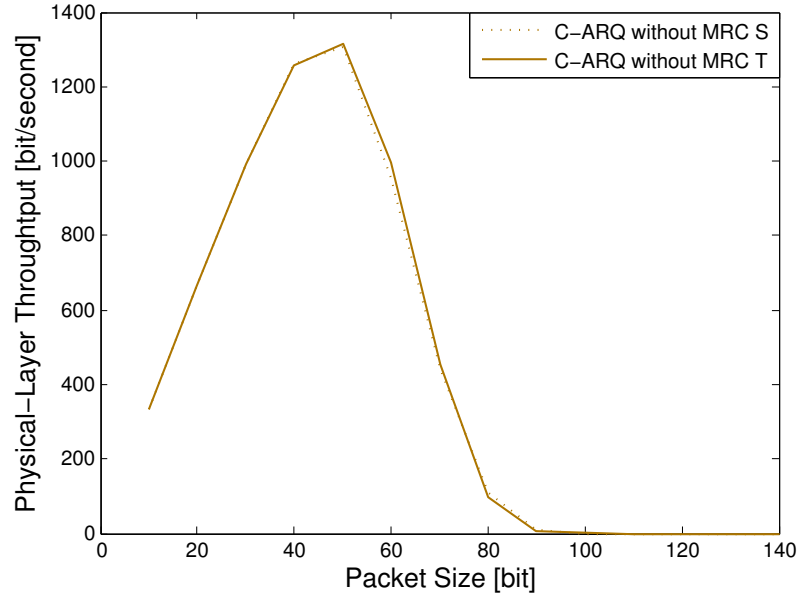


Abbildung 4.10: Physical-layer throughputs of the C-ARQ protocol without MRC.

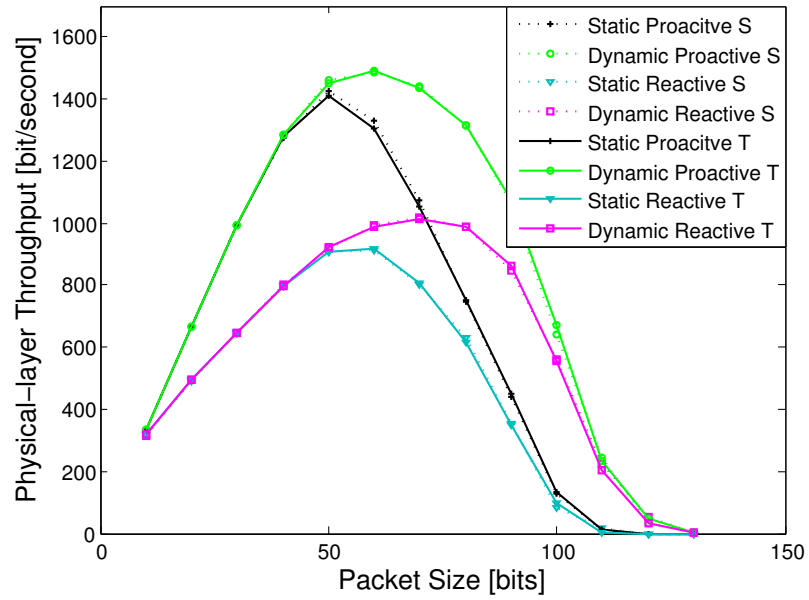


Abbildung 4.11: Physical-layer throughputs of C-ARQ protocols with MRC

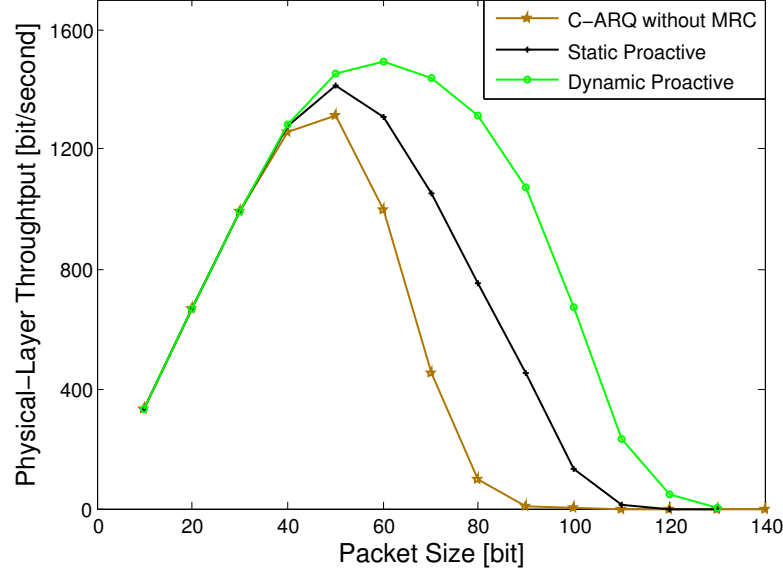


Abbildung 4.12: Physical-layer throughputs of C-ARQ protocols with different assumptions of MRC

4.4.2 Higher-layer Effective Capacity

As introduced in Section 2.2, in this work we refer to MSDR as a metric for the higher-layer queuing performance. Moreover, based on (5.30) the major challenge for determining the MSDR is to obtain the mean and variance of the increment of the service process. Recall that the mean of service process increments of each protocol is equal to the corresponding physical-layer throughput and has been derived in the previous section. So far, only variance expressions of service process increments are missing.

In the following, we first provide variance expressions of considered C-ARQ protocols under different MRC consumptions. Subsequently, we validate these variance expressions and evaluate their corresponding MSDR performance.

Simplified Case – C-ARQ without MRC

The variance of service process increments of the C-ARQ protocol without MRC is given by:

$$\text{Var}[s_i] = \rho P^{\text{out}}(1 - P^{\text{out}})/M^2, \quad (4.22)$$

where ρ is the packet size, and recall that M is the frames cost of each transmission/retransmission under this C-ARQ protocol without MRC (in the studied relaying system $M = 3$).

General Case – C-ARQ with MRC

Similar to (4.21), based on the PMF of the transmission delay τ the variance of service process increments of these C-ARQ protocols with MRC is given by:

$$\text{Var}[s_i] = \frac{\rho^2}{\mathbb{E}[\tau_{k_1, k_2}]} \sum_{k_1, k_2 \in \mathbb{N}} \frac{\Pr[\tau_{k_1, k_2}]}{\tau_{k_1, k_2}} - (\mathbb{E}[s_i])^2. \quad (4.23)$$

Validation

We validate the variance expressions and evaluate the MSDR performance of these studied C-ARQ protocols. Provided simulation results are based on the model parameterization which is same as the simulation in Section 4.3.

In Figure 4.13 and Figure 4.14 we show the variance of service process increments and the MSDR of the C-ARQ protocol without MRC, respectively. Again, we observe that simulation results match well with our theoretical analysis derived in (4.22).

The variance of service process increments and MSDRs of C-ARQ protocols with MRC are shown in Figure 4.15 and Figure 4.16, respectively. First, we observe in Figure 4.15 that the dynamic protocols have a lower variance of the service process increment in comparison to the static protocols.

Second, it is also observed from Figure 4.15 that dynamic protocols have relatively lower variance values. In addition, we find that all the variance curves have global maxima for distinct packet sizes ($\rho > 50$ bits). The reason is as follows. As the variance of ρ/τ is increasing in ρ^2 , the variance decreases for smaller packet sizes. However,

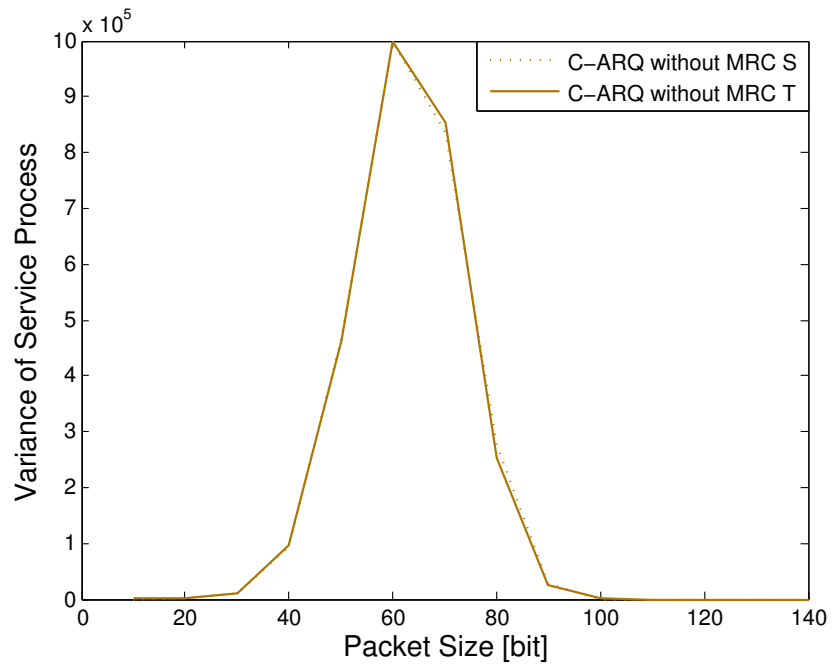


Abbildung 4.13: Variance of the C-ARQ protocol without MRC.

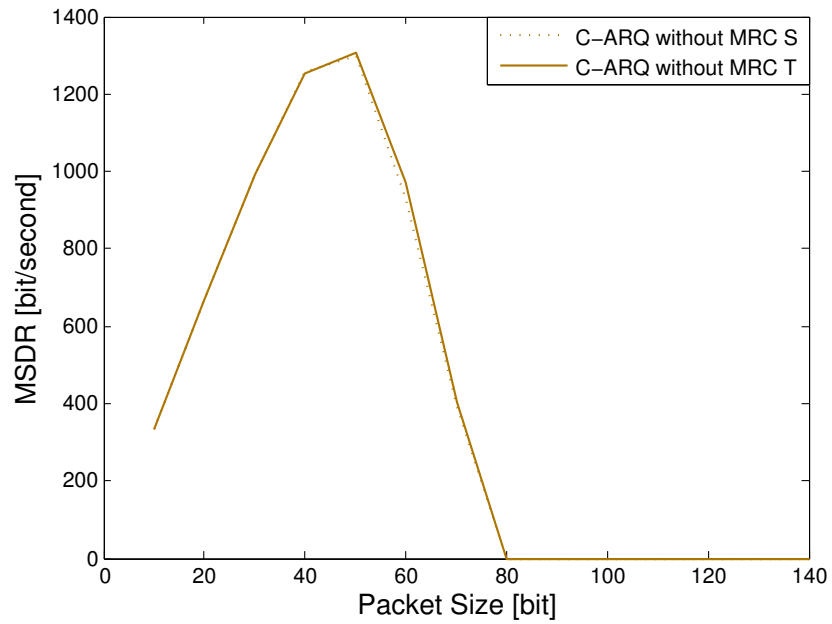


Abbildung 4.14: MSDR of the C-ARQ protocol without MRC.

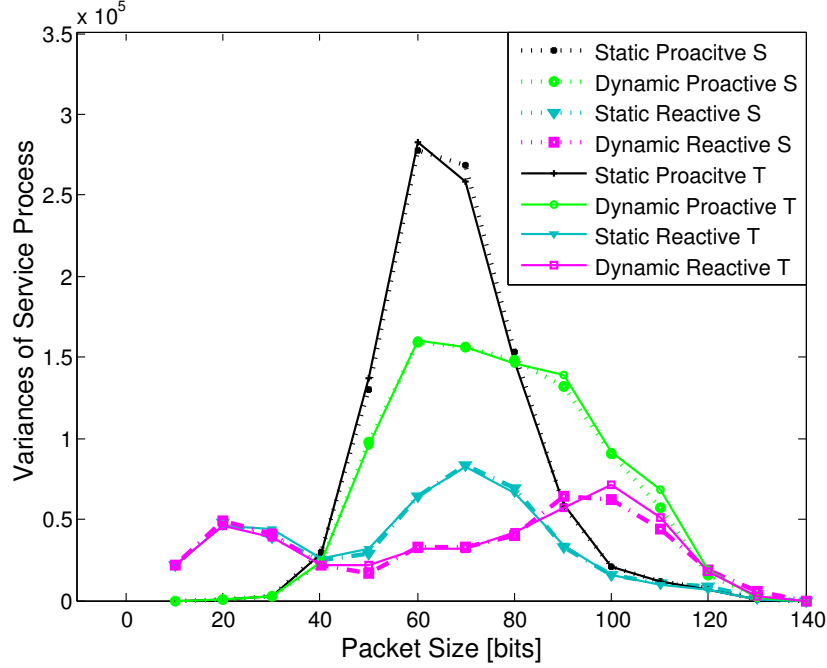


Abbildung 4.15: Variance of C-ARQ protocols with MRC.

small packets also lead to a higher transmission success probability which results in a low expected value of τ . As the packet size increases, the (expected) value of τ increases since the destination needs more retransmissions to finally decode the packet successfully. This finally leads to a low value of the variance of ρ/τ . Summarizing, the variance of ρ/τ actually demonstrates a tradeoff between ρ and τ . This is the reason why the variance curves first increase and then decrease as the packet size increases. Note that the variance of each reactive protocol has also a local maximum for small packet sizes (of about 20 bits). These local maxima are caused by the direct link transmission. The initial transmission of a reactive protocol is a direct link transmission which has a much lower SNR in comparison to relay-assisted transmissions. Small packet sizes let the packet be decoded successfully at the destination relying on just a direct link transmission. Hence, the probability increases strongly that $\tau = 2$. As ρ slightly increases, the direct link transmission still very likely leads to a successfully transmission, which means that τ does not increase. Therefore, the variance of ρ/τ slightly increases. However, as ρ continues to increase, the probability of a successful direct link transmission significantly decreases. Hence, more and more packets need

to be retransmitted by the relays, which leads to a much larger τ . As a result, the variance of ρ/τ then decreases.

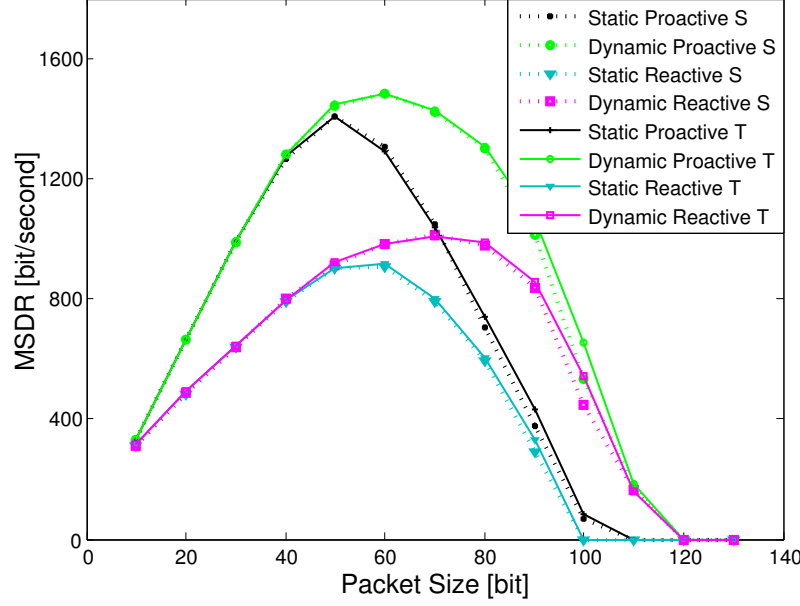


Abbildung 4.16: MSDR of C-ARQ protocols with MRC.

Then, we compare the C-ARQ protocol without MRC and proactive C-ARQ protocols (with MRC) regarding the service process variance and the MSDR. We show the comparisons in Figure 4.17 and Figure 4.18, respectively.

Obviously, applying MRC reduces the variance of service process increments and increases the higher-layer performance (MSDR) of C-ARQ protocols. In addition, Figure 4.18 shows that the higher-layer performance of C-ARQ protocols become more stable for dynamic packet size by applying MRC.

Summary

In Section 4.4, we derived the physical-layer throughput as well as the higher-layer effective capacity of each studied C-ARQ protocol. We validated our analytical models by simulation. In addition, we concluded several guidelines for the design of efficient C-ARQ protocols from our simulation from the perspective of either throughput or effective capacity: (1) Applying MRC significantly increases both the throughput and

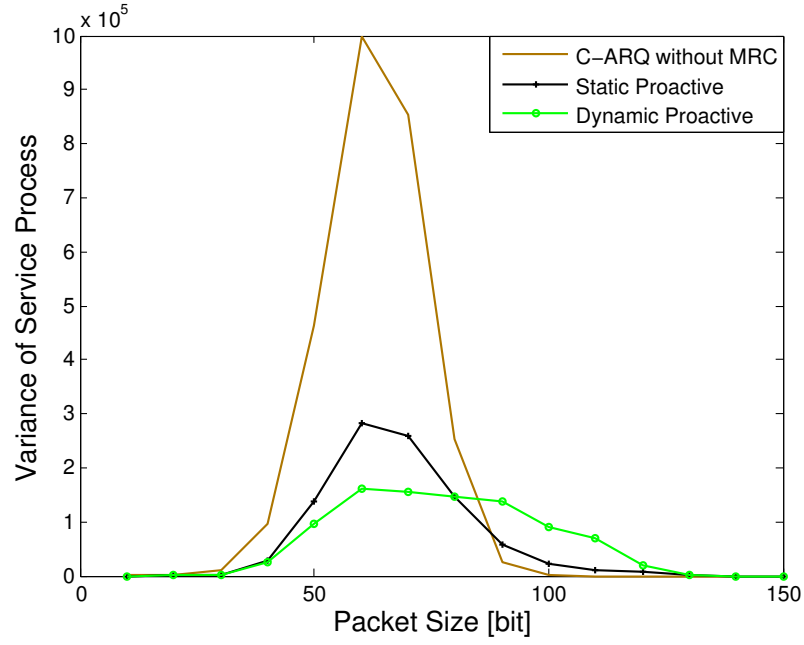


Abbildung 4.17: Variance of service process increments under studied C-ARQ protocols with different MRC assumptions.

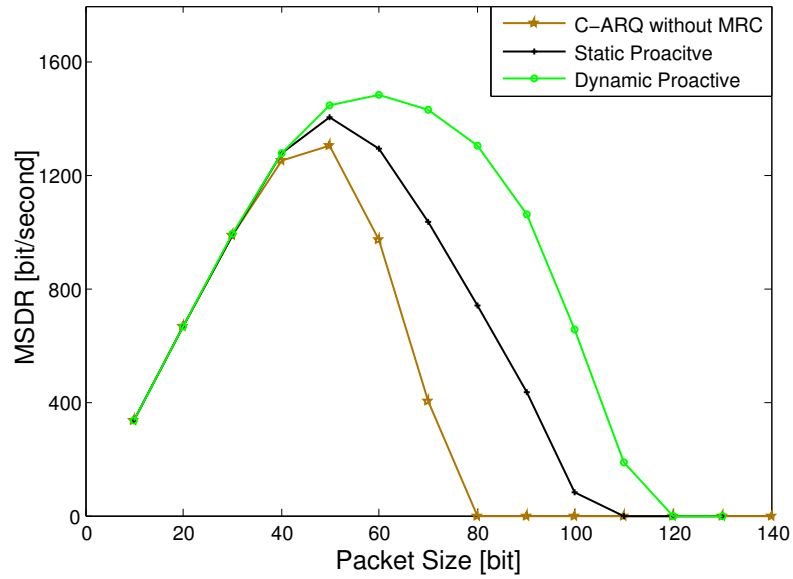


Abbildung 4.18: MSDR of the C-ARQ protocols under different MRC assumptions.

the effective capacity of studied C-ARQ protocols. (2) Direct link transmission can usually be ignored by studied C-ARQ protocols. (3) There exists an optimal packet size for maximizing either the throughput or the effective capacity of each C-ARQ protocol.

4.5 Energy Efficiency

In this section, we first analyze the energy consumption model of each studied C-ARQ protocol based on the distribution of τ . Subsequently, we combine the analysis of the throughput and effective capacity in the previous section together with the energy consumption to investigate the corresponding energy efficiency performance.

4.5.1 Energy Consumption

Recall that the energy consumption of the system contains two parts: The basic energy consumption and the energy consumption due to transmission. The average total consumed energy per frame is then obtained by (recall that K is the probability that the queue is non-empty and p_c is the power of basic energy consumption at each node):

$$\bar{\Psi} = K \cdot \bar{\Psi}_T + (J + 2) \cdot p_c T_f, \quad (4.24)$$

where $(J + 2) \cdot p_c T_f$ is the basic energy consumption over the duration of one frame of all nodes including the source, the destination and J relays. Obviously, all the studied C-ARQ protocols have the same basic energy consumption. Furthermore, $\bar{\Psi}_T$ is the average energy consumption over one frame due to data transmission (when the queue is non-empty). More importantly, $\bar{\Psi}_T$ totally depends on the C-ARQ protocol. In other words, $\bar{\Psi}_T$ likely has different value under different C-ARQ protocols. Therefore, the major challenge for determining the energy consumption $\bar{\Psi}$ becomes to obtain the $\bar{\Psi}_T$ for each protocol. In the following, we first derive $\bar{\Psi}_T$ for C-ARQ protocols with MRC. Then, based on this we provide the expression of $\bar{\Psi}_T$ for the C-ARQ protocol without MRC.

General Case – C-ARQ with MRC

In general, C-ARQ protocols require different numbers of broadcasting frames, relaying frames and ACK/NACK frames for transmitting/retransmitting a packet. From an asymptotic perspective, these different types of frames appear alternately with different probabilities. In other words, these frame types occupy different proportions over a longer time span of operation.

Recall that static and dynamic protocols have different numbers of relays participating in the retransmission process. Hence, energy consumption (over one frame due to data transmission) $\bar{\Psi}_T$ has different expressions under these protocols. For static C-ARQ protocols, we denote by P_S^{ct} , P_R^{ct} and P_D^{ct} percentages of broadcasting frame, relaying frame (including the initial transmission and retransmissions) and ACK/NACK frame. Therefore, the average transmit energy consumption per frame is given by $(P_S^{\text{ct}}p_{\text{tx}} + E[\Omega]P_R^{\text{ct}}p_{\text{tx}} + P_D^{\text{ct}}p_{\text{fb}})T_f$, where Ω denotes the random number of active relays with expectation $E[\Omega]$.

Under a dynamic protocol, the number of active relays varies in the retransmission process. Recall that under a dynamic protocol, initially some relays become active by decoding the broadcasted packet (from the source) successfully while the other relays (passive relays) become active later on by overhearing the forwarded packet from the initial active relays. As distances between relays are assumed to be much shorter than the distance from the source to each relay, all the links between relays are significantly more reliable than these source-relay links. Hence, all the passive relays are very likely to decode the packet successfully by overhearing one retransmission. Then, all the relays become active for retransmission. Therefore, the percentage of relaying frames of a dynamic protocol can be modeled by P_{i-R}^{ct} and P_{a-R}^{ct} which are percentages of initial active relays retransmitting frames and all relays retransmitting frames. Hence, the average transmit energy consumption³ (per frame) of dynamic protocols can be modeled as $(P_S^{\text{ct}}p_{\text{tx}} + E[\Omega]P_{i-R}^{\text{ct}}p_{\text{tx}} + JP_{a-R}^{\text{ct}}p_{\text{tx}} + P_D^{\text{ct}}p_{\text{fb}})T_f$, where Ω is the number of the initial active relays.

³ This expression is rather realistic under the assumed scenario where the relays are very close to each other. It will be an upper bound for the energy consumption if the relays are more spread out.

Summarizing, the average energy consumption due to transmission of the C-ARQ protocols is given by:

$$\bar{\Psi}_T = \begin{cases} (P_S^{\text{ct}} + E[\Omega]P_R^{\text{ct}})p_{\text{tx}}T_f + P_D^{\text{ct}}p_{\text{fb}}T_f, & \text{static;} \\ (P_S^{\text{ct}} + E[\Omega]P_{\text{i-R}}^{\text{ct}} + JP_{\text{a-R}}^{\text{ct}})p_{\text{tx}}T_f + P_D^{\text{ct}}P_{\text{fb}}T_f, & \text{dynamic.} \end{cases} \quad (4.25)$$

As each node is assumed to operate with the same transmit power, $\bar{\Psi}$ is fully determined by variables P_S^{ct} , P_R^{ct} (or $P_{\text{i-R}}^{\text{ct}}$ and $P_{\text{a-R}}^{\text{ct}}$), P_D^{ct} and $E[\Omega]$. In the following, we derive these variables for these C-ARQ protocols with MRC.

First, the expected size of the initial active relay set $E[\Omega]$ is equivalent to the expected number of relays which decode the packet correctly after a single broadcasting frame. In addition, Ω is a binomially distributed random variable [57]. Hence, $E[\Omega]$ is obtained by:

$$E[\Omega] = J \cdot (1 - P_1). \quad (4.26)$$

Second, for all C-ARQ protocols with MRC the percentage of the broadcasting frames is given by:

$$P_S^{\text{ct}} = \sum_{k_1, k_2 \in \mathbb{N}} \frac{k_1}{\tau_{k_1, k_2}} \cdot P[\tau_{k_1, k_2}]. \quad (4.27)$$

Third, percentages of relaying frames and ACK/NACK frames under proactive protocols and reactive protocols need to be analyzed separately. From Figure 4.2 we observe that there is a one-to-one correspondence between a relaying frames and an ACK/NACK frame in each proactive protocol. Hence, for a proactive protocol percentages of relaying frames and ACK/NACK frames are the same, and are given by:

$$P_R^{\text{ct}} = P_D^{\text{ct}} = (1 - P_S^{\text{ct}})/2. \quad (4.28)$$

For these two reactive C-ARQ protocols, corresponding percentages of relaying frames with different relay sets are given by:

$$P_{\text{i-R}}^{\text{ct}} = \sum_{k_1, k_2 \in \mathbb{N}} \frac{1}{\tau_{k_1, k_2}} \cdot \Pr[\tau_{k_1, k_2}], \quad (4.29)$$

$$P_{\text{a-R}}^{\text{ct}} = \sum_{k_1, k_2 \in \mathbb{N}} \frac{k_2 - 1}{\tau_{k_1, k_2}} \cdot \Pr[\tau_{k_1, k_2}]. \quad (4.30)$$

Finally, the percentage of ACK/NACK frames of reactive protocols P_D^{ct} is obtained by:

$$P_D^{\text{ct}} = 1 - P_S^{\text{ct}} - P_{\text{i-R}}^{\text{ct}} - P_{\text{a-R}}^{\text{ct}}. \quad (4.31)$$

Simplified Case – C-ARQ without MRC

For the simplified case, each retransmission can be treated a new independent transmission. Therefore, the expected/average energy consumptions of transmitting and retransmitting a packet are the same. The average transmit energy consumption per frame is given by $(P_S^{\text{ct}} p_{\text{tx}} + \text{E}[\Omega] P_R^{\text{ct}} p_{\text{tx}} + P_D^{\text{ct}} p_{\text{fb}}) T_f$, which is the same to a static C-ARQ protocol with MRC. Regarding the variables in the expression, the rest percentage variables have the same value $P_S^{\text{ct}} = P_R^{\text{ct}} = P_D^{\text{ct}} = 1/M = 1/3$. In addition, $\text{E}[\Omega]$ is also the same to (4.26). Hence, the average transmit energy consumption of the C-ARQ without MRC is given by:

$$\bar{\Psi}_T = \frac{T_f}{3} (p_{\text{tx}} + \text{E}[\Omega] p_{\text{tx}} + p_{\text{fb}}). \quad (4.32)$$

Validation

By simulation we validate above derivations of energy consumption. The simulation is based on the model parameterization which is same as the simulations in Section 4.3.

First, the energy consumption of the C-ARQ protocol without MRC is shown in Figure 4.19, where the theoretical curve matches the simulation curve well. In addition, we find that energy consumption curves in the figure are decreasing in the packet size. The reason is as follows. As packet size increases, the outage probability of each transmission and retransmission increases. Hence, the expected number of active relays (which decode the packet successfully) is decreasing in the packet size. Moreover, the energy consumed by transmitting packet at relays is linearly increasing in the number of active relays. Therefore, this part of energy consumption is reduced as the packet increases. In particular, this part of energy consumption is zero for an extremely big packet which makes no relay decode it successfully. At the same time, the basic energy

consumption and the transmit power consumption at the source are constants. As a result, the total energy consumption is decreasing in the packet size.

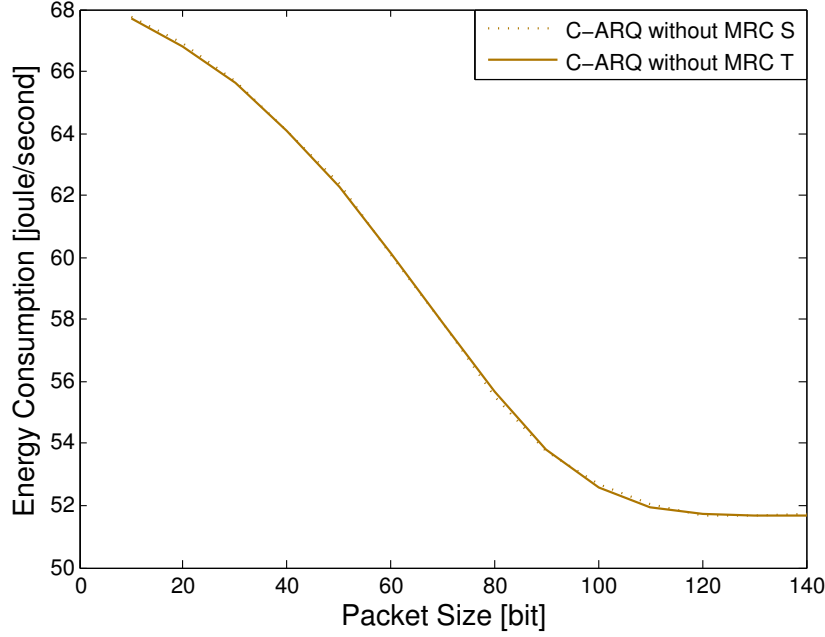


Abbildung 4.19: Energy consumption of the C-ARQ protocol without MRC.

Second, we show the energy consumption of C-ARQ protocols with MRC in Figure 4.20. We observe that proactive protocols consume more energy than reactive protocols as they make more relays join in the retransmissions. In addition, we find that energy consumptions of static protocols first decrease and then increase as the packet size increases. This is due to the fact that in the simulation the transmit power at each relay is much lower than the one at the source. Recall that under a static protocol the forwarding relay number in retransmissions is fixed as the number of relays which initially decode the packet (transmitted from the source) successfully. The minimal energy consumption is achieved when the packet is not extremely big and there is only one relay decodes the packet successfully and retransmits the packet to the destination again and again. If the packet is extremely big which makes none of the relays decode it correctly, not a relay but the source retransmits the packet and consumes more transmit power than a relay. This is a special characteristic of reactive protocols.

4 Relaying Performance in the Infinite Blocklength Regime

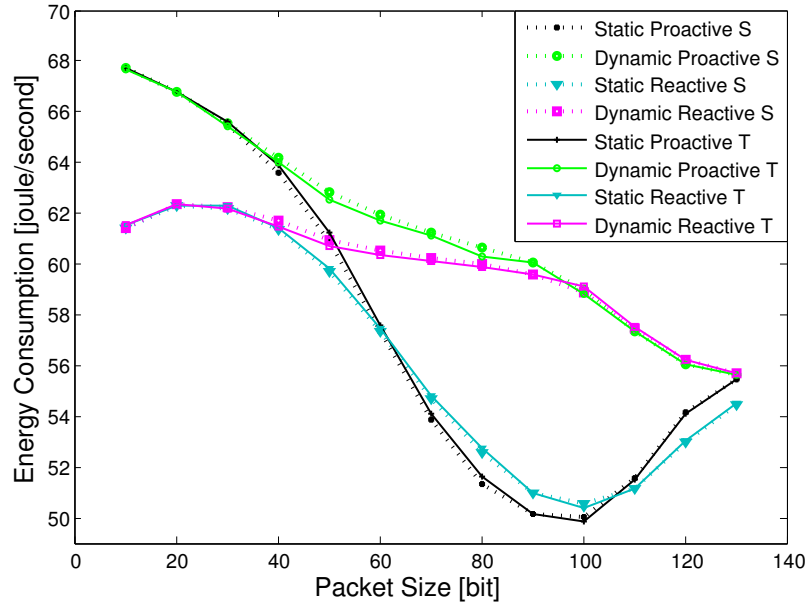


Abbildung 4.20: Energy consumptions of C-ARQ protocols with MRC.

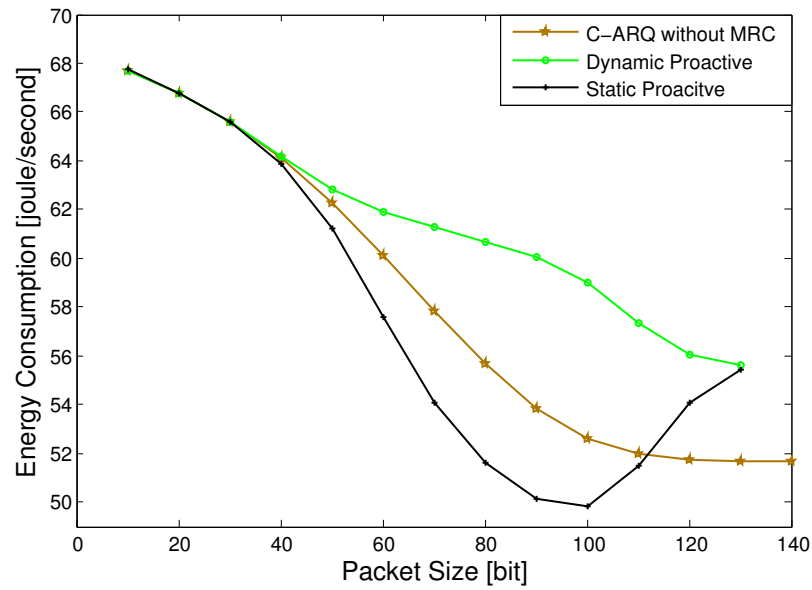


Abbildung 4.21: Energy consumptions of C-ARQ protocols under different MRC assumptions.

Then, we compare the energy consumptions of the C-ARQ protocol without MRC and proactive C-ARQ protocols (with MRC) in Figure 4.21. We observe that the dynamic proactive protocol consume more energy than the C-ARQ protocol without MRC. Recall that it is shown in Figure 4.12 and Figure 4.18 that the dynamic proactive protocol has higher throughput and effective capacity than the C-ARQ protocol without MRC. Then a question is raised: which of these protocols are more energy-efficient? We address the this question in Section 4.5.2 and Section 5.5.3 from perspectives of physical-layer energy efficiency and the higher-layer energy efficiency, respectively.

4.5.2 Physical-layer Energy Efficiency (PEE)

In this subsection, we investigate the physical-layer energy efficiency of studied C-ARQ protocols.

4.5.2.1 The Derivation of Physical-layer Energy Efficiency

The physical-layer performance does not consider the queuing effect at the source, i.e., the source buffer is assumed to be non-empty. In this work, the physical-layer energy efficiency (PEE) is defined as the ratio of the average throughput to the average energy consumption. This ratio is also equal to the ratio mean of service process increment (which is actually the throughput per frame) to the energy consumption per frame:

$$\Phi_p = \frac{\rho(1 - P^{\text{out}})}{\bar{\Psi}_T + (J + 2) \cdot p_c T_f}. \quad (4.33)$$

4.5.2.2 The Performance Evaluation of PEE

In the following, we evaluate physical-layer energy efficiencies of the studied C-ARQ protocols. The same system model and scenario parameterization are used as in Section 4.3. In Figure 4.22 simulation results are shown for these C-ARQ protocols with MRC. Here it is clearly seen that with non-extremely small packets proactive protocols outperform reactive protocols. This illustrates that making use of the direct link for the initial transmission is usually not a good choice. In addition, the PEE performance of each protocol can be maximized by choosing an appropriate packet size. However,

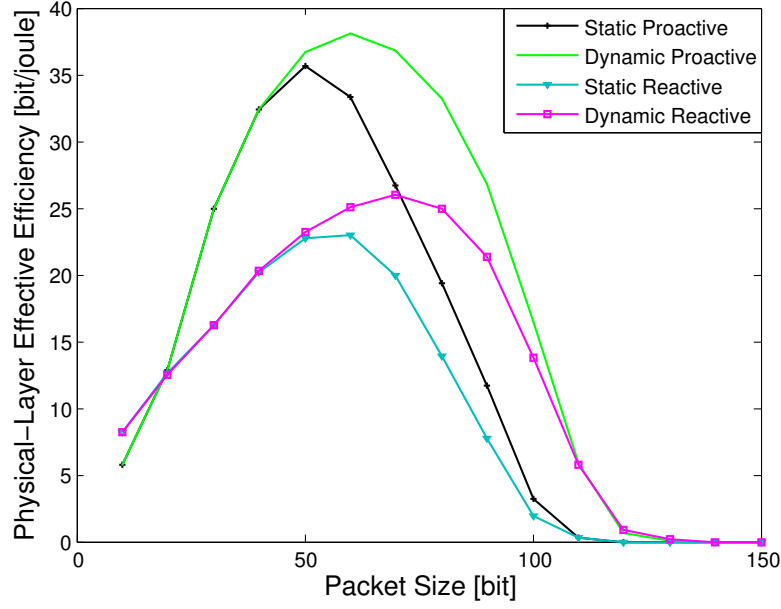


Abbildung 4.22: Physical-layer energy efficiencies of C-ARQ protocols with MRC.

regarding the optimization problem which maximizes the PEE these protocols have different optimal solutions and different optimal values. In particular, proactive protocols have higher optimal values of PEE than reactive protocols. In addition, optimal packet sizes for dynamic protocols (to maximize the PEE) are bigger than static protocols.

The physical-layer energy efficiency of the C-ARQ protocol without MRC is shown in Figure 4.23, where proactive protocols (with MRC) are plotted as references. We observe a significant performance enhancement due to applying MRC to a C-ARQ protocol.

4.5.2.3 The Performance Improvement of PEE

In the following, we consider to improve the PEE of a C-ARQ system with QoS constraint, i.e., the outage probability constraint. The system model is the same as the previous sections in this Chapter, where the channels experience Rayleigh block-fading while only average CSI is available at the source or relays. Moreover, to simplify the

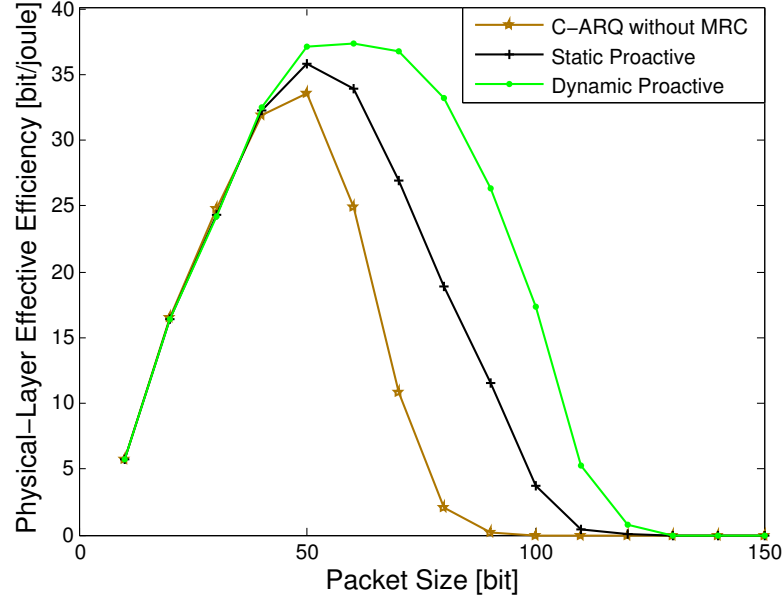


Abbildung 4.23: Physical-layer energy efficiencies of C-ARQ protocols under different MRC assumptions.

problem, here we don't consider MRC. Hence, the outage probabilities of a (initial) transmission and retransmissions are the same (assume the packet size to be fixed).

For a given outage probability constraint for each transmission or retransmission, it is possible that the number of active relays is higher than what the system really needs to satisfy the constraint. Hence the system is possible to waste energy due to having overmuch number of relays for forwarding. The idea to improve the PEE is motivated by reducing this potential waste of energy.

Tradeoff Problem Statement

In order to reduce this waste of energy, we propose a simple 'protocol' solution by introducing a weight factor ω . The factor ω might reduce the number of forwarding relays by moderating the relationship between two probabilities of following events: fail decoding the packet and keeping silence in relaying phase. Denote probabilities of these two events by P_1 and P_X . Unlike previous sections which treat the two probabilities as

4 Relaying Performance in the Infinite Blocklength Regime

the same, now we propose to increase the probability of a relay keeping silent during the relaying frame by the factor ω .

$$P_X(\omega) = \omega \cdot P_1; \quad \omega \in [1, 1/P_1]. \quad (4.34)$$

Therefore, the number of forwarding relays n becomes the binomial distributed based on probability P_X : i.e., $n \sim B(J, P_X)$. Thus, the expected value of n is given by:

$$E[n] = [1 - P_X(\omega)] \cdot J. \quad (4.35)$$

As a result, the outage probability of a two-hop transmission/retransmission changes from equation (4.10) into:

$$P^{\text{out}}(\omega) = \sum_{n=0}^J P_2(n) P_B(n; J, P_X(\omega)). \quad (4.36)$$

Denote the QoS requirement of outage probability as P_{QoS} . Based on (5.37) as well as (4.32), our tradeoff problem statement can be given by:

$$\begin{aligned} \max \Phi_p(\omega) &= \frac{\rho(1 - P^{\text{out}})}{\frac{T_i}{3} (p_{\text{tx}} + E[n]p_{\text{tx}} + p_{\text{fb}}) + (J + 2) \cdot P_c T_f}, \\ \text{s.t. } P^{\text{out}}(\omega) &\leq P_{\text{QoS}}; \quad \omega \in [1, 1/P_1] \end{aligned} \quad (4.37)$$

where $E[n]$ and P^{out} are function of ω .

Tradeoff Analysis

Now we analyze how the tradeoff factor ω affects the system's energy efficiency and how to solve the tradeoff problem.

Based on the definition of ω , a high value of ω leads to a small (average) number of forwarding relays. This results in a low received SNR at the destination. In other words, the outage probability of a (single) transmission/retransmission is strictly increasing in ω . Thus, the minimum of outage probability $P_{\text{min}}^{\text{out}}$ can be achieved when $\omega = 1$, i.e., $P_{\text{min}}^{\text{out}} = P^{\text{out}}(\omega = 1)$. With a outage probability constraint P_{QoS} (per each transmis-

sion/retransmission) of the served transmission, there are two different cases that need to be distinguished regarding the constraint of the tradeoff problem:

- $P_{\text{QoS}} < P_{\text{min}}^{\text{out}}$, the system can't support the transmission already. There is no energy waste. Hence, the tradeoff process is refused.
- $P_{\text{QoS}} \geq P_{\text{min}}^{\text{out}}$, the energy efficiency of the system is possible to be improved by the tradeoff process.

Regarding the first case, we cannot save any energy from a system which is already not able to support the QoS of the transmission by the maximum power consumption. In the following, we analyze our tradeoff problem for the second case.

Based on (4.37), we observe that both the numerator and denominator of the objective function are monotonic decreasing in ω while differing their slopes. With a derivative $-P_1 p_{\text{tx}} J T_f / 3$ the denominator decreases in a linear way. In addition, the derivative of the numerator is $-\frac{\partial P^{\text{out}}}{\partial \omega}$, where $\frac{\partial P^{\text{out}}}{\partial \omega}$ is given by:

$$\frac{\partial P^{\text{out}}}{\partial \omega} = \frac{\partial P^{\text{out}}}{\partial P_X} \frac{\partial P_X}{\partial \omega} = \frac{\partial P^{\text{out}}}{\partial P_X} P_1, \quad (4.38)$$

where $\frac{\partial P^{\text{out}}}{\partial P_X}$ can be obtained by equation (4.39):

$$\frac{\partial P^{\text{out}}}{\partial P_X} = \sum_{n=0}^J \left\{ P_2(n) \binom{J}{n} P_X^{J-n} (1 - P_X)^n \left[1 + \left(\frac{J-n}{P_X} - \frac{n}{1-P_X} \right) \left(\frac{P^\circ}{J p_{\text{tx}}} + 1 - P_X \right) \right] \right\} \quad (4.39)$$

Based on the comparison between derivatives of the numerator and the denominator of Φ with respect to ω , we have some preliminary conclusions:

- if $\frac{\partial P^{\text{out}}}{\partial \omega} > P_1 p_{\text{tx}} J T_f / 3$, then Φ is strictly decreasing.
- if $\frac{\partial P^{\text{out}}}{\partial \omega} < P_1 p_{\text{tx}} J T_f / 3$, then Φ is strictly increasing.
- if $\frac{\partial P^{\text{out}}}{\partial \omega} = P_1 p_{\text{tx}} J T_f / 3$, then Φ achieves an extreme value by one point of ω or a stationary value over a continuous interval of ω .

Denote for short that $\ell(P_X, n) = 1 + \left(\frac{J-n}{P_X} - \frac{n}{1-P_X} \right) \left(\frac{P^\circ}{J P_R} + 1 - P_X \right)$, then we can simplify equation (4.39) into:

$$\frac{\partial P^{\text{out}}}{\partial P_X} = \mathbb{E}_{n \sim B(J, P_X)} [P_2(n) \ell(P_X, n)], \quad (4.40)$$

where $\mathbb{E}_{n \sim B(J, P_X)}[X(n)]$ is the expectation of $X(n)$ while the random variable n is binomially distributed. Similarly, P^{out} can be expressed as $\mathbb{E}_{n \sim B(J, P_X)}[P_2(n)]$.

Consider the following proposition:

Proposition 1. *Let $g_1: \mathbb{R} \rightarrow \mathbb{R}$, $y \rightarrow g_1(y)$, with $g_1(y) \geq 0 \forall y$. Let $g_2: \mathbb{R}^2 \rightarrow \mathbb{R}$, $(x, y) \rightarrow g_2(x, y)$ be monotonically decreasing in x . If $f: \mathbb{R}^2 \rightarrow \mathbb{R}$, $(x, z) \rightarrow f(x, z)$ is monotonically decreasing in z and strictly increasing in x , then $f(x, g_1(y)g_2(x, y))$ is also strictly decreasing in x .*

Beweis. See Appendix A.2. □

Now let us apply the above Proposition into our tradeoff problem. First, as a probability, $P_2(n)$ is non-negative. Second, according to the earlier analysis, $\mathbb{E}_{n \sim B(J, P_X)}[P_2(n)]$ is a strictly increasing function of P_X . For a given P_X , $\mathbb{E}_{n \sim B(J, P_X)}[P_2(n)]$ has a fixed value, which can be treated as special monotonically decreasing function of $\text{Pr}_2(n)$. Third, $\frac{\partial \ell}{\partial P_X} = -(J-n) - \left(\frac{P^\circ}{Jp_{\text{tx}}} + 1\right) \frac{J-n}{(P_X)^2} - \frac{nP^\circ}{JP_R(1-P_X)^2} \leq 0$ since the forwarding relay number n satisfies that: $0 \leq n \leq J$. Hence $\ell(P_X, n)$ is a monotonically decreasing function of P_X . Therefore, based on the proposition above, $\frac{\partial P^{\text{out}}}{\partial P_X}$ is a strictly increasing function of P_X . Consider that the relationship between P_X and ω in equation (4.38), we can further figure out a very important analysis result that $\frac{\partial P^{\text{out}}}{\partial \omega}$ is also strictly increasing in ω . This means that if equation $\frac{\partial P^{\text{out}}}{\partial \omega} = P_1 p_{\text{tx}} J T_f / 3$ has a solution ω° , then the solution ω° is unique. Combined with our former analysis, we have:

- if $\omega > \omega^\circ$, Φ is strictly decreasing in ω .
- if $\omega < \omega^\circ$, Φ is strictly increasing in ω .
- if $\omega = \omega^\circ$, Φ achieves the unique extreme value: the global maximum.

Take into account that our tradeoff is subject to the target statistical outage probability, the extreme point ω° is the solution of the tradeoff only if the constraint is satisfied by the extreme point, i.e., $P^{\text{out}}(\omega^\circ) \leq P_{\text{QoS}}$. Otherwise, the solution of the tradeoff problem is not the extreme point anymore but the boundary solution $\omega^\bullet$ which can be obtained by the QoS edge equation:

$$P^{\text{out}}(\omega^\bullet) = P_{\text{QoS}}; \omega^\bullet \in [1, \omega^\circ] . \quad (4.41)$$

And $\omega^\bullet$ is easy to be searched out since the objective function (energy efficiency) is strictly increasing in the variable ω during the interval $[1, \omega^\circ]$. The process of solving the tradeoff problem is also summarized in Figure 4.24.

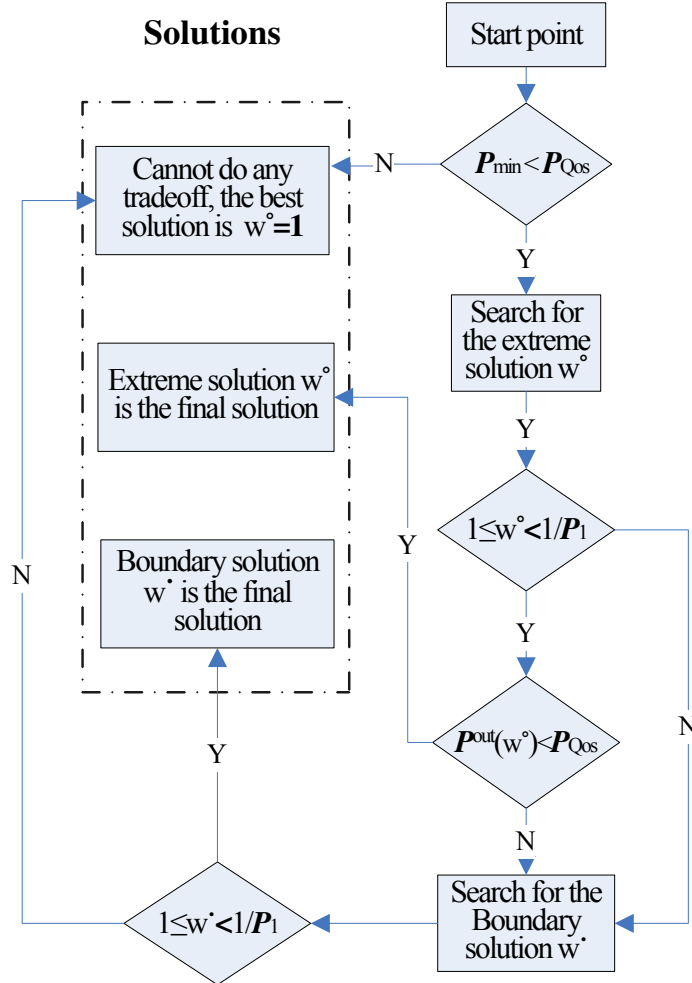


Abbildung 4.24: The flow chart of the process of solving the tradeoff problem

Evaluation

In order to validate the above analysis of the tradeoff problem, we show by simulation how the factor ω makes tradeoff between the energy efficiency and the outage probability. The simulation is based on the model parameterization which is same as the simulations in Section 4.3.

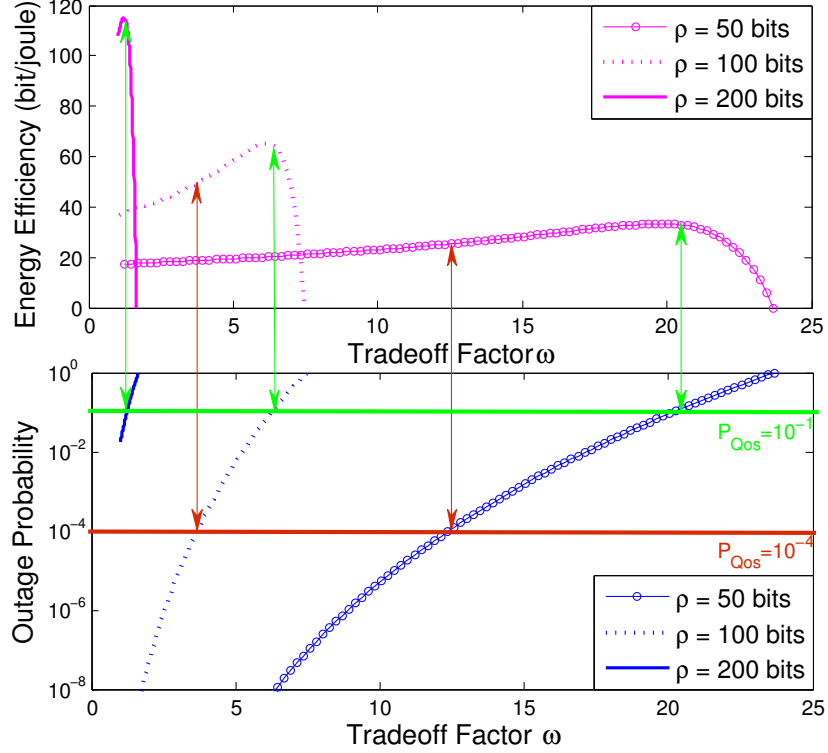


Abbildung 4.25: The energy efficiency and the outage probability on tradeoff factor

Figure 4.25 presents energy efficiency (top) and outage probability (bottom) versus the factor ω under three cases with different packet sizes ρ . We observe that every energy efficiency curve has an extreme value over ω , and that each curve first increases and then decreases after the extreme value. This perfectly matches our theoretical analysis. In the bottom figure these outage probability curves also validate our analysis that outage probability is strictly increasing in ω .

Moreover, in Figure 4.25 we also consider two outage probability constraints: A red threshold line with outage constraint $P_{\text{QoS}} = 10^{-4}$ and a green threshold line with outage constraint $P_{\text{QoS}} = 10^{-1}$. These two threshold lines intersect three outage

probability curves. For the case $\rho = 200$ bits with a strict QoS constraint $P_{\text{QoS}} = 10^{-4}$, there is no intersection between the probability curve and the threshold line. This means that with such a big packet size the system cannot satisfy this QoS constraint at all no matter which value of ω is set. Apart from this, there are 5 intersections in the bottom part, and we marked the corresponding energy efficiency value of the intersections by additional arrows.

For each packet size, the feasible interval of ω is from 1 to the intersection of the outage probability curve of the case and the threshold line of the constraint (in order to satisfy a QoS constraint). However, we find that not all the extreme points of ω are contained by this feasible interval. For example, with $\rho = 100$ bits the energy efficiency achieves its maximum when $\omega = 6$ which is out of the effective interval $[1, 3.7]$ under the strict QoS constraint (red line and arrows) but in the feasible interval $[1, 6.3]$ under the loose QoS constraint (green line and arrows). In fact, for all the cases the maximum of energy efficiency can be achieved only if the QoS constraint is not strict, otherwise the tradeoff only has boundary solutions (intersections of the objective function and the constraint function). This actually matches our analytical model and also the flow chart in Figure 4.24.

The solution and the relationship between outage probability and energy efficiency are shown in Figure 4.26. It shows more clearly that the maximum of energy efficiency can only be achieved when the QoS constraint is not strict.

Summary

In this section 4.5.2.3, we introduced a weight factor based on which we reduce the probability of a relay joining forwarding. We showed that this operation is able to improve the energy efficiency of the studied relaying system when the QoS constraint is not strict. To simplify the problem, we did not consider the MRC in this section. However, we will extend the idea of above operation into C-ARQ protocols with MRC in Section 4.5.3.3 where we improve the higher-layer energy efficiency of the system.

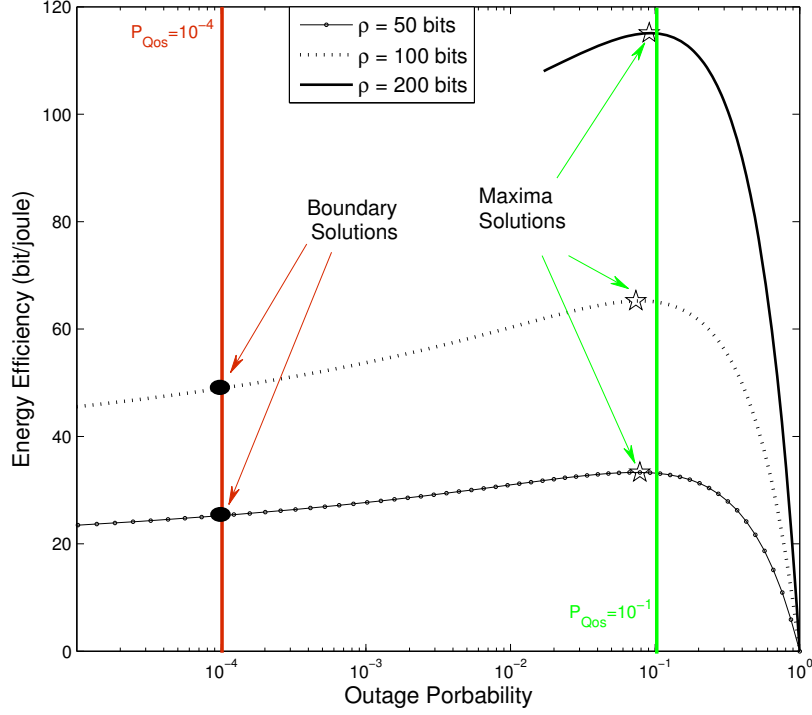


Abbildung 4.26: The relationship between energy efficiency and outage probability

4.5.3 Higher-layer Energy Efficiency– Effective Energy Efficiency

The higher-layer (QoS-constrained) energy efficiency of C-ARQ protocols are investigated in this section by considering the ratio between the MSDR (under QoS constraints $\{d, \Pr_d\}$) and the related average energy consumption. We refer to this ratio as Effective Energy Efficiency (EEE).

4.5.3.1 The Definition of Effective Energy Efficiency (EEE)

Recall that the MSDR is the maximum arrival rate at the source that can be supported by the system under certain QoS constraints and is derived by upper bounding K (the probability of non-empty buffer) by 1. Hence, the related energy consumption of the system when the arrival rate at the source equals the MSDR can be obtained based on (4.24) by considering $K = 1$. Therefore, the EEE is given based on (5.30) and (4.24)

by:

$$\Phi_h = \frac{0.5 \mathbb{E}[s_i] + 0.5 \sqrt{(\mathbb{E}[s_i])^2 + \frac{2 \ln(\text{Pr}_d)}{d} \cdot \text{Var}[s_i]}}{\bar{\Psi}_T + (J + 2) \cdot p_c T_f}. \quad (4.42)$$

The EEE differs from physical-layer energy efficiency metrics in that it directly demonstrates how effective the consumed energy is on providing the delay-constrained service. For example, let us consider a system which consumes low energy but supports a high data rate for a service with certain delay constraint. However, delays of most transmitted packets violate the QoS constraint. Therefore, the EEE of the system is rather low while the physical-layer energy efficiency has a more higher value. In fact, if system A has a higher EEE than system B, this means either that system A is able to use less energy to support the same service (with the same data arrival rate and the same QoS requirements) compared to system B or that while consuming the same energy system A is able to transmit more bits under the QoS requirements.

4.5.3.2 Performance Evaluation regarding EEE

In the previous sections, we have shown that C-ARQ protocols significantly improve the reliability and the throughput while introducing additional delay and energy consumption. Therefore, a central question regarding C-ARQ protocols is how a C-ARQ protocol performs and makes tradeoff among throughput, delay and energy consumption. By introducing EEE which combines these metrics, we facilitate evaluating the behavior of a C-ARQ protocol regarding the above complex multivariate tradeoff. We show the EEE performance of studied protocols in the following. These simulation results are based on the model parameterization which is same as the simulation in Section 4.3.

Firstly, the EEE performance of each C-ARQ protocol with MRC is shown in Figure 4.27. Recall that (in these C-ARQ protocols with MRC) proactive protocols consume more energy (shown in Figure 4.20) but have higher MSDRs (shown in Figure 4.16) in comparison to reactive protocols. The combination of these two observations raises the question which of these protocols is more energy-efficient? This question is addressed now in Figure 4.27 with respect to a scenario where distances from the source to relays and from relays to the destination are the same. We find that the dynamic proactive protocol is the most energy-efficient one. In addition, a dyna-

mic protocol always significantly outperforms a static protocol no matter whether it is combined with a reactive protocol or a proactive protocol.

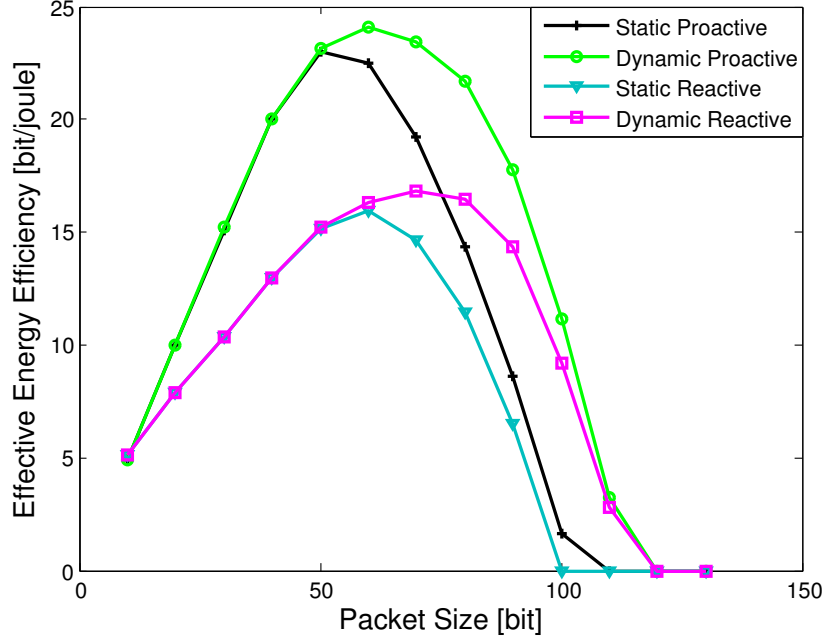


Abbildung 4.27: Higher-layer energy efficiencies of C-ARQ protocols with MRC.

Subsequently, we provide the EEE performance of the C-ARQ protocol without MRC and these two proactive C-ARQ protocols (with MRC) in Figure 4.28. The figure shows that applying MRC significantly improves the C-ARQ performance regarding EEE. In addition, based on Figure 4.27 and Figure 4.28 we find that the EEE performance of each C-ARQ protocol is first increasing and then decreasing in the packet size. In other words, the EEE performance can be maximized by choosing an appropriate packet size. Moreover, by comparing Figure 4.27 and Figure 4.28 with Figure 4.22 and Figure 4.23 we observe that for each C-ARQ protocol the PEE is much higher than the EEE. This actually matches with our analysis that in comparison to the PEE the EEE is subject to additional QoS constraints, i.e., $\{d, P_d\}$ delay and delay violation probability. In particular, if these QoS constraints are arbitrarily loose, i.e., $\{d = +\infty, P_d = 1\}$, then the PEE equals the EEE. However, this is obviously not the case in practice.

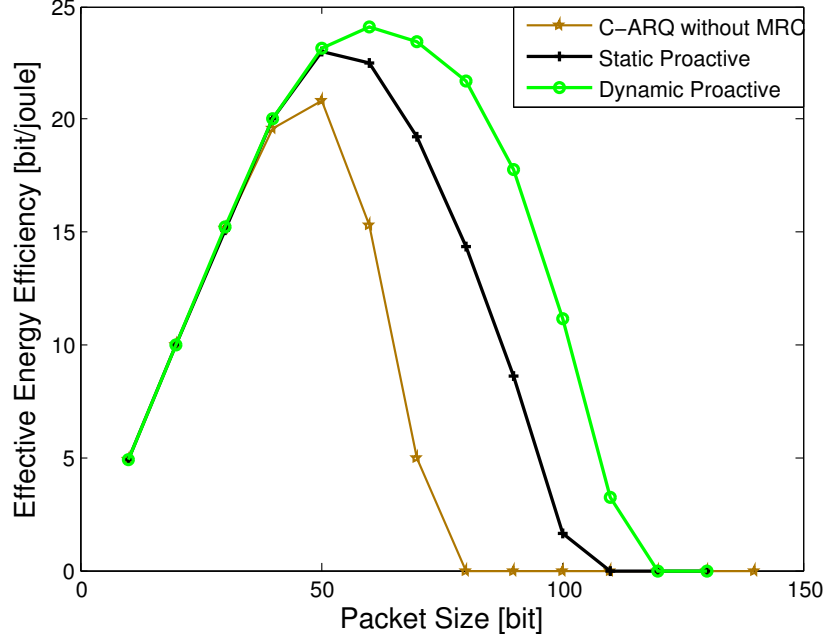


Abbildung 4.28: Higher-layer energy efficiencies of C-ARQ protocols under different MRC assumptions.

4.5.3.3 The Performance Improvement of EEE

Based on the evaluation of these four protocols with MRC, we observed that static protocols consume less energy while dynamic protocols lead to higher MSDRs. This observation indicates that the EEE is likely to be improved by introducing a tradeoff between the two schemes. We therefore propose a simple protocol which considers this tradeoff. The idea of this tradeoff protocol is similar with the improvement of PEE in Section 4.5.2.3.

The proposed tradeoff protocol is described as follows. Different from the dynamic protocols which require all the passive relays to overhear the forwarding frame and then participate in the retransmission, we introduce a probability ω , ($\omega \in [0, 1]$). In the proposed protocol, if a passive relay decodes the packet successfully by overhearing, it participates in the retransmission only with probability $1 - \omega$. Hence, the performance of the proposed protocol is similar to a static protocol when the value of ω is close to 1, and similar to a dynamic protocol if ω is close to 0. So it makes a tradeoff between static protocols and dynamic protocols while varying ω in the interval $[0, 1]$.

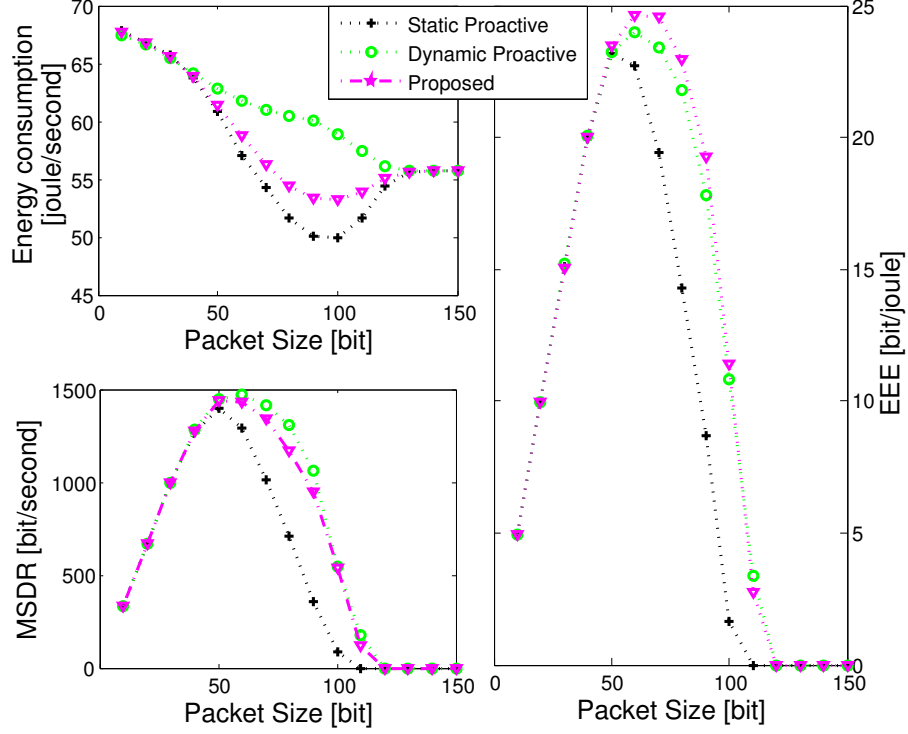


Abbildung 4.29: The energy consumption, MSDR and EEE performance of proposed protocol with $\omega = 0.2$.

In numerical experiments (which is not shown here), we observe that while varying the ω from 1 to 0 the EEE performance of the tradeoff protocol initially outperforms the static protocols and later on even exceeds the dynamic protocol performance, before the EEE decreases again for small values of ω . In Figure 4.29 where ω is set to 0.2, we show that the proposed protocol has slightly lower MSDR but much lower energy consumption than the proactive protocols. As a result, the proposed protocol improves the EEE performance by 4% (when the packet size is around 70 bits).

4.6 Performance Evaluation of C-ARQ Protocols

In this section, we continue to evaluate these C-ARQ protocols while the focus is to derive guidelines for the design of efficient C-ARQ protocols. As we have shown in previous sections that applying MRC is essential for C-ARQ protocols, in this section

we only consider these C-ARQ protocols with MRC. As multi-relay C-ARQ protocols exploit the diversity of multiple relays, the performance of these protocols are subject to the number of relays present in the system. In addition, the performance of reactive protocols are influenced by the channel quality of the direct link which is further subject to the distance between the source and the destination. Hence, these considered C-ARQ protocols are expected to show different performances while varying the system topology. To evaluate these performances is important for the design of effective C-ARQ systems.

4.6.1 Evaluation of the System Performance on Relay Numbers:

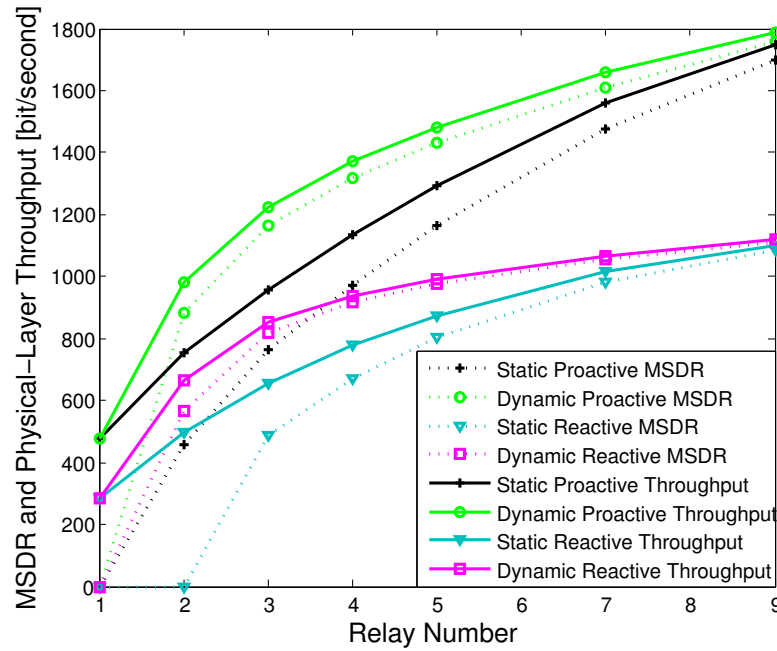


Abbildung 4.30: MSDR and physical-layer throughput with dynamic number of relays deployed in the system. The QoS parameters are $\{d, P_d\} = \{20T_f, 10^{-3}\}$.

We show in Figure 4.30 MSDRs and physical-layer throughputs of C-ARQ protocols with a varying number of relays present in the system. It is observed that deploying more relays in the system monotonically increases the throughput and the MSDR for each protocol. In addition, we show the PEE and EEE performance of these protocols in

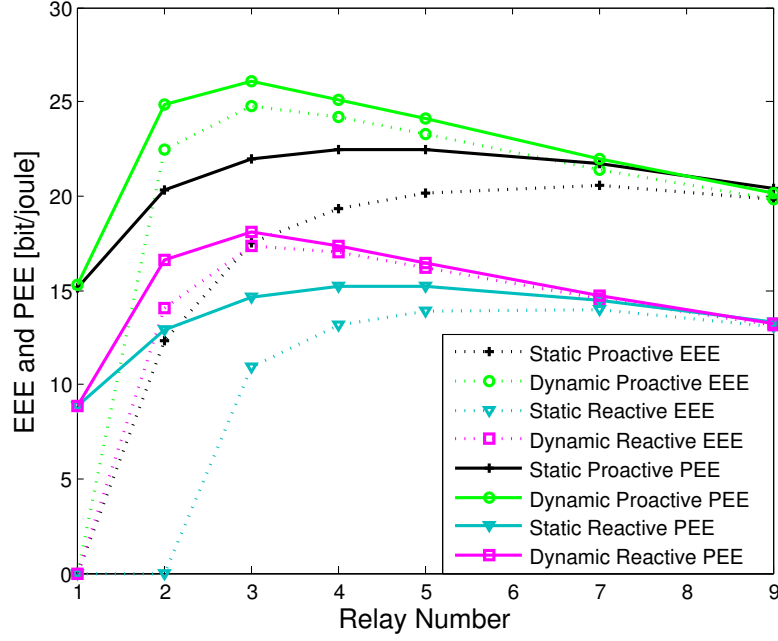


Abbildung 4.31: EEE and PEE of the protocols with dynamic number of relays deployed in the system. The QoS parameters are $\{d, P_d\} = \{20T_f, 10^{-3}\}$.

Figure 4.31 where we also vary the number of relays deployed the system. We observe that deploying more relays is not always beneficial for EEE. Instead, as shown in Figure 4.31 EEE curves are quasi-convex in the number of relays deployed the system.

Moreover, from Figure 4.30 and Figure 4.31 we observe significant gaps between MSDRs and throughputs as well as between EEEs and PEEs. In particular, all the EEE curves (dotted lines) are zero when the system only has one relay, as shown in Figure 4.31. This means that the system cannot support corresponding QoS requirements at all and hence the MSDR becomes 0. In other words, with only one relay the system consumes energy but cannot transmit any packets under these QoS requirements. At the same time, none of PEE curves (solid lines) has value zero when the relay number is one. Hence, for a QoS-support system, the PEE is not a good metric to evaluate the efficiency of the system. Instead, the EEE has to be considered, as the proposed EEE directly demonstrates how efficient the system spends energy for these QoS requirements of interest.

4.6.2 Performance Loss due to Relay Number Uncertainty

Previous evaluations (e.g., in Figure 4.27) show that the packet size is one of the major concerns for maximizing the EEE. Actually, when the number of relays in the system varies, we find that the optimal packet size changes accordingly. However, if

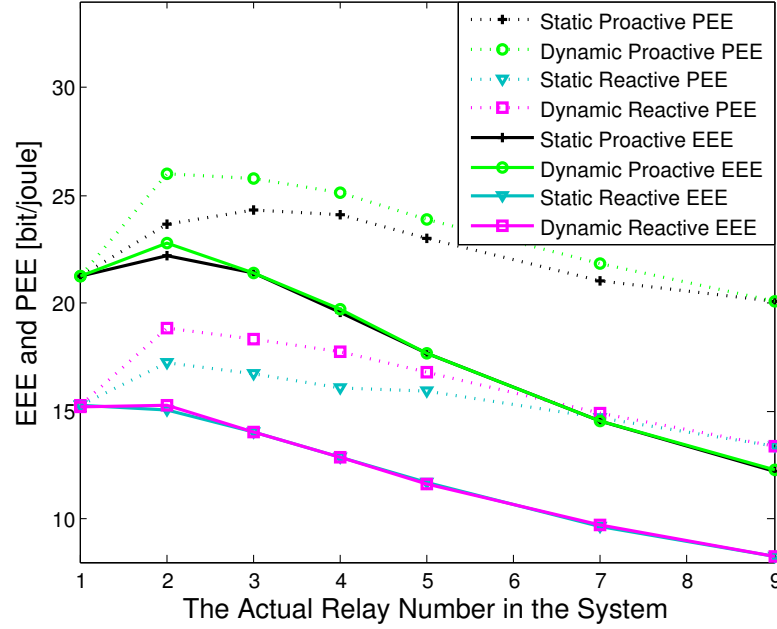


Abbildung 4.32: EEE Performance Loss due to relay number uncertainty.

the source only knows that there is a relay group (at least one relay) in the system but does not know the exact number of these relays, the source could determine the packet size according to the pessimistic assumption that there is only one relay in the system in order to make the transmission reliable. This leads to a performance loss of the system, if there are in fact more relays. We are interested in this performance loss and illustrate this in Figure 4.32. Curves with solid lines represent the situation that the source determines the packet size according to the pessimistic assumption while the other curves with dotted lines represent the situation that the source knows the exact relay number. The figure shows that the performance loss is about 20% for a system with two relays and is about 40% for a system which has nine relays. Therefore, we conclude that a C-ARQ protocol should spend the required overhead to determine the exact number of relays in the system first.

4.6.3 Optimality of Dynamic Protocols for Different Topologies

Based on above numerical results, dynamic protocols always show better EEE performance than static protocols. Hence, we further compare the two dynamic protocols to determine the optimal one in these studied protocols.

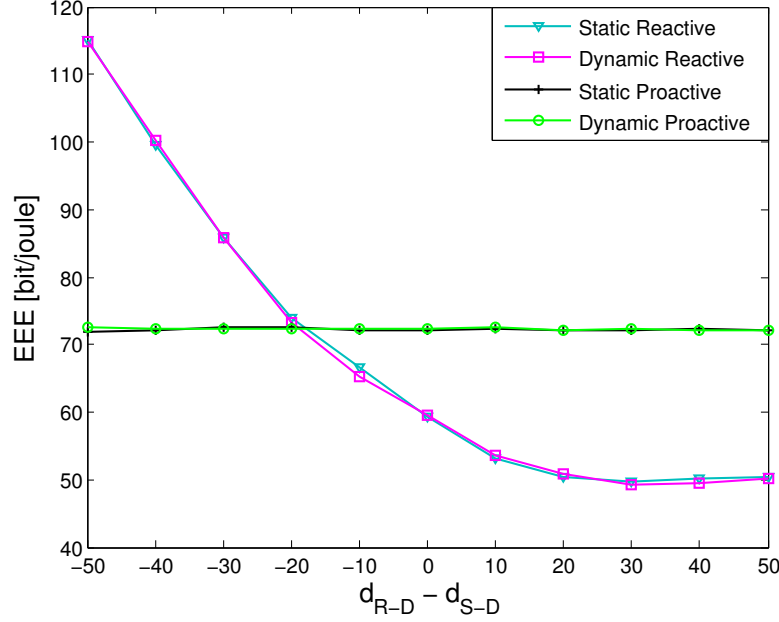


Abbildung 4.33: The EEE comparison between the dynamic reactive protocol and the dynamic proactive protocol while varying the location of the destination.

Since a reactive protocol uses the direct link transmission as the initial transmission, the performance of it is subject to the system topology, especially the distance of the direct link d_{S-D} . To compare the dynamic reactive protocol with the dynamic proactive protocol, we vary the system topology in the following way: While fixing locations of the source and relays and fixing the distance d_{R-D} from the center of relay group to the destination (in the simulation, d_{S-R} and d_{R-D} are set to 50m and 100m), we vary the location of the destination by circularly moving it around the relay group. Then, we obtain the comparison between the dynamic proactive protocol and the dynamic reactive protocol, as shown in Figure 4.33. In the figure, the x-axis is the distance

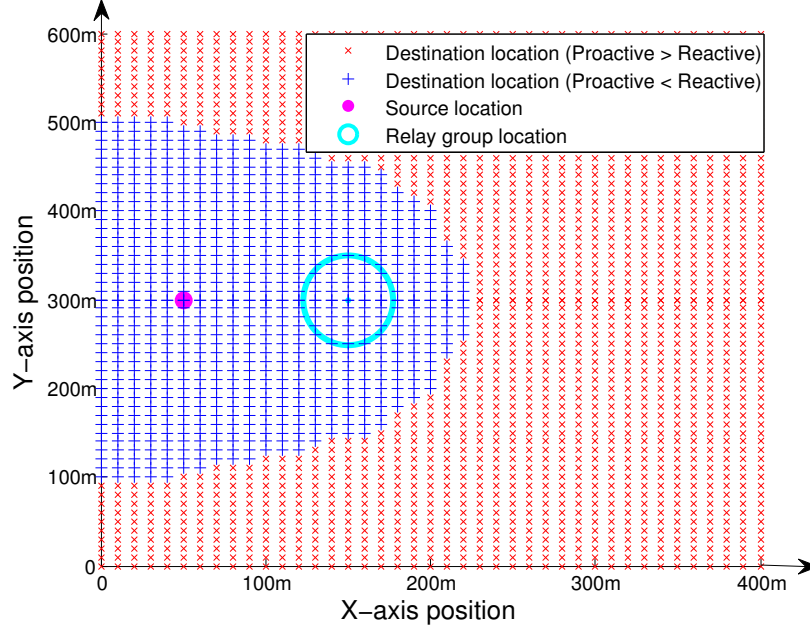


Abbildung 4.34: The EEE comparison between proactive protocols and reactive protocols, where locations of the source and relays are fixed.

gap between d_{R-D} and d_{S-D} , which indicates whether the destination is more close to relays or to the source.

We observe that even at the zero point of the X-axis ($d_{R-D} - d_{S-D} = 0$), where the source-destination distance is the same as the relays-destination distance, using direct link as the initial transmission attempt (reactive protocols) has still a lower performance than using relaying (proactive protocols).

Spatially, we show the comparison between proactive protocols and reactive protocols again in Figure 4.34. The reactive protocols are superior to the proactive protocols only if the destination is located in a rather small range around the source.

Summary

In this section, we combined the results of capacity and throughput (in the previous section) with energy consumption to study the energy efficiency. By simulation, we concluded several guidelines for the design of efficient C-ARQ protocols by maximizing the energy efficiency.

4.7 Summary and Discussion

In this section, we briefly summarize the work that has been done in this Chapter and provide discussions regarding the work.

Summary

In this chapter, in the infinite blocklength regime we developed a systematic investigation model of relaying networks with retransmissions. Initially, we investigated the outage probability and derived the distribution of transmission delay. Subsequently, the physical-layer throughput and the higher-layer effective capacity were analyzed. Later on, we analyzed the physical-layer as well as the higher-layer energy efficiency by considering the PEE and the EEE which are ratios of the throughput to the energy consumption and of the effective capacity to the energy consumption. In particular, the proposed EEE is proved to be a good metric in investigating the tradeoff among throughput, delay, energy consumption". Moreover, by simulation we validated our analytical model. These are our contributions regarding the system performance analysis.

We also made several contributions to the system design. First, by introducing weight factors to the probabilities of a relay joining in forwarding and re-forwarding, we improved the PEE as well as the EEE, respectively. More importantly, we evaluated these considered C-ARQ protocols and concluded several guidelines for the design of efficient C-ARQ protocols in the infinite blocklength regime, which are as follows:

- MRC should be applied in C-ARQ protocols.
- Direct link transmission can usually be ignored by C-ARQ protocols unless a specific topology of the nodes is given.
- There exists an optimal packet size to maximize the EEE or the PEE of a C-ARQ protocol.
- Dynamic relay selection should be enabled as it leads to a better energy efficiency, although they consume more energy in comparison to static protocols.
- The energy efficiency of C-ARQ can be improved by introducing a tradeoff between the dynamic protocol and the static protocol.

- There also exists an optimal number of relays deployed in the system when maximizing the EEE or the PEE under a C-ARQ protocol.
- Overhead should be spent to acquire the exact number of relays present in the system.

Discussion

The work in this chapter is an important component of the thesis. As introduced in Chapter 1, in this thesis we develop fundamental investigation and evaluation models for relaying networks under various scenarios with different assumptions regarding CSI and blocklength. The infinite blocklength regime we considered in this chapter corresponds the assumption of infinite blocklengths and average CSI. Although infinite blocklengths vanishes noise errors, the transmission with only the average CSI is likely unsuccessful due to outage. However, we found that the probability of this outage can be significantly reduced by relay-assisted retransmissions. By making more relays join in the retransmission or by deploying more relays in the system, the throughput is boosted. These are some general findings in this chapter regarding relaying in the infinite blocklength regime. Then, some questions immediately arise: What is the error model of relaying in the finite blocklength regime? In particular, what is the error model if there is only average CSI available at the source, which means the transmission suffers from both noise and the unpredictable random channel fading? With a finite block, does relaying pay off more or pay off less? Moreover, what is the blocklength-limited performance of relaying if we introduce more relays to the system? These questions will be addressed in the following two chapters.

5 Relaying Performance in the Finite Blocklength Regime

The relaying performance with finite blocklengths are considered in this chapter as well as the next chapter. Similar to the previous chapter which considers the infinite blocklength regime, in this chapter we also develop an analytical performance model of relaying with finite blocklengths. Moreover, we investigate the relaying performance under several scenarios which differ in the channel model and in the assumption of CSI. Under these scenarios, we first derive the overall error probability of relaying with a finite blocklength. Subsequently, we investigate the Blocklength-Limited capacity (BL-capacity) and the effective capacity of relaying. Finally, the related energy efficiencies PEE and EEE are studied. As mentioned in Section 2.3, the derivations regarding the relaying performance in the finite blocklength regime in this work is facilitated by the accurate approximation of the coding rate under the p2p scenario which initially provided in reference [8].

5.1 System Model

5.1.1 System General Description

We consider a simple relaying scenario with a source, a destination and a decode-and-forward (DF) relay as schematically shown in Figure 5.1. Links between the above transceivers are referred to as the direct link (from the source to the destination), the backhaul link (from the source to the relay) and the relaying link (from the relay to the destination). In general, we assume the direct link to be much weaker than the backhaul link as well as the relaying link. The entire system operates in a slotted fashion

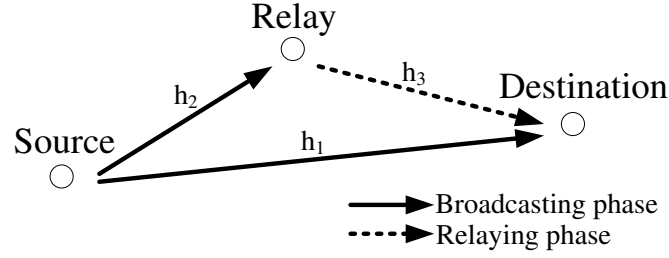


Abbildung 5.1: Example of the considered single relay system scenario.

where time is divided into transmission periods. Each transmission period has length $2m$ (symbols) and contains two frames (each frame with length m) which correspond to two hops of relaying. The blocklength of the coding over the channel in each frame is as long as the frame length m .

5.1.2 Channel Model and CSI Assumptions

We consider several scenarios which differ in if channels are static and in if the perfect CSI or the average CSI is available at the source. In total three scenarios are actually considered in this chapter as shown in Figure 5.2 (note that under static channels the average CSI and the perfect CSI are the same). Next, we describe these two channel

		Channel model	
		Static channel model	Quasi-static channel model
CSI assumption	Perfect CSI at the source	Perfect CSI scenario with static channels	Perfect CSI scenario with quasi-static channels
	Average CSI at the source		Average CSI scenario with quasi-static channels

Abbildung 5.2: Three scenarios considered in this chapter.

models of the studied relaying system respectively.

Static Channel model

Under the static model, channels at different transmission periods are the same. We denote these static channel vectors of the S-D direct link, the S-R backhaul link and the R-D access link by h_1 , h_2 and h_3 . In addition, the transmit power at the relay and the source is denoted by p_{tx} . Hence, received signals at the destination and the relay in a broadcasting frame are given by:

$$\mathbf{y}_1 = h_1 \mathbf{x} + \mathbf{n}_1, \quad (5.1)$$

$$\mathbf{y}_2 = h_2 \mathbf{x} + \mathbf{n}_2. \quad (5.2)$$

Next, if the data is decoded correctly and forwarded by the relay, the received signal at the destination in a relaying frame is given by:

$$\mathbf{y}_3 = h_3 \mathbf{x} + \mathbf{n}_3. \quad (5.3)$$

The transmitted signal $\mathbf{x}$ and received signals $\mathbf{y}_1$, $\mathbf{y}_2$ and $\mathbf{y}_3$ are complex m -dimensional vectors. Furthermore, $\mathbf{n}_k, k = 1, 2, 3$ represents the Gaussian noise vector $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$, where $\mathbf{I}_m$ denotes an $m \times m$ identity matrix. Denote the gains of these channels as z_1 , z_2 and z_3 , where $z_k = |h_k|^2, k = 1, 2, 3$. In addition, we assume the destination to apply maximum ratio combining (MRC), where the combined channel gain is given by $z_1 + z_3$.

Quasi-Static Channel model

Under the quasi-static channel model, channels vary from one transmission period to the next. Hence, different from the static channel model the denotation under the quasi-static model must take the index of the transmission period into account. During a transmission period i , first a broadcasting frame is employed, followed by a relaying frame. Under quasi-static channel model, received signals at the destination and the relay in the broadcasting frame of transmission period i are given by:

$$\mathbf{y}_{1,i} = p_{\text{tx}} h_{1,i} \mathbf{x}_i + \mathbf{n}_{1,i}, \quad (5.4)$$

$$\mathbf{y}_{2,i} = p_{\text{tx}} h_{2,i} \mathbf{x}_i + \mathbf{n}_{2,i}. \quad (5.5)$$

Next, if the data is decoded correctly and forwarded by the relay, the received signal at the destination in the relaying frame of transmission period i is given by

$$\mathbf{y}_{3,i} = p_{\text{tx}} h_{3,i} \mathbf{x}_i + \mathbf{n}_{3,i}. \quad (5.6)$$

The transmitted signal $\mathbf{x}_i$ and received signals $\mathbf{y}_{1,i}$, $\mathbf{y}_{2,i}$ and $\mathbf{y}_{3,i}$ are also complex m -dimensional vectors. In addition, the noise vectors of these links in transmission period i are denoted by $\mathbf{n}_{1,i}$, $\mathbf{n}_{2,i}$ and $\mathbf{n}_{3,i}$, which are independent and identically distributed (i.i.d.) complex Gaussian vectors: $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m)$, $\mathbf{n} \in \{\mathbf{n}_{1,i}, \mathbf{n}_{2,i}, \mathbf{n}_{3,i}\}$. Moreover, $h_{1,i}$, $h_{2,i}$ and $h_{3,i}$ are the channels (scalars) of the direct link, backhaul link and relaying link during transmission period i , respectively. Under quasi-static Rayleigh fading channels, the channels remain constant within each transmission period (includes two hops/frames) and vary independently from one period to the next. Hence, the instantaneous channel gain of each link has two components, i.e., the average channel gain and the random fading. On one hand, we denote average channel gains (e.g., due to the path loss) of these three links over the transmission periods by $\bar{z}_1$, $\bar{z}_2$ and $\bar{z}_3$. On the other hand, we assume these channels to experience Rayleigh fading where the envelope $\sqrt{z}$ of the channel fading (response) is Rayleigh distributed with the probability density function (PDF):

$$f(\sqrt{z}) = 2 \frac{\sqrt{z}}{\sigma^2} e^{-z/\sigma^2}. \quad (5.7)$$

Typically, the channel fading is modeled as having unit power gain. This requires that $\sigma^2 = 1$. Hence, the PDF of z , the gain due to Rayleigh fading, is given by the exponential distribution: $f(z) = e^{-z}$. We denote by $z_{j,i}$ ($j = 1, 2, 3$) the gains due to Rayleigh fading (during transmission period i) of the direct link, the backhaul link and the relaying link. Hence, we have the instantaneous channel gain $|h_{j,i}|^2 = z_{j,i} \bar{z}_j$, $j = 1, 2, 3$. Moreover, these channel fading gains at different links during the same transmission period are assumed to be i.i.d.

At last, the destination is also assumed to apply maximal ratio combining (MRC) where the combined channel gain is given by $|h_{1,i}|^2 + |h_{3,i}|^2$. Thus, the received signal-to-noise ratio (SNR) at the relay and the combined SNR at the destination in trans-

mission period i are given by $\gamma_{2,i} = \frac{|h_{2,i}|^2 p_{\text{tx}}}{\sigma^2} = \frac{z_{2,i} \bar{z}_2 p_{\text{tx}}}{\sigma^2}$ and $\gamma_{\text{MRC},i} = \frac{(|h_{1,i}|^2 + |h_{3,i}|^2) p_{\text{tx}}}{\sigma^2} = \frac{(z_{1,i} \bar{z}_1 + z_{3,i} \bar{z}_3) p_{\text{tx}}}{\sigma^2}$.

5.2 Methodology

In the finite blocklength regime, the error probability of communication is no longer negligible (in comparison to the infinite blocklength regime). It is possible that decoding errors occur at the destination even if the source has the perfect CSI. Those errors are due to the noise as this noise can not be perfectly averaged based on a finite blocklength. Moreover, if the source does not have the perfect CSI, e.g., only the average CSI is available at the source, it is not able to determine an appropriate coding rate to fit the instantaneous channel. Therefore, it is likely that the coding rate set by the source based on the average CSI is higher than the Shannon capacity of the instantaneous channel. This results in a more significant error probability in comparison to the scenario with perfect CSI.

The error probability is actually an interface between system operation and performance in the finite blocklength regime. On the one hand, the blocklength as well as the coding rate (determined by the source) directly influence the error probability of the transmission. On the other hand, the system performance is strongly subject to the error probability. For example, the BL-capacity is directly subject to the error probability as it is defined as the successfully decoded bits per channel use. Moreover, a data block should be retransmitted after an error. This introduces an additional delay to this block (which influences the higher-layer performance) and consumes more energy. Hence, the error probability plays a very important role in the analysis of this Chapter. The methodology of this chapter is shown in Figure 5.3. We start with the error probability of relaying and then derive the capacity and energy efficiency of relaying. Both the physical-layer performance and the higher-layer performance of relaying are studied.

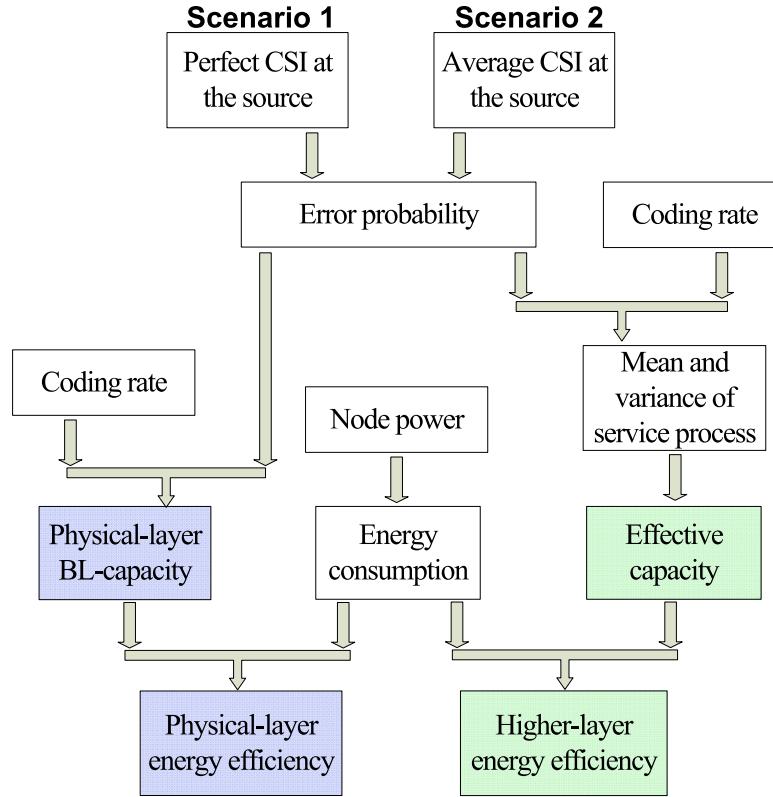


Abbildung 5.3: Methodology of developing investigation models of relaying in the finite blocklength regime.

5.3 Reliability Issue in the Finite Blocklength Regime – Error Probability

5.3.1 Perfect CSI Scenario with Static Channels

For the real additive white Gaussian noise (AWGN) channel, [8, Theorem 54] derives a tight bound for the coding rate of a single-hop transmission system. With blocklength m , block error probability ε and SNR γ , the coding rate (in bits per channel use) is given by: $\frac{1}{2}\log_2(1 + \gamma) - \sqrt{\frac{1}{2m} \left(1 - \frac{1}{(1+\gamma)^2}\right)} Q^{-1}(\varepsilon) \log_2 e + \frac{O(\log_2 m)}{m}$, where $Q^{-1}(\cdot)$ is the inverse Q-function, and as usual, the Q-function is given by:

$$Q(w) = \int_w^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt. \quad (5.8)$$

5.3 Reliability Issue in the Finite Blocklength Regime – Error Probability

In [26], the above result has been extended to a complex channel model with channel gain $|h|^2$, where the coding rate (in bits per channel use) is given by:

$$r = R(|h|^2, \varepsilon, m) \approx C(|h|^2) - \sqrt{\frac{1}{m} \left(1 - 2^{-2C(|h|^2)}\right)} Q^{-1}(\varepsilon) \log_2 e, \quad (5.9)$$

where $C(|h|^2)$ is the Shannon capacity function of a complex channel with gain $|h|^2$:

$$C(|h|^2) = \log_2 \left(1 + \frac{|h|^2 p_{\text{tx}}}{\sigma^2}\right) = \log_2 \left(1 + \frac{z \bar{h}^2 p_{\text{tx}}}{\sigma^2}\right). \quad (5.10)$$

Hence, for a single hop transmission with blocklength m and coding rate r , the decoding (block) error probability at the receiver is given by:

$$\varepsilon = P(|h|^2, r, m) = Q \left(\frac{C(|h|^2) - r}{\sqrt{\frac{1}{m} \left(1 - 2^{-2C(|h|^2)}\right)} \log_2 e} \right). \quad (5.11)$$

As the Q -function is a strictly decreasing function, obviously P is strictly increasing in r and strictly decreasing in either m or $|h|^2$.

In the studied relaying system, as we consider maximum ratio combining at the receiver, the coding rate on different links needs to be the same. Hence, with coding rate r the decoding error probability at the relay is given by $\varepsilon_2 = P(|h_2|^2, r, m) = P(z_2, r, m)$ while the decoding error probability at the destination under maximum ratio combining is given by $\varepsilon_{\text{MRC}} = P(|h_1|^2 + |h_3|^2, r, m) = P(z_1 + z_3, r, m)$.

The coding rate is determined by the source based on the CSI of all the links and in particular the bottleneck link of the system which is either the backhaul link or the combined link. Recall that we assume both of these links to be significantly stronger than the direct link: $\min\{z_2, z_1 + z_3\} \gg z_1$. As the coding rate of a block is chosen based on z_2 or $z_1 + z_3$, the probability of decoding the block correctly at the destination just relying on the much lower (direct link) channel gain z_1 is negligible, which means $P^e(z_1, r, m) \approx 1$. Therefore, the overall error probability ε_{R} of the two-frame relaying equals the probability of the intersection of the following two events: "an error occurs at the relay" and "an error occurs at the destination after maximum ratio combining". This leads to the following proposition:

Proposition 2. *For a relaying network under static channels, with coding rate r and blocklength m in each frame/hop, the overall error probability of relaying is given by:*

$$\begin{aligned}\varepsilon_R &= \varepsilon_2 + (1 - \varepsilon_2) \varepsilon_{\text{MRC}} \\ &= P(z_2, r, m) + [1 - P(z_2, r, m)]P(z_1 + z_3, r, m).\end{aligned}\tag{5.12}$$

Hence, we immediately have a lower bound for ε_R :

$$\varepsilon_R \geq \max\{\varepsilon_{\text{MRC}}, \varepsilon_2\}.\tag{5.13}$$

Equations (5.12) and (5.13) indicate that the overall error probability of relaying is directly subject to the error probabilities of the backhaul link and the combined link. Based on (5.11), the error probability of the two links are subject to the coding rate. As the coding rate is determined by the source based on the channel conditions of the bottleneck link, we further study an upper bound of the overall error probability by investigating the following two scenarios with different bottleneck links:

- **Case** $z_2 \geq z_1 + z_3$: The backhaul link is stronger than the combined link. Hence, the bottleneck link is the combined link. Therefore, $P(z_2, r, m) \leq P(z_1 + z_3, r, m)$. We have the following relationship between the overall error probability ε_R and the error probability of the combined link ε_{MRC} :

$$\varepsilon_R \leq 2\varepsilon_{\text{MRC}} - (\varepsilon_{\text{MRC}})^2.\tag{5.14}$$

When $z_2 \gg z_1 + z_3$, we have $\varepsilon_R \approx \varepsilon_{\text{MRC}}$.

- **Case** $z_2 < z_1 + z_3$: The bottleneck link is the backhaul link. Similarly, the upper bound of ε_R is:

$$\varepsilon_R \leq 2\varepsilon_2 - (\varepsilon_2)^2.\tag{5.15}$$

When $z_1 + z_3 \gg z_2$, we have $\varepsilon_R \approx \varepsilon_2$.

Combining (5.14) and (5.15) with (5.13), the overall error probability of relaying is bounded by:

$$2\varepsilon_\star - (\varepsilon_\star)^2 \geq \varepsilon_R \geq \varepsilon_\star,\tag{5.16}$$

where $\varepsilon_\star = \max\{\varepsilon_{\text{MRC}}, \varepsilon_2\}$ is the error probability of the bottleneck link. Hence, we have $\varepsilon_{\text{R}} \approx \varepsilon_\star$ if $z_1 + z_3 \gg z_2$ or $z_1 + z_3 \ll z_2$. On the other hand, if $z_1 + z_3 \approx z_2$, this results in $\varepsilon_{\text{MRC}} \approx \varepsilon_2$. Therefore, based on (5.12) we have $\varepsilon_{\text{R}} \approx 2\varepsilon_\star - (\varepsilon_\star)^2$ and in particular we have $\varepsilon_{\text{R}} \approx 2\varepsilon_\star$ with a low $\varepsilon_\star$, e.g., $\varepsilon_\star < 0.1$.

For given error probabilities of the backhaul link and the combined link, the coding rate can be determined by the source based on (5.9) and is given by:

$$r = R(\min\{z_2, z_1 + z_3\}, \max\{\varepsilon_2, \varepsilon_{\text{MRC}}\}, m). \quad (5.17)$$

According to (5.17) (with given channel gains and blocklength) r is only influenced by the error probability of the bottleneck link $\varepsilon_\star$. Moreover, (5.11) and (5.12) show that the overall error probability ε_{R} is fully determined by r . As a result, ε_{R} is totally determined by $\varepsilon_\star$. The relationship between the overall error probability ε_{R} and the bottleneck link error probability $\varepsilon_\star$ is given in the following proposition:

Proposition 3. *For a relaying network under static channels, the overall error probability ε_{R} is strictly increasing in the error probability of the bottleneck link $\varepsilon_\star$.*

Beweis. See Appendix B.1. □

5.3.2 Perfect CSI Scenario with Quasi-Static Channels

Under the quasi-static channel model, channels are fixed in during each relaying transmission period and vary from one period to the next. In this sub-section, we assume that instantaneous CSI of all the links to be available at the source at the beginning of each transmission period. Therefore, for each transmission period the source is able to determine the coding rate r_i by optimizing the expected transmission performance¹ according to $|h_{j,i}|^2$, $j = 1, 2, 3$. Recall that $|h_{j,i}|^2 = z_{j,i}\bar{z}_j$, $j = 1, 2, 3$ are the channel gains at transmission period i of the direct link, backhaul link and relaying link. Hence, the coding rate (and therefore the error probability) varies along with the channels from one transmission period to the next.

¹ For example, r_i can be determined by optimizing BL-capacity based on the CSI of transmission period i . Later on, we show in Proposition 4 that the BL-capacity is roughly concave in the coding rate.

In addition, denoted by $\varepsilon_{2,i}$, $\varepsilon_{\text{MRC},i}$ and $\varepsilon_{\text{R},i}$ are error probabilities of backhaul link, error probability of combined link and the overall error probability of relaying at transmission period i . Hence, we have

$$\varepsilon_{\text{R},i} = \varepsilon_{2,i} + (1 - \varepsilon_{2,i}) \cdot \varepsilon_{\text{MRC},i}, \quad (5.18)$$

where $\varepsilon_{\text{MRC},i} = \text{P}(z_{1,i}\bar{z}_1 + z_{3,i}\bar{z}_3, r_i, m)$ and $\varepsilon_{2,i} = \text{P}(z_{2,i}\bar{z}_2, r_i, m)$.

5.3.3 Average CSI Scenario with Quasi-Static Channels

Now we move on to consider the error probability of a relaying network where the source only has the information of average channel gains. With only average CSI, if the source determines the coding rate directly based on it, this likely results in that the instantaneous channel gain at transmission period i is lower than the average, i.e., $|h_{j,i}|^2 < \bar{z}_j, j = 1, 2, 3$. Moreover, this leads to a significant error probability which is higher than 0.5. Therefore, we propose the source to choose a relatively lower coding rate which is obtained by the weighted average channel gains $\eta\bar{z}_j$ ($j = 1, 2, 3$), where η is a weight factor. In addition, we assume that $0 < \eta \leq \hat{z}$, where the $\hat{z}$ is the median of z . Then, although the instantaneous channel gain is still possible to be lower than the average, the probability of this becomes much lower when $\eta < \hat{z}$ and is bounded by 0.5 (achieved when $\eta = \hat{z}$). The motivation as this setup (letting $\hat{z}$ be the upper limit of η) is as follows. The setup facilitates the proof of Theorem 1 and Theorem 2. In addition, in practice it is unreasonable to choose a coding rate which likely exceeds the Shannon limit with a probability higher than 50%. And our setup just reduces this probability and bounds it by 50%. Hence, this setup is reasonable. Moreover, we will also show in the simulation that this setup is not impacting the system operation/optimization as the optimal value of η is about 0.2 which is much lower than $\hat{z} \approx 0.7$ (under Rayleigh channels).

Recall that as applying MRC at the destination coding rates at different hops of relaying are required to be the same. This coding rate r is determined by the source based on the weighted average CSI. As the average CSI of each link is constant, r is actually fully-influenced by the weight factor η . In particular, a big η means a high

expectation on the channel quality and results in a high coding rate. In other words, r is strictly increasing in η .

Once the coding rate r is determined, it will not be changed during transmission periods. This is different from the perfect CSI scenario in the Section 5.3.2. In other words, from one transmission period to the next the coding rate is fixed and therefore the error probability varies along with the channel fading. The overall error probability of relaying during transmission period i is given by:

$$\varepsilon_{R,i} = \varepsilon_{2,i} + (1 - \varepsilon_{2,i}) \cdot \varepsilon_{\text{MRC},i}, \quad (5.19)$$

where $\varepsilon_{\text{MRC},i} = P(z_{1,i}\bar{z}_1 + z_{3,i}\bar{z}_3, r, m)$ and $\varepsilon_{2,i} = P(z_{2,i}\bar{z}_2, r, m)$.

Recall that all the channels are independent from each other, hence the expected value of the overall error probability of relaying is further given by:

$$\mathbb{E}_i[\varepsilon_{R,i}] = \mathbb{E}_{z_1, z_2, z_3}[\varepsilon_R] = \mathbb{E}_{z_2}[\varepsilon_2] + \left(1 - \mathbb{E}_{z_2}[\varepsilon_2]\right) \mathbb{E}_{z_1, z_3}[\varepsilon_{\text{MRC}}], \quad (5.20)$$

where $\mathbb{E}_{z_2}[\varepsilon_2]$ and $\mathbb{E}_{z_1, z_3}[\varepsilon_{\text{MRC}}]$ are the expectation values (over fading) of error probabilities of the backhaul link and the combined link:

$$\begin{aligned} \mathbb{E}_{z_2}[\varepsilon_2] &= \int_0^\infty e^{-z_2} \varepsilon_2 dz_2 \\ &= \int_0^\infty P(z_2\bar{z}_2, r, m) e^{-z_2} dz_2 \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \int_{\sqrt{mw}(z_2)}^\infty e^{-\frac{t^2 + 2z_2}{2}} dt dz_2, \end{aligned} \quad (5.21)$$

$$\begin{aligned} \mathbb{E}_{z_1, z_3}[\varepsilon_{\text{MRC}}] &= \int_0^\infty \int_0^\infty \varepsilon_{\text{MRC}} e^{-z_1 - z_3} dz_1 dz_3 \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \int_0^\infty P(z_1\bar{z}_1 + z_3\bar{z}_3, r, m) e^{-z_1 - z_3} dz_1 dz_3 \\ &= \frac{1}{\sqrt{2\pi}} \int_0^\infty \int_0^\infty \int_{\sqrt{mw}(z_1, z_3)}^\infty e^{-\frac{t^2 + 2z_1 + 2z_3}{2}} dt dz_1 dz_3, \end{aligned} \quad (5.22)$$

where $w(z_2) = \frac{C(z_2\bar{z}_2) - r}{\sqrt{\frac{1}{m}(1 - 2^{-2C(z_2\bar{z}_2)})} \log_2 e}$ and $w(z_1, z_3) = \frac{C(z_1\bar{z}_1 + z_3\bar{z}_3) - r}{\sqrt{\frac{1}{m}(1 - 2^{-2C(z_1\bar{z}_1 + z_3\bar{z}_3)})} \log_2 e}$.

Summary

In Section 5.3.1, we derived the overall error probability of two-hop relaying. We first studies static channels with perfect CSI. Subsequently, we extended the model to quasi-static channels while both perfect CSI and average CSI assumptions were considered. In the following, move on to study the capacity and throughput of relaying under these channel assumptions.

5.4 BL-Capacity (Physical-Layer Throughput) and Effective Capacity

5.4.1 The Blocklength-Limited Capacity: BL-Capacity

Recall that in the finite blocklength regime decoding errors may occur and that in this work we assume that both the relay and the destination reliably detect errors. Based on the DF relaying protocol, the relay does not forward the block to the destination when a decoding error occurs at the relay. In addition, if a decoding error occurs at the destination during transmission period i , the service process increment s_i of the two-frame relaying equals zero. On the contrary, if no error occurs, the service process increment equals $s_i = m \cdot r$, where r is the coding rate (in bits per channel use) employed over a block of m symbols in each hop/frame of relaying. Hence, the mean of service process increments over time (transmission periods) is given by $\mathbb{E}_i[s_i]$. Finally, the BL-capacity C_{BL} of two-frame relaying is defined as the average effectively transmitted information per channel use, given by:

$$C_{\text{BL}} = \mathbb{E}_i[s_i] / 2m. \quad (5.23)$$

In the following, we respectively study the BL-capacity under scenarios with different assumptions of CSI and channel fading model.

5.4.1.1 Perfect CSI Scenario with Static Channels

As channels are static, for a given coding rate error probabilities of different transmission periods are the same. In other words, service process increments $s_1, s_2, \dots, s_i, \dots$ are Bernoulli-distributed and i.i.d. Hence we have:

$$\mathbb{E}_i[s_i] = \mathbb{E}[s_i] = (1 - \varepsilon_R) mr. \quad (5.24)$$

Therefore, under the perfect CSI scenario with static channels C_{BL} is given by:

$$C_{BL} = \frac{(1 - \varepsilon_R) mr}{2m} = \frac{(1 - \varepsilon_R) r}{2}. \quad (5.25)$$

We then have the following proposition:

Proposition 4. *For a relaying network under static channels, if $\varepsilon_R < 0.5$, the BL-capacity C_{BL} is concave in the coding rate r .*

Beweis. See Appendix B.2. □

Note that according to Equation (5.9) the coding rate r is strictly increasing in the error probability of the bottleneck link $\varepsilon_\star$, and that according to Proposition 3 ε_R is strictly increasing in $\varepsilon_\star$. Hence, r is strictly increasing in $\varepsilon_\star$ or ε_R . Therefore, we have the following corollary of Proposition 4:

Proposition 5. *For a relaying network under static channels, if $\varepsilon_R < 0.5$, the BL-capacity C_{BL} is quasi-concave in ε_R or $\varepsilon_\star$.*

Beweis. See Appendix B.3. □

Proposition 5 shows that the BL-capacity of relaying has a global maximum which can be achieved by choosing an appropriate error probability of ε_R or $\varepsilon_\star$.

Simulation

In the following, we present some numerical results obtained by simulations to show the appropriateness of our analytical models regarding the error probability as well as

the BL-capacity under static channels. In the simulation, we consider an outdoor urban scenario and the distances of the backhaul, relaying and direct links are set to 200 m, 200 m and 360 m (we vary the system topology only in Fig. 6.9). We set the transmit power p_{tx} equal to 22 dBm and noise power to -95 dBm, respectively. In addition, we utilize the well-known COST [63] model for calculating the path-loss while the center frequency is set to equal 2 GHz. As we consider static channels, in the simulation the channel gains are assumed to be fully subject to the path-loss. Based on the above topology setting, we have $z_2 = z_3 \gg z_1$.

The relationship between the overall error probability and the error probability of the bottleneck link is shown in Fig. 6.9. It is shown that ε_R is increasing in $\varepsilon_\star$, which

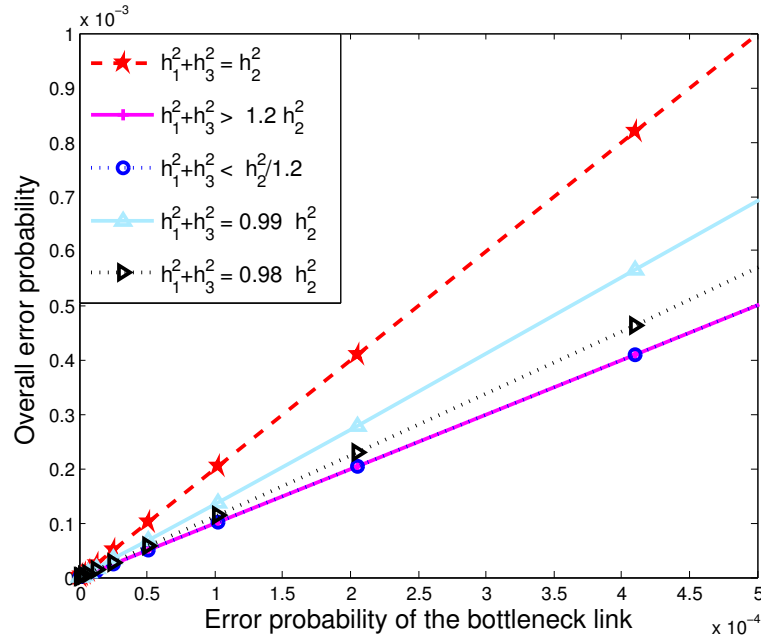


Abbildung 5.4: The relationship between overall error probability and error probability of the bottleneck link.

matches Proposition 3. Moreover, it matches our analysis that ε_R is approximately linearly increasing in $\varepsilon_\star$, i.e., $\varepsilon_R \approx k\varepsilon_\star$. More specifically, $k = 2$ when $z_1 + z_3 = z_2$ which means both of the two links equally limit the system performance. In addition, $k = 1$ if the gap between channel gains of the backhaul link and the combined link is considerable, e.g., one is 20% higher than the other one. Finally, $1 < k < 2$ when

the gap between channel gains of these two links is neither negligible nor significantly high.

The comparison between the BL-capacity and the Shannon capacity of relaying is shown in Fig. 5.5. We plot the corresponding capacities of (non-relay) direct transmission as references. Firstly, Shannon capacities are not influenced by the error probability

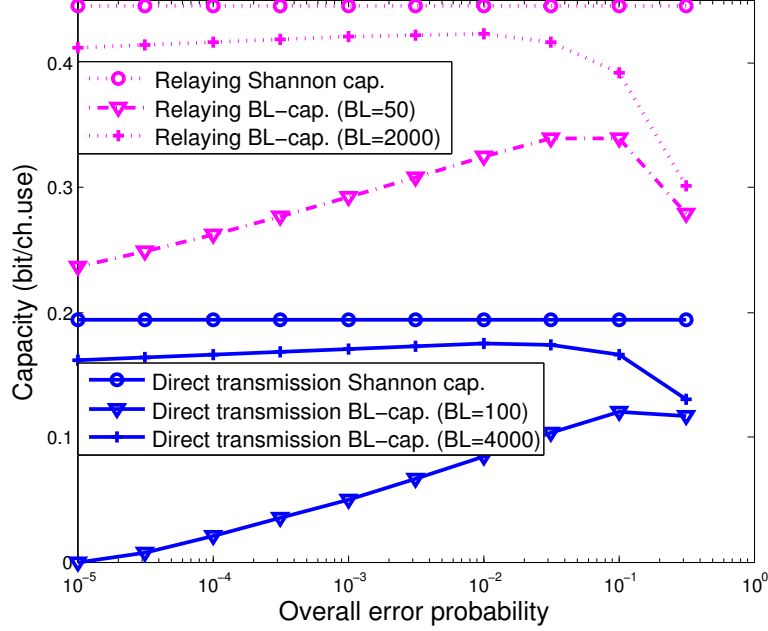


Abbildung 5.5: BL-capacities under static channels with $\gamma_1 = -5$ dB, $\gamma_2 = \gamma_3 = 4$ dB and $\gamma_{\text{MRC}} = 4.5$ dB.

lity. Secondly, we observe from the plot that the BL-capacity curves are quasi-concave in the overall error probability, which shows the appropriateness of Proposition 4. In particular, the BL-capacities with a short blocklength are more sensitive to the overall error probability. Therefore, the optimization of the BL-capacity is more important for short blocklength systems. Thirdly, maximal values of BL-capacity of relaying with a short blocklength and a long blocklength are achieved by different overall error probabilities. Hence, the optimal solution of maximizing the BL-capacity of relaying is also influenced by the blocklength. Finally, relaying has significant higher Shannon capacity as well as BL-capacity in comparison to direct transmission. In other words, in comparison to direct transmission, relaying improves the performance in both the infinite blocklength regime and the finite blocklength regime.

5.4.1.2 Perfect CSI Scenario with Quasi-Static Channels

Recall that under this case the overall error probability and the coding rate vary along with the instantaneous CSI from one transmission period to the next. Hence, the mean/expected value of service process increment in transmission period i is given by:

$$\bar{s}_i = (1 - \varepsilon_{R,i}) m \cdot r_i, \quad (5.26)$$

where r_i is determined based on Proposition 4 at the beginning of transmission period i . In addition, the overall error probability can be obtained from (5.18) based on r_i . Therefore, the BL-capacity of relaying under this scenario is given by:

$$C_{BL} = \frac{\mathbb{E}_i[\bar{s}_i]}{2m}. \quad (5.27)$$

5.4.1.3 Average CSI Scenario with Quasi-Static Channels

Recall that under the average CSI scenario the service process increment of a transmission period is either zero or rm . In other words, it is Bernoulli-distributed. Moreover, the probability of the error event of this Bernoulli distribution is actually the expected overall error probability of relaying which was given in the previous section by (5.20). Based on the characteristic of the Bernoulli distribution, we immediately have the mean of service process increments:

$$\mathbb{E}_i[s_i] = rm \cdot (1 - \mathbb{E}_{z_1, z_2, z_3}[\varepsilon_R]), \quad (5.28)$$

Therefore, the (average) BL-capacity of relaying is given by:

$$C_{BL} = \frac{\mathbb{E}_i[s_i]}{2m} = r(1 - \mathbb{E}_i[\varepsilon_{R,i}])/2, \quad (5.29)$$

where $\mathbb{E}_i[\varepsilon_{R,i}]$ has been given in (5.20).

So far, we derived the BL-capacity of relaying under the average CSI scenario. We then have the following proposition regarding the BL-capacity under this scenario:

Proposition 6. *Under a relaying scenario with quasi-static Rayleigh channels where only the average CSI is available at the source, the BL-capacity is quasi-concave in the coding rate (determined by the source).*

Beweis. See Appendix B.4. □

Recall that the higher η is the higher the coding rate the source chooses. In other words, the coding rate chosen by the source is strictly increasing in η . Combined with Proposition 6, we have:

Proposition 7. *Consider a relaying scenario with quasi-static Rayleigh channels where only the average CSI is available at the source. If the source determines the coding rate according to the weighted average CSI, the BL-capacity is quasi-concave in the weight factor η .*

Beweis. See Appendix B.5. □

Therefore, with only the average CSI of quasi-static channels there is an unique optimal value of η that maximizes this BL-capacity of relaying.

Numerical results

In the following, we provide numerical results regarding the BL-capacity performance of relaying under Quasi-static channels. The model parameterization of the simulation is same as the simulation in Section 5.4.1.1.

First of all, we show BL-capacities (in bit/ch.use) of relaying under both the perfect CSI scenario and the average CSI scenario in Fig. 5.6. Obviously, the case with perfect CSI outperforms the case with only average SCI. In addition, BL-capacities of both these two cases are increasing in the blocklength. At the same time, the BL-capacity of the average SCI case increase slowly (in comparison to the perfect SCI case) as blocklength increase.

In comparison to the perfect CSI scenario, we are more interested in the scenario with only average CSI at the source. Hence, we specifically consider the average CSI scenario and show the relationship between the BL-capacity and the coding rate (both

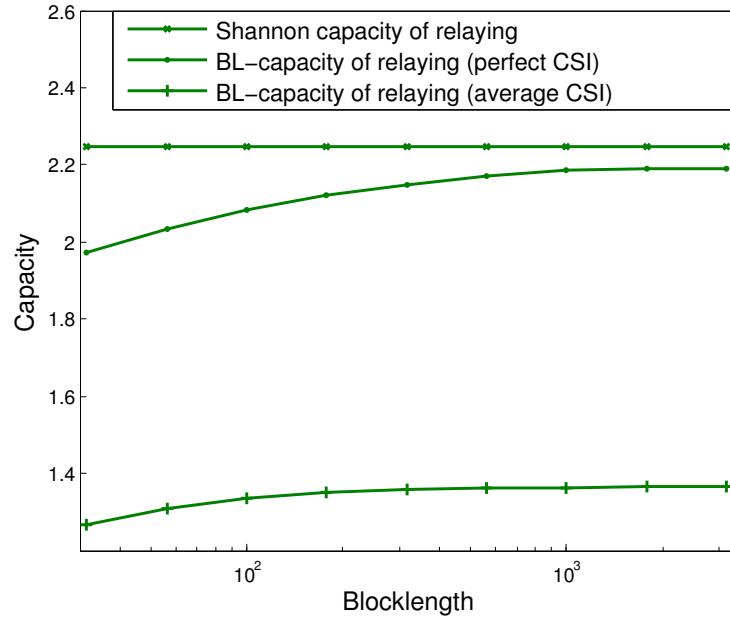


Abbildung 5.6: The performance comparison between relaying with average CSI and relaying with perfect CSI while varying blocklength.

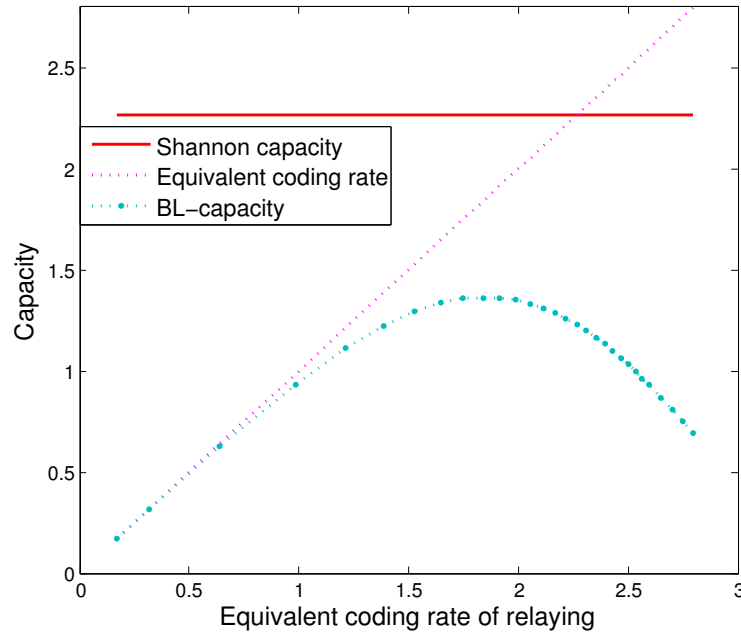


Abbildung 5.7: The relationship between BL-capacity and coding rate under the average CSI scenario.

in bit/ch.use) in Fig. 5.7. As shown in the figure, the BL-capacity is concave in the coding rate, which matches with Proposition 6. Subsequently, the relationship between the BL-capacity (in bit/ch.use) and the weight factor under the average CSI scenario is shown in Fig. 5.8. We observe that the coding rate of relaying is strictly increasing in the weight factor. Moreover, the BL-capacity is quasi-concave in the weight factor, which shows the appropriateness of Proposition 7.

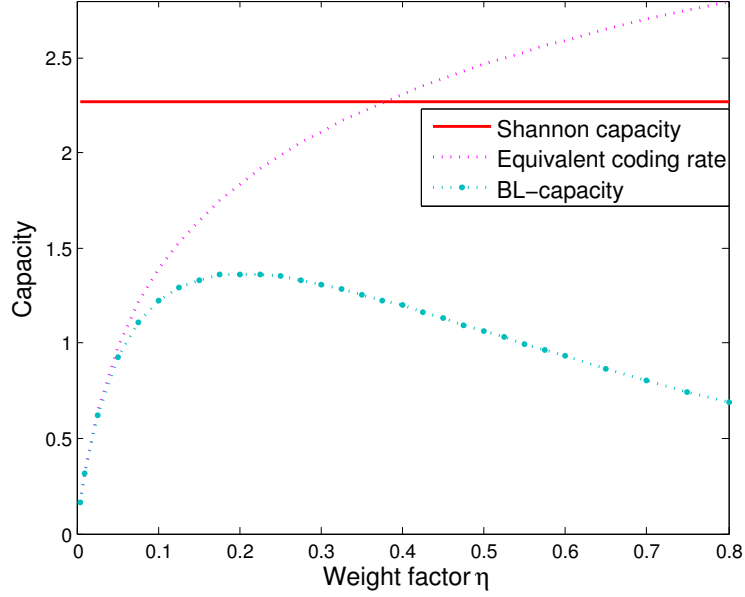


Abbildung 5.8: BL-capacity (in bit/ch.use) vs. weight factor under the average CSI scenario.

Moreover, in Fig. 5.9 we provide intuitive parallel sub-figures to show how the weight factor introduces a tradeoff between the coding rate and the error probability and further influences the BL-capacity. By increasing the weight factor, the coding rate increases as well as the error probability. In particular, the coding rate increases rapidly at the beginning but slowly later on. On the other hand, the error probability increases approximately linearly. As a result, the BL-capacity first increases and then decreases. In other words, the BL-capacity is quasi-concave in the weight factor.

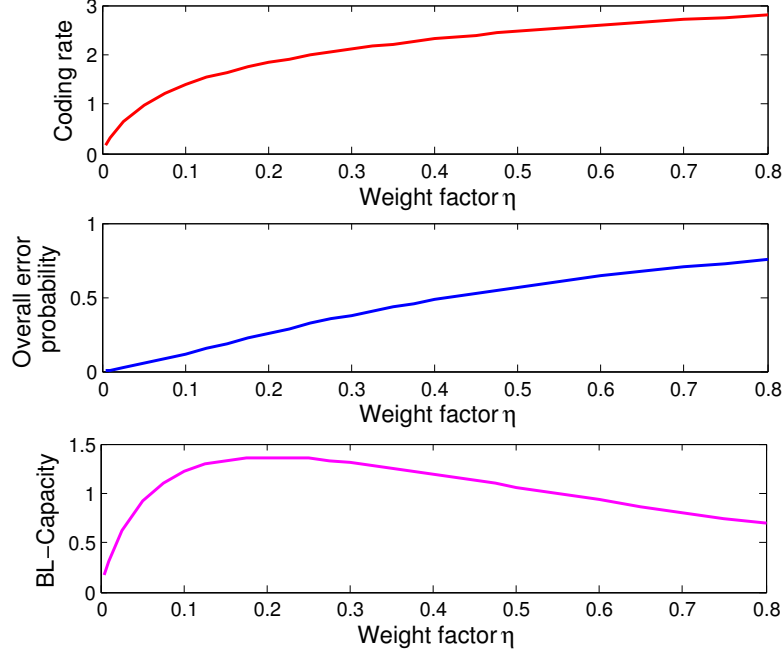


Abbildung 5.9: How the weight factor introduces tradeoff between error probability and coding rate and further influences the BL-capacity.

5.4.2 Effective Capacity

Recall that in this thesis we refer to MSDR as a metric for measuring the higher-layer queuing performance. In the previous chapter, we have studied the MSDR of relaying system in the infinite blocklength regime. Now we extend the MSDR model into the finite blocklength regime by redefining the unit of the delay constraint d and let d in symbols. Recall that in the studied relaying system the length of each (two-hop relaying) transmission period is $2m$. Therefore, the delay constraint is $d/2m$ times of a transmission period. Then, the MSDR (bits per channel use) of the studied two-hop relaying system is given by:

$$R_{\text{MS}} \approx \frac{\mathbb{E}[s_i]}{4m} + \frac{1}{4m} \sqrt{(\mathbb{E}[s_i])^2 + \frac{4m \ln(P_d)}{d} \cdot \text{Var}[s_i]}, \quad (5.30)$$

Based on (5.30), the major challenge for determining the MSDR is to obtain the mean and variance of service process increments of the studied relaying system.

In previous sections, means of service process increments (under different scenarios regarding the channel and CSI assumption) have been given by (5.24), (5.26) and (5.28), respectively. Therefore, in this section we mainly study variances of service process increments under these scenarios.

5.4.2.1 Perfect CSI Scenario with Static Channels

We have shown that with static channels service process increments in different transmission periods $s_1, s_2, \dots, s_i, \dots$ are Bernoulli-distributed and i.i.d.. Hence, the variance of service process increments is given by:

$$\text{Var}[s_i] = \varepsilon_R (1 - \varepsilon_R) (2m \cdot r_R)^2. \quad (5.31)$$

Simulation

Numerical results of the MSDR performance of relaying under static channels are shown in Figure 5.10, where the model parameterization of the simulation is same as the simulation in Section 5.4.1.1. From the figure, we observe that the MSDR of relaying under static channels is strongly subject to the blocklength as well as the error probability. In particular, the MSDR is quasi-concave in the error probability. Moreover, a significant blocklength results in a high MSDR.

5.4.2.2 Perfect CSI Scenario with Quasi-Static Channels

Recall that under this case the overall error probability and the coding rate are vary along with the instantaneous CSI during from one transmission period to the next. The variance of service process increments is given by:

$$\text{Var}[s_i] = \sum_i (s_i - \text{E}[s_i])^2 = \sum_i (s_i - (1 - \varepsilon_{R,i}) m \cdot r_i)^2. \quad (5.32)$$

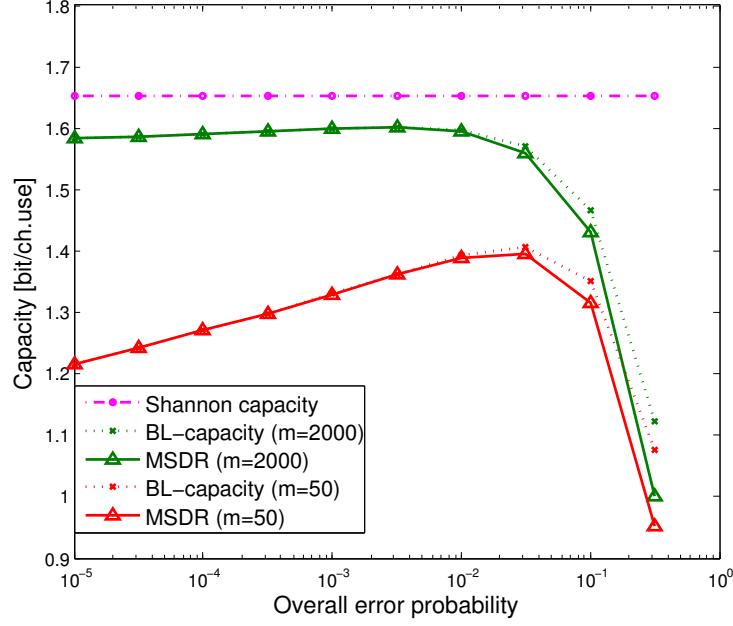


Abbildung 5.10: MSDR performance of relaying under static channels.

5.4.2.3 Average CSI Scenario with Quasi-Static Channels

Under quasi-static channels, we are more interested in the scenario with only the average CSI at the source. In Section 5.4.1.3, we have shown that the service process increment under this scenario is Bernoulli-distributed with error probability $\mathbb{E}_{z_1, z_2, z_3} [\varepsilon_R]$. Hence, the variance of the service process increment s_i are given by:

$$\text{Var}[s_i] = r^2 m^2 \mathbb{E}_{z_1, z_2, z_3} [\varepsilon_R] \cdot (1 - \mathbb{E}_{z_1, z_2, z_3} [\varepsilon_R]), \quad (5.33)$$

where the ergodic overall error probability (of relaying over channel fading) $\mathbb{E}_{z_1, z_2, z_3} [\varepsilon_R]$ was given in the previous section by (5.20).

Substituting (5.28) and (5.33) into (5.30), we have

$$R_{\text{MS}} = \frac{r(1 - \mathbb{E}_{z_1, z_2, z_3} [\varepsilon_R])}{4} + \frac{r}{4} \sqrt{\left(1 - \mathbb{E}_{z_1, z_2, z_3} [\varepsilon_R]\right)^2 + \frac{4m \ln(P_d)}{d} \cdot \mathbb{E}_{z_1, z_2, z_3} [\varepsilon_R] \cdot (1 - \mathbb{E}_{z_1, z_2, z_3} [\varepsilon_R])}. \quad (5.34)$$

Obviously, the performance of MSDR R_{MS} is subject to the coding rate r , as $\mathbb{E}_i[\varepsilon_{\text{R},i}]$ is also a function of r . In fact, the following proposition holds regarding the relationship between MSDR and the coding rate.

Proposition 8. *Under a relaying scenario with quasi-static Rayleigh channels where only the average CSI is available at the source, the MSDR is concave in the coding rate r .*

Beweis. See Appendix B.6. □

Similar to Proposition 6, Proposition 8 has the following corollary regarding the weight factor ² η :

Proposition 9. *Consider a relaying scenario with quasi-static Rayleigh channels where only the average CSI is available at the source. If the source determines the coding rate according to the weighted average CSI, the MSDR is quasi-concave in the weight factor η .*

Beweis. The proof of Proposition 9 (based on Proposition 8) is similar to the proof of Proposition 7 (based on Proposition 6). □

Simulation

In this part, we show our simulation results regarding the MSDR of relaying under quasi-static channels. Both the average CSI scenario and the perfect CSI scenario are considered in the simulation. Moreover, these provided simulation results are based on the model parameterization which is same as the simulation in Section 5.4.1.1.

Firstly, we show the ergodic MSDR (over channel fading) of relaying versus weight factor in Fig. 5.11. Obviously, the (ergodic) MSDR of relaying with perfect CSI is not influenced by the weight factor. In addition, the perfect CSI case has higher performance than the average CSI case.

Then, we further focus on the average CSI scenario to compare the coding rate, BL-capacity and MSDR of relaying in Figure 5.12. The figure clearly shows that setting

² Recall that η is an SNR margin of an operation we introduced to reduce error probability for the scenario where the source only has average CSI. See Section 5.3.3

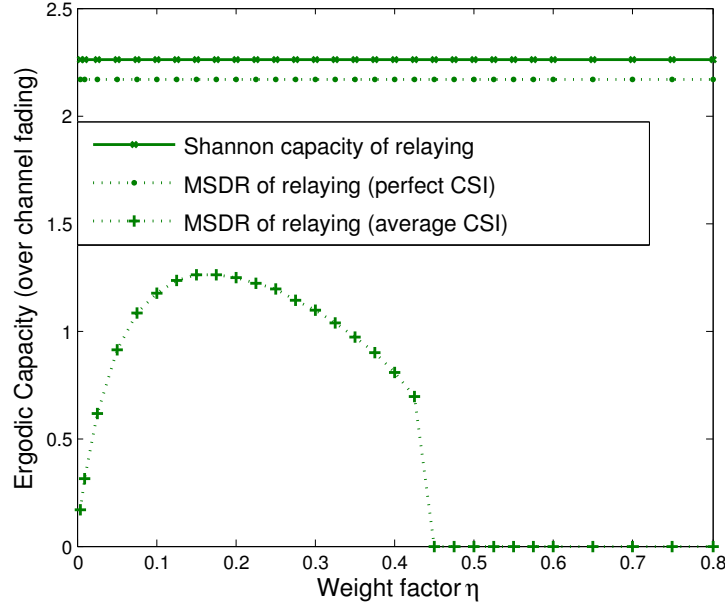


Abbildung 5.11: The relationship between MSDR and coding rate under both the perfect CSI scenario and the average CSI scenario.

an overaggressive coding rate, e.g., to be higher than the Shannon capacity, does not benefit the BL-capacity and the MSDR as an overaggressive coding rate leads to a significant error probability. In fact, Figure 5.12 shows that both BL-capacity and MSDR are concave in the coding rate while the optimal coding rate (which maximize the BL-capacity or MSDR) is significantly lower than the Shannon capacity.

Thirdly, the relationship between the MSDR and the weight factor under the average CSI scenario is shown in Fig. 5.14. It matches with Proposition 7 and Proposition 9 well that both BL-capacity and MSDR are quasi-concave in the proposed weight factor. This is an important guideline for the design of QoS-support relaying systems.

Finally, we show in Figure 5.14 the relationship between the relaying performance and the blocklength under the average CSI scenario with quasi-static channels. We observe that the BL-capacity of the studied relaying system is increasing in the blocklength while the MSDR is concave in the blocklength. The explanation is as follows. On one hand, based on (5.11) and (5.9) both error probability and coding rate are influenced by the blocklength m . In particular, with a fixed coding rate at each link a long blocklength leads to a low error probability of each link. Obviously, this results

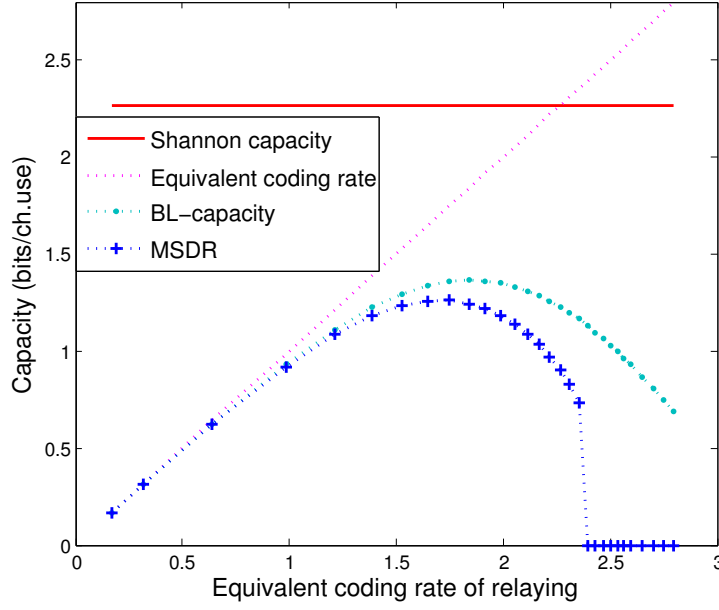


Abbildung 5.12: The relationship between MSDR and coding rate under the average CSI scenario.

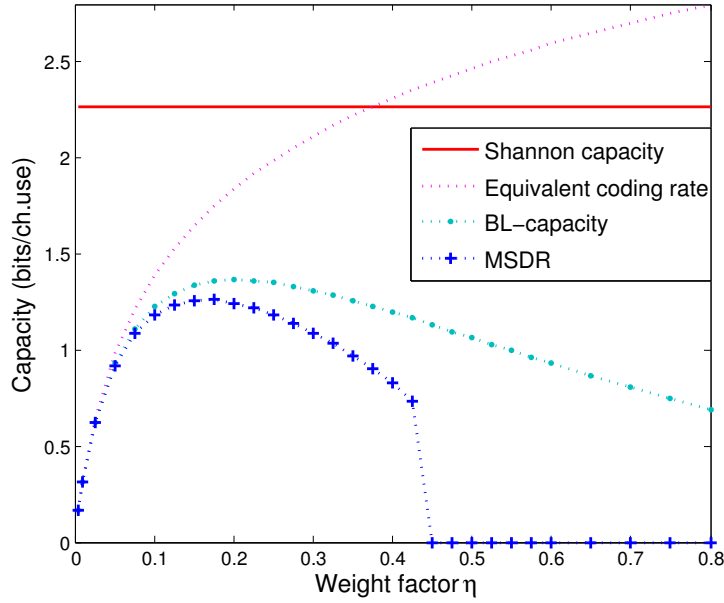


Abbildung 5.13: The relationship between MSDR and weight factor under the average CSI scenario.

in a low overall error probability of relaying and therefore a high BL-capacity. This is the reason why the BL-capacity is increasing in the blocklength. On the other hand,

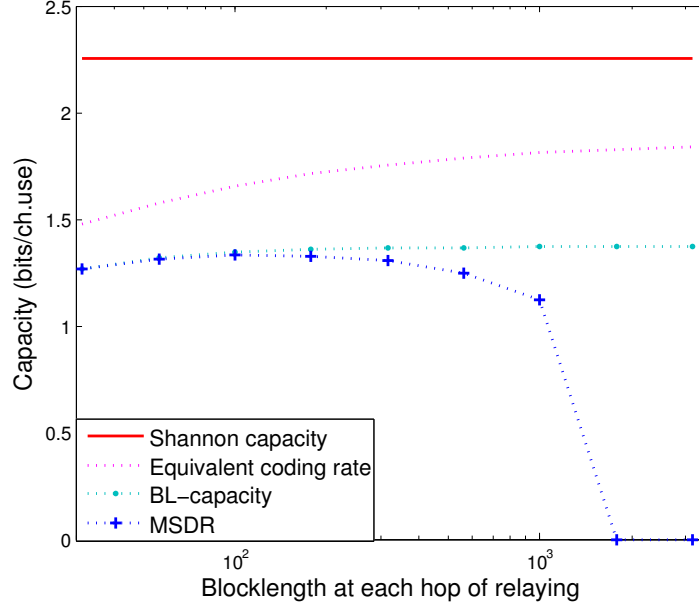


Abbildung 5.14: The relaying performance versus blocklength under the average CSI scenario.

a long blocklength means that a single transmission (i.e., two-hop relaying) costs a long time. This reduces the number of allowed retransmission attempts under a given delay constraint. For example, consider a relaying system where the blocklength of each hop of relaying is 500 symbols (two-hop relaying transmission period is 10^3 symbols). To support a service with a delay constraint $d = 10^4$ symbols, the maximal number of transmission attempts (including an initial transmission and retransmissions), which not violates the delay constraint, is 10. This number would be 5 if the blocklength is doubled. In other words, a long blocklength reduces the flexibility of a QoS-support system due to limiting the number of transmission attempts. At the same time, the gain from increasing blocklength becomes tiny as the blocklength increases (e.g., even regarding the physical-layer performance, the BL-capacity is approximately a constant in the long blocklength region). As the result, the higher-layer performance MSDR reduces in the long blocklength region. For an extreme example, the MSDR is definitely zero if the blocklength is longer than the delay constraint.

It should be mentioned that the above observation MSDR is concave in the blocklength is an important guideline for determining the blocklength in the design of QoS-support relaying systems.

Summary

In this section, we study the capacity of two-hop relaying. Both static channels and quasi-static channels are considered. Under static channels with perfect CSI, we proved that the BL-capacity is concave in the coding rate and quasi-concave in the overall error probability of relaying. In addition, under the scenario with only the average CSI we proposed a simple system operation by introducing a weight factor based on which we weight the average CSI and let the source determine the coding rate based on the weighted CSI. Closed-form expression of the related BL-capacity was derived. In addition, the distribution of the service process increment as well as the MSDR were investigated. Moreover, we proved that both the BL-capacity and the MSDR are concave in the coding rate. More importantly, it was proved that the BL-capacity and the MSDR are quasi-concave in the weight factor.

5.5 Energy Efficiency

In this section, we first briefly introduce our energy consumption model. Subsequently, we combine this model with previous derivations of the BL-capacity and the MSDR to investigate their corresponding energy efficiencies.

5.5.1 Energy Consumption

We consider the same energy consumption model in Section 4.1.2, where the energy consumption contains two parts: the basic energy consumption and the energy consumption due to transmission. The average total consumed energy per two-hop relaying transmission period is then given by (recall that K is the probability that the queue is non-empty and p_c is the power of basic energy consumption at each node):

$$\bar{\Psi} = 2mKp_{\text{tx}} + 6mp_c, \quad (5.35)$$

where $6mp_c$ is the basic energy consumption over the duration of a transmission period (with length $2m$) of all the three nodes including the source, the destination and the relay.

5.5.2 The Physical-Layer Energy Efficiency (PEE)

In this subsection, we investigate the PEE of relaying. Recall than PEE is defined as the ratio of the average throughput to the average energy consumption. Hence, the PEE of relaying is equal to the ratio of mean of service process increments (which is actually the throughput per transmission period) to the energy consumption per transmission period:

$$\Phi_p = \frac{2mC_{BL}}{2mp_{tx} + 6mp_c} = \frac{C_{BL}}{p_{tx} + 3p_c}. \quad (5.36)$$

5.5.3 The Higher-Layer Energy Efficiency (EEE)

Similarly, the EEE of relaying in the finite blocklength regime is given by:

$$\Phi_h = \frac{2mR_{MS}}{2mp_{tx} + 6mp_c} = \frac{R_{MS}}{p_{tx} + 3p_c}. \quad (5.37)$$

Numerical Results

Simulation results regarding the PEE and EEE of relaying under both static channels and quasi-static channels are provided, while these results are based on the model parameterization which is same as the simulation in Section 5.4.1.1.

First, energy efficiencies of relaying under static channels (with perfect CSI) are shown in Figure 5.15. We plot the Shannon capacity of the relaying as a reference. As shown in the figure, the PEE is increasing in the blocklength while the EEE is quasi-concave in the blocklength. This matches well with our previous observation that BL-capacity is increasing in blocklength while MSDR is quasi-concave in the blocklength.

Second, the PEE and the EEE relaying under quasi-static channels are presented in Figure 5.16 where both perfect CSI scenario and average CSI scenario are considered.

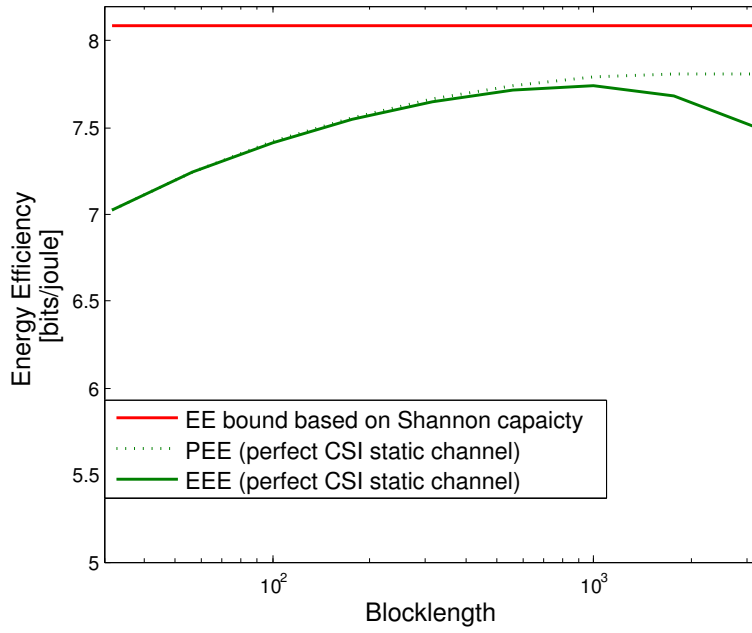


Abbildung 5.15: Energy efficiencies of relaying under perfect CSI scenario (with static channels).

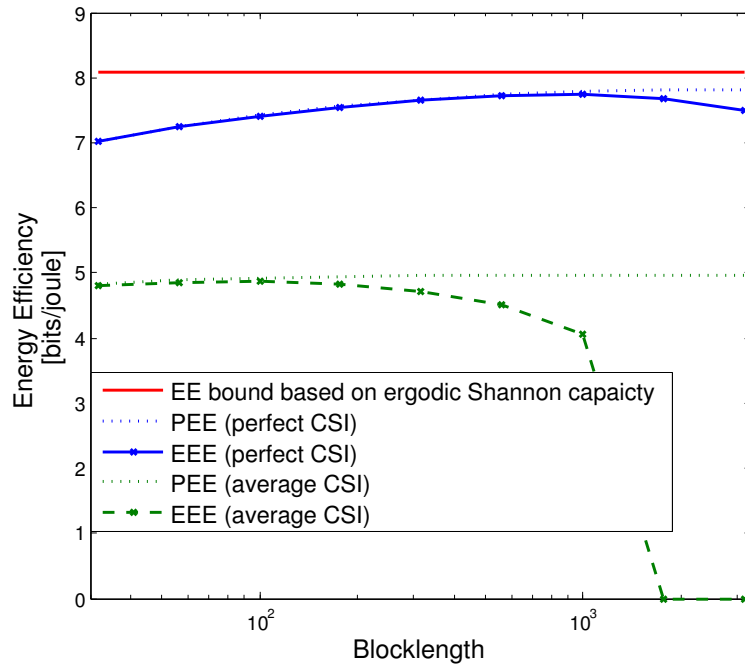


Abbildung 5.16: energy efficiencies of relaying under quasi-static channels.

In the figure, the energy efficiency based on the ergodic Shannon capacity over the channel fading is plotted as a reference. From the figure, both the PEE and the EEE under the perfect CSI scenario outperform the ones under the average CSI scenario. More importantly, it is observed that the EEE performance under both these two scenarios are quasi-concave in the blocklength.

Summary

In this section, we investigated the energy efficiency of relaying. So far, we have studied the error probability, capacity and energy efficiency of relaying in the finite blocklength regime. We will summarize and discuss the whole work of this chapter in the next section.

5.6 Summary and Discussion

In the following, we summarize the work in this chapter before the discussion.

Summary

In the finite blocklength regime, we investigated the physical-layer performance (BL-capacity and PEE) as well as the higher-layer performance (MSDR and EEE) of a relaying system under different scenarios which differ in the channel fading model and CSI assumption.

Under the static channel model, the overall error probability as well as the BL-capacity of relaying are derived. We proved that the overall error probability is strictly increasing in the error probability of the bottleneck link. In addition, we proved that the BL-capacity is concave in the coding rate and quasi-concave in the overall error probability of relaying. Therefore, the BL-capacity of relaying has a global maximum which can be achieved by choosing an appropriate overall error probability.

Moreover, we extended the results under static channels into quasi-static Rayleigh channels. In particular, under the scenario with only the average CSI we proposed a simple system operation by introducing a weight factor based on which we weight the average CSI and let the source determine the coding rate based the weighted

CSI. Closed-form expression of the related BL-capacity was derived. In addition, the distribution of the service process increment as well as the MSDR were investigated. Moreover, we proved that both the BL-capacity and the MSDR are concave in the coding rate. More importantly, it was proved that the BL-capacity and the MSDR are quasi-concave in the weight factor. In other words, the performance of the studied system can be easily optimized based on the proposed operation.

Through numerical results, we validated our analytical models and evaluated the relaying performance in the finite blocklength regime. We concluded a set of guideline for designing efficient relaying networks with finite blocklengths:

- Under static channels, the overall error probability is approximately linearly increasing in the error probability of the bottleneck link.
- Under a quasi-static Rayleigh scenario with only the average CSI, the MSDR and BL-capacity are quasi-concave in the proposed weight factor.
- Under a quasi-static Rayleigh scenario with only the average CSI, MSDR is concave in the blocklength while BL-capacity is increasing in blocklength.
- Under a quasi-static Rayleigh scenario (no matter with average CSI or perfect CSI), EEE performance of both these two cases are quasi-concave in the blocklength.

Discussion

The relaying performance with finite blocklengths has investigated in this chapter. Recall that the relaying performance model with infinite blocklengths has been studied in the previous chapter. So far, we have addressed all the scenarios (with interests) shown in Figure 1.1. Combining the findings of these two chapters, we have several corresponding observations as follows:

- In the infinite blocklength regime, with the perfect CSI the source perfectly adjusts the packet size based on the CSI to achieve an error-free transmission. However, even with the perfect CSI, a finite blocklength introduces error probability.

- If there is only the average CSI available at the source, outages occur in the infinite blocklength regime while errors due to both noise and channel fading occur in the finite blocklength regime.
- In the infinite blocklength regime (with infinite blocklengths), both the physical-layer energy efficiency and the higher-layer energy efficiency of relaying are quasi-concave in the packet size. In the finite blocklength regime, the physical-layer energy efficiency is increasing in the blocklength while the higher-layer energy efficiency is quasi-concave in the blocklength. Moreover, with a finite blocklength both these two energy efficiencies are concave or quasi-concave in the coding rate of the block.

So far, we observed that relaying is efficient under both the infinite blocklength regime or the finite blocklength regime. However, it is still unclear that under which regime or which scenario relaying is more efficient. In particular, in this chapter we studied the blocklength-limited performance of a single relay system. It is unknown the impact on the blocklength-limited performance if multiple relays are deployed in the system. Moreover, the performance difference of relaying with finite blocklengths and with infinite blocklengths under a multi-relay scenario is also open. These questions are going to be addressed in the next chapter.

6 Relaying Performance Comparison: Infinite Blocklength Regime vs. Finite Blocklength Regime

In this chapter, by simulation we investigate the performance difference between relaying in the finite blocklength regime and the one in the infinite blocklength regime. First, we consider this performance difference under a single relay scenario. Later on, we investigate the impact on this performance difference if multiple relays are deployed in the system.

6.1 Performance Advantage of Relaying in a Single Relay System

In this section, we consider a single relay system. We first provide our simulation results to show the performance advantage of relaying in the finite blocklength regime (in comparison to the infinite blocklength regime). Afterwards, we analytically explain why relaying has this advantage in the finite blocklength regime

6.1.1 Simulative Investigation

Methodology

For a non-relaying transmission (i.e., direct transmission), a finite blocklength introduces certain performance loss to the (blocklength-limited) performance in comparison

to the Shannon capacity. This is well-described in [8]. In this thesis, our focus is on if the performance loss (introduced by finite blocklengths) under the relaying scheme differs from the performance loss under the direct transmission scheme? If relaying helps reduce the performance loss in comparison to direct transmission?

To find answers for above questions, in our simulations the direct transmission scheme is treated as a reference/benchmark. In particular, by adjusting the transmit power of direct transmission we set the average Shannon capacities of either relaying or direct transmission are the same. Our aim can be achieved by observing if these two schemes (relaying and direct transmission) have different performance in the finite blocklength regime. For example, if relaying has same Shannon capacity but higher BL-capacity in comparison to direct transmission, this indicates that relaying is more efficient in the finite blocklength regime. In addition, to make a fair comparison, we set the coding rate of relaying to be equal to the equivalent coding rate of relaying. Moreover, we set the blocklength of direct transmission to be twice as large as the blocklength at each hop of relaying. Hence, the length of each transmission period of relaying equals that of direct transmission.

Under the above setup, we first show the performance advantage of relaying in comparison to direct transmission in the finite blocklength regime. Then, we compare these performance advantages of relaying under the perfect CSI scenario and under the average CSI scenario. Finally, we investigate the performance loss (introduced by finite blocklengths) for a relaying system under these two scenarios.

Parameterization

In the simulative comparison, we consider the following parameterization of the system model: Firstly, in the simulation we consider an outdoor urban scenario and the distances of the backhaul, relaying and direct links are set to 200 m, 200 m and 360 m. Secondly, we set the transmit power p_{tx} equal to 30 dBm and noise power to -90 dBm, respectively. Thirdly, we utilize the well-known COST [63] model for calculating the path-loss while the center frequency is set to equal 2 GHz. Regarding the channel, we only consider quasi-static channel model in the simulation. Hence, the numerical results for all validations and evaluations are based on the average/ergodic

performance over the random channel fading. Moreover, as we consider the higher-layer performance, in the simulation the QoS constraints {delay, delay violation probability} are set to $\{10^4$ symbols, $10^{-2}\}$. Finally, to observe the relaying performance we mainly vary the following parameters in the simulation: Blocklength and weight factor η (for average SCI scenarios).

Under the above parameterization, we show the performance advantage of relaying in the finite blocklength regime in the following subsections.

6.1.1.1 Performance Advantage of Relaying in the Finite Blocklength Regime: Relaying vs. Direct Transmission.

Comparisons under the average CSI scenario are shown in Figure 6.1 and Figure 6.2 where we vary the coding rate and blocklength, respectively.

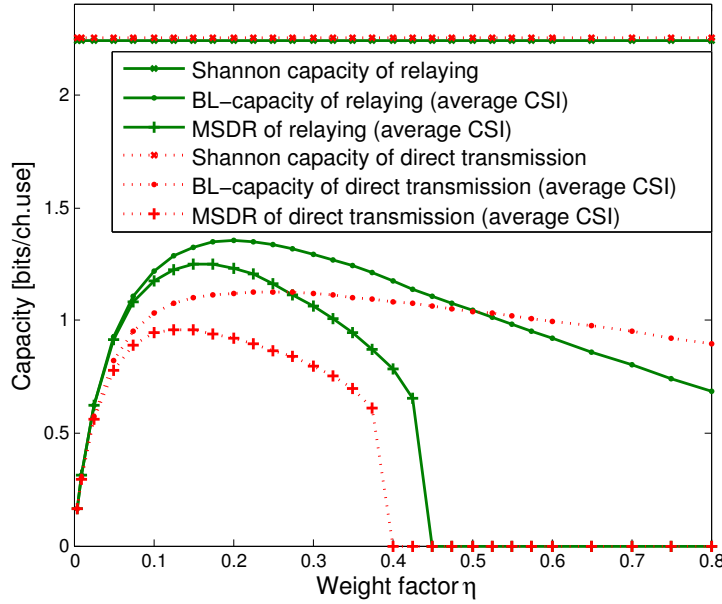


Abbildung 6.1: The performance comparison between relaying (with average CSI) and direct transmission (with average CSI) while varying the weight factor η . In the simulation, the blocklength at each hop of relaying is 500 symbols.

In the simulation the blocklength of direct transmission is twice as long as the blocklength in each hop of relaying. Recall that shorter blocklength introduces more signifi-

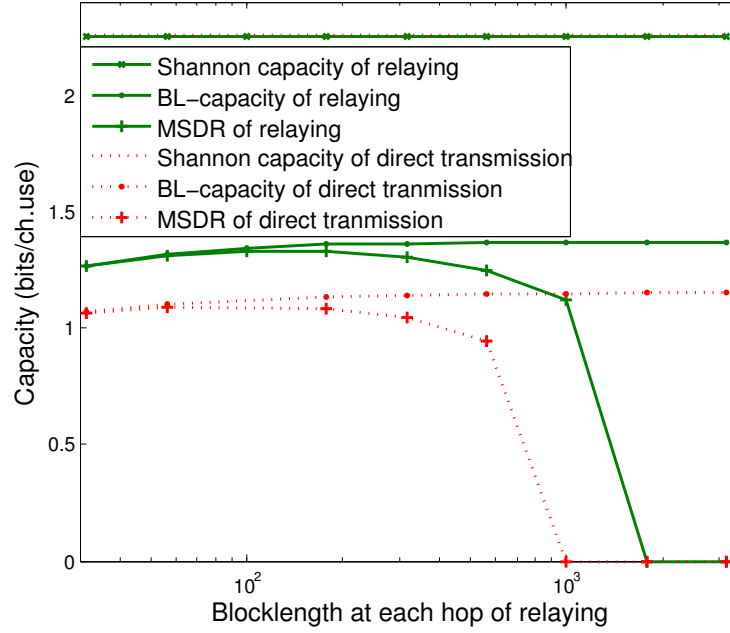


Abbildung 6.2: The performance comparison between relaying (with average CSI) and direct transmission (with average CSI) under dynamic blocklength ($\eta = 0.2$).

cant performance loss to the BL-capacity performance in comparison to the Shannon capacity. Hence, one could suspects that relaying pays off less in the finite blocklength regime compared to the infinite blocklength regime. From these two figures, we observe that both the BL-capacity and the MSDR of relaying significantly outperform direct transmission, although these two schemes have similar Shannon capacity. It is very surprising that the performance loss due to a finite (relatively shorter) blocklength m in the relaying case is much smaller than expected, and the performance loss due to a finite (relatively longer) blocklength $2m$ in the direct transmission scheme is larger than that. This is actually the performance advantage of relaying in the finite blocklength regime (in comparison to the infinite blocklength regime).

6.1.1.2 The Performance Advantage of Relaying under Perfect CSI Scenario vs. Average CSI Scenario.

In this subsection, we further compare the performance advantages of relaying under perfect CSI scenario and average CSI scenario. We show the comparison in Figure 6.3 and Figure 6.4 where the weight factor and the blocklength are varied respectively.

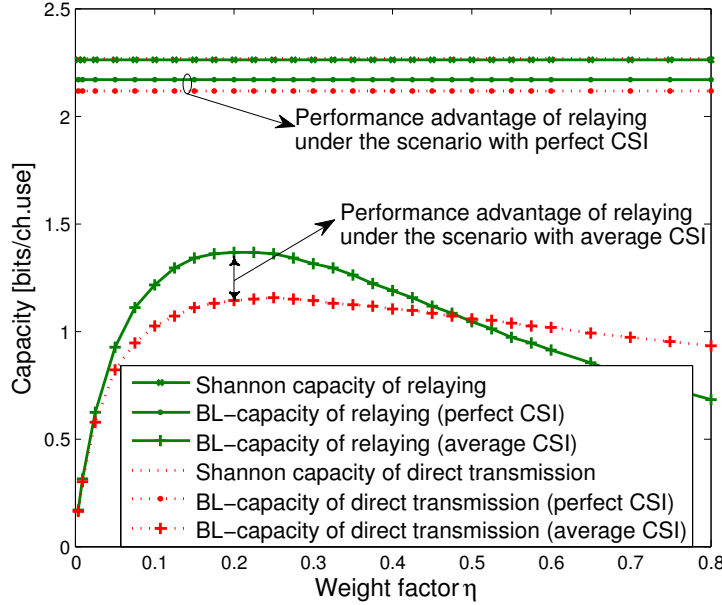


Abbildung 6.3: The performance comparison between relaying with average CSI and relaying with perfect CSI while varying the weight factor.

In general, these figures show that relaying has better blocklength-limited performance than direct transmission under both the perfect CSI scenario and the average CSI scenario. However, we further observe that the performance advantage of relaying under the average CSI scenario is more significant than under the perfect CSI scenario, i.e., the performance improvement by relaying (comparing relaying with direct transmission) under the average CSI scenario is significantly higher than the one under the perfect CSI scenario. The observation from Figure 6.3 and Figure 6.4 indicates that in the finite blocklength regime relaying is more beneficial for the average CSI scenario in comparison to the perfect CSI scenario.

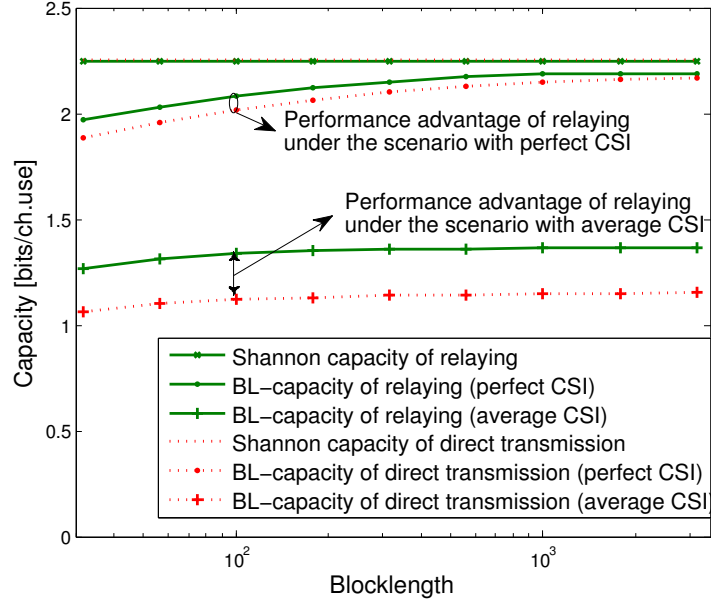


Abbildung 6.4: The performance comparison between relaying with average CSI and relaying with perfect CSI under dynamic blocklength ($\eta = 0.2$).

6.1.1.3 The Performance Loss Due to a Finite Blocklength: Perfect CSI Scenario vs. Average CSI Scenario

Finally, we compare performance losses due to finite blocklengths under the perfect CSI scenario and under the average CSI scenario.

Under the perfect CSI scenario, the performance loss due to a finite blocklength is actually the performance gap between the Shannon capacity and the BL-capacity (with perfect CSI). In addition, the performance loss under the average CSI scenario is observed by comparing the BL-capacity (with average CSI) with the outage capacity. In particular, the outage capacity is a performance metric under infinite blocklength regime for the average CSI scenario. It is given by $\rho(1 - \text{Pr}_{\text{out}}(\rho))$, where ρ is the packet size and $\text{Pr}_{\text{out}}(\rho)$ is the outage probability due to this packet size. To calculate the outage capacity, the weighting CSI operation is also considered in the simulation, i.e., the packet size is chosen based on the Shannon capacity of the weighted CSI.

Under the above setup, we first show numerical results of the comparison in Figure 6.5 where we fix the blocklength and vary the weight factor. The figure shows

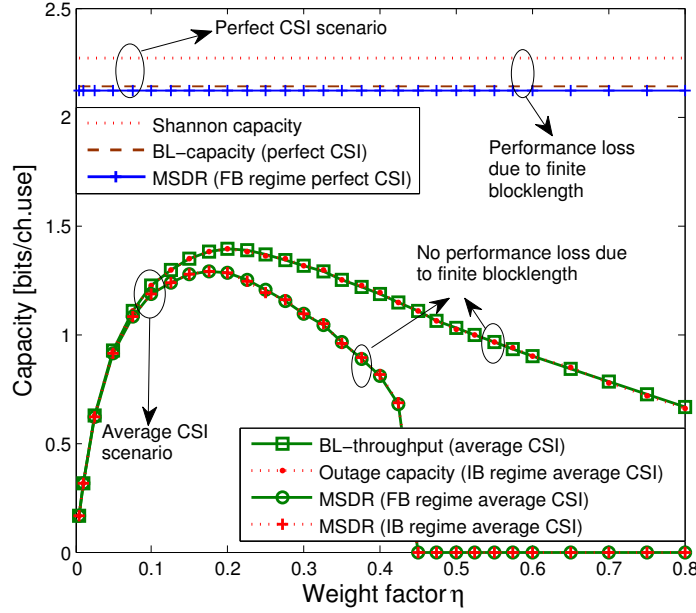


Abbildung 6.5: The performance comparison between relaying in the infinite blocklength regime and relaying in the finite blocklength regime (blocklength at each hop of relaying is 500 symbols).

that the performance loss due to a finite blocklength is considerable under the perfect CSI scenario. However, we find that under the average CSI scenario the performance loss due to a finite blocklength is negligible. In the simulation, we also observe that (not shown here) with only average CSI at the source the outage probability (in the infinite blocklength regime) and the average error probability (in the finite blocklength regime) have similar performance.

The observation from Figure 6.5 are based on the setup that the blocklength $m = 500$ symbols. Recall that the BL-capacity is limited by the blocklength and the performance in the infinite blocklength regime is not influenced by the blocklength. Hence, the performance loss due to a finite blocklength should also be subject to the blocklength. Then, we further investigate this performance loss due to a finite blocklength in Figure 6.14 where we vary the blocklength. Again, we observe that the performance loss under the average CSI scenario is much smaller than the one under the perfect CSI scenario. In particular, with a non-extremely short blocklength (e.g., blocklength

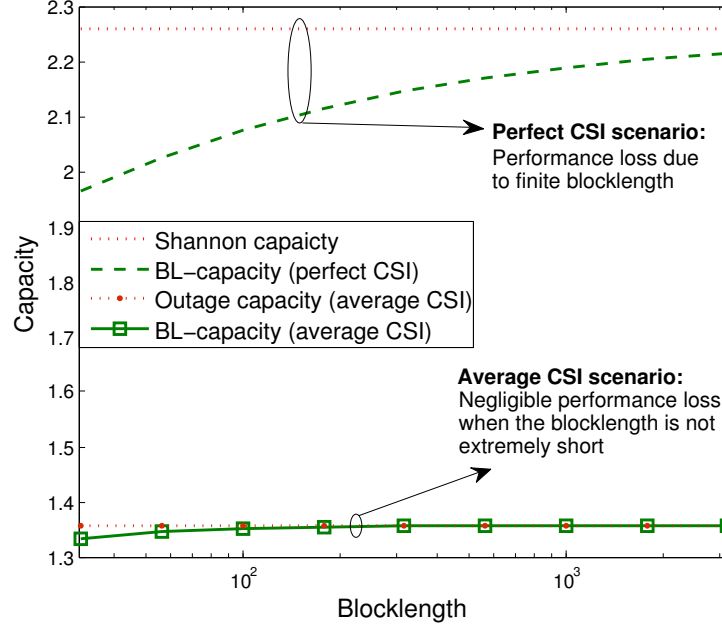


Abbildung 6.6: The performance comparison between relaying in the infinite blocklength regime and relaying in the finite blocklength regime while varying the blocklength. In the simulation, $\eta = 0.2$.

$m > 100$) the performance loss due to a finite blocklength is negligible under the average CSI scenario.

Summary

In this section, the performance advantage of relaying in the finite blocklength regime was shown. An important and surprising observation is that the performance loss due to a finite (relatively shorter) blocklength m in the relaying scheme is much smaller than expected, and the performance loss due to a finite (relatively longer) blocklength $2m$ in the direct transmission scheme is larger than that. This clearly shows the performance advantage of relaying in the finite blocklength regime (in comparison to the infinite blocklength regime). In other words, relaying is more efficient in systems with short latency and short blocklengths. Moreover, we found that this performance advantage of relaying under the average CSI scenario is more significant than under the perfect CSI scenario.

In the next section, we provide a theoretical explanation of this performance advantage of relaying.

6.1.2 A Theoretical Explanation of the Performance Advantage of Relaying

In this section, we theoretically study the performance advantage of relaying in the finite blocklength regime. Firstly, we continue considering the scenario introduced in the previous section where relaying and direct transmission have equivalent Shannon capacity. Based on this scenario we analyze the conditions under which relaying is more efficient in the finite blocklength regime than in the infinite blocklength regime, i.e., relaying has a higher blocklength-limited performance than direct transmission. Then, we discuss the performance advantage of relaying under more general scenarios.

The equivalent Shannon capacity of two-phase relaying is $C_R = C(z_*)/2$, where z_* is the gain of the bottleneck link and $C(z_*)$ is the corresponding Shannon capacity. Accordingly, the Shannon capacity of direct transmission is $C_D = C(z_1)$, where z_1 is the gain of the direct link. Therefore, if relaying and direct transmission have the same performance in the infinite blocklength regime, we have $C(z_*)/2 = C(z_1)$. Next, we study the performance difference (between relaying in the finite blocklength regime and relaying in the infinite blocklength regime) by comparing the (equivalent) coding rates between relaying and direct transmission. To make a fair comparison, the error probability of direct transmission and the overall error probability of relaying are assumed to be the same, i.e., $\varepsilon_D = \varepsilon_R$.

Therefore, the performance gap between the coding rate of relaying with blocklength m at each phase and the coding rate of direct transmission with blocklength

$2m$ is given by:

$$\begin{aligned}
 r_R - r_D &= R(z_*, \varepsilon_*, m)/2 - R(z_1, \varepsilon_D, 2m) \\
 &= C(z_*)/2 - \sqrt{(1 - 2^{-4C(z_*)})/8m} \cdot Q^{-1}(\varepsilon_*) \log_2 e \\
 &\quad - C(z_1) + \sqrt{(1 - 2^{-4C(z_1)})/4m} \cdot Q^{-1}(\varepsilon_D) \log_2 e \\
 &= C(z_1) - \sqrt{(1 - 2^{-8C(z_1)})/8m} \cdot Q^{-1}(\varepsilon_*) \log_2 e \\
 &\quad - C(z_1) + \sqrt{(1 - 2^{-4C(z_1)})/4m} \cdot Q^{-1}(\varepsilon_R) \log_2 e \\
 &= \sqrt{(1 - 2^{-4C(z_1)})/8m} \log_2 e \left(\sqrt{2} Q^{-1}(\varepsilon_R) - \sqrt{1 + 2^{-4C(z_1)}} \cdot Q^{-1}(\varepsilon_*) \right) \\
 &= A \cdot B,
 \end{aligned} \tag{6.1}$$

where $A = \sqrt{(1 - 2^{-4C(z_1)})/8m} \log_2 e$ and $B = \sqrt{2} Q^{-1}(\varepsilon_R) - \sqrt{1 + 2^{-4C(z_1)}} \cdot Q^{-1}(\varepsilon_*)$. Under the above assumptions, it follows that for $A \cdot B \neq 0$ there is a performance difference between the two schemes. In particular, if $A \cdot B > 0$, this indicates that relaying is more efficient in the finite blocklength regime than in the infinite blocklength regime. Obviously, A is positive and decreasing in m . On the other hand, B is not dependent on m . Hence, we have the following proposition:

Proposition 10. *In the finite blocklength regime, the performance gap between relaying and direct transmission, i.e., the absolute value of $r_R - r_D$, increases as the blocklength decreases. However, the blocklength does not influence if relaying or direct transmission has a better performance.*

Proposition 10 implies that if relaying outperforms direct transmission, i.e., $A \cdot B > 0$, the shorter the blocklength is, the bigger the performance gap is. This is an unexpected insight that relaying is more beneficial with shorter blocklengths. We further investigate the condition where relaying outperforms direct transmission based on the following error scenarios:

- **Common error scenario¹:** The (block) error probability of each link of relaying is lower than 0.5, i.e., $\max\{\varepsilon_1, \varepsilon_*\} < 0.5$. Hence, $Q^{-1}(\varepsilon_*) > 0$ and $\frac{\sqrt{2}Q^{-1}(\varepsilon_R)}{Q^{-1}(\varepsilon_*)}$ is

¹ In practice, the (block) error probabilities of transmissions are generally expected to be much lower than 0.5, e.g., an error probability in the range of 0.3-0.5 is normally regarded as relatively high. The common error scenario in this paper refers to error probabilities in the range of 0-0.5. It is actually the complement of the extreme error scenario.

decreasing in ε_R . Next, we prove that $A \cdot B > 0$ to show that relaying is definitely superior to direct transmission under this error scenario.

Beweis. $\max\{\varepsilon_1, \varepsilon_\star\} < 0.5$

$$\Rightarrow \varepsilon_R < \varepsilon_\star < 0.5$$

$$\Leftrightarrow Q^{-1}(\varepsilon_R) > Q^{-1}(\varepsilon_\star) > 0$$

$$\Rightarrow \frac{\sqrt{2}Q^{-1}(\varepsilon_R)}{Q^{-1}(\varepsilon_\star)} > \sqrt{2} > \sqrt{1 + \frac{1}{[1+\text{SNR}(z_1)]^2}} = \sqrt{1 + 2^{-4C(z_1)}}$$

$$\Rightarrow B > 0 \text{ and } A \cdot B > 0. \quad \square$$

- **Extreme error scenario²:** The error probability of at least one link of relaying is higher than 0.5, i.e., $\max\{\varepsilon_1, \varepsilon_\star\} > 0.5$. In the following, we prove that the condition $\varepsilon_R < \varepsilon_\star$ is sufficient to have $A \cdot B > 0$ under this scenario.

Beweis. Under the condition $\varepsilon_R < \varepsilon_\star$, if $\varepsilon_\star < 0.5$, we have $\varepsilon_R < \varepsilon_\star < 0.5$. Therefore, the rest of the proof is similar to the one above (starting from the second step) for the common error scenario. Hence, here we mainly consider the other situation where $\varepsilon_\star > 0.5$:

$$\varepsilon_R < \varepsilon_\star \text{ and } \varepsilon_\star > 0.5$$

$$\Leftrightarrow \varepsilon_R < \varepsilon_\star \text{ and } \frac{\sqrt{2}Q^{-1}(\varepsilon_R)}{Q^{-1}(\varepsilon_\star)} \text{ is increasing in } \varepsilon_R$$

$$\Rightarrow \frac{\sqrt{2}Q^{-1}(\varepsilon_R)}{Q^{-1}(\varepsilon_\star)} < \sqrt{2}$$

$$\Rightarrow \sqrt{2}Q^{-1}(\varepsilon_R) > \sqrt{2}Q^{-1}(\varepsilon_\star) > \sqrt{1 + \frac{1}{[1+\text{SNR}(z_1)]^2}}Q^{-1}(\varepsilon_\star) \\ = \sqrt{1 + 2^{-4C(z_1)}}Q^{-1}(\varepsilon_\star)$$

$$\Rightarrow B > 0 \text{ and } A \cdot B > 0. \quad \square$$

Summarizing, under the common error scenario (or under the extreme error scenario with $\varepsilon_R < \varepsilon_\star$) relaying is definitely superior to direct transmission in the finite blocklength regime even if their performance is the same in the infinite blocklength regime, i.e., $C_R = C_D$. Note that under the common error scenario we definitely have $\varepsilon_R < \varepsilon_\star$.

² Based on (5.11), the extreme error scenario corresponds to the case where the Shannon capacity of at least one link of relaying is lower than the coding rate (at each phase). We provide an example to clarify that this scenario is also feasible in practice. Recall that the source determines the coding rate based on the channel gain of the bottleneck link (either h_2^2 or $h_1^2 + h_3^2$). Therefore, it is possible that this coding rate is higher than the Shannon capacity of the direct link (with channel gain h_1^2).

Hence, $\varepsilon_R < \varepsilon_\star$ is actually a sufficient but not necessary condition to make relaying outperform direct transmission.

Based on Proposition 10, (if $A \cdot B > 0$) the performance advantage of relaying increases as the blocklength decreases. This actually can be explained based on the condition $\varepsilon_R < \varepsilon_\star$. The above analysis shows that with the same (overall) error probability ($\varepsilon_D = \varepsilon_R$) relaying can operate with a higher error probability per phase (i.e., $\varepsilon_R < \varepsilon_\star$) and hence set the coding rate more aggressively. Note that the fundamental difference between Shannon capacity and finite blocklength regime is the consideration of error probabilities in the latter case. Therefore, the performance loss (if comparing the finite blocklength coding rate to the Shannon capacity) is caused by the error probability. However, relaying can compensate the performance loss/decrease by operating with a higher error probability than direct transmission. One could also say that relaying simply has a lower performance decrease, and that leads to a relative performance advantage of relaying in comparison to direct transmission. If the blocklength is long and only introduces a slight performance loss in the finite blocklength regime, the advantage of relaying is not significant. However, a relatively short blocklength leads to a significant performance loss. Then, in case of relaying the performance loss is lower which leads to a higher coding rate in comparison to direct transmission.

So far we analyzed the advantage of relaying under the condition $C_R = C_D$, i.e., $C(z_\star)/2 = C(z_1)$. In the following, we consider the other channel conditions. Based on (5.9), r_D is decreasing in $C(z_1)$. Therefore, if $\varepsilon_R < \varepsilon_\star$ and $C(z_\star)/2 > C(z_1)$ (which means that $C(h_1^2)$ becomes smaller in comparison to the scenario $C(z_\star)/2 = C(z_1)$), in the finite blocklength regime relaying is definitely superior to direct transmission, too. Moreover, recall that the condition $\varepsilon_R < \varepsilon_\star$ is sufficient but not necessary to let $A \cdot B > 0$. This indicates the possibility that relaying could still be a better choice even if $C(z_\star)/2 < C(z_1)$. For example, relaying is more likely to be superior to direct transmission if $C(z_\star)/2$ is just slightly lower than $C(z_1)$.

Therefore, the performance comparison between relaying and direct transmission can be summarized in the following proposition with respect to various channel conditions:

Proposition 11. *Under the common error scenario (or under the extreme error scenario with $\varepsilon_R < \varepsilon_\star$), we have:*

- $C_R > C_D \Rightarrow r_R > r_D$.

- $C_R = C_D \Rightarrow r_R > r_D$.
- *Even if $C_R < C_D$, it is possible that $r_R > r_D$.*

Proposition 11 clearly shows the performance advantage of relaying in the finite blocklength regime in comparison to the infinite blocklength regime. In fact, based on Proposition 11 the performance advantage of relaying leads to a broader spatial area for selecting/deploying a relay in the finite blocklength regime as shown in Fig. 6.7. In the figure, if a relay is deployed in region A, relaying outperforms direct transmission under both the finite blocklength regime and the infinite blocklength regime. On the other hand, if a relay is deployed in the (ring shaped) region B, relaying has a

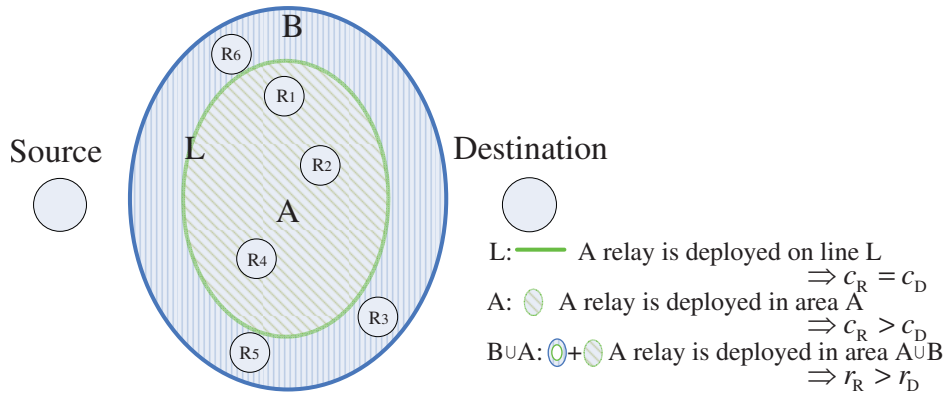


Abbildung 6.7: The performance advantage of relaying in the finite blocklength regime (from a perspective of topology).

higher blocklength-limited performance but a lower Shannon capacity in comparison to direct transmission. From a perspective of topology, our explanation of the performance advantage of relaying can be understood as that we show when and why region B exists.

Numerical Results and Discussion

We present numerical results to illustrate our analytical model. As both the finite blocklength coding rates and Shannon capacities vary in SNR, to investigate the difference between the performances (of either relaying or direct transmission) in the finite blocklength regime and in the infinite blocklength regime, we observe the ratio (ρ) of the (equivalent) coding rate to the (equivalent) Shannon capacity: $\rho_R = r_R/C_R$ for

relaying and $\rho_D = r_D/C_D$ for direct transmission. In Fig. 6.8 we compare the performance ratios ρ_R and ρ_D for different error probabilities and different transmit powers while setting $C_R = C_D$ and $\varepsilon_R = \varepsilon_D$. The figure demonstrates our analytical finding that relaying is superior to direct transmission if the error probability is not extremely high. From the figure, we observe that only under extreme error scenario, e.g., $\varepsilon_R = \varepsilon_D = 10^{-0.3} \approx 0.501$, the performance ratio of direct transmission is slightly higher.

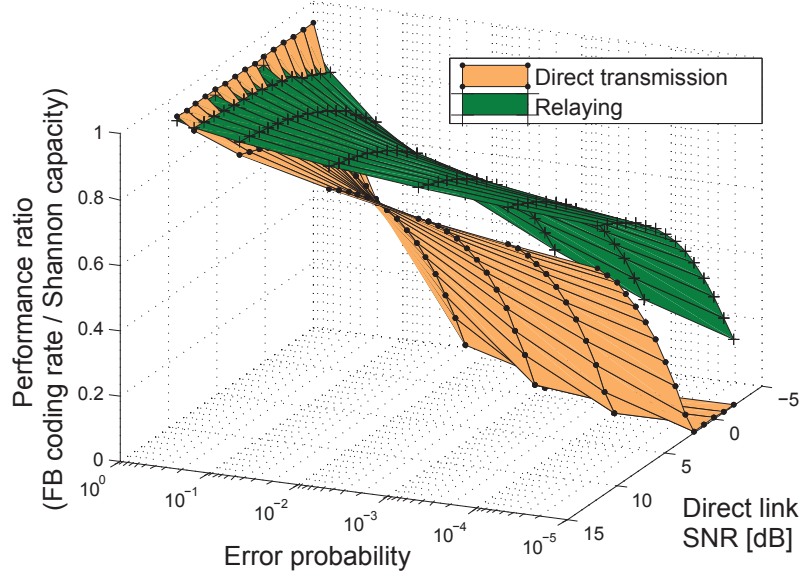


Abbildung 6.8: Performance ratios of relaying and direct transmission while the blocklength of each phase of relaying and the blocklength of direct transmission are 100 and 200. We vary p_{tx} to obtain different SNR while setting $C_R = C_D$.

We further investigate the performance advantage of relaying by observing ρ_R/ρ_D in Fig. 6.9. In the figure, the abscissa is the blocklength at each phase of relaying while the blocklength of direct transmission is twice as large. The figure illustrates Proposition 10, i.e., the performance advantage of relaying in the finite blocklength regime is more significant with short blocklengths in comparison to direct transmission. Moreover, the performance advantage of relaying is higher under poor channel conditions.

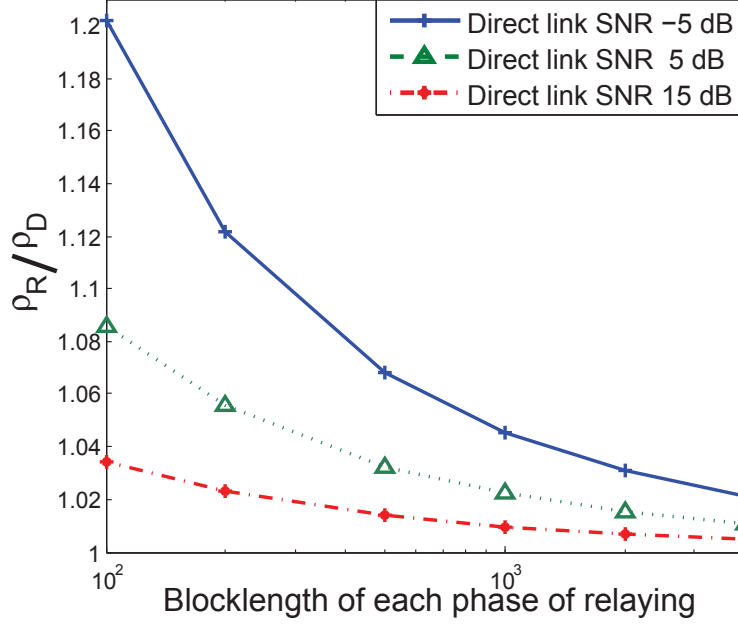


Abbildung 6.9: Performance advantage ratio of relaying to direct transmission ($\varepsilon_R = \varepsilon_D = 10^{-3}$).

In addition, we provide the corresponding absolute values of the second case in Fig. 6.9 where the SNR of the direct link is about 5 dB. We show these values in Fig. 6.10. This figure illustrates our analysis in Proposition 11 that (although having a slightly lower Shannon capacity) the coding rate of relaying in Fig. 6.10 shows a significantly higher performance than direct transmission. For example, when the blocklength of each phase of relaying equals 500 (at the same time, the blocklength of direct transmission is 1000), the coding rate of direct transmission is only about 89% of the corresponding Shannon capacity while for relaying this ratio increases to 94%.

Summary

In this section, in a single relay system we analyzed the performance advantage of relaying analytically. We proved that relaying is definitely superior to direct transmission in the finite blocklength regime if the error probability of the bottleneck link of relaying is higher than the overall error probability of relaying (even relaying and direct transmission have the same Shannon capacity). In particular, for the common error

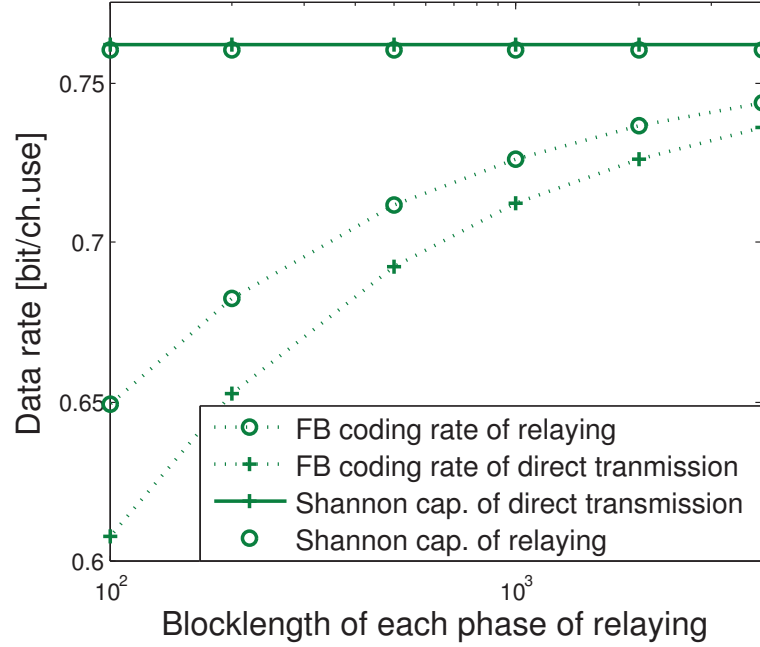


Abbildung 6.10: The comparison between relaying and direct transmission with finite blocklengths while the SNR of the direct link is 5 dB and $\varepsilon_R = \varepsilon_D = 10^{-3}$.

scenario, relaying is definitely superior to direct transmission in the finite blocklength regime. This performance advantage of relaying in the finite blocklength regime is more significant with short blocklengths. From a perspective of topology, this performance advantage of relaying leads to a broader area for selecting/deploying a relay.

6.2 Performance Advantage of Relaying under a Multi-Relay Scenario

In the previous section, by comparing with direct transmission we have shown the performance advantage of relaying in the finite blocklength regime. In this section, the focus is to evaluate the impact on this performance advantage of relaying while multiple relays are deployed in the network.

6.2.1 Simulative Investigation

We first provide the performance of relaying in the infinite blocklength regime as well as in the finite blocklength regime. By doing so, we observe the performance advantage of multi-relay in the finite blocklength regime in comparison to the infinite blocklength regime. Then, we further evaluate this performance advantage of multi-relay under different scenarios.

In this section, simulations are based on a multi-relay system with a model same as Section 4.1.1. Recall that after the source transmitting a packet, all the relays try to decode it. Only these relays which decode the packet successfully participate in the forwarding. If the source has perfect CSI of all the links, it is able to determine an optimal packet size to have a set of appropriate forwarding relays to maximize the overall capacity in the infinite blocklength regime. This maximal overall capacity is actually the equivalent Shannon capacity of multi-relay transmission. We denoted this equivalent Shannon capacity by C_{MR} . Hence, C_{MR} is given by $C_{\text{MR}} = \max_{\Omega \subseteq \{R_j, j=1,2,\dots,J\}} \{C(\Omega)\}$, where $R_j, j = 1, 2, \dots, J$ denote all the relays deployed in the system. In addition, $C(\Omega)$ is equivalent Shannon capacity of multi-relay transmission with the forwarding relay set Ω . In other words, C_{MR} is the maximum of all the possible equivalent Shannon capacities with feasible set of forwarding relays. Moreover, $C(\Omega)$ is actually subject to the bottleneck hop (either the first hop or the second hop) of this two-hop multi-relay-assisted transmission. It equals the half of the Shannon capacity of the bottleneck hop. On one hand, the Shannon capacity of the first hop is subject to the weakest one among links from the source to all the relays in set Ω . On the other hand, by combining (i.e., by MRC) these signals from all relays in Ω to the destination, the received SNR at the destination is given by $\sum_{R_j \in \Omega} \gamma_{R_j-D}$, where γ_{R_j-D} is the SNR of the link from relay j to the destination. Therefore, we have $C(\Omega) = \min \left\{ \min_{R_j \in \Omega} \{C(\gamma_{S-R_j})\}, C\left(\sum_{R_j \in \Omega} \gamma_{R_j-D}\right) \right\} / 2$, where γ_{D-R_j} is the SNR of the link from the source to relay j .

Regarding the model parameterization, we consider a similar one as presented in Section 6.1. The only difference is that here not a single relay but multiple relays are deployed in a group (with radius 10 m) between the source and the destination.

6.2.1.1 The Impact of Relay Number on the Performance Advantage of Relaying

As shown in Figure 6.11, we investigate the relationship between the relaying performance and the number of relays deployed in the system. Both the perfect CSI scenario and the average CSI scenario are considered in the simulation.

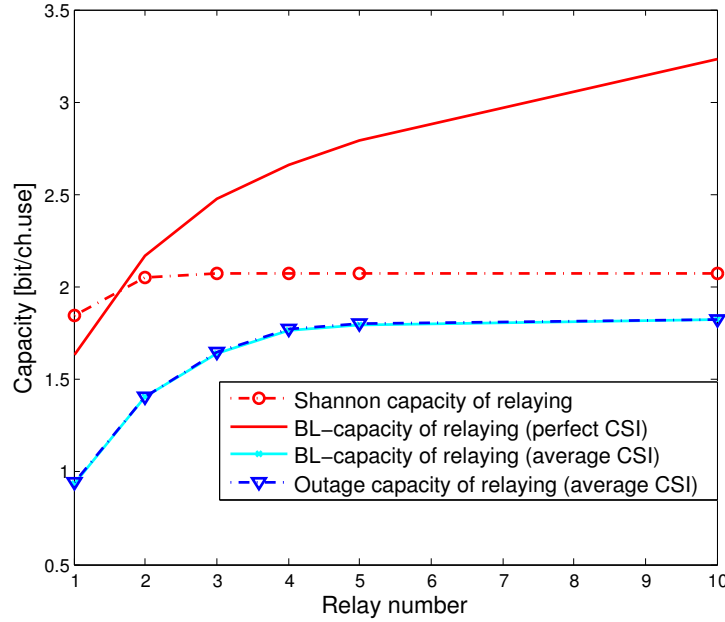


Abbildung 6.11: Shannon capacity vs. BL-capacity while varying the number of relays deployed in the system.

Firstly, as expected the figure shows that as more relays are deployed in the system both BL-capacities and the Shannon capacity or outage capacity are improved.

Secondly, with only the average CSI at the source multi-relay transmission has similar performance under these two regimes. This observation matches our previous observations (Figure 6.5) in a single relay scenario.

Thirdly and surprisingly, the figure shows that with perfect CSI the BL-capacity of multi-relay transmission exceeds the Shannon capacity if there are two or more relays deployed in the system! This is really surprising as the Shannon capacity is always expected to be a upper limit of the throughput/capacity of a transmission. However,

our results show that under the studied multi-relay system the upper performance limit is not the Shannon limit anymore.

Finally, (under the perfect CSI scenario) the mismatch between the BL-capacity and Shannon capacity is considerable and it is significantly increasing in the number of relays deployed in the system. This is another important observation regarding the performance advantage of relaying with multiple relays in the finite blocklength regime.

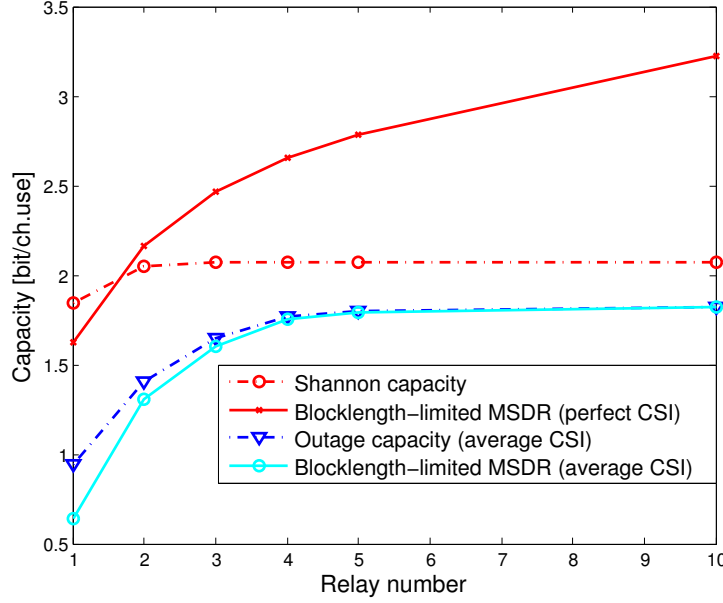


Abbildung 6.12: Shannon capacity vs. blocklength-limited MSDR while varying the number of relays deployed in the system.

The corresponding blocklength-limited MSDR performance of relaying with multi-relay is shown in Figure 6.12. We observe similar advantage of relaying that with perfect CSI the blocklength-limited MSDR violates the corresponding Shannon capacity limit.

6.2.1.2 Evaluating the Performance Advantage of Multi-Relay in the Finite Blocklength Regime

We continue to evaluate the performance advantage of multi-relay in the finite blocklength regime. In Figure 6.13, we vary the weight factor which influences the coding rate under the average CSI scenario. We observe again that the BL-capacity matches

outage capacity under the average CSI scenario but higher than the Shannon capacity under the perfect CSI scenario. More importantly, we find that the optimal weight factor (to maximize the BL-capacity) is about 2. Recall that with a single relay in the system the optimal value of this weight factor is about 0.2 (see Figure 6.5). This is a significant difference. In fact, if the source set this weight factor to 1, this results in a coding rate which has similar value as the average Shannon capacity (over fading). Hence, the observation that the optimal weight factor is about 2 actually indicates that the coding rate determined by the source is much higher than the average Shannon capacity. More importantly, in the finite blocklength regime this significant coding rate fits well with the considered multi-relay system. Therefore, (in comparison to single relay) multiple relays introduce more gain to the relaying performance in the finite blocklength regime.

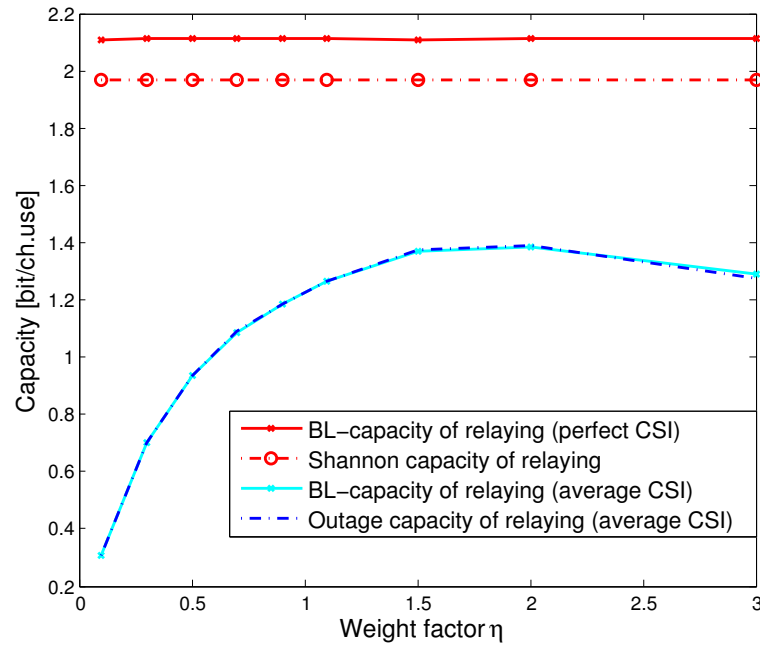


Abbildung 6.13: Relaying performance under two-relay scenario.

In Figure 6.14, we further show that (under the perfect CSI scenario) the mismatch between the BL-capacity and the Shannon capacity is also subject to the blocklength. In particular, the BL-capacity is increasing in the blocklength while the Shannon capacity is not influenced by the blocklength. As a result, the BL-capacity is lower than

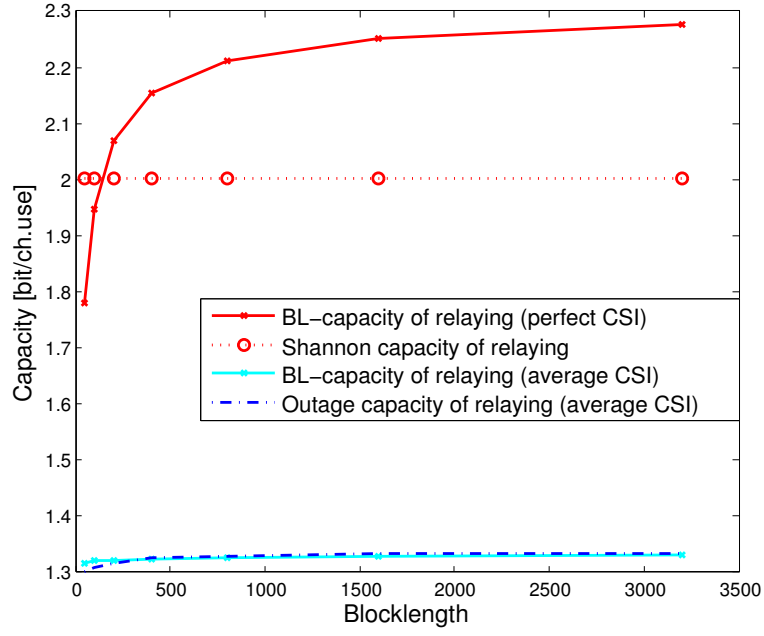


Abbildung 6.14: The performance of two-relay transmission while varying the blocklength ($\eta = 2$).

the Shannon capacity in the extreme short blocklength region but significantly higher than the Shannon capacity with a long block.

Summary

In this section, we observe the performance advantage of multi-relay in the finite blocklength regime. In particular, with perfect CSI at the source the BL-capacity is significantly higher than the Shannon capacity when multiple relays are deployed in the system, which violates the common sense regarding the Shannon limit. In the next section, we continue to discuss the reason of this observation.

6.2.2 An Explanation of the Performance Advantage of Multi-Relay in the Finite Blocklength Regime

Shannon capacity theory describes the upper capacity limit of a transmission without errors. It indicates an SNR threshold (subject to packet size) based on which the transmission in the infinite blocklength regime can be determined as a success or an outage. In other words, with given instantaneous CSI (SNR) and packet size it is predictable if the transmission succeeds. However, in the finite blocklength the transmission is suffered by noise. In particular, even the instantaneous SNR is significantly higher than the above SNR threshold, it is possible that an error occurs. On the other hand, even the instantaneous SNR is lower than the SNR threshold (which definitely leads to an error in the infinite blocklength regime) the transmission is possible to succeed in the finite blocklength regime. And the probability of this is up to 50%. These are fundamental differences between the Shannon capacity theory and the finite blocklength model. These fundamental differences do not change the relationship between the BL-capacity and the Shannon capacity (i.e., the Shannon capacity is the upper limit of BL-capacity) until considering a network where a successful (overall) transmission allows errors to occur at partial links, e.g, the considered multi-relay system.

In the considered multi-relay system, it is possible that some of relays fail to decode the packet and only the rest relays forward the packet to the destination. It is also possible that the overall transmission is successful while only partial relays decode the packet correctly (from the source) and join the forwarding. The Shannon bound does not hold anymore for such a transmission which is not a completely error-free communication.

For an extreme example, consider a system with 1000 relays. These relays are divided into two groups based on their channel conditions: One group has 999 relays while the other group has only one relay. Relays in the first group have the same Shannon capacity of their backhaul links, which is 0.99 bits/ch.use . At the same time, they also have the same Shannon capacity of forwarding links, which is 10 bits/ch.use . For the second group, the only relay has the same Shannon capacities of its backhaul link and forwarding link, which is 1 bits/ch.use . For such a scenario, the first hop is obviously the bottleneck hop of the transmission. In particular, in the infinite blocklength regime there are two options when the source determines the coding rate, i.e. either 1 bits/ch.use or 0.99 bit/ch.use . On the one hand, determining the coding rate as 1 bits/ch.use results in that only the relay in the second group decodes data packets successfully and forwards packets to the destination. Then, Shannon capacities of first and second hops are the same as 1 bits/ch.use . Hence, the equivalent Shannon capacity of the two-hop relaying is 0.5 bits/ch.use . On the other hand, setting the coding rate as 0.99 bits/ch.use makes all these 1000 relays decode the packet correctly. Under this case, the Shannon capacity of the first hop is 0.99 bits/ch.use while the Shannon capacity of the second hop is significantly higher than the first hop. As a result, the equivalent Shannon capacity of the two-hop relaying is 0.495 bit/ch.use . Therefore, in the infinite blocklength regime (with perfect CSI) the source definitely chooses the first option which has a better performance. However, in the finite blocklength regime even the source chooses coding rate based on 1 bits/ch.use , these 999 relays in the first group are also possible to decode the packet successfully while the probability of this is just slightly lower than 0.5 (as backhaul links of these 999 relays are just slightly poorer than the last relay's). Hence, the expected number of forwarding relays is not 1 but about 500. Obviously, 1 bits/ch.use is not the optimal coding rate in the finite blocklength regime. The source should set the coding rate more aggressively. For example, setting the coding rate as 4 bits/ch.use makes the expected number of forwarding relays be about 75, which is generally quite enough for a successful transmission at the second hop. Under this case, the BL-capacity is 2 bits/ch.use which is significantly higher than the Shannon capacity 0.5 bits/ch.use .

In a word, multiple relays introduce a kind of potential and opportunistic diversity gain in the finite blocklength regime (in comparison to in the infinite blocklength regime) by potentially making more relays forward data packets.

6.3 Summary

The work in this chapter is the last step of this thesis for developing investigation and evaluation models of relaying. In this chapter, by simulation we compared the relaying performance in the infinite blocklength regime and the one in the finite blocklength regime. We showed the performance advantages of relaying in the finite blocklength regime (in comparison to the infinite blocklength regime) in a single relay scenario as well as a multi-relay scenario. In particular, if the system only has one relay, relaying helps reduce the performance loss (due to a finite blocklength). Second, under a multi-relay scenario, relaying introduces a significantly high blocklength-limited performance which is more likely higher than the Shannon limit. All these are very surprising and interesting observations. Moreover, our work shows that relaying behaves differently under these two regimes. This may motivate researchers to analyze and improve the blocklength-limited performance of various cooperative communications networks as well as design efficient relaying protocols in the finite blocklength regime.

7 Conclusions and Future Work

Relaying has been adopted as a key technology for the future wireless communication. At the same time, latency, reliability and energy efficiency are considered as three major concerns for designing future wireless networks. Hence, for designing efficient future wireless networks it is essential to investigate how much improvement can be introduced by relaying regarding these performance metrics. However, most existing research on investigating the relaying performance typically does not only focus on the physical-layer performance but also rely on simplifying assumptions, such as perfect CSI and infinite blocklength, which may be overoptimistic, especially for systems of applications with critical QoS constraints.

In this thesis, systematical investigation models of relaying are developed. The above performance metrics are addressed from both the physical-layer and the higher-layer perspectives. Moreover, various scenarios with different assumptions regarding CSI, blocklength and topology (e.g., perfect or imperfect CSI, finite or infinite blocklengths and single-relay or multi-relay) are studied. In this chapter, we first summarize this thesis and present its contributions. Then, we discuss possible future research directions.

7.1 Summary and Discussion

In Chapter 4, in the infinite blocklength regime we made a systematic analysis of a relaying network where relays participate retransmissions to improve the reliability. Initially, we investigated the outage probability and derived the distribution of transmission delay. Subsequently, the physical-layer throughput and the higher-layer effective capacity were analyzed. Later on, we analyzed the physical-layer as well as the higher-layer energy efficiency by considering the PEE and the EEE which are the

ratios of the throughput to the energy consumption and of the effective capacity to the energy consumption. In particular, the proposed EEE was proved to be a good metric in investigating the tradeoff among throughput, delay, energy consumption". Moreover, by simulation we validated our analytical model. Besides, we proposed two simple protocols to improve the PEE and the EEE, respectively. Finally and more importantly, we evaluated the considered relaying network and concluded several guidelines for system design, e.g., the PEE and EEE are quasi-concave in the packet size and the number of relays deployed in the system.

Chapter 5 addressed in general the blocklength-limited relaying performance. The overall error probability, BL-capacity, effective capacity, PEE and EEE were derived under various scenarios which differ in the channel model and in the assumption of CSI. In particular, under the perfect CSI scenario with static channels we proved that the BL-capacity is concave in the coding rate and quasi-concave in the overall error probability of relaying. Under the average CSI scenario with quasi-static channels, we proposed a simple system operation by introducing a weight factor based on which we weight the average CSI and let the source determine the coding rate based the weighted CSI. We proved that both the BL-capacity and the MSDR are concave in the coding rate. More importantly, it was proved that the BL-capacity and the MSDR are quasi-concave in the weight factor. In other words, the relaying performance of all the studied scenarios can be easily optimized based by choosing an appreciate coding rate or weight factor.

In Chapter 6, by simulation we investigated the performance difference between relaying in the finite blocklength regime and the one in the infinite blocklength regime. We showed the performance advantage of relaying in the finite blocklength regime (in comparison to the infinite blocklength regime). In particular, this performance advantage of relaying is mainly reflected in two aspects. First, if the system only has one relay, relaying helps reduces the performance loss (which is due to a finite blocklength). Second, under a multi-relay scenario, relaying introduces a significant blocklength-limited performance which is more likely higher than the Shannon limit.

Discussion

Combing analytically investigations as well as observations from simulative evaluations of the above three Chapters, we have the following findings:

- The fundamental difference between assumptions of finite blocklengths and infinite blocklengths is regarding the error model, as shown in Figure 7.1. In parti-

	Infinite blocklength regime	Finite blocklength regime
Perfect CSI	Error-free	Errors due to noise
Average CSI	Outages due to channel fading	Errors due to noise and channel fading

Abbildung 7.1: Error models under different scenarios/assumptions

cular, in the infinite blocklength regime (infinite blocklengths) with the perfect CSI the source perfectly adjusts the packet size based on the CSI to achieve a error-free transmission. However, even with the perfect CSI, a finite blocklength introduces errors. If there is only the average CSI available at the source, outages may occur in the infinite blocklength regime while errors due to both noise and channel fading likely occur in the finite blocklength regime.

- Considering the above error models, the relaying system is able to improve the reliability of the transmission by various operations, e.g., applying C-ARQ protocols, reducing packet size (coding rate) and introducing more relays to the system. We have following recommendations for these operations:
 - Regarding C-ARQ: i. MRC should be applied in C-ARQ protocols. ii. Direct link transmission can usually be ignored by C-ARQ protocols unless a specific topology of the nodes is given. iii. Dynamic relay selection should be enabled as it leads to a better energy efficiency, although they consume more energy in comparison to static protocols.
 - Regarding reducing the packet size: Under average CSI scenario we introduce a weight factor for the source to determining the packet size (coding

rate) based on the weighted average CSI. More importantly, it is shown that the BL-capacity and blocklength-limited MSDR as well as the outage capacity and MSDR with infinite blocklengths are quasi-concave in the proposed weight factor.

- Regarding deploying multiple relays: i. This boots the system performance significantly while overhead should be spent to acquire the exact number of relays present in the system. ii. As more relays joining the system and the transmission, additional energy is consumed. In fact, both the EEE and PEE are quasi-concave in the relay number.
- Based on the investigations of above operations of relaying, we conclude that relaying is an efficient way to improve the system performance under all the scenarios with different assumptions regarding CSI and blocklength. In particular, we find that relaying is more efficient for finite blocklengths and average CSI scenarios:
 - A finite blocklength introduces performance loss in comparison to the Shannon capacity. However, relaying (which halves the blocklength) is able to mitigate this loss by setting the coding rate at each hop aggressively. Moreover, we found that this performance advantage of relaying (with a single relay) under the average CSI scenario is more significant than under the perfect CSI scenario. In other words, relaying (with a single relay) is more efficient in systems with short latency/blocklengths and average CSI.
 - Surprisingly, under a multi-relay scenario relaying introduces a significant high blocklength-limited performance, e.g., the BL-capacity. In particular, with only average CSI relaying has similar performance in the infinite blocklength regime and in the finite blocklength regime. However, if perfect CSI is assumed to be available at the source, the BL-capacity of relaying is possible and more likely to be higher than the Shannon capacity limit.
- After observation significant performance gaps between different assumptions, we recognized that it should be recommend to the literature the importance to consider the more realistic assumptions regarding CSI and blocklengths, especially when a low-latency or/and low error probability scenario is considered.
- Finally, we conclude a set of guidelines for designing efficient relaying networks:

- For relaying with infinite blocklengths and average CSI: i. The packet size can be optimized by maximizing the EEE or the PEE. ii. The EEE and PEE can be further improved by introducing a tradeoff between dynamic protocols and static protocols. iii. Deploying relays further boosts the throughput and effective capacity. Moreover, an optimal number of relays deployed in the system can be determined by maximizing the EEE or the PEE.
- For relaying with finite blocklengths and perfect CSI: i. The PEE is increasing in the blocklength while EEE is quasi-concave in the blocklength. Hence, for a QoS-support system the blocklength can be determined by maximizing the EEE. ii. The BL-capacity and the blocklength-limited effective capacity are concave in the coding rate. This is an important guideline for the source to determine the coding rate based on the perfect CSI. iii. Under this scenario, deploying multiple relays boosts the blocklength-limited performance more significant than the rest two scenarios.
- For relaying with finite blocklengths and average CSI: i. The blocklength-limited effective capacity and BL-capacity are quasi-concave in the proposed weight factor. ii. The MSDR is concave in the blocklength while BL-capacity is increasing in blocklength. iii. EEE performance is quasi-concave in the blocklength.

7.2 Novelty and Contributions

Novelty and Discussion

In this work, we developed a set of investigation models of relaying, while both the infinite blocklength regime and the finite blocklength regime are considered. Most of the scenarios of relaying we considered (especially from a higher-layer perspective) are novel to the literature. In the following we discuss the novelty of this work in detail.

In the infinite blocklength regime, we considered the impact of relay-assisted retransmissions on the reliability, queuing performance and the higher-layer energy efficiency. Different from existing articles in the literature, we considered a relaying scenario

where the channel knowledge is not perfect and ARQ feedback overhead of time and energy is considered. In particular, the tripartite tradeoff among throughput, delay and energy consumption is studied from both physical-layer and higher-layer perspectives. These are key components of the novelty of this thesis regarding relaying in the infinite blocklength regime.

We also study the relaying performance in the finite blocklength regime in Chapter 5 and compare it with the infinite blocklength regime in Chapter 6. To the best of our knowledge, this part of work is original (regarding investigating relaying in the finite blocklength regime). Hence, expressions we derived and relationships we proved (e.g., the BL-capacity is concave in the coding rate) are novel to the literature. Moreover, we show the performance advantage of relaying in the finite blocklength regime. In particular, in a single-relay system the performance loss due to a finite (relatively shorter) blocklength is much smaller than expected. And more surprisingly, in a multi-relay system the BL-capacity violates the Shannon capacity limit. These interesting observations and explanations we provided are also an important component of the novelty of this thesis.

Contributions and Discussion

Discussions regarding contribution actually have been provided in summary sections at the end of previous three chapters. However, these are mainly regarding the guidelines we concluded for designing a relaying system under the same/similar assumptions as this work. In this part, we would like to discuss several theoretical contributions that can be used in investigating and designing a general relaying system or even a non-relaying system.

First of all, the topology approximation for a multi-relay scenario should be an important one. It indeed significantly facilitates the analytical model of a multi-relay system. At the same time, it is an appropriate approximations even if the relays are a bit separated.

Second, the higher-layer energy efficiency EEE we proposed should be also mentioned. This metric can be considered in the design of any QoS-support networks, as it

directly demonstrates how effective the system spends energy for the QoS requirements of interest.

Third, the methodology we considered to show the performance difference between infinite blocklength regime and the finite blocklength regime is also can be considered in the study of blocklength-limited performance of non-relaying networks. By determining the single-hop transmission as a reference scheme and making the subject transmission scheme and single-hop scheme have the same Shannon capacity, the performance difference of this subject transmission scheme under these two regimes can be easily compared and observed.

7.3 Future Research Directions

Future work regarding relaying and C-ARQ protocols in the infinite blocklength regime

Several issues remain as future work. The first issue is to develop an optimization model for a C-ARQ protocol with truncation by determining the optimal delay threshold for packet dropping. For example, in a system the source has packets which need to be transmitted to the destination under certain delay constraint (including the delay of transmission/retransmissions and the waiting time in the queue at the source). Unfortunately, if the related channel SNRs are very low, it is possible that a packet has been retransmitted many times but still cannot be decoded correctly at the destination. This situation is very bad, it introduces more delays to all the other packets waiting in the queue buffer at the source. Therefore, a packet dropping mechanism should be enabled. In this thesis, we do not consider the dropping mechanism in these studied C-ARQ protocols. However, based on the delay distribution and the outage probability of C-ARQ in (4.18) we derived, one can consider to design an optimal dropping mechanism to maximize the MSDR or the EEE.

Another potential issue is associated with and motivated by the first one. It is to improve the effective capacity model by taking into account the packet dropping/missing ratio. Recall that the effective capacity is modeled based on QoS cons-

straints of the delay and the delay violation probability. It is assumed under the service process model that each packet only be removed from the source buffer if the destination has successfully decoded it. In other words, all the data packets in the buffer at the source needs to be transmitted to the destination. However, if we consider a packet dropping mechanism, it is possible that a packet is removed from the source buffer although it has not been correctly decoded at the destination. Moreover, sometimes we have the situation of data transmission in practice that a data packet or some data packets are missing during the transmission. Hence, the effective capacity model does not fit these situations well and therefore one could consider to improve it.

Moreover, the topology approximation proposed in this work can be extended to other networks (with multi-relay), e.g., sensor networks and cognitive networks. For example, the it can be applied in a multi-relay cognitive network to simplify the investigation of either physical-layer or higher-layer performance of the secondary transmission. We have some primary results regarding the outage probability performance in a multi-relay cognitive network with average CSI, which have been published in [65]. Owing to the topology approximation, the analysis in [65] is facilitated, which results in some compact and closed-form derivations. The work in [65] is only a specific application of the approximation. There are still a large number of issues regarding multi-relay transmission/retransmission (without perfect CSI) under various networks aiming at various performance metrics. Most of these issues are quite open since under the multi-relay scenario (without perfect CSI) the number of forwarding relays is unpredictable (without applying our approximation). However, these tough problems can be significantly facilitated and more likely become solvable by applying this approximation.

Finally, it should be mentioned that relay-assisted network coding (RANC) ARQ shares similar objectives with the C-ARQ. RANC ARQ protocols require the relay to forward a new coded packet which is generated by combining several existing data packets in the relay buffer. Several RANC ARQ protocols are proposed and related throughput and delay performances are evaluated under two-way relay scenarios [66, 67], multi-cast scenarios [68, 69] and single one-way relay scenarios [70, 71]. However, to the best of our knowledge the related higher-layer queuing performance regarding RANC ARQ protocols has not been studied so far. Hence, it is also a potential future work to extend our analysis regarding C-ARQ protocols to the RCNC ARQ systems.

Future work regarding relaying in the finite blocklength regime

In the finite blocklength regime, research issues regarding relaying or cooperative communication are quite open. In this work, we in general addressed the relaying performance with a finite blocklength in several scenarios: static channel with perfect CSI as well as quasi-static channels with either perfect or average CSI. Several closely issues remain as follows.

Firstly, we have analytically proved that the BL-capacity of relaying is generally concave in the coding rate. It will be a further step if we can derive a closed-form expression of the optimal solution as well as the optimal value while considering to maximize the BL-capacity).

Secondly, under a fading scenario the CSI can be obtained at the destination by sampling the channel a slot prior to each upcoming transmission period. Owing to feedback mechanism, the instantaneous CSI could be available at the source. However, in practice the instantaneous CSI is subject to delay [72, 73, 74]. In particular, in a relaying network the second link (hop) of relaying is expected to have a longer delay (of the outdated CSI) than the first hop. The blocklength-limited performance of relaying with outdated instantaneous CSI would be a very interesting issue. In particular, it is interesting to investigate if relaying still has advantage over direct transmission as the delay (of the outdated CSI) at the second hop of relaying is expected to be longer than that of direct transmission.

Thirdly, in the infinite blocklength regime there is a hybrid relaying protocol combining AF and DF called DAF relaying [75, 76], i.e., if the relay decode the packet successfully it works under a DF mode, otherwise it works under a AF mode. DF relaying removes the noise but only suits in the high SNR region. Af relaying forwards data packet in the full SNR region but does not introduce significantly high SNR as it amplifies data together with noise. This hybrid relaying protocol successes advantages [77, 78] from both AF and DF protocols, i.e., it shift to DF relaying if the SNR is good enough to let the relay decode the data correctly and shift to AF relaying on the contrary. In the infinite blocklength regime, there exists a SNR threshold for determining the work mode of relaying. If we consider the finite blocklength regime, does this threshold still hold? Note that the fundamental difference between the infi-

nite blocklength regime and the finite blocklength regime is regarding errors due to noise. And the same time, AF relaying amplifies the noise. Hence, it is not clear if this kind of hybrid relaying protocol still make sense? In this dissertation, we show that DF relaying is more efficient in the finite blocklength regime in comparison to the infinite blocklength regime. It is suspected that AF relaying also introduces performance advantage like DF relaying. Therefore, a careful analysis is needed and also interesting.

Fourthly, in the finite blocklength regime we have shown the performance advantage of multi-relay. We also provide an explanation regarding this advantage. However, it should a further step to study the fundamental relationship between the blocklength-limited performance and the number of relays deployed in the system. In particular, it is essential to study the achievable diversity order of the multi-relay system in the finite blocklength regime, which is expected to be a function of the number of relays.

In addition, our study of relaying in the finite blocklength regime has focused on the capacity issue, e.g., BL-capacity and MSDR. It is also interesting to design a relaying system while minimizing the overall error probability. For example, if the coding rate is given the overall error probability of relay can be optimized by time and power allocation (for two hops) based on either perfect CSI or imperfect CSI.

Moreover, in this work when we considered the multi-relay network we assume all the active relays forward the data packet together (at the same time) in the relaying hop. In fact, there is another well-known operation of the relaying hop, which divides this hop into several slots. Each active relay takes one slot to forward a part of the data packet. It has been shown in [79] that in the infinite blocklength regime by time allocation the system performance can be optimized. In this paper, it is assumed that for a active relay the length the part of the data packet (needs to be forwarded) is proportional to the length of the slot. However, this optimization problem should be different if we consider the finite blocklength regime. Time allocation directly influences the blocklength of each slot. As we know, the shorter the blocklength is the more significant the performance loss (in comparison to the Shannon capacity) is. Hence, the BL-capacity of a slot is not linearly associated with the length of the slot anymore. Therefore, this time allocation problem becomes more interesting and challenging.

Lastly, one of the most important challenges of wireless communications is the path-loss. As we know, the pathloss is increasing in the transmission distance not linearly

but logarithmically. Hence, by separating the total transmission distance into several segments the total pathloss is significantly reduced. This is actually the advantage of multi-hop relaying [80, 81, 82]. Therefore, in the infinite blocklength regime introducing multiple hops likely leads to a better performance. However, in the finite blocklength regime the story is different. Introducing more hops shortens the blocklength of each hop, which has negative impact to the system performance in the finite blocklength regime. In other words, increasing the number of hops introduces a tradeoff to the system performance, as it reduces the total pathloss and the frame length/blocklength of each hop simultaneously. As a results, in the finite blocklength regime determining optimal number of hops becomes an interesting optimization problem.

Anhang A

Proof of Proposition in Chapter 4

A.1 The Proof of Equation (4.6)

Beweis. A Rayleigh fading channel from relay j to the destination satisfies $\sqrt{2}|h_{j,D}|/|\bar{h}_{R,D}| \sim \text{Rayleigh}$

$\Rightarrow$ the channel can be modeled as $|h_{j,D}|/|\bar{h}_{R,D}| = \sqrt{(X_j^2 + Y_j^2)/2}$, where $X_j, Y_j \sim N(0, 1)$

$$\Rightarrow \frac{1}{|\bar{h}_{R,D}|^2} \sum_{j=1}^n |h_{j,D}|^2 = \frac{1}{2} \sum_{j=1}^n [X_j^2 + Y_j^2] \sim \chi^2(n) \Leftrightarrow \Gamma(n, 1)$$

$$\Rightarrow \frac{1}{|\bar{h}_{R,D}|^2} \sum_{j=1}^n |h_{j,D}|^2 \sim \Gamma(n, 1)$$

$$\Rightarrow \gamma_D \cdot \frac{\sigma^2}{p_{\text{tx}}|\bar{h}_{R,D}|^2} \sim \Gamma(n, 1)$$

$$\Rightarrow \gamma_D \sim \Gamma(n, \beta) \text{ with the mean } \beta = p_{\text{tx}}|\bar{h}_{R,D}|^2/\sigma^2. \quad \square$$

A.2 The Proof of Proposition 1

Beweis. $g_2(x, y)$ is monotonically decreasing in x ,

$$\Rightarrow \forall x_1 < x_2, g_2(x_1, y) \geq g_2(x_2, y).$$

$$\forall y, g_1(y) \geq 0,$$

$$\Rightarrow g_1(y)g_2(x_1, y) \geq g_1(y)g_2(x_2, y).$$

$f(x, z)$ is monotonically decreasing in z and strictly increasing in x ,

$$\Rightarrow f(x_1, g_1(y)g_2(x_1, y)) \leq f(x_1, g_1(y)g_2(x_2, y)) < f(x_2, g_1(y)g_2(x_2, y)),$$

$$\Rightarrow f(x, g_1(y)g_2(x, y)) \text{ is strictly increasing in } x.$$

□

Anhang B

Proofs of Propositions in Chapter 5

B.1 The poof of Proposition 3

We provide here the proof under the situation $z_2^2 \geq z_1^2 + z_3^2$, the other case can be proved similarly.

Beweis. $z_2^2 \geq z_1^2 + z_3^2$,

$\Rightarrow$ the coding rate is decided by the source based on the combined channel, giving:
 $r = R(z_1^2 + z_3^2, \varepsilon_{\text{MRC}}, m)$.

$$\Rightarrow \varepsilon_{\text{R}} = P[h_2^2, R(z_1^2 + z_3^2, \varepsilon_{\text{MRC}}, m), m] + (1 - P[h_2^2, R(z_1^2 + z_3^2, \varepsilon_{\text{MRC}}, m), m]) \cdot \varepsilon_{\text{MRC}}.$$

$\Rightarrow$ the derivative of ε_{R} with respect to ε_{MRC} is given by:

$$\frac{\partial \varepsilon_{\text{R}}}{\partial \varepsilon_{\text{MRC}}} = \frac{\partial \varepsilon_{\text{R}}}{\partial r} \frac{\partial r}{\partial \varepsilon_{\text{MRC}}} = \left[\frac{\partial \varepsilon_{\text{MRC}}}{\partial r} (1 - \varepsilon_2) + \frac{\partial \varepsilon_2}{\partial r} (1 - \varepsilon_{\text{MRC}}) \right] \frac{\partial r}{\partial \varepsilon_{\text{MRC}}}. \quad (\text{B.1})$$

$R(z_1^2 + z_3^2, \varepsilon_{\text{MRC}}, m)$ is the increasing in ε_{MRC} as well as $P(h^2, r, m)$ is increasing in r ,

$$\Rightarrow \frac{\partial \varepsilon_{\text{MRC}}}{\partial r} > 0, \frac{\partial \varepsilon_2}{\partial r} > 0 \text{ and } \frac{\partial r}{\partial \varepsilon_{\text{MRC}}} > 0.$$

$$0 \leq \varepsilon_2 \leq 1 \text{ and } 0 \leq \varepsilon_{\text{MRC}} \leq 1,$$

$$\Rightarrow \text{based on (5.12), } \frac{\partial \varepsilon_{\text{R}}}{\partial \varepsilon_{\text{MRC}}} > 0.$$

$$\Rightarrow \varepsilon_{\text{R}} \text{ is strictly increasing in } \varepsilon_{\text{MRC}}. \quad \square$$

B.2 The poof of Proposition 4

Here we provide the proof under the situation that the bottleneck link is the backhaul link, the other situation can be analyzed similarly.

Beweis. The first derivative of C_{BL} with respect to r is:

$$\frac{\partial C_{\text{BL}}}{\partial r} = \frac{1 - \varepsilon_{\text{R}}}{2} - \frac{1}{2} \frac{\partial \varepsilon_{\text{R}}}{\partial r}, \quad (\text{B.2})$$

where $\frac{\partial \varepsilon_{\text{R}}}{\partial r} = \frac{1 - \varepsilon_2}{2} \frac{\partial \varepsilon_{\text{MRC}}}{\partial r} + \frac{1 - \varepsilon_{\text{MRC}}}{2} \frac{\partial \varepsilon_2}{\partial r} > 0$.

$\Rightarrow$ the second derivative of C_{BL} with respect to r_3 :

$$\begin{aligned} \frac{\partial^2 C_{\text{BL}}}{\partial^2 r} &= \frac{r}{2} \frac{\partial \varepsilon_{\text{MRC}}}{\partial r} \frac{\partial \varepsilon_2}{\partial r} - \frac{r(1 - \varepsilon_{\text{MRC}})}{4} \frac{\partial^2 \varepsilon_2}{\partial^2 r} \\ &\quad - \frac{r(1 - \varepsilon_2)}{4} \frac{\partial^2 \varepsilon_{\text{MRC}}}{\partial^2 r} - \frac{1 - \varepsilon_{\text{MRC}}}{2} \frac{\partial \varepsilon_2}{\partial r} \\ &\quad - \frac{1 - \varepsilon_2}{2} \frac{\partial \varepsilon_{\text{MRC}}}{\partial r}. \end{aligned} \quad (\text{B.3})$$

$\Rightarrow$ Proposition 3 holds if the right side of (B.3) is negative.

We further obtain the first and second derivative of $\varepsilon_{\star}$ with respect to the coding rate r based on (5.11):

$$\frac{\partial \varepsilon_{\star}}{\partial r} = \frac{e^{-\frac{m(C(z_{\star}) - r)^2}{(1 - 2^{-2C(z_{\star})})(\log_2 e)^2}}}{2\sqrt{\pi}(1 - 2^{-2C(z_{\star})})(\log_2 e)^2}, \quad (\text{B.4})$$

$$\frac{\partial^2 \varepsilon_{\star}}{\partial^2 r} = \frac{m(C(z_{\star}) - r) e^{-\frac{m(C(z_{\star}) - r)^2}{(1 - 2^{-2C(z_{\star})})(\log_2 e)^2}}}{\sqrt{\pi}(1 - 2^{-2C(z_{\star})})^2 (\log_2 e)^4}, \quad (\text{B.5})$$

where $\varepsilon_{\star} \in \{\varepsilon_{\text{MRC}}, \varepsilon_2\}$ and $z_{\star}$ is the channel gain of the bottleneck link: $z_{\star} \in \{z_2^2, z_1^2 + z_3^2\}$.

Note that we consider a non-extreme error transmission where $\varepsilon_2 < 0.5$. Therefore, $\varepsilon_{\star} \leq \varepsilon_2 < 0.5$. Based on (5.11), the coding rate of each frame r determined by the source based on the channel quality of bottleneck link is lower than the Shannon capacity of the link: $C(z_{\star}) > r$.

$\Rightarrow$ both $\frac{\partial \varepsilon_\star}{\partial r}$ and $\frac{\partial^2 \varepsilon_\star}{\partial^2 r}$ are positive.

The bottleneck link is the backhaul link

$\Rightarrow \forall r, \varepsilon_2 = P^e(z_2, r, m) > \varepsilon_{\text{MRC}} = P^e(z_1 + z_3, r, m)$ and $\frac{\partial \varepsilon_2}{\partial r} > \frac{\partial \varepsilon_{\text{MRC}}}{\partial r}$.

$\Rightarrow$ based on (B.3), we have:

$$\begin{aligned}
 4 \frac{\partial^2 C_{\text{BL}}}{\partial^2 r} &= 2r \frac{\partial \varepsilon_{\text{MRC}}}{\partial r} \frac{\partial \varepsilon_2}{\partial r} - (1 - \varepsilon_{\text{MRC}}) \left[r \frac{\partial^2 \varepsilon_2}{\partial^2 r} + 2 \frac{\partial \varepsilon_2}{\partial r} \right] \\
 &\quad - (1 - \varepsilon_2) \left[r \frac{\partial^2 \varepsilon_{\text{MRC}}}{\partial^2 r} + 2 \frac{\partial \varepsilon_{\text{MRC}}}{\partial r} \right] \\
 &< 2r \frac{\partial \varepsilon_{\text{MRC}}}{\partial r} \frac{\partial \varepsilon_2}{\partial r} - r (1 - \varepsilon_{\text{MRC}}) \frac{\partial^2 \varepsilon_2}{\partial^2 r} \\
 &< 2r \left(\frac{\partial \varepsilon_2}{\partial r} \right)^2 - r (1 - \varepsilon_2) \frac{\partial^2 \varepsilon_2}{\partial^2 r}.
 \end{aligned} \tag{B.6}$$

$\Rightarrow \frac{\partial^2 C_{\text{BL}}}{\partial^2 r} < 0$ if $2 \left(\frac{\partial \varepsilon_2}{\partial r} \right)^2 / \frac{\partial^2 \varepsilon_2}{\partial^2 r} < 1 - \varepsilon_2$, where $1 - \varepsilon_2 \geq 1 - \varepsilon_{\text{R}} > 0.5$.

$2 \left(\frac{\partial \varepsilon_2}{\partial r} \right)^2 / \frac{\partial^2 \varepsilon_2}{\partial^2 r} = \frac{\exp\{-\frac{m(C(z_2)-r)^2}{(1-2^{-2C(z_2)})(\log_2 e)^2}\}}{2\sqrt{\pi}m(C(z_2)-r)} \ll 0.5$. Especially, for a relative big m or/and a big gap between $C(h_2^2)$ and r , $2\sqrt{\pi}m(C(z_2)-r) \gg 1$ while $\exp\{-\frac{m(C(z_2)-r)^2}{(1-2^{-2C(z_2)})(\log_2 e)^2}\} \ll 1$. For example, if $m = 100$ and $r = \frac{4}{5}C(z_2) = 0.4$ bit/ch.use, $2 \left(\frac{\partial \varepsilon_2}{\partial r} \right)^2 / \frac{\partial^2 \varepsilon_2}{\partial^2 r} \approx 8.2 \times 10^{-3} \ll 0.5$.

$\Rightarrow \frac{\partial^2 C_{\text{BL}}}{\partial^2 r} < 0$ for the non-extreme error transmissions. Hence, the BL-capacity C_{BL} is a concave function of the coding rate r . $\square$

B.3 The poof of Proposition 5

Beweis. r is strictly increasing in ε_{R} , $\varepsilon_{\text{R}} \in [0, 1]$.

$\Rightarrow \forall x < y, x, y \in [0, 1]$ and $\lambda \in [0, 1]$, we have $r|_{\varepsilon_2=x} < r|_{\varepsilon_2=\lambda x+(1-\lambda)y} < r|_{\varepsilon_2=y}$.

C_{BL} is concave in r ,

$\Rightarrow \min \{C_{\text{BL}}(r|_{\varepsilon_2=x}), C_{\text{BL}}(r|_{\varepsilon_2=y})\} \leq C_{\text{BL}}(r|_{\varepsilon_2=\lambda x+(1-\lambda)y})$.

$\Rightarrow C_{\text{BL}}$ is quasiconcave in ε_{R} , $\varepsilon_{\text{R}} \in [0, 1]$. And similarly, we can prove that C_{BL} is quasiconcave in $\varepsilon_\star$. $\square$

B.4 The poof of Proposition 6

Beweis. Based on Equation (5.29) we immediately have

$$\frac{\partial C_{\text{BL}}}{\partial r} = \frac{1}{2}(1 - \mathbb{E}_i[\varepsilon_{\text{R},i}]) - \frac{r}{2} \frac{\partial \mathbb{E}_i[\varepsilon_{\text{R},i}]}{\partial r}, \quad (\text{B.7})$$

$$\frac{\partial^2 C_{\text{BL}}}{\partial^2 r} = -\frac{\partial \mathbb{E}_i[\varepsilon_{\text{R},i}]}{\partial r} - \frac{r}{2} \frac{\partial^2 \mathbb{E}_i[\varepsilon_{\text{R},i}]}{\partial^2 r}. \quad (\text{B.8})$$

In the following, we prove Proposition 1 by proving $\frac{\partial^2 C_{\text{BL}}}{\partial^2 r} < 0$.

According to Equation (5.20), we have:

$$\begin{aligned} \frac{\partial \mathbb{E}_i[\varepsilon_{\text{R},i}]}{\partial r} &= \frac{\partial \mathbb{E}_{z_2}[\varepsilon_2]}{\partial r} \left(1 - \mathbb{E}_{z_1, z_3}[\varepsilon_{\text{MRC}}]\right) \\ &\quad + \frac{\partial \mathbb{E}_{z_1, z_3}[\varepsilon_{\text{MRC}}]}{\partial r} \left(1 - \mathbb{E}_{z_2}[\varepsilon_2]\right) \geq 0, \end{aligned} \quad (\text{B.9})$$

$$\begin{aligned} \frac{\partial^2 \mathbb{E}_i[\varepsilon_{\text{R},i}]}{\partial^2 r} &= \frac{\partial^2 \mathbb{E}_{z_2}[\varepsilon_2]}{\partial^2 r} \left(1 - \mathbb{E}_{z_1, z_3}[\varepsilon_{\text{MRC}}]\right) \\ &\quad - 2 \frac{\partial \mathbb{E}_{z_2}[\varepsilon_2]}{\partial r} \frac{\partial \mathbb{E}_{z_1, z_3}[\varepsilon_{\text{MRC}}]}{\partial r} \\ &\quad + \frac{\partial^2 \mathbb{E}_{z_1, z_3}[\varepsilon_{\text{MRC}}]}{\partial^2 r} \left(1 - \mathbb{E}_{z_2}[\varepsilon_2]\right). \end{aligned} \quad (\text{B.10})$$

Hence, $\frac{\partial^2 C_{\text{BL}}}{\partial^2 r} < 0$ if $\frac{\partial^2 \mathbb{E}_i[\varepsilon_{\text{R},i}]}{\partial^2 r} > 0$.

Recall that the coding rate is determined based on the weighted average channel gains $\eta \bar{z}_j$ ($j = 1, 2, 3$), where $0 < \eta \leq \bar{z}$. In other words, the coding rate is definitely lower than the Shannon capacity of a link (either the backhaul link or the combined link) if the fading gain of this link is higher than the η . Consider for example the backhaul link with fading z_2 . We have $r < C(z_2 \bar{z}_2)$, $z_2 \in (\eta, +\infty)$ and $r > C(z_2 \bar{z}_2)$, $z_2 \in (0, \eta)$. Denote the integral part in (B.14) as $\Delta(z_2)$ for short. Hence we have:

$$\frac{\partial^2 \mathbb{E}_{z_2}[\varepsilon_2]}{\partial^2 r} = \int_0^\eta \Delta(z_2) dz_2 + \int_\eta^{\bar{z}} \Delta(z_2) dz_2 + \int_{\bar{z}}^{+\infty} \Delta(z_2) dz_2, \quad (\text{B.11})$$

where $\int_0^\eta \Delta(z_2) dz_2 < 0$ $\int_{\hat{z}}^{\hat{z}} \Delta(z_2) dz_2 > 0$ and $\int_{\hat{z}}^{+\infty} \Delta(z_2) dz_2 > 0$.

As median, $\hat{z}$ satisfies:

$$\int_0^{\hat{z}} e^{-z_2} dz_2 = \int_{\hat{z}}^{+\infty} e^{-z_2} dz_2 = \frac{1}{2}. \quad (\text{B.12})$$

The error probability ε_2 satisfies $0 \leq \varepsilon_2(z_2) \leq 1$, and in particular $0 \leq \varepsilon_2(z_2) \leq \frac{1}{2}$ if the instantaneous channel gain z_2 higher than the weighted average channel gain (based on which the source determines the coding rate), i.e., $z_2 > \hat{z} \geq \eta$.

$\Rightarrow$ following relationship holds:

$$\begin{aligned} \mathbb{E}_{z_2}[\varepsilon_2] &= \int_0^{\hat{z}} e^{-z_2} \varepsilon_2(z_2) dz_2 + \int_{\hat{z}}^{\infty} e^{-z_2} \varepsilon_2(z_2) dz_2 \\ &< \int_0^{\hat{z}} e^{-z_2} dz_2 + \int_{\hat{z}}^{\infty} e^{-z_2} \frac{1}{2} dz_2 \\ &= \frac{1}{2} + \frac{1}{4} = \frac{3}{4}. \end{aligned} \quad (\text{B.13})$$

In addition, $\frac{\partial^2 \mathbb{E}[\varepsilon_2]}{\partial^2 r}$ can be given as:

$$\begin{aligned} \frac{\partial^2 \mathbb{E}[\varepsilon_2]}{\partial^2 r} &= \mathbb{E}_{z_2} \left[\frac{\partial^2 \varepsilon_2}{\partial^2 r} \right] \\ &= \int_0^{+\infty} \frac{m^{\frac{3}{2}} (C(z_2 \bar{z}_2) - r) e^{-\frac{m(C(z_2 \bar{z}_2) - r)^2}{2(1 - 2^{-2C(z_2 \bar{z}_2)})(\log_2 e)^2}}}{\sqrt{2\pi} (1 - 2^{-2C(z_2 \bar{z}_2)})^{\frac{3}{2}} (\log_2 e)^3} e^{-z_2} dz_2. \end{aligned} \quad (\text{B.14})$$

Moreover, the following relationship holds based on $\hat{z}$:

$$\begin{aligned}
 & \int_{\widehat{z}}^{+\infty} \frac{(C(z_2 \bar{z}_2) - r) e^{-\frac{(C(z_2 \bar{z}_2) - r)^2}{(1 - 2^{-4C(z_2 \bar{z}_2)})}}}{(1 - 2^{-4C(z_2 \bar{z}_2)})^{\frac{3}{2}}} e^{-z_2} dz_2 \\
 & > \int_0^{\widehat{z}} \frac{|C(z_2 \bar{z}_2) - r| e^{-\frac{(C(z_2 \bar{z}_2) - r)^2}{(1 - 2^{-4C(z_2 \bar{z}_2)})}}}{(1 - 2^{-4C(z_2 \bar{z}_2)})^{\frac{3}{2}}} e^{-z_2} dz_2 \\
 & \geq \int_0^{\eta} \frac{|C(z_2 \bar{z}_2) - r| e^{-\frac{(C(z_2 \bar{z}_2) - r)^2}{(1 - 2^{-4C(z_2 \bar{z}_2)})}}}{(1 - 2^{-4C(z_2 \bar{z}_2)})^{\frac{3}{2}}} e^{-z_2} dz_2 > 0
 \end{aligned} \tag{B.15}$$

$$\Rightarrow \int_{\widehat{z}}^{+\infty} \Delta(z_2) dz_2 - |\int_0^{\eta} \Delta(z_2) dz_2| > 0.$$

$$\Rightarrow \text{we have } \frac{\partial^2 \mathbb{E}[\varepsilon_2]}{\partial^2 r} > 0 \text{ for the backhaul link.}$$

Regarding the combined link, we have:

$$\begin{aligned}
 & \int_{\widehat{z}}^{+\infty} \int_{\widehat{z}}^{+\infty} \frac{\sqrt{2\pi} (C(z_1 \bar{z}_1 + z_3 \bar{z}_3) - r) e^{-\frac{m(C(z_1 \bar{z}_1 + z_3 \bar{z}_3) - r)^2}{(1 - 2^{-4C(z_1 \bar{z}_1 + z_3 \bar{z}_3)}) (\log_2 e)^2} - z_1 - z_3}}{\left(\frac{1 - 2^{-4C(z_1 \bar{z}_1 + z_3 \bar{z}_3)}}{m}\right)^{\frac{3}{2}} (\log_2 e)^3} dz_1 dz_3 \\
 & \geq \int_0^{\eta} \int_0^{\eta} \frac{\sqrt{2\pi} |C(z_1 \bar{z}_1 + z_3 \bar{z}_3) - r| e^{-\frac{m(C(z_1 \bar{z}_1 + z_3 \bar{z}_3) - r)^2}{2(1 - 2^{-4C(z_1 \bar{z}_1 + z_3 \bar{z}_3)}) (\log_2 e)^2} - z_1 - z_3}}{4 \left(\frac{1 - 2^{-4C(z_1 \bar{z}_1 + z_3 \bar{z}_3)}}{m}\right)^{\frac{3}{2}} (\log_2 e)^3} dz_d dz_r \\
 & > 0
 \end{aligned} \tag{B.16}$$

$$\text{Similarly, we have } \frac{\partial^2 \mathbb{E}[\varepsilon_{\text{MRC}}]}{\partial^2 r} > 0.$$

Then, to prove $\frac{\partial^2 \mathbb{E}[\varepsilon_{\text{R},i}]}{\partial^2 r} > 0$, the following two cases are considered, which differ in the relationship between the average channel gains of the backhaul link and the combined link are considered. The first case is $\bar{z}_2 \leq \bar{z}_1 + \bar{z}_3$, where average channel gain of the backhaul link is lower than the combined link. As the fading of different links are i.i.d., under this case the instantaneous channel gain of the backhaul link is more likely to be lower than the combined link.

$$\bar{z}_2 \leq \bar{z}_1 + \bar{z}_3$$

$\Rightarrow 3/4 > \mathbb{E}_{z_2} [\varepsilon_2] \geq \mathbb{E}_{z_1, z_3} [\varepsilon_{\text{MRC}}]$ it is easy to prove $\frac{\partial^i \mathbb{E}_{z_2} [\varepsilon_2]}{\partial^i r} \geq \frac{\partial^i \mathbb{E}_{z_1, z_3} [\varepsilon_{\text{MRC}}]}{\partial^i r}, i = 1, 2, \dots + \infty$ based on (5.11), (5.21) and (5.22).

$\Rightarrow$ based on (B.10), $\frac{\partial^2 \mathbb{E}_i [\varepsilon_{\text{R}, i}]}{\partial^2 r}$ is bound by:

$$\frac{\partial^2 \mathbb{E}_i [\varepsilon_{\text{R}, i}]}{\partial^2 r} > \frac{1}{4} \frac{\partial^2 \mathbb{E}_{z_2} [\varepsilon_2]}{\partial^2 r} - 2 \left(\frac{\partial \mathbb{E}_{z_2} [\varepsilon_2]}{\partial r} \right)^2. \quad (\text{B.17})$$

Considering that m is the blocklength and $m \gg 1$, we have:

$$\begin{aligned} \frac{\partial^2 \mathbb{E}_i [\varepsilon_{\text{R}, i}]}{\partial^2 r} &> \frac{m^{\frac{3}{2}}}{8\sqrt{2\pi}(\log_2 e)^3} \int_0^{+\infty} \frac{(C(z_2 \bar{z}_2) - r) e^{-\frac{m(C(z_2 \bar{z}_2) - r)^2}{2(1-2^{-2C(z_2 \bar{z}_2)})(\log_2 e)^2}}}{(1 - 2^{-2C(z_2 \bar{z}_2)})^{\frac{3}{2}}} e^{-z_2} dz_2 \\ &= \frac{m^{\frac{3}{2}}}{8\sqrt{2\pi}(\log_2 e)^3} \int_0^{+\infty} \Phi(z_2) e^{-z_2} dz_2. \end{aligned} \quad (\text{B.18})$$

Based on the same idea during the above proof of $\frac{\partial^2 \mathbb{E}_{z_2} [\varepsilon_2]}{\partial^2 r} > 0$, it is easy to have $\int_z^{+\infty} \Phi(z_2) dz_2 - |\int_0^\eta \Phi(z_2) dz_2| > 0$ and $\int_\eta^{\widehat{z}} \Phi(z_2) dz_2 > 0$

$$\Rightarrow \int_0^\infty \Phi(z_2) dz_2 > 0$$

$$\Rightarrow \frac{\partial^2 \mathbb{E}_i [\varepsilon_{\text{R}, i}]}{\partial^2 r} > 0 \text{ under the case } \bar{z}_2 \leq \bar{z}_1 + \bar{z}_3.$$

Under the other case $\bar{z}_2 > \bar{z}_1 + \bar{z}_3$, it can be proved similarly (based on $\frac{\partial^2 \mathbb{E}_{z_1, z_3} [\varepsilon_{\text{MRC}}]}{\partial^2 r} >$

$$0) \text{ that } \frac{\partial^2 \mathbb{E}_i [\varepsilon_{\text{R}, i}]}{\partial^2 r} > 0.$$

$$\Rightarrow \frac{\partial^2 C_{\text{BL}}}{\partial^2 r} < 0$$

$\Rightarrow C_{\text{BL}}$ is a strictly concave in the coding rate.

□

B.5 The poof of Proposition 7

Beweis. r is strictly increasing in η , $0 < \eta \leq \widehat{z}$.

$\Rightarrow \forall x < y, x, y \in (0, \widehat{z})$ and $\lambda \in [0, 1]$, we have $r|_{\eta=x} < r|_{\eta=\lambda x+(1-\lambda)y} < r|_{\eta=y}$.

Based on Proposition 1, C_{BL} is concave in r ,

$\Rightarrow \min \left\{ C_{\text{BL}} \left(r|_{\eta=x} \right), C_{\text{BL}} \left(r|_{\eta=y} \right) \right\} \leq C_{\text{BL}} \left(r|_{\eta=\lambda x+(1-\lambda)y} \right)$.

$\Rightarrow C_{\text{BL}}$ is quasi-concave in η , $0 < \eta \leq \widehat{z}$.

□

B.6 The poof of Proposition 8

Beweis. Comparing (5.34) and (5.29), we have:

$$R_{\text{MS}} = \frac{C_{\text{BL}}}{2} + R^*, \quad (\text{B.19})$$

where $R^* = \frac{r}{4} \sqrt{(1 - \mathbb{E}_{z_1, z_2, z_3} [\varepsilon_{\text{R}}])^2 + \frac{4m \ln(P_{\text{d}})}{d} \cdot \mathbb{E}_{z_1, z_2, z_3} [\varepsilon_{\text{R}}] \cdot (1 - \mathbb{E}_{z_1, z_2, z_3} [\varepsilon_{\text{R}}])}$.

As we have proved that C_{BL} is concave in r in Proposition 1, Proposition 3 holds if R^* is concave, too.

Note that $\frac{4m \ln(P_{\text{d}})}{d}$ is not influenced by r . It is actually a negative constant as $\ln(P_{\text{d}}) < 0$. We denote this constant by φ ($\varphi < 0$) for short. To facilitate the proof we also denote $\mathbb{E}_{z_1, z_2, z_3} [\varepsilon_{\text{R}}]$ by $\bar{\varepsilon}_{\text{R}}$ for short.

Then, we have:

$$R^* = \frac{r}{4} \sqrt{(1 - \bar{\varepsilon}_{\text{R}})^2 + \varphi \bar{\varepsilon}_{\text{R}} (1 - \bar{\varepsilon}_{\text{R}})} = \frac{r}{4} \sqrt{1 + (\varphi - 2) \bar{\varepsilon}_{\text{R}} + (1 - \varphi) \bar{\varepsilon}_{\text{R}}^2}. \quad (\text{B.20})$$

As R^* is a square root function, it should be satisfied that:

$$(1 - \bar{\varepsilon}_{\text{R}})^2 + \varphi \bar{\varepsilon}_{\text{R}} \cdot (1 - \bar{\varepsilon}_{\text{R}}) \geq 0. \quad (\text{B.21})$$

$\bar{\varepsilon}_{\text{R}}$ is the expectation of the error probability over fading

$$\Rightarrow 0 < \bar{\varepsilon}_{\text{R}} < 1$$

$\Rightarrow 1 - \bar{\varepsilon}_R > 0$. Combining this with (B.21)

$\Rightarrow 1 - \bar{\varepsilon}_R + \varphi \bar{\varepsilon}_R \geq 0$,

$\Rightarrow 1 \geq (1 - \varphi) \bar{\varepsilon}_R$.

$$\frac{\partial R^*}{\partial r} = \frac{1}{4} \sqrt{1 + (\varphi - 2) \bar{\varepsilon}_R + (1 - \varphi) \bar{\varepsilon}_R^2} + \frac{r}{8} \frac{(\varphi - 2) + 2(1 - \varphi) \bar{\varepsilon}_R}{\sqrt{1 + (\varphi - 2) \bar{\varepsilon}_R + (1 - \varphi) \bar{\varepsilon}_R^2}} \frac{\partial \bar{\varepsilon}_R}{\partial r}$$

$$\frac{\partial^2 R^*}{\partial^2 r} = \frac{1}{8} \frac{(\varphi - 2) + 2(1 - \varphi) \bar{\varepsilon}_R + r(1 - \varphi)}{\left((1 - \bar{\varepsilon}_R)^2 + \varphi \bar{\varepsilon}_R(1 - \bar{\varepsilon}_R)\right)^{\frac{1}{2}}} + \frac{r}{16} \frac{-((\varphi - 2) + 2(1 - \varphi) \bar{\varepsilon}_R)^2}{\left((1 - \bar{\varepsilon}_R)^2 + \varphi \bar{\varepsilon}_R(1 - \bar{\varepsilon}_R)\right)^{\frac{3}{2}}} \left(\frac{\partial \bar{\varepsilon}_R}{\partial r}\right)^2 + \frac{r}{8} \frac{(\varphi - 2) + 2(1 - \varphi) \bar{\varepsilon}_R}{\left((1 - \bar{\varepsilon}_R)^2 + \varphi \bar{\varepsilon}_R(1 - \bar{\varepsilon}_R)\right)^{\frac{1}{2}}} \frac{\partial^2 \bar{\varepsilon}_R}{\partial^2 r}$$

As we have $\varphi < 0$ and $0 < \bar{\varepsilon}_R < 1$

$\Rightarrow 0 < (1 - \bar{\varepsilon}_R)^2 + \varphi \bar{\varepsilon}_R(1 - \bar{\varepsilon}_R) < (1 - \bar{\varepsilon}_R)^2 < 1$

$\Rightarrow 0 < \left((1 - \bar{\varepsilon}_R)^2 + \varphi \bar{\varepsilon}_R(1 - \bar{\varepsilon}_R)\right)^{\frac{3}{2}} < \left((1 - \bar{\varepsilon}_R)^2 + \varphi \bar{\varepsilon}_R(1 - \bar{\varepsilon}_R)\right)^{\frac{1}{2}} < 1$

$\Rightarrow$ we have:

$$\begin{aligned} \frac{\partial^2 R^*}{\partial^2 r} &< \frac{1}{8} \frac{(\varphi - 2) + 2(1 - \varphi) \bar{\varepsilon}_R + r(1 - \varphi)}{\left((1 - \bar{\varepsilon}_R)^2 + \varphi \bar{\varepsilon}_R(1 - \bar{\varepsilon}_R)\right)^{\frac{1}{2}}} + \frac{r}{16} \frac{-((\varphi - 2) + 2(1 - \varphi) \bar{\varepsilon}_R)^2}{\left((1 - \bar{\varepsilon}_R)^2 + \varphi \bar{\varepsilon}_R(1 - \bar{\varepsilon}_R)\right)^{\frac{1}{2}}} \left(\frac{\partial \bar{\varepsilon}_R}{\partial r}\right)^2 \\ &\quad + \frac{r}{8} \frac{(\varphi - 2) + 2(1 - \varphi) \bar{\varepsilon}_R}{\left((1 - \bar{\varepsilon}_R)^2 + \varphi \bar{\varepsilon}_R(1 - \bar{\varepsilon}_R)\right)^{\frac{1}{2}}} \frac{\partial^2 \bar{\varepsilon}_R}{\partial^2 r}. \end{aligned} \tag{B.22}$$

As $1 \geq (1 - \varphi) \bar{\varepsilon}_R$, we have:

$$\frac{\partial^2 R^*}{\partial^2 r} < \frac{2\varphi + 2r \left(1 - \varphi - \frac{\varphi^2}{2} \left(\frac{\partial \bar{\varepsilon}_R}{\partial r}\right)^2 + \varphi \frac{\partial^2 \bar{\varepsilon}_R}{\partial^2 r}\right)}{16 \left((1 - \bar{\varepsilon}_R)^2 + \varphi \bar{\varepsilon}_R(1 - \bar{\varepsilon}_R)\right)^{\frac{1}{2}}}. \tag{B.23}$$

As shown in Appendix A, $\frac{\partial^2 E[\varepsilon_{R,i}]}{\partial^2 r} > 0$. In addition, $\left(\frac{\partial \bar{\varepsilon}_R}{\partial r}\right)^2 \sim O(m)$ and $\frac{\partial^2 \bar{\varepsilon}_R}{\partial^2 r} \sim O(m^{3/2})$. Moreover, φ is a constant with reasonable value¹ $\varphi \in (-30, -0.05)$. Obviously, $\frac{\partial^2 R^*}{\partial^2 r} < 0$ holds if $m \gg 1$.

Hence, R_{MS} is concave in r .

□

¹ Recall that φ is totally subject to the QoS constraints {delay (in symbols), delay violation probability}. Considering extremely loose constraints $\{100m, 10^{-0.5}\}$ and extremely strict constraints $\{2m, 10^{-7}\}$ (recall that the length of a transmission period of relaying is $2m$), we have $\varphi \in (-30, -0.05)$.

Glossary

3GPP 3rd generation partnership project

5G 5th generation of wireless communication

ACK acknowledgment

AF amplify-and-forward

ARQ automatic repeat request

AWGN additive white Gaussian noise

BL-capacity blocklength-limited capacity

C-ARQ cooperative-ARQ

CDF cumulative distribution function

CSI channel state information

DF decode-and-forward

EEE effective energy efficiency — a higher-layer energy efficiency metric introduced in this work

MRC maximal ratio combining

MSDR maximum sustainable data rate

M2M machine-to-machine

NACK negative-acknowledgment

PEE physical-layer energy efficiency

PMF probability mass function

QoS quality of service

SNR power to noise ratio

Glossary

VoIP in voice over IP

i.i.d. independent and identically distributed

List of Symbols and Notations

General Notations	
$\mathcal{C}$	set
$\mathbb{R}$	real number
$\mathbb{N}$	integer
$\mathbf{x}$	transmitted signal
$\mathbf{y}$	received signal
h	channel
γ	SNR
$\mathbf{I}_m$	an $m \times m$ identity matrix
Notations Regarding Effective Capacity Model	
d	delay
P_d	delay violation probability
Q	queue length
R	data arrive rate at the source buffer
R_{MS}	MSDR
s_i	service process increment
S_i	cumulative service process
K	the probability that the queue is non-empty
θ^*	QoS exponent

Notations Regarding Relaying Performance in the Infinite Blocklength Regime	
J	number of relays deployed in the system
j	index of relay
Ω	set of active relays
n	index of active relay
i	index of time/frame
$z_{S,D,i} = h_{S,D,i} ^2$	channel gain of link from the source to the destination at frame i
$z_{S,j,i} = h_{S,j,i} ^2$	channel gain of link from the source to relay j at frame i
$z_{j,D,i} = h_{j,D,i} ^2$	channel gain of link from relay j to the destination at frame i
σ^2	noise power
p_{tx}	transmit power at the source or a relay
p_{fb}	feedback power at the destination
p_c	the power of basic energy consumption at each node
ρ	packet size
N	number of channel use for transmitting the packet
τ	length of random transmission delay (in frame)
β	average of the signal strength
M	the number of frames in each transmission/retransmission
T_f	frame length (in second)
k_1	number of broadcasting frames during a retransmission process
k_2	number of forwarding frames during a retransmission process
P_S^{ct}	percentage of broadcasting frames
P_R^{ct}	percentage of relaying frames (fixed protocols)
P_{i-R}^{ct}	percentage of initial active relays retransmitting frames (under dynamic protocols)
P_{a-R}^{ct}	percentage of all relays retransmitting frames (dynamic protocols)
P_D^{ct}	percentage of ACK/NACK frames
P_0	outage probability of the direct link
P_1	outage probability of a backhaul link
P_X	the probability of a relay keeping silent during the relaying frame
ω	weight factor which is associated with P_X

P^{out}	overall outage probability of relaying
$\bar{\Psi}_{\text{T}}$	average transmit energy consumption
Φ_{p}	PEE
Φ_{h}	EEE

Notations Regarding Relaying Performance in the Finite Blocklength Regime	
V	channel dispersion
m	blocklength at each frame of relaying
r	data coding rate for transmission at each hop of relaying
σ^2	noise power
p_{tx}	transmit power at the source or a relay
z_1	channel gain of the direct link under a static channel
z_2	channel gain of the backhaul link under a static channel
z_3	channel gain of the relaying link under a static channel
ε_2	error probability of the backhaul link under a static channel
ε_{MRC}	error probability of the combined link under static channels
$\varepsilon_{\star}$	error probability of the bottleneck link under static channels
ε_{R}	overall error probability of relaying under static channels
C_{BL}	BL-capacity of relaying under static channels
$z_{1,i}$	channel gain of the direct link at transmission period i under a quasi-static channel
$z_{2,i}$	channel gain of the backhaul link at transmission period i under a quasi-static channel
$z_{3,i}$	channel gain of the relaying link at transmission period i under a quasi-static channel
$\bar{z}_1$	average channel gain of the direct link (over time) under a quasi-static channel
$\bar{z}_2$	average channel gain of the backhaul link (over time) under a quasi-static channel
$\bar{z}_3$	average channel gain of the relaying link (over time) under a quasi-static channel
$\varepsilon_{2,i}$	error probability of the backhaul link

List of Symbols and Notations

	at transmission period i under a quasi-static channel
$\varepsilon_{\text{MRC},i}$	error probability of the combined link
	at transmission period i under quasi-static channels
$\varepsilon_{\text{R},i}$	overall error probability of relaying
	at transmission period i under quasi-static channels
$C_{\text{BL},i}$	BL-capacity of relaying
	at transmission period i under quasi-static channels
ω	weight factor for the weighting the average SNR
	under quasi-static channels
ε_{D}	error probability of direct transmission
	(under a quasi-static channel)
C_{D}	Shannon capacity of direct transmission
	(under a quasi-static channel)
C_{R}	equivalent Shannon capacity of relaying
	(under quasi-static channels)

Literaturverzeichnis

- [1] RP-110911, *3GPP work programme: Relays for LTE - core part*. June 2011. 1
- [2] Huawei's METIS project team, *METIS: Striding towards 5G*. Dec. 2014. 1
- [3] Ericsson, *5G radio access – Ericsson White paper*. Feb. 2015. 1
- [4] J. Laneman, D. Tse, and G. W. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *Information Theory, IEEE Transactions on*, vol. 50, no. 12, pp. 3062–3080, Dec. 2004. 1, 24, 26
- [5] S. Karmakar and M. Varanasi, "The diversity-multiplexing tradeoff of the dynamic decode-and-forward protocol on a MIMO half-duplex relay channel," *Information Theory, IEEE Transactions on*, vol. 57, pp. 6569–6590, Oct. 2011. 1, 26
- [6] Y. Huang, F. Al-Qahtani, C. Zhong, Q. Wu, J. Wang, and H. Alnuweiri, "Performance analysis of multiuser multiple antenna relaying networks with co-channel interference and feedback delay," *Communications, IEEE Transactions on*, vol. 62, pp. 59–73, Jan. 2014. 1, 26
- [7] Y. Wang and K. Liu, "Statistical delay QoS protection for primary users in cooperative cognitive radio networks," *Communications Letters, IEEE*, vol. 19, pp. 835–838, May 2015. 1
- [8] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite block-length regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, May 2010. 2, 3, 13, 14, 15, 27, 89, 122
- [9] G. Neonakis Aggelou and R. Tafazolli, "On the relaying capability of next-generation GSM cellular networks," *Personal Communications, IEEE*, vol. 8, pp. 40–47, Feb. 2001. 5, 23
- [10] R. Pabst, B. H. Walke, et al., "Relay-based deployment concepts for wireless and mobile broadband radio deployment concepts for wireless and mobile broadband

- radio,” *Communications Magazine, IEEE*, vol. 42, pp. 80–89, Sep. 2004. 5, 23
- [11] B. Rankov and A. Wittneben, “Spectral efficient protocols for half-duplex fading relay channels,” *Selected Areas in Communications, IEEE Journal on*, vol. 25, pp. 379–389, Feb. 2007. 5
- [12] T. Riihonen, S. Werner, and R. Wichman, “Mitigation of loopback self-interference in full-duplex MIMO relays,” *Signal Processing, IEEE Transactions on*, vol. 59, pp. 5983–5993, Dec. 2011. 5
- [13] J. Laneman, D. Tse, and G. W. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” *Information Theory, IEEE Transactions on*, vol. 50, pp. 3062–3080, Dec. 2004. 5
- [14] R. Pabst, B. Walke, D. Schultz, P. Herhold, H. Yanikomeroglu, S. Mukherjee, H. Viswanathan, M. Lott, W. Zirwas, M. Dohler, H. Aghvami, D. Falconer, and G. Fettweis, “Relay-based deployment concepts for wireless and mobile broadband radio,” *Communications Magazine, IEEE*, vol. 42, pp. 80–89, Sep. 2004. 5
- [15] S. rae Cho, W. Choi, and K. Huang, “QoS provisioning relay selection in random relay networks,” *Vehicular Technology, IEEE Transactions on*, vol. 60, pp. 2680–2689, July 2011. 6
- [16] Y. Zou, B. Champagne, W.-P. Zhu, and L. Hanzo, “Relay-selection improves the security-reliability trade-off in cognitive radio systems,” *Communications, IEEE Transactions on*, vol. 63, pp. 215–228, Jan. 2015. 6
- [17] H. Eghbali, S. Muhaidat, S. Hejazi, and Y. Ding, “Relay selection strategies for single-carrier frequency-domain equalization multi-relay cooperative networks,” *Wireless Communications, IEEE Transactions on*, vol. 12, pp. 2034–2045, May 2013. 6
- [18] J. Lee and N. Al-Dhahir, “Exploiting sparsity for multiple relay selection with relay gain control in large af relay networks,” *Wireless Communications Letters, IEEE*, vol. 2, pp. 347–350, June 2013. 6
- [19] Y. Jing and H. Jafarkhani, “Single and multiple relay selection schemes and their achievable diversity orders,” *Wireless Communications, IEEE Transactions on*, vol. 8, pp. 1414–1423, March 2009. 6

- [20] C.-H. Yu and O. Tirkkonen, "Opportunistic multiple relay selection with diverse mean channel gains," *Wireless Communications, IEEE Transactions on*, vol. 11, pp. 885–891, March 2012. 6
- [21] C.-S. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *Automatic Control, IEEE Transactions on*, vol. 39, pp. 913–931, May 1994. 6, 7
- [22] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *Wireless Communications, IEEE Transactions on*, vol. 2, no. 4, pp. 630–643, July 2003. XIII, 6, 7, 10, 11
- [23] B. Soret, M. Aguayo-Torres, and J. Entrambasaguas, "Capacity with explicit delay guarantees for generic sources over correlated Rayleigh channel," *Wireless Communications, IEEE Transactions on*, vol. 9, pp. 1901–1911, June 2010. 11
- [24] C. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, July 1948. 13
- [25] J. Wolfowitz, *Coding Theorems of Information Theory*, Springer Berlin Heidelberg, 1962. 14
- [26] M. Gursoy, "Throughput analysis of buffer-constrained wireless systems in the finite blocklength regime," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, no. 1, 2013. 15, 95
- [27] J. Cui and A. Sheikh, "Outage probability of cellular radio systems using maximal ratio combining in the presence of multiple interferers," *Communications, IEEE Transactions on*, vol. 47, pp. 1121–1124, Aug. 1999. 16
- [28] Jean Paul Linnartz' Reference Website, "Wireless Communication", <http://www.wirelesscommunication.nl/reference/chaptr05/diversit/mrc.htm>
16
- [29] Stephen Boyd and Lieven Vandenberghe, *Convex Optimization*. Cambridge University Press 1962. 18
- [30] A. Nosratinia, T. Hunter, and A. Hedayat, "Cooperative communication in wireless networks," *Communications Magazine, IEEE*, vol. 42, pp. 74–80, Oct. 2004. 23

- [31] J. Laneman and G. W. Wornell, “Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks,” *Information Theory, IEEE Transactions on*, vol. 49, pp. 2415–2425, Oct. 2003. 24
- [32] C. Sun, J. Cheng, and M. Taromaru, “Achieving the lowest outage probability for multi-phase cooperative wireless networks,” in *Information Theory for Wireless Networks, 2007 IEEE Information Theory Workshop on*, pp. 1–5, July 2007. 24
- [33] P. Elia, K. Vinodh, M. Anand, and P. Vijay Kumar, “D-MG tradeoff and optimal codes for a class of AF and DF cooperative communication protocols,” pp. 681–685, June 2007. 24
- [34] G. Kramer, M. Gastpar, and P. Gupta, “Cooperative strategies and capacity theorems for relay networks,” *Information Theory, IEEE Transactions on*, vol. 51, no. 9, pp. 3037–3063, Sep. 2005. 24
- [35] R. Tutgun and E. Aktas, “Cooperative network coded ARQ strategies for two-way relay channels,” *Vehicular Technology, IEEE Transactions on*, vol. 64, pp. 3205–3217, July 2015. 24
- [36] J. Alonso-Zarate, E. Stavrou, A. Stamou, P. Angelidis, L. Alonso, and C. Verikoukis, “Energy-efficiency evaluation of a medium access control protocol for cooperative ARQ,” in *Communications (ICC), 2011 IEEE International Conference on*, pp. 1–5, June 2011. 24
- [37] I. Cerutti, A. Fumagalli, and P. Gupta, “Delay models of single-source single-relay cooperative arq protocols in slotted radio networks with Poisson frame arrivals,” *Networking, IEEE/ACM Transactions on*, vol. 16, no. 2, pp. 371–382, April 2008. 24
- [38] L. Dai and K. Letaief, “Throughput maximization of ad-hoc wireless networks using adaptive cooperative diversity and truncated ARQ,” *Communications, IEEE Transactions on*, vol. 56, pp. 1907–1918, Nov. 2008. 24
- [39] M. Dianati, X. Ling, K. Naik, and X. Shen, “A node-cooperative ARQ scheme for wireless ad hoc networks,” *Vehicular Technology, IEEE Transactions on*, vol. 55, no. 3, pp. 1032–1044, 2006. 24
- [40] G. F. Ernesto Zimmermann, Patrick Herhold, “On the performance of cooperative relaying protocols in wireless networks,” *European Transactions on Telecommu-*

- nications*, vol. 16, no. 2, pp. 5–16, 2005. 24
- [41] S. Lee, W. Su, S. Batalama, and J. Matyjas, “Cooperative decode-and-forward ARQ relaying: Performance analysis and power optimization,” *Wireless Communications, IEEE Transactions on*, vol. 9, no. 8, pp. 2632–2642, Aug. 2010. 24
- [42] J. Ramis and G. Femenias, “Cross-layer QoS-constrained optimization of adaptive multi-rate wireless systems using infrastructure-based cooperative ARQ,” *Wireless Communications, IEEE Transactions on*, vol. 12, no. 5, pp. 2424–2435, May 2013. 24
- [43] I. Stanojev, O. Simeone, Y. Bar-Ness, and D. H. Kim, “Energy efficiency of non-collaborative and collaborative hybrid-ARQ protocols,” *Wireless Communications, IEEE Transactions on*, vol. 8, pp. 326–335, Jan. 2009. 24
- [44] M. Maaz, P. Mary, and M. Helard, “Energy efficiency analysis in relay assisted hybrid-ARQ communications,” in *Personal Indoor and Mobile Radio Communications (PIMRC), 2012 IEEE 23rd International Symposium on*, pp. 2263–2268, Sep. 2012. 24
- [45] H. Boujemaa, “Delay analysis of cooperative truncated harq with opportunistic relaying,” *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 9, pp. 4795–4804, 2009. 24
- [46] M. Maaz, P. Mary, and M. Helard, “Delay outage probability in block fading channel and relay-assisted hybrid-ARQ network,” *Wireless Communications Letters, IEEE*, vol. 3, pp. 129–132, April 2014. 24
- [47] M. Mardani, J. Harsini, F. Lahouti, and B. Eliasi, “Link-adaptive and QoS-provisioning cooperative ARQ applications to relay-assisted land mobile satellite communications,” *Vehicular Technology, IEEE Transactions on*, vol. 60, pp. 3192–3206, Sep. 2011. 25, 31
- [48] L. Musavian, S. Aissa, and S. Lambotharan, “Effective capacity for interference and delay constrained cognitive radio relay channels,” *Wireless Communications, IEEE Transactions on*, vol. 9, pp. 1698–1707, May 2010. 25, 31
- [49] S. rae Cho, W. Choi, and K. Huang, “QoS provisioning relay selection in random relay networks,” *Vehicular Technology, IEEE Transactions on*, vol. 60, pp. 2680–2689, July 2011. 25, 31

- [50] W. Cheng, X. Zhang, and H. Zhang, “QoS-aware power allocations for maximizing effective capacity over virtual-MIMO wireless networks,” *Selected Areas in Communications, IEEE Journal on*, vol. 31, pp. 2043–2057, Oct. 2013. 25, 31
- [51] D. Qiao, M. Gursoy, and S. Velipasalar, “Effective capacity of two-hop wireless communication systems,” *Information Theory, IEEE Transactions on*, vol. 59, pp. 873–885, Feb. 2013. 25, 31
- [52] C. Lin, Y. Liu, and M. Tao, “Cross-layer optimization of two-way relaying for statistical QoS guarantees,” *Selected Areas in Communications, IEEE Journal on*, vol. 31, pp. 1583–1596, Aug. 2013. 25, 31
- [53] Y. Wang, P. Ren, and F. Gao, “Power allocation for statistical QoS provisioning in opportunistic multi-relay df cognitive networks,” *Signal Processing Letters, IEEE*, vol. 20, pp. 43–46, Jan. 2013. 25, 31
- [54] S. Ren and K. Letaief, “Maximizing the effective capacity for wireless cooperative relay networks with QoS guarantees,” *Communications, IEEE Transactions on*, vol. 57, pp. 2148–2159, July 2009. 25, 31
- [55] Y. Hu and L. Qiu, “A novel multiple relay selection strategy for LTE-advanced relay systems,” in *Vehicular Technology Conference, IEEE*, (Budapest,Hungary), pp. 1–5, May 2011. 26
- [56] Y. Hu, J. Gross, and A. Schmeink, “QoS-constrained energy efficiency of cooperative ARQ in multiple df relay systems,” *Vehicular Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2015. 26
- [57] Y. Hu and J. Gross, “On the outage probability and effective capacity of multiple decode-and-forward relay system,” in *2012 IFIP Wireless Days conference (WD)*, Dublin, Ireland, pp. 1–8, Nov. 2012. 26, 63
- [58] M. Bhatnagar, “On the capacity of decode-and-forward relaying over Rician fading channels,” *Communications Letters, IEEE*, vol. 17, pp. 1100–1103, June 2013. 26
- [59] Y.-M. Wang and S. Lin, “A modified selective-repeat type-ii hybrid ARQ system and its performance analysis,” *Communications, IEEE Transactions on*, vol. 31, pp. 593–608, May 1983. 33, 44
- [60] F. Richter, A. Fehske, and G. Fettweis, “Energy efficiency aspects of base station deployment strategies for cellular networks,” in *Vehicular Technology Conference*

- Fall, IEEE*, (Anchorage, USA), pp. 1–5, Sep. 2009. 34
- [61] H. Kim, C.-B. Chae, G. de Veciana, and R. Heath, “A cross-layer approach to energy efficiency for adaptive MIMO systems exploiting spare capacity,” vol. 8, pp. 4264–4275, Aug. 2009. 34
 - [62] A. Papoulis and S. Pillai, *Probability, random variables, and stochastic processes*. 2002. 41
 - [63] A. F. Molisch, *Wireless Communications*, Wiley, 2011. 42, 102, 122
 - [64] A. Kumar, D. Manjunath, and J. Kuri, *Communication networking: An analytical approach*. The Morgan Kaufmann Series in Networking, 2004. 51
 - [65] Y. Hu, J. Gross, and A. Schmeink, “Outage probability of a multi-relay cognitive network with an uncertain number of forwarding relays,” in *Wireless Communications Systems (ISWCS), 2014 11th International Symposium on*, pp. 350–354, Aug 2014. 152
 - [66] A. Antonopoulos and C. Verikoukis, “Network coding-based cooperative ARQ scheme,” in *Communications (ICC), 2011 IEEE International Conference on*, pp. 1–5, June 2011. 152
 - [67] A. Antonopoulos, C. Skianis, and C. Verikoukis, “Network coding-based medium access control protocol for cooperative wireless networks,” in *Communications (ICC), 2012 IEEE International Conference on*, pp. 1933–1937, June 2012. 152
 - [68] P. Fan, C. Zhi, C. Wei, and K. Ben Letaief, “Reliable relay assisted wireless multicast using network coding,” *Selected Areas in Communications, IEEE Journal on*, vol. 27, pp. 749–762, June 2009. 152
 - [69] Q. Song, Y. Li, Z. He, and J. Lin, “On reliable multicast with network coding-ARQ for relay cooperation cells,” pp. 1–5, May 2012. 152
 - [70] N. Abuzainab and A. Ephremides, “Energy efficiency of cooperative relaying over a wireless link,” *Wireless Communications, IEEE Transactions on*, vol. 11, pp. 2076–2083, June 2012. 152
 - [71] J.-C. Kao and F.-W. Chen, “On RANC ARQ for wireless relay networks: From the transmission perspective,” *Wireless Communications, IEEE Transactions on*, vol. 12, no. 6, pp. 2962–2976, 2013. 152

- [72] J. Vicario, A. Bel, J. Lopez-Salcedo, and G. Seco, “Opportunistic relay selection with outdated CSI: outage probability and diversity analysis,” *Wireless Communications, IEEE Transactions on*, vol. 8, pp. 2872–2876, June 2009. 153
- [73] Y. Li, Q. Yin, W. Xu, and H.-M. Wang, “On the design of relay selection strategies in regenerative cooperative networks with outdated CSI,” *Wireless Communications, IEEE Transactions on*, vol. 10, pp. 3086–3097, Sep. 2011. 153
- [74] J. Moualeu, W. Hamouda, and F. Takawira, “Relay selection for coded cooperative networks with outdated CSI over Nakagami-m fading channels,” *Wireless Communications, IEEE Transactions on*, vol. 13, pp. 2362–2373, May 2014. 153
- [75] X. Bao and J. Li, “Efficient message relaying for wireless user cooperation: Decode-amplify-forward (DAF) and hybrid DAF and coded-cooperation,” *Wireless Communications, IEEE Transactions on*, vol. 6, pp. 3975–3984, Nov. 2007. 153
- [76] E. Olfat and A. Olfat, “Performance of hybrid decode-amplify-forward protocol for multiple relay networks over independent and non-identical flat fading channels,” *Communications, IET*, vol. 5, pp. 2018–2027, Sep. 2011. 153
- [77] H. Ngo, T. Nguyen, and L. Hanzo, “Amplify-forward and decode-forward cooperation relying on systematic luby transform coded hybrid automatic-repeat-request,” *Communications, IET*, vol. 5, pp. 1096–1106, May 2011. 153
- [78] Z. Bai, J. Jia, C.-X. Wang, and D. Yuan, “Performance analysis of SNR-based incremental hybrid decode-amplify-forward cooperative relaying protocol,” *Communications, IEEE Transactions on*, vol. 63, pp. 2094–2106, June 2015. 153
- [79] S. Ren and K. Letaief, “Maximizing the effective capacity for wireless cooperative relay networks with QoS guarantees,” *Communications, IEEE Transactions on*, vol. 57, pp. 2148–2159, July 2009. 154
- [80] G. Farhadi and N. Beaulieu, “On the ergodic capacity of multi-hop wireless relaying systems,” *Wireless Communications, IEEE Transactions on*, vol. 8, pp. 2286–2291, May 2009. 155
- [81] I. Trigui, S. Affes, and A. Stephenne, “Ergodic capacity analysis for interference-limited af multi-hop relaying channels in Nakagami-m fading,” *Communications, IEEE Transactions on*, vol. 61, pp. 2726–2734, July 2013. 155

- [82] K. Lee, W. Sung, and J. W. Jang, “Application of network coding to IEEE 802.16j mobile multi-hop relay network for throughput enhancement,” *Communications and Networks, Journal of*, vol. 10, pp. 412–421, Dec. 2008. 155