Design and Analysis of Coded Cooperation in Relay Networks

by

Josephine Pui Kwan Chu

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Electrical and Computer Engineering
University of Toronto

Copyright © 2010 by Josephine Pui Kwan Chu
Abstract

Design and Analysis of Coded Cooperation in Relay Networks

Josephine Pui Kwan Chu

Doctor of Philosophy

Graduate Department of Electrical and Computer Engineering

University of Toronto

2010

This dissertation deals with wireless communications using cooperating relay nodes. Specifically, this dissertation relaxes two restrictive conditions ubiquitous in the current literature. First, the assumption that relay nodes can perform complex calculations is lifted. Demodulate-encode-forward (DEF) is a low-complexity relaying scheme where the relay is asked only to demodulate, not decode, a source transmission. The implementation of DEF and various methods that can be used with DEF to improve the performance while satisfying the hardware complexity limitations are detailed here. Second, we remove the assumption that the relays either transmit the complete source codeword or not transmit at all. When relays have limited resources, each relay may only be able to transmit part of the source codeword. Fractional cooperation, which allows nodes to transmit a fraction of the source codeword, is proposed and analyzed. Fractional cooperation is also very flexible because coordination between relaying nodes is not required.

A third contribution of this dissertation is the use of the union-Bhattacharyya bound (UBB) to analyze relay networks. The bound has the significant advantage of accounting for the specifics of the system parameters and coding scheme used. The UBB is shown here to provide an effective and efficient scheme for relay selection, performance prediction, and system design. It can also be used to distribute relay resources in order to optimize the total energy consumed and error rate performance. A sub-optimal distributed algorithm that can be used to solve the optimization problems is introduced.
Acknowledgements

I would like to take this opportunity to express my gratitude towards all the support without which this research would not have been possible. First and foremost, I would like to thank my supervisor, Prof. Raviraj Adve, for his constant encouragement and guidance throughout my studies. Not only has he helped me learn how to perform research, he has also taught me invaluable life lessons. I feel very fortunate to have Prof. Adve as a mentor. I would also like to thank Prof. Andrew Eckford, for acting as my unofficial co-supervisor. His inspiring ideas have not only helped me make advancement in my research, but have also helped steer me in the right direction when I was unable to make progress.

I would like to thank all the professors and students in the Department of Electrical and Computer Engineering at University of Toronto for all the interesting discussions and for providing valuable advice throughout my studies. They have helped me overcome setbacks I had encountered during my studies. I would also like to thank my friends for providing company throughout this journey. They have made sure that I know I was never alone. Finally, and most importantly, I would like to thank my parents for their patience and for the continuous support they have provided throughout my life.
## Contents

1 Introduction .......................... 1
   1.1 Motivation .......................... 1
   1.2 Overview of Relay Channels ........... 3
   1.3 Literature Review .......................... 4
   1.4 Thesis Contribution ....................... 13
   1.5 Thesis Outline .......................... 15

2 System Model and Background .......... 16
   2.1 System Model .......................... 16
      2.1.1 Decode-and-forward ................... 19
      2.1.2 Amplify-and-forward .......................... 19
   2.2 Diversity Order .......................... 20
   2.3 Factor Graphs and Sum-Product Algorithm ...................... 21
   2.4 Union-Bhattacharyya Bound .......................... 26

3 Demodulate-and-Forward ............... 31
   3.1 Encoding .......................... 32
      3.1.1 DEF with Mod-2 Parity ................... 33
      3.1.2 DEF with LDGM Codes ....................... 33
      3.1.3 DEF with RA Codes\(^1\) ..................... 36

---

\(^1\)Professors Andrew W. Eckford and Raviraj S. Adve have presented this relaying scheme in [1]. The
3.2 Decoding

3.2.1 Relay-to-Source Messages

3.2.2 Source-to-Relay Messages

3.3 Simulation Results

3.3.1 DEF with Mod-2 Parity

3.3.2 DEF with LDGM Codes

3.3.3 DEF with RA Codes

3.4 Discussion

4 Fractional Cooperation

4.1 Fractional Cooperation with One Relay

4.1.1 Analysis Results

4.1.2 Simulation Results

4.2 Fractional Cooperation with Multiple Relays

4.2.1 Diversity Order Analysis for Fractional Cooperation

4.2.2 Outage Set Analysis for Fractional Cooperation

4.2.3 Simulation Results

4.3 Discussion

5 Analysis and Design using the Union-Bhattacharyya Bound

5.1 Union-Bhattacharyya Bound for Relay Channels

5.2 Relay Selection for DEF

5.2.1 Bhattacharyya Parameter for DEF

The relaying scheme is detailed here as it is used extensively in the simulations.

Professor Andrew W. Eckford is the main contributor to the theoretical analysis in this section, and the results were published in [2, 3]. However, additional contributions by the dissertation author has been made based on the results from these analysis and the dissertation is organized to provide a continuous flow of presentation. The contribution made by the dissertation author will be made clear in the contents.
D Proof of Border Points

Bibliography
List of Tables

1.1 Full-duplex and half-duplex transmission schemes. .......................... 4

4.1 Lower bound on $r_c$ for fractional cooperation using rate-1/2 punctured systematic RA code for various $\epsilon$ derived from outage set analysis. ... 70

5.1 Lower bound on $r_c$ for fractional cooperation using rate-1/2 punctured systematic RA code for various $\epsilon$ derived from the UBB. .............. 95

6.1 Comparison of error performance and energy consumed with SQP20 and the distributed optimization scheme. ................................. 111
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>System model of a three-node relay channel.</td>
<td>3</td>
</tr>
<tr>
<td>2.1</td>
<td>Multiple-source multiple-relay system.</td>
<td>17</td>
</tr>
<tr>
<td>2.2</td>
<td>Factor graph of a (7,4) Hamming code.</td>
<td>23</td>
</tr>
<tr>
<td>2.3</td>
<td>Messages passed (a) from variable node to factor node and (b) from factor node to variable node.</td>
<td>25</td>
</tr>
<tr>
<td>3.1</td>
<td>Factor graph of DEF with simple parities formed at the relay.</td>
<td>34</td>
</tr>
<tr>
<td>3.2</td>
<td>Factor graph of DEF with LDGM code.</td>
<td>36</td>
</tr>
<tr>
<td>3.3</td>
<td>Factor graph of DEF with RA code.</td>
<td>37</td>
</tr>
<tr>
<td>3.4</td>
<td>Illustration of messages passed between source’s and relay’s variable nodes.</td>
<td>41</td>
</tr>
<tr>
<td>3.5</td>
<td>BER <em>(dashed lines)</em> and FER <em>(solid lines)</em> of DEF with mod-2 parity in AWGN channels.</td>
<td>43</td>
</tr>
<tr>
<td>3.6</td>
<td>BER <em>(dashed lines)</em> and FER <em>(solid lines)</em> of DEF with mod-2 parity in fading channels.</td>
<td>43</td>
</tr>
<tr>
<td>3.7</td>
<td>FER of DEF with LDGM code in AWGN channels, where (\epsilon_{\text{par}} = 0) <em>(solid lines)</em> or (\epsilon_{\text{par}} = 1) <em>(dashed lines)</em>.</td>
<td>45</td>
</tr>
<tr>
<td>3.8</td>
<td>FER of DEF with LDGM code in fading channels, where (\epsilon_{\text{par}} = 0) <em>(solid lines)</em> or (\epsilon_{\text{par}} = 1) <em>(dashed lines)</em>.</td>
<td>45</td>
</tr>
<tr>
<td>3.9</td>
<td>FER of DEF with RA code in AWGN channel, where (\epsilon_{\text{par}} = 0) <em>(solid lines)</em> or (\epsilon_{\text{par}} = 1) <em>(dashed lines)</em>.</td>
<td>47</td>
</tr>
</tbody>
</table>
3.10 FER of DEF with RA code in fading channel, where $\epsilon_{\text{par}} = 0$ (solid lines) or $\epsilon_{\text{par}} = 1$ (dashed lines). ................................. 47

3.11 Comparison of no relaying, DF, AF and DEF with RA code in fading channels. ................................................................. 48

4.1 FER of source’s (solid lines) and relay’s (dashed lines) codewords for fractional cooperation with one relay, where R is closer to S. ............... 60

4.2 FER of source’s (solid lines) and relay’s (dashed lines) codewords for fractional cooperation with one relay, where R is closer to D. ............... 60

4.3 FER of DEF with RA code under regular fractional cooperation for various $r$ and $\epsilon$ values. .......................................................... 68

4.4 FER of DEF with RA code under irregular fractional cooperation for $r = 11$, where $\bar{\epsilon} = 0.125$. .......................................................... 68

5.1 Plot of equivalent channel noise variance before and after decoding for the rate-1/4 systematic RA code. .............................................. 78

5.2 Contours of BER from density evolution (Markers) and from lookup table and (5.14) (Lines). .......................................................... 79

5.3 FER for relay selection based on the BP, max-min over the S-R and R-D channel, and exhaustive search. ........................................... 81

5.4 FER of fractional cooperation with relay selection, where $r_a = 8$ and $\epsilon_j = 0.2$. ................................................................. 83

5.5 Outage probability from simulation and upper bound on outage probability, where $r = 4$ and $\epsilon_j = 0.3$, and $r = 6$ and $\epsilon_j = 0.2$. ............... 85

5.6 Relationship between $p_{\text{SR,in}}$ and $p_{\text{SR,out}}$, the equivalent bit-flip probability after a given number of decoding iterations for the rate-1/2 punctured systematic RA code. ........................................... 88
5.7 BER and upper bound on BER are shown for various number of SPA iterations performed at the relay with $\gamma_{S,D} = -6$ dB, $\gamma_{S,R} = -1$ dB and $\gamma_{R,D} = -4$ dB. ........................................... 89

5.8 FER for relay selection based on the BP with 10 iterations of SPA. .... 90

5.9 Contours of $\beta_j$ for AF (solid) and DEF (dash-dot). .................. 91

5.10 FER for AF and DEF with blocklength $l_S = 16000$ and their corresponding SNR thresholds (dashed lines) derived from $c_0$ of the rate-1/2 RA code. 94

5.11 FER for AF and DEF with blocklength $l_S = 16000$ and $P_{out}$ of the rate-1/2 RA code with $m = 6$ and $\epsilon_j = 0.2$. ................................. 94

6.1 Contours of maximum FER for a two-source two-relay network with AWGN. 107

6.2 Contours of maximum FER for two-source two-relay network with fading channels using optimization (solid) and equal allocation (dash-dot). .... 108

6.3 Plot of $\sum_{i,j} \epsilon_{j,i}$ under optimization for a two-source two-relay network in fading channels. .................................................. 109

6.4 Distribution of number of random points used to generate an optimal solution coinciding with that generated from SQP200. ............... 110

D.1 Optimal points for $S_1$ with given $\bar{\epsilon}_{1,1}$ and $\bar{\epsilon}_{2,1}$. ......................... 129

D.2 Optimal points for two-source two-relay system with reserved resources $\bar{\epsilon}_{1,1}, \bar{\epsilon}_{1,2}, \bar{\epsilon}_{2,1}$ and $\bar{\epsilon}_{2,2}$. ................................. 130
Chapter 1

Introduction

1.1 Motivation

Fading is one major obstacle to the transmission of data over wireless links. In fading channels, the amplitude of the received signal is random, significantly increasing the probability of decoding error. In multiple-input multiple-output (MIMO) systems, where multiple antennas are available at the transmitter and/or receiver, path diversity helps decrease the probability of error. However, in networks where nodes are expected to be small and simple, implementing multiple antennas proves difficult. In such networks, path diversity can be made available if nodes cooperate to relay data for one another.

In the simplest relay channel, a source node (S), after broadcasting its own data, asks for help from other nodes in the network. A relay node (R) receives this data and upon hearing the request for assistance from S, forwards a message to the destination node (D), which uses both received messages to decode the source data. In addition to relaying, cooperative coding can be used, where channel codes are used at the relay to further improve system performance.

Wireless sensor networks (WSNs) are one type of network where relaying can be used to drastically improve the performance of the communication system. WSNs usually
Chapter 1. Introduction

comprise many autonomous nodes used, for example, to monitor environmental conditions. Depending on the application, these nodes can vary widely in size and cost [4]. It is generally accepted that large-scale networks will comprise relatively low-cost nodes, often operating on battery power. In such networks it is therefore important to conserve energy by limiting the transmission power and complexity of the processes executed on the nodes.

In most of the existing work on relay channels, assumptions were made that limit the potential of relaying in such low-complexity networks. In this dissertation, we focus on two of the most common assumptions. The first assumption is that the network comprises sufficiently complex nodes. Based on this assumption, two popular relaying schemes are decode-and-forward (DF), wherein the relay must decode the source’s codeword, and amplify-and-forward (AF), wherein the relay must be capable of transmitting signals with large amplitude variations. In this dissertation, a low-complexity relaying scheme called demodulate-encode-forward (DEF) is introduced to allow for relaying without complex hardware. DEF only requires the relay to demodulate the source transmission, a task any digital communication receiver should be capable of. The second assumption is that the relay either forwards all the source data, or does not assist the source node at all. However, given the available energy resources and data payload, a relay may only be able to devote a part of its resources to assist other nodes. Fractional cooperation, where relays are allowed to forward a fraction of the source’s codeword, is introduced and analyzed here. In a WSN, this would be beneficial to the extension of network lifetime as this helps to distribute the workload among nodes in the network.

As channel codes are used at the source node and potentially the relay node, it is quite difficult to characterize the performance of the relay channel without simulations. This dissertation develops the union-Bhattacharyya bound (UBB) as an analysis and design tool for relay channels. The UBB can be used for relay selection, and as an efficient method to find an upper bound on the outage performance and characterize the relay
channel. In addition, it can be used as a system design tool in relay channels, where it is used to optimize resources, e.g., to minimize frame error rate (FER) or the energy consumed for a target FER.

1.2 Overview of Relay Channels

The simplest relay channel is the 3-node network illustrated in Fig. 1.1 comprising a source (S), a relay (R) and a destination (D) node. In the figure $h_{SD}$, $h_{SR}$ and $h_{RD}$ are the channel coefficients between (for now) S and D, S and R, and R and D respectively. In the system, S is transmitting information to D, and the sole purpose of R is to assist S in transmitting to D.

If the relay is *full-duplex*, it can transmit and receive simultaneously. However, it is difficult, and generally expensive, to build full-duplex relays because of the required isolation between the transmit and receive circuits. In some networks, therefore, only *half-duplex* nodes are available, where the relay node cannot transmit and receive simultaneously on the same time/frequency channel. If *orthogonal* channels are required, then S and R cannot simultaneously transmit to D on the same time/frequency channel, such that the channels used by different nodes do not interfere with each other. If *non-orthogonal* channels are available, both S and R can transmit to D simultaneously. The signals are transmitted on the same time/frequency channel, and algorithms must
be implemented to separate the signals from the two nodes. Higher transmission rates can be achieved with non-orthogonal channels, but synchronization and/or interference cancellation must be implemented at D, increasing the complexity of the system.

With half-duplex channels, the transmission can be separated into two phases. In the first phase, S broadcasts its data. This is received by both R and D. After receiving the signal from S, R processes the data. There are many ways to process this data, which is dependent on the channel quality and the hardware available; some of these schemes are described below. In the second phase, R transmits the relayed data to D, and S either stays quiet (if the channels are orthogonal) or transmits simultaneously to D (if the channels are non-orthogonal). If full-duplex channels are available, R transmits to D at the same time S is transmitting, and the broadcast transmission in the first phase of next transmission can overlap with the second phase of the transmission. These three possible scenarios are summarized in Table 1.1.

### Table 1.1: Full-duplex and half-duplex transmission schemes.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Full-duplex</th>
<th>Half-duplex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time Slot 2n</td>
<td>S → (R, D), R → D</td>
<td>S → (R, D)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Orthogonal</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(S, R) → D</td>
</tr>
<tr>
<td>Time Slot 2n + 1</td>
<td>S → (R, D), R → D</td>
<td>R → D</td>
</tr>
</tbody>
</table>

This section reviews the key contributions in the, now vast, body of literature in the area of relay networks. The focus here is on publications related to the topics considered in this dissertation. The reader is referred to [5] for the background material on multiple-input multiple-output (MIMO) systems assumed in the discussion below.

In the understanding of relay networks, one of the most important contributions is the
work of Cover and El Gamal [6], where an upper bound on the capacity of the three-node full-duplex relay channel was presented. In addition, two well-known relaying schemes, DF and compress-and-forward (CF), were introduced, and their associated achievable rates were presented. With DF, R attempts to decode the source’s codeword after the transmission. After decoding, R transmits additional data to D to assist in the decoding of the source’s codeword at D. As correct decoding is required at R, the capacity of DF is limited by quality of the S-R channel, and it performs well when the S-R link quality is better than the S-D link quality. It was shown that DF achieves the capacity for the degraded relay channel, where the source’s codeword can be recovered given the information available at R.

With CF, which is also known as estimate-and-forward in other publications such as [7, 8], the relay compresses the received signal from the source and transmits the compressed signal to the destination. For example, the received signal space can be separated into a finite number of regions, and the index of the region to which the received signal belongs can be transmitted to D. R does not attempt to decode, and hence the transmission rate is not limited by the correct decoding of the source’s codeword at R. In addition to DF and CF, the relaying scheme partial DF (PDF) is also presented in [6]. With PDF, the source’s codeword consists of two components. The first component is a codeword designed to ensure correct decoding at R, and the second component is only decoded by D. After correct decoding, R transmits additional data to D to assist in the decoding of the first component. After decoding the first component using signals from both S and R, D then attempts to decode the second component of the source’s codeword.

In [9], the capacity of half-duplex non-orthogonal relay channels is derived. Here the transmission is divided into two phases: The relay receives the source transmission in the first phase and transmits to D in the second phase. The capacity is found by optimizing the division of time between the two phases. The achievable rates of networks comprising multiple nodes, each equipped with a single antenna, can be found in [10,11].
In these papers, DF, CF and a combination of the two relaying schemes are used by the relays. Instead of studying the transmission rates for large networks where a defined relaying scheme is used, the authors in [12] analyzed the transport capacity, defined as the product of transmission rate and distance, for a fixed power constraint using the max-flow min-cut bound. Upper bounds on the transport capacity were found for networks with different topologies, independent of the relaying scheme used. In [8], the achievable rates of CF using various quantization schemes was found for multi-relay networks. A study on achievable rates for networks with nodes that possess more than one antenna can be found in [13,14].

The work in [15–17] developed the diversity-multiplexing tradeoff for relay channels. Essentially, the DMT describes the tradeoff between the reliability of the data and the data rate when multiple channels are available [18], and the DMT of a relay channel describes the best possible diversity-multiplexing tradeoff achievable by a given relaying scheme. In [15], the relaying scheme AF was introduced. As the name implies, the relay amplifies the received signal and transmits it to the destination node. In theory AF is a simple scheme, but if the relay nodes are half-duplex, they must store the complex received signal for amplification in the second phase. If full-duplex channels are available, no storage is required, but an extra set of RF components is needed to allow simultaneous transmission and reception of signals, increasing the complexity and cost of the relays. If the range of the transmitted signal is allowed to be continuous, conventional channel coding cannot be used to improve the error performance. Furthermore, because the amplitude of the transmitted signal can have large variation, this complicates the design of amplifiers in the transmission modules. The DMT of DF and AF are found under the assumption that the channels are orthogonal in [15]. It was shown that AF has better DMT performance than DF, mainly because with DF, the relay cannot contribute when correct decoding is not achieved. The DMT of various relaying schemes under the assumption of non-orthogonal channels is studied in [16]. A study of the DMT for
systems comprising multiple-antenna relays can be found in [17].

Instead of requesting a node to selflessly assist another node, nodes can pair up or form groups to assist each other, as suggested in [19, 20] for cellular networks. This is denoted as user cooperation. Not only does user cooperation increase the capacity for both users, it also increases the cellular coverage. Examples of implementation in a code-division multiple access (CDMA) system with full-duplex nodes were shown in [20]. To account for the probability of estimation error by the partnering node, a sub-optimal receiver structure, denoted as $\lambda$ maximal-ratio combining ($\lambda$-MRC), was suggested. The signals received by the destination from the partner is weighted by $\lambda \in [0, 1]$, which is an indication of the destination’s confidence in the contributions of the partner.

Space-time codes (STC) have also been applied to relay channels to obtain diversity gain without sacrificing the achievable transmission rate. With distributed STC (DSTC), relays transmits simultaneously to the destination node using STC. In [21], the outage probability and DMT of STC for DF were derived. There are several publications studying DSTC, including [22–24]. In [22], the diversity order of the FER was approximated for relay networks with good and poor inter-user channel quality. The use of Alamouti’s space-time block code (STBC) [25] was presented in [23], where two different detectors were shown and their corresponding error performance was given. Noticing that the diversity order of Alamouti’s scheme is limited by the errors in decoding at the relay, the authors in [24] introduced a improved transmission scheme with distributed STBC. In this scheme, the S-R channel quality is taken into account while implementing the STBC: the transmission power at R is adjusted to optimize the signal to interference plus noise ratio (SINR) or error rate at D. It was shown that a diversity order of 2 can be achieved with a single relay.

The pairwise error probability of STC in networks comprising single-antenna relays was found in [26], and the results were extended to networks consisting of relays with multiple antennas in [27]. The use of DSTC allows the network to achieve full diver-
sity order with limited overhead in terms of bandwidth. However, DSTC requires strict symbol-level synchronization between relaying nodes. How this can be achieved and maintained in a distributed network potentially with mobile nodes is unclear. Instead of requiring synchronization between nodes, an asynchronous cooperative scheme was introduced in [28], where decision feedback equalizers are used. The feedforward and feedback filter coefficients were found to minimize the mean square error such that diversity can still be introduced without requiring the relays to be synchronized. However, merely increasing the number of relays does not increase the diversity order; in order to increase the diversity order, additional artificial delays must also be introduced.

In networks where symbol synchronization between nodes is unavailable, relay selection, instead of DSTC, can be used to increase the diversity order. With relay selection, only the best relay is chosen out of a pool of available relays to assist the source. Different criteria, such as mutual information [29–31], can be used to choose the best relay. Relay selection for DF was studied in [29,30]. For DF the best relay is the relay in the decoding set with the optimal R-D channel, where the \textit{decoding set} is the set of relays that are able to decode the source’s codeword. In [31], relay selection for AF was studied, where the best relay is the relay with the best S-R-D compound channel. Relay selection was shown in [29] to be more power-efficient than distributed STBC while also avoiding the associated problems of synchronization. In both cases, it has been shown that \textit{full diversity} can be achieved with the use of only one relay, i.e., the outage probability of a system with \( r \) relays has diversity order \((r + 1)\). One drawback with relay selection is that the transmitting node must possess knowledge of the channel in order to choose the best relay.

In addition to studies performed on the achievable rates and DMT of relay channels, there are works available that study coding schemes that can be used to minimize the outage probability and error probability. Implementations of DF using various codes have been presented. One of the earliest works on coded cooperation is [32], where
users cooperate while using convolutional codes (CC). After receiving the signal from its partner node, a node attempts to decode its partner’s codeword. Successful decoding is declared when the cyclic redundancy check (CRC) is satisfied, upon which the node forms a new codeword using the decoded bits and transmits this to the destination. If decoding is unsuccessful at the node, it retransmits a codeword based on its own data and ignores the data from its partner. By providing redundancy for its partner’s data, the error rate improves significantly. Furthermore, by eliminating transmissions of incorrect bits, the propagation of errors during decoding at the relay is minimized.

Building on a similar concept is [33], where a turbo code (TC) is used. However, instead of transmitting a complete codeword, the relay transmits only parity bits formed using the interleaved decoded information bits. Hence, the source’s codeword, together with the parity bits formed at the relay, forms a complete TC codeword, and is known as parallel concatenated convolutional code (PCCC) in the paper. A different relaying scheme, where serial concatenated convolutional code (SCCC) is used, is also introduced in the paper, where the relay forms a new codeword on the decoded codeword. Contours of SNR over the S-R and R-D links required to achieve a FER of $10^{-2}$ were shown to compare a rate-1/4 cooperative code from [32] and a rate-1/4 distributed TC from [33], and for the sample scenario given, where R is 1 m from S and 9 m from D, and S is 10 m from D, the code from [32] performs better than the PCCC shown, but worse than SCCC. Variations of relay coding schemes using TCs were introduced in [34] for full-duplex relay channels, and the performance of the various codes was compared to the theoretical limits. For example, for a relay channel with a perfect S-R link, and S-D and R-D links with same SNR, the distance between the SNR required to achieve a BER of $10^{-5}$ and the theoretical limits ranges from 1.08 dB to 1.13 dB for code rates ranging from 1/2 to 1. Since the relaying schemes proposed in [34] are distributed codes over both time and space, it is difficult to compare the them to those presented in [33], where half-duplex channels are used. Similar to the concept presented in [33], the authors in [35]
suggested a coding scheme using CCs where the relay transmits additional parity bits if the source’s codeword is successfully decoded. Even though the underlying concept behind this scheme is quite similar to that in [33], no comparison is given in [35]. Since TC is a more powerful code than CC, we believe that the code from [33] performs better than [35]. The relaying schemes introduced in the above publications are different than that detailed in this dissertation as the low-complexity coding scheme introduced here does not require correct decoding at the relay, and quantized values of the received signal, in addition to the parity bits formed using the received signal, are transmitted by R to D.

To improve the chance of successful decoding by both R and D, the source’s codeword and the codeword formed by the concatenation of the source’s codeword and parity bits transmitted by the relay must be a strong code [36–38]. Examples of literature that focus on such designs using LDPC codes include [36] and [37], where the former paper optimized the check node and variable node edge degrees of the LDPC codewords for non-orthogonal half-duplex channels, while the latter focused on LDPC codes where the relay forms additional parity bits to implement the “binning” concept for DF introduced in [6]. An implementation of [36] can be found in [38], where the convergence behavior of LDPC coded relay systems was analyzed using extrinsic information transfer (EXIT) charts [39]. Similar to the relaying schemes using TC introduced above, the relaying schemes introduced in [36–38] assume that successful decoding can be achieved at the relay, and is different than the low-complexity relaying scheme introduced here in this dissertation. As stated in [37], the design of the bilayer LDPC code from [37] employs structured expurgating and lengthening techniques, whereas the LDPC codes in [36, 38] uses random puncturing techniques to achieve the desired rate. It is believed that the bilayer LDPC code supersedes the LDPC codes from [36,38] as it is designed specifically for the relay channels. This, however, comes at a cost of higher complexity.

In additional to code optimization for DF, studies on implementation of CF using
channel codes are available [40–42]. An implementation of CF from [6] can be found in [40], where a Gaussian quantizer is used at R. In [41], the authors took advantage of the correlation between the source bits and the received bits at the relay, and suggested using Slepian-Wolf coding [43] to obtain a relaying scheme that minimizes the number of transmission bits while ensuring a diversity order of 2 is achieved with one relay. This concept is further extended to the case where, instead of performing a hard detection on the received bits, the relay utilizes the received signal to encode using Wyner-Ziv coding [42]. Since Wyner-Ziv is a generalization of Slepian-Wolf coding, the relaying scheme from [42] outperforms that from [41], but at a cost of higher complexity. CF is also studied in [7] and [8]. In both papers, since relaying is mostly likely used when the S-D channel quality is poor, CF relaying schemes using binary modulation at S was studied, as it provides the greatest improvement in capacity in the low signal-to-noise ratio (SNR) regime. Based on numerical power allocation results, the authors in [7] concluded that in the second phase of transmission, where both S and R can be transmitting to D in half-duplex channels, the rate is maximized by allocating all available power to R. This translates to orthogonal channels, where S stays silent in the second phase. It is difficult to compare the relaying scheme from [42] and that from [7] as the goal of the former scheme is to improve the error performance, whereas the latter relay scheme aims to increase the achievable data rate.

Since it is unclear which type of quantization provides the best achievable rates, the authors in [8] studied various quantization schemes when CF is used, and showed that depending on the channel quality of the various links, Gaussian quantization is not necessarily the optimal quantization method. It was pointed out that when the S-R channel quality is poor, a Gaussian quantizer should be used at the relay to provide the best performance, where as a binary quantizer should be used when the S-R channel quality is slightly better. One of the relaying schemes studied in [8] is very similar to that presented in this dissertation. The main difference between the work detailed in [8]
and here is that instead of analyzing the achievable data rate, we focus on the \textit{practical implementation} of such relaying schemes.

Instead of using fixed-rate codes, rateless codes can be used in relay channels, as suggested in \cite{44,45}, where fountain codes are used to implement DF. In \cite{44}, the source and relay nodes use rateless codes, and two different relaying protocols are introduced. With the first protocol, which is synchronous, the number of relays that have accumulated enough information to decode must exceed a threshold value before the relays are allowed to transmit to D. With the second asynchronous protocol, each relay can start transmitting to D as soon as enough information has been collected to decode the source’s codeword. In addition to D, other relays that have not decoded yet can also benefit from the transmissions from relays that have decoded correctly. It was shown that this is an energy efficient scheme, in that no excess transmissions occur. In \cite{45}, a one-relay system was studied. The source node transmits a rateless code, and after listening to the source’s broadcast for long enough to decode the source’s codeword, R attempts to decode the source’s codeword. After decoding the codeword, R cooperates with S using Alamouti’s code to continue to transmit to D until enough information is received in order to decode the source’s codeword. As with all rateless codes, feedback is needed to inform the transmitters when decoding is successful.

The use of a cluster of relays, each relaying only a fraction of the source’s codeword, was explored in \cite{46}. The relaying scheme, called incremental redundancy (IR), divides the source’s codeword into non-overlapping sections, where each relay is assigned one of the sections. After S has transmitted, each relay attempts to decode the source’s codeword. If decoding is successful, then the relay transmits the section of codeword it is responsible for; otherwise, it stays quiet while S takes over the task of transmitting that section of the codeword. Coordination between all the relays beforehand is required such that each node knows the section it is responsible for. Our work on fractional cooperation can also be used to share the relaying burden, however as we will see, with
little coordination required.

Instead of limiting coding schemes to DF or CF, innovative coding schemes that combine DF and AF can be found in [47, 48]. In [47], a soft-input soft-output decoder is used at R to encoding the received signal from S. The log-likelihood ratio (LLR) of the encoder output is scaled and transmitted to D. The received signals from both S and R are combined before being fed into a decoder, and hence error propagation may occur since the probability of a decoding error is not taken into account in the decoder at D. It was shown that the equivalent SNR is enhanced when the received SNR at R lies between 0 and 15 dB. Independently, decode-amplify-forward (DAF) [48], a relaying scheme where the LLR from the relay decoder is amplified was presented. The LLR is scaled before being transmitted to D to satisfy the power constraints. In addition, a hybrid scheme that allows DF to be used if correct decoding is accomplished at R and DAF to be used otherwise was introduced. Since DAF allows the use of DF when the CRC is satisfied, it is more flexible and outperforms the soft-DF relaying scheme from [47]. These schemes, however, suffer from the same issues as AF, as the transmitted power is not constant and can have a large variation. This section has reviewed the relevant subset of a vast body of work in relay channels. The specific relevance to, and differences from, this dissertation will be made clear in the next section and subsequent chapters.

1.4 Thesis Contribution

This dissertation focuses on aspects of practical implementation of relay channels. First, DEF, a robust low-complexity coded cooperation scheme is presented. In DEF the relay must only demodulate, not decode, the source’s transmission. The relay forms its own codeword using the demodulated bits. This relaying scheme can therefore be implemented in any digital transceiver, without requirements of complex hardware associated with some of the other relaying scheme such as DF. The one drawback is that the destination
must be informed of the reliability of the demodulated bits. Here channel codes are utilized to improve the error performance. The encoding and decoding procedures are defined, and the error performance of DEF in additive white Gaussian noise (AWGN) and fading channels is presented.

Second, the concept of fractional cooperation is introduced wherein the requirement of relaying the complete codeword is removed. Instead, each relay can relay only a fraction of the source’s codeword to the destination node. Unlike the work in [46], no coordination between relaying nodes is required. It is shown that, depending on the relaying fraction and channel condition, there is a critical number of relays required to obtain a diversity of 2. Analysis results on fractional cooperation are also presented.

As seen from the literature review in Sec. 1.3, most of the analysis on relay channels is limited to information theoretic capacity results. There has been little work on analyzing system performance while accounting for the specifics of the coding scheme used and/or the resources available. The published works resort to simulations to illustrate system behavior. In this dissertation the UBB is applied to relay networks to analyze system behaviour. It is shown that the UBB can be used as a system design and analysis tool in relay channels where low-complexity coded cooperation and/or fractional cooperation is used. The applications presented include relay selection for DEF, relay channel characterization, and system analysis and design for fractional cooperation.

Finally, optimization problem formulations with the use of UBB are presented. Two goals are achieved through the optimization: minimize the energy consumption while ensuring the FER is below a given threshold, and in the case where the resources cannot guarantee the FER threshold is achieved, the maximum FER over all the source nodes is minimized. A distributed algorithm is introduced to solve the non-convex optimization problems.
1.5 Thesis Outline

The dissertation is organized as follows. The system model and background information useful in understanding this dissertation is presented in Ch. 2. In Ch. 3, the relaying scheme DEF is introduced, and details of the encoding, decoding and relay processing involved are presented. In Ch. 4, the concept of fractional cooperation is introduced, and analysis of fractional cooperation is presented. In Ch. 5, various applications of the UBB for analysis and system design of the relay channels are presented. The optimization problems for multiple-source multiple-relay systems are presented in Ch. 6, where the energy consumption or maximum FER are minimized. Finally, concluding remarks and possible future work are presented in Ch. 7.
Chapter 2

System Model and Background

In this chapter, the system model is presented. In addition, some background information that is useful to understand the topics in this thesis is presented: an overview of diversity order, factor graphs, the sum-product algorithm (SPA), density evolution and the union-Bhattacharyya bound. The review here is brief and the reader is referred to the literature for further details.

For the rest of the thesis, the following notation is used: a vector is represented by boldface, e.g., \( \mathbf{z} \), and its components are represented as \( z(n) \), such that if \( \mathbf{z} \) is of length \( l \), \( \mathbf{z} = [z(1), z(2), \ldots, z(l)]^T \).

2.1 System Model

The multiple-relay system under consideration is shown in Fig. 2.1. The system comprises a source node S, \( r \) relay nodes \( R_j, j = 1, \ldots, r \), and a destination node D. It is assumed that the channels between the various nodes can be modeled as either additive white Gaussian (AWGN) or quasi-static Rayleigh fading channels [49] as specified. If a quasi-static Rayleigh fading channel is assumed, then the channel coefficients over the fading channels remain constant for a block, and change from block to block. It is also assumed that the channel coefficients changes slowly (slow-fading), such that the implementation
of time diversity introduces large delays in the receiving of data and hence is not used to provide diversity. In addition, we assume that only half-duplex nodes are available, so the relays cannot transmit and receive simultaneously. Transmission synchronization among various nodes, where the transmitted signal is synchronized, is not available, and orthogonal channels are used for communications between the nodes. For example, if time division multiple access (TDMA) or frequency division multiple access (FDMA) is used, then these resources are shared among the nodes such that only one node is transmitting at one time in a given frequency band. Channel allocation must be decided ahead of time if TDMA is used to coordinate communication between nodes to eliminate or minimize the probability of collisions. In addition, it is assumed that channel state information (CSI) of a link is available to the corresponding receiver. Finally, it is assumed that the S-R$_j$ channel SNR is known at D, and this information is used in the decoding process. For simplicity, it is assumed that binary phase-shift keying (BPSK) is used. However, the concepts developed here can be extended to other types of modulation schemes, and such extensions is further discussed in Ch. 7.

As stated in Ch. 1, the transmission is divided into two phases. In the first phase, S first forms a codeword $c_S$ of length $l_S$ and rate $R_S$, and the codeword is modulated to
the corresponding BPSK signal $x_S$. Let $\xi$ be the function that maps $\{0, 1\}$ to $\{+1, -1\}$, where $\xi(z) = (-1)^z$; then $x_S(n) = \xi(c_S(n))$. The baseband representation of the received at $R_j$ and $D$ respectively is given by

$$y_{S,Rj} = h_{S,Rj}x_S + n_{S,Rj}, \quad (2.1)$$

$$y_{S,D} = h_{S,D}x_S + n_{S,D}, \quad (2.2)$$

where $h_{S,Rj}$ and $h_{S,D}$ are channel coefficients between $S$ and $R_j$, and $S$ and $D$ respectively, which are zero-mean, independent, circularly symmetric complex Gaussian random variables (RVs) with variances 1, and $n_{S,Rj}$ and $n_{S,D}$ are zero-mean, independent, circularly symmetric complex Gaussian RV vectors with variances $N_{Rj}$ and $N_D$ respectively. Relay node $R_j$ processes the signal $y_{S,Rj}$ to form codeword $c_{Rj}$. Details of this processing are given later. After modulating the codeword to form $x_{Rj}$ the relay transmits the modulated codeword in the second phase and the destination receives

$$y_{Rj,D} = h_{Rj,D}x_{Rj} + n_{Rj,D}, \quad (2.3)$$

where $h_{Rj,D}$ is the channel coefficient between the $R_j$ and $D$, which is a zero-mean, independent, circularly symmetric complex Gaussian RV with variance 1, and $n_{Rj,D}$ is a zero-mean, independent, circularly symmetric complex Gaussian RV vector with variance $N_D$. Various relaying schemes are available, such as DF and AF, where the relays can process the received source signal in different ways. The low-complexity relaying scheme, DEF, is detailed in Chapter 3.

In this system, each transmission is assumed to use unit average power. The average received SNR of the channel between $S$ and $R_j$ is given by

$$\gamma_{S,Rj} = \mathbb{E}[\gamma_{S,Rj}] = \mathbb{E}[|h_{S,Rj}|^2]/N_{Rj}, \quad (2.4)$$

where $\gamma_{S,Rj}$ is the instantaneous SNR of the channel between $S$ and the $R_j$ relay and $\mathbb{E}[\cdot]$ denotes statistical expectation. The average received SNR of the S-D and $R_j$-D channel, $\gamma_{SD}$ and $\gamma_{Rj,D}$, are defined in a similar fashion. In the next sections, two well-studied relaying schemes, DF and AF are discussed.
2.1.1 Decode-and-forward

Assuming encoding and decoding is performed block-by-block, the available DF schemes can be separated into two types. The first type is based on *repetition codes*, where the relay simply recreates and transmits the source’s codeword if decoding is successful [21]. After receiving the signals from S and all the relays transmitting for S, the signals are summed after they are multiplied by the complex conjugate of the corresponding channel coefficients $h_{S,D}$ and $h_{R_j,D}$. If $r$ relays have decoded the source’s codeword correctly, the combined signal for S is given by

$$y_D = h_{S,D}^* y_{S,D} + \sum_{j=1}^{r} h_{R_j,D}^* y_{R_j,D}, \quad (2.5)$$

where $z^*$ is the complex conjugate of $z$. This is known as maximal ratio combining (MRC). It can be shown that $y_D$ is a Gaussian random signal, with equivalent instantaneous SNR for S given by

$$\gamma_D = \gamma_{S,D} + \sum_{j=1}^{r} \gamma_{R_j,D}. \quad (2.6)$$

The second type of DF assumes that independent codebooks are used at the relays. The symbols transmitted by the relay are used to help clarify confusion at D about the bits transmitted by S [6]. One method of implementing this is through the use of linear codes as suggested in [37]. As the performance of DF with different codebooks is dependent on the coding scheme, it is very difficult to derive an equivalent SNR. In general, the use of independent codebooks provides better performance than repetition code. Furthermore, the repetition factor scales as $\frac{1}{r+1}$, so repetition codes become very inefficient as the number of relays increases. However, they are much simpler to implement. Note that in both cases the relays must fully decode the source’s codeword.

2.1.2 Amplify-and-forward

AF is based on repetition coding where each relay $R_j$ amplifies the received signal and transmits it to D, and these signals are combined using MRC before it is used in
the decoder. The amplification factor can be determined by either the amplitude limit imposed by the range of the transmitting antenna or the average transmitted power. If the amplification factor is limited by the range of the transmitting antenna, the transmitted signal must be normalized to fall within the range of the transmitting antenna. When the average transmitted power constraint is enforced, the normalization factor is $E[|y_{S,R_j}|^2]$ and the transmitted signal at $R_j$ is given by

$$x_{R_j} = \frac{y_{S,R_j}}{\sqrt{E[|y_{S,R_j}|^2]}} = \frac{y_{S,R_j}}{\sqrt{|h_{S,R_j}|^2 + N_{R_j}}}.$$  

After receiving the signals from $S$ and all the signals from the $r$ relays that are transmitting for $S$ at $D$, the signals are matched to their corresponding channel coefficients and normalized by their associated noise variance before they are summed. The combined signal $y_D$ is given by

$$y_D = \frac{h_{S,D}^* y_{S,D}}{N_D} + \sum_{j=1}^{r} \frac{h_{S,R_j}^* h_{R_j,D} y_{R_j,D}}{N_{SRD,j} \sqrt{|h_{S,R_j}|^2 + N_{R_j}}},$$  

where $N_{SRD,j} = N_D + \frac{|h_{R_j,D}|^2 N_{R_j}}{|h_{S,R_j}|^2 + N_{R_j}}$. After some manipulation, it can be shown that $y_D$ is a Gaussian signal, with equivalent SNR given by [31]

$$\gamma_{AF} = \gamma_{S,D} + \sum_{j=1}^{r} \frac{\gamma_{S,R_j} \gamma_{R_j,D}}{\gamma_{S,R_j} + \gamma_{R_j,D} + 1}.$$  

### 2.2 Diversity Order

When transmitting data over a wireless channel, the transmitted signal is reflected by various objects in the environment, resulting in multiple copies of the original signal arriving at the receiver with different delays and amplitudes. Multipath fading occurs as a consequence of these out-of-phase signals combining to form a signal with random amplitude. In a slow-fading channel, where the amplitude changes slowly, one method to combat the destructive effects of fading is by providing path diversity. This can be implemented with the use of multiple antennas at the transmitter and/or receiver to
provide multiple paths. With multiple independent paths, the probability of error is reduced significantly, as errors occur only if all paths are in a deep fade. For example, for a system where the BPSK signal is transmitted over \( r \) independent Rayleigh fading channels, the error rate is inversely proportional to the \( r \)th power of the average SNR.

The performance of a system in fading channels can be measured with probability of error \( P_e \) or probability of outage \( P_{out} \). An outage is declared when either the average \( P_e \) for the realized AWGN channel is above a given threshold, or when the mutual information over the realized channels is less than a chosen threshold rate. One measure that can be used to assess the performance improvement brought along with path diversity is \textit{diversity order}. The diversity order is an asymptotic concept that describes the relationship between the error rate or outage probability to the average SNR \( \bar{\gamma} \) as \( \bar{\gamma} \to \infty \) under Rayleigh fading. The diversity order is defined as

\[
    d = \lim_{\bar{\gamma} \to \infty} -\frac{\log_{10}(P_e(\bar{\gamma}))}{\log_{10}(\bar{\gamma})}.
\]

For example, if the error rate has a diversity order of \( c \), then for every increment of 10 dB in \( \bar{\gamma} \), \( P_e \) decreases by a order of \( 10^c \). In this expression, the outage probability \( P_{out} \) can replace \( P_e \). For the rest of the thesis, we use the notation \( \Theta(\cdot) \) where if \( g(z) = \Theta(h(z)) \), then

\[
    \lim_{z \to \infty} \frac{g(z)}{h(z)} = c,
\]

where \( c \) is a constant. For example, in a Rayleigh fading channel, if the system has an outage probability with diversity order \( c \), then \( P_{out}(\bar{\gamma}) = \Theta(\bar{\gamma}^{-c}) \).

### 2.3 Factor Graphs and Sum-Product Algorithm

Factor graphs are bipartite graphs consisting of two types of nodes: variable and factor nodes. They are used for visual representations of functions which are products of “local” functions [50]. Here it is illustrated how they can be used to detail the relationship
between the symbols in a codeword. Let \( p(c|y) \) be the \textit{a posteriori} probability (APP) that modulated signal \( x = \xi(c) \) is transmitted given that signal \( y \) is received. From Bayes’ rule,

\[
p(c|y) = \frac{f(y|c)p(c)}{f(y)},
\]

where \( f(y|c) \) is the likelihood probability and \( p(c) \) is the probability that codeword \( c \) is selected for transmission. For a fixed received signal \( y \), \( p(c|y) \propto f(y|c)p(c) \), and is a function of \( c \). Let \( g(c) = f(y|c)p(c) \). For a memoryless channel and length-\( l \) channel code, \( f(y|c) = \prod_{n=1}^{l} f(y(n)|c(n)) \). Assuming that all codewords have equal probability, \( g(c) \) can be written as [50]

\[
g(c) = \frac{1}{|C|} \chi_C(c) \prod_{n=1}^{l} f(y(n)|c(n)),
\]

where \( C \) denotes the codebook and \( \chi_C(c) = 1 \) if \( c \) is a valid codeword in \( C \), and 0 otherwise. For all linear codes, \( \chi_C(c) \) can be factored into “local” functions where the local function output is 1 if even parity function is satisfied, and 0 otherwise. For example, for a binary code where \( c(1) \oplus c(2) \oplus c(3) \oplus c(5) = 0 \) must be satisfied for a vector \( c \) to be a valid codeword, the local function \( I_1 \) is given by

\[
I_1(c) = \begin{cases} 
1 & \text{if } c(1) \oplus c(2) \oplus c(3) \oplus c(5) = 0, \\
0 & \text{otherwise}.
\end{cases}
\]

(2.9)

A Hamming code where (2.9) is true is represented as a factor graph in Fig. 2.2. In the figure, the circles represent variable nodes, and the squares represent factor nodes. Each factor node represents one parity constraint that must be satisfied for \( c \) to be a valid codeword, such as that in (2.9). Hence all the code bits connected to a factor node must satisfy the even parity check.

When a code is represented as a factor graph, the sum-product algorithm (SPA), which is a message-passing algorithm, can be used to efficiently calculate the exact or approximate marginal probabilities of the transmitted code bits. If the factor graph is
cycle-free, the marginal probability obtained through SPA is exact. For a factor graph
that has cycles, an *approximation* to the marginal probability of the transmitted bits
can be obtained through iterations. The factor graph of most codes that perform well in
AWGN channels contain cycles, and the iterative SPA is described here.

The iterative SPA is initialized by setting the likelihood probability $p(y(n)|c(n))$ as
the channel message $\lambda_{ch}$ at each variable node representing $c(n)$, where the conditional
probability is calculated for each possible realization of $c(n)$. For example, if $c(n) \in \{0, 1\}$,
then $f(y(n)|c(n) = 0)$ and $f(y(n)|c(n) = 1)$ are calculated. If the associated variable
is not transmitted (e.g., parity bits that are punctured), then $f(y(n)|c(n))$ is set to
have equal probability over all possible values, as this is the *a priori* probability of
$c(n)$. Let $\lambda_{f\rightarrow v}$ represent the message from factor node $f$ to variable node $v$, and during
initialization, $\lambda_{f\rightarrow v} = 1/|\mathcal{V}|$, where $\mathcal{V}$ is the alphabet of $v$.

After initialization, the SPA involves iterating two steps: calculating the variable-to-
factor-node messages and calculating the factor-to-variable-node messages, as described
below:

1. Variable-to-factor node messages: After receiving the messages from all its neigh-
bours, each variable node calculates the message to send to its neighbours. The

Figure 2.2: Factor graph of a (7,4) Hamming code.
message from variable node \( v \) to factor node \( f \) is given by

\[
\lambda_{v \rightarrow f} = \lambda_{ch} \prod_{f' \in \mathcal{N}(v) \setminus f} \lambda_{f' \rightarrow v}, \tag{2.10}
\]

where \( \mathcal{N}(v) \) represent the set of neighbours of \( v \), or equivalently, factor nodes that are connected to \( v \). The product term includes all the neighbours except the factor node where \( v \) is sending these messages to. These messages are then normalized to sum to 1 to represent a probability function.

2. Factor-to-variable-node messages: After receiving the messages from all its neighbours, each factor node calculates the message to send to its neighbours. The message \( \lambda_{f \rightarrow v} \) from factor node \( f \) carries information pertaining to the factor node’s belief, or probability, associated with each possible realization of \( v \), where this belief is derived from the messages from its neighbours. The message from factor node \( f \) to variable node \( v \) is given by

\[
\lambda_{f \rightarrow v} = \sum_{v' \in \mathcal{N}(f) \setminus v} \prod_{v'} \lambda_{v' \rightarrow f} I_f(v, v'), \tag{2.11}
\]

where \( \mathcal{N}(f) \) is the set of neighbours of \( f \) and \( I_f(v, v') \) is the local function as illustrated in (2.9) such that if the constraint is satisfied, \( I_f(v, v') = 1 \), and \( I_f(v, v') = 0 \) otherwise. As with \( \lambda_{v \rightarrow f} \), these messages are normalized to 1.

The computation of messages \( \lambda_{v \rightarrow f} \) and \( \lambda_{f \rightarrow v} \) is illustrated in Fig. 2.3. The two steps described above are iterated until the maximum number of iterations (or some other stopping criterion) has been reached. Then the approximate APP of \( c(n) \) associated with \( v \) is given by

\[
f(c(n)|y) = \lambda_{ch} \prod_{f \in \mathcal{N}(v)} \lambda_{f \rightarrow v}. \tag{2.12}
\]

The estimated value of \( c(n) \) is given by

\[
c(n) = \arg \max_{\hat{c}(n)} f(c(n)|y). \tag{2.13}
\]
Figure 2.3: Messages passed (a) from variable node to factor node and (b) from factor node to variable node.

If $c \in \{0, 1\}^l$, then instead of passing probabilities associated with each possible value, messages passed between the various nodes can be the log-likelihood ratios (LLRs) to simplify the required calculations. The LLR is given by

$$L = \log \frac{f(y|0)}{f(y|1)},$$

where $f(y|0)$ and $f(y|1)$ are, respectively, the likelihood probability of received signal $y$ given the code bit is 0 or 1. To initiate the SPA, channel message is initiated for each variable nodes with

$$L_{ch} = \log \left( \frac{f(y(n)|c(n) = 0)}{f(y(n)|c(n) = 1)} \right). \quad (2.14)$$

After some manipulation [51], it can be shown that the variable-to-factor-node messages $L_{v\rightarrow f}$ and check-to-variable-node messages $L_{f\rightarrow v}$ in the LLR format are given by

$$L_{v\rightarrow f} = \sum_{f' \in \mathcal{N}(v) \setminus f} L_{f'\rightarrow v}, \quad (2.15)$$

$$L_{f\rightarrow v} = 2 \tanh^{-1} \left( \prod_{v' \in \mathcal{N}(f) \setminus v} \tanh \left( \frac{L_{v'\rightarrow f}}{2} \right) \right). \quad (2.16)$$

The calculations of $L_{v\rightarrow f}$ and $L_{f\rightarrow v}$ are performed iteratively until the maximum number of iteration is reached. The approximate LLR of the APP for $c$ that is associated with $v$
is formed from collecting the messages from all the neighbours

\[ L = L_{ch} + \sum_{f \in N(v)} L_{f \rightarrow v}. \]  (2.17)

If \( L \geq 0 \), then it is assumed that \( c = 0 \) was transmitted; otherwise it is assumed that \( c = 1 \) was transmitted.

In general, it is very difficult to know the achievable rate of a channel when SPA is used for decoding. Density evolution was introduced in [51], which can be used to determine whether the bit error rate (BER) goes to 0 for a binary code in a memoryless channel. With density evolution, the probability density function of the messages passed between the variable and factor nodes are tracked for each iteration. Assuming that an all-zero codeword is transmitted, the BER is equal to the probability of a negative LLR in (2.17). Hence, the BER goes to 0 if the probability of a negative LLR goes to 0. Density evolution is usually performed by averaging over all possible codes with the same check and variable degree distribution, but it was shown in [52] that the probability that the SNR threshold at which the BER goes to 0 found using density evolution is very close for almost any code with the same check and variable node degree distribution and large blocklength. Even though density evolution is an asymptotic concept that assumes a large blocklength and independence of the messages passed between the variable and factor nodes, it can be used to provide insight into the performance of a finite-length code with the same check and variable degree distribution.

### 2.4 Union-Bhattacharyya Bound

The union-Bhattacharyya bound (UBB) provides a convenient upper bound on the maximum likelihood (ML) FER [53], and can be applied to any linear code. Let \( C = \{ c_0, c_1, \ldots, c_{|C|-1} \} \) be a codebook, where \( c_i \) is the \( i \)th codeword. For simplicity, we assume the code is binary, i.e. \( c_i \in \{0,1\}^l \). We also assume each codeword is equally likely to be sent. Assume that the modulated signal \( \mathbf{x} = \xi(c_j) \) is transmitted over the
channel, and $y$ is the received signal. Let $f(y|c_i)$ be the likelihood function of $y$ given $c_i$ is transmitted, and let $Y_i = \{y : f(y|c_j) < f(y|c_i), i \neq j\}$, represent the set of received signals that would lead us to decode $c_i$ as the correct codeword under ML decoding given $c_j$ is sent. A frame error occurs whenever the received signal belongs to any of the set $Y_i$ for $i = 0, \ldots, |C| - 1$, $i \neq j$, and the FER $(P_f)$ is upper-bounded by the sum of all pairwise error probabilities

$$P_f \leq \sum_{i=0, i \neq j}^{|C|-1} \Pr(y \in Y_i). \tag{2.18}$$

It has been shown in [54] that if $c_i$ and $c_j$ differs by $h_i$ positions, then the pairwise error probability with ML decoding is given by

$$\Pr(y \in Y_i) = \frac{1}{2} \text{erfc}\left(\sqrt{h_i \gamma}\right), \tag{2.19}$$

where $\gamma$ is the channel SNR. Calculating this value, however, involves invoking the error function $\text{erfc}(\cdot)$ where the Hamming distance between the two codewords, $h_i$, is used as a parameter. Since it is difficult to characterize the error function as a function of $h_i$, this makes it difficult to calculate the upper bound in (2.18). Instead, as illustrated in [53], an upper bound on the pairwise error probability can be used:

$$\Pr(y \in Y_i) = \int_{y \in Y_i} f(y|c_j)dy$$

$$\leq \int_{y \in Y_i} f(y|c_j) \sqrt{f(y|c_i)/f(y|c_j)}dy$$

$$= \int_{y \in Y_i} \sqrt{f(y|c_j)f(y|c_i)}dy, \tag{2.20}$$

where the inequality results from the fact that for all $y \in Y_i$, $f(y|c_i) > f(y|c_j)$ by definition. Assuming identically and independently distributed (i.i.d.) memoryless channels, after some manipulation (2.20) becomes [54]

$$\Pr(y \in Y_i) \leq \prod_{j=1}^{h_i} \int_{y \in Y} \sqrt{f(y|0)f(y|1)}dy = \left(\int_{y \in Y} \sqrt{f(y|0)f(y|1)}dy\right)^{h_i}, \tag{2.21}$$
where $\mathcal{Y}$ is the alphabet of the output, and $f(y|0)$ and $f(y|1)$ are the probability of receiving $y$ given that 0 and 1 are sent respectively. Here, the upper bound on the pairwise probability of error can be characterized as a function of $h_i$, and is used to obtain an upper bound on the FER.

For all linear codes, the XOR of any two codewords is also a codeword. Since the pairwise error probability in (2.21) is only dependent on the weight of the difference between two codewords, i.e. the Hamming distance between two codewords, it can be assumed that the all-zero codeword is transmitted without loss of generality, i.e. $c_j = c_0 = 0$, and let $h_i$ be the weight of $c_i$.

The Bhattacharyya parameter (BP) associated with a channel is defined as

$$\beta \triangleq \int_{y \in \mathcal{Y}} \sqrt{f(y|0)f(y|1)} dy. \quad (2.22)$$

For example, for the binary symmetric channel (BSC) with bit-flip probability $p$, $\beta = 2\sqrt{p(1-p)}$, and for the AWGN channel with noise variance $\sigma^2$ and received SNR $\gamma = \frac{1}{2\sigma^2}$, $\beta = e^{-\gamma}$. The BP therefore provides an indication of the quality of the channel between the transmitted symbol, $x$, and the data used for processing, $y$.

Let $k$ be the length of the encoder input, and $l$ be the length of the encoder output. Then by summing over all pairwise probabilities of error, for a given codebook, the UBB, obtained by substituting (2.21) into (2.18), can be used to provide an upper bound on the FER:

$$P_f \leq \sum_{h=1}^{l} A_h \beta^h = \sum_{h=1}^{l} \left( \sum_{w=1}^{k} A_{w,h} \right) \beta^h. \quad (2.23)$$

where $A_{w,h}$ is the input-output weight enumerator (IOWE) representing the number of codewords with input weight $w$ and output weight $h$, and $A_h$ is the weight enumerator (WE) representing the number of codewords with weight $h$. For simplicity, in the above equation, it is assumed that the sequence of all-zero information bits fed into the encoder is mapped to the codeword $0$, i.e. $A_{0,0} = 1$. If this is not the case, then $A_{w,h}$ represents the number of codewords with input weight $w$, and $h$ being the Hamming
distance between the codeword and the codeword corresponding to the input sequence of all-zero information bits. If the channel code is systematic, $A_{w,h}$ represents the number of codewords with $w$ as the weight of the information bits, and $h$ as the weight of the codeword. This is useful when a distinction is made between codewords with the same output weight but different input weight, as illustrated in Ch. 5.

For codes such as turbo codes and repeat-accumulate (RA) [55] codes where interleavers are used, the IOWE and WE are different when the interleaver used is changed. In general it is very difficult to obtain the IOWE and WE for a particular interleaver. Instead, the IOWE and WE averaged over all possible interleavers can be used for analysis instead as they can be obtained more readily. Let $\bar{A}_{w,h}$ and $\bar{A}_h$ be the IOWE and WE averaged over all possible codebooks generated from using all possible interleavers. The derivation of average IOWE and WE for “turbo-like” codes, such as RA codes, can be found in [56].

Since deriving the IOWE and WE for finite blocklengths can be time-consuming, asymptotic results can be used instead for analysis when the blocklength $l$ is large. Let $\delta \triangleq h/l$, and

$$r(\delta) \triangleq \limsup_{l \to \infty} \frac{1}{l} \log \bar{A}_h.$$ 

Then from [57],

$$P_f \leq \sum_{\delta} \exp\{lr(\delta) + h \log \beta\} = \sum_{\delta} \exp\{l(r(\delta) + \delta \log \beta)\}.$$ 

For large $l$, the upper bound on $P_f$ goes to 0 if $(r(\delta) + \delta \log \beta) < 0$ for all $\delta$. Let the asymptotic UBB threshold be

$$c_0 \triangleq \max_{0 < \delta \leq 1} \frac{r(\delta)}{\delta}.$$ 

Then if $\log(\beta) < -c_0$, the upper bound on $P_f$ decreases at an exponential rate with respect to $l$, and $\exp\{-c_0\}$ is an upper bound on the BP, $\beta$, associated with the channel such that if $\beta < \exp\{-c_0\}$ then for large blocklengths $l$ the FER goes to 0.
Importantly, the UBB for the FER can be extended to scenarios where the codeword is sent through parallel channels, each with a different BP [58]. Let the number of parallel channels be $J$. Assume that the codeword has $l$ bits and each bit is transmitted through channel $j$ with probability $\alpha_j$, where $\sum_j \alpha_j = 1$. By averaging over all possible bit assignments to all the channels, the authors in [58] derived the UBB for the parallel channels as

$$P_f \leq \sum_{h=1}^{n} \bar{A}_h \bar{\beta}^h,$$  \hspace{1cm} (2.24)

where

$$\bar{\beta} = \sum_{j=1}^{J} \alpha_j \beta_j,$$  \hspace{1cm} (2.25)

and $\beta_j$ is the BP associated with the $j$th channel. As with single channel transmissions, if

$$\bar{\beta} < \exp\{-c_0\}$$  \hspace{1cm} (2.26)

then for large blocklength the FER goes to 0. This concept has also been extended to analyze incremental redundancy in [46], a cooperative scheme used in relay channels. More details on the UBB can be found in [53,54].
Chapter 3

Demodulate-and-Forward

In some networks, such as sensor networks, nodes only possess simple hardware, and the well-studied relaying schemes DF and AF may not be practical. Even though DF can provide good performance when the source-relay channel is good, the decoding process may be too complicated for the nodes. AF may not be suitable in sensor networks because full-duplex channels are required to implement AF. If only half-duplex channels are available, the relaying nodes must store the received signal before transmitting. This requires the relays to possess large storage. Furthermore, a wide dynamic range on the transmitter radio frequency components is needed, which again, may not be available.

In this chapter, demodulate-encode-forward (DEF) is introduced as a practical relaying scheme in such a scenario. Clearly every digital communication device is able to demodulate the received signals. Only one bit is required to represent each demodulated bit and demodulation is extremely simple compared to the decoding required with DF. In most WSNs, it is assumed that the destination D has hardware to support more complex calculations, and there is no limitation on the battery power. In this chapter we assume that the decoding complexity can be transferred to the destination. Here, DEF is introduced and described in detail to illustrate the simplicity of the scheme and its performance is tested in fading channels.
3.1 Encoding

Instead of only repeating the demodulated bits from the source’s codeword as in [59], channel codes can be used at the relay to improve the performance of the system. In this section, DEF with channel codes is discussed in detail. Even though channel codes illustrated here only include simple parity, low-density generator matrix (LDGM) and RA codes, the implementation is not limited to these codes. LDGM and RA codes are chosen because the encoding of these codes is extremely simple, reducing the complexity required at the relaying nodes. In addition, they have excellent performance in AWGN channels. Finally, the code rate can be changed very easily, making these relaying schemes very flexible.

DEF involves two steps at the relay: demodulation and encoding. Let $y_{S,R_j}$ be the signal received by $R_j$, and let

$$\tilde{y}_{S,R_j} = h^*_{S,R_j} y_{S,R_j}.$$  

The demodulated bits at $R_j$, $\hat{c}_{S,R_j}$ are obtained from

$$\xi^{-1}(\text{sgn}(\tilde{y}_{S,R_j})),$$

where $\text{sgn}(x) = 1$ if $x \geq 0$ and $\text{sgn}(x) = -1$ if $x < 0$, and $\xi^{-1}$ is the reverse mapping function. These are often referred to as hard-decisions in coding theory.

DEF is extremely flexible in that the relays do not necessarily have to forward all the source bits. Instead, a relay can choose independently the bits that it wants to relay. Let $\epsilon_{info,j}$ and $\epsilon_{par,j}$ be the fraction of source’s information and parity bits relayed by $R_j$ respectively. Then $R_j$ can choose, either randomly or dependent on the received SNR, $\epsilon_{info,j}$ of the source’s information bits and $\epsilon_{par,j}$ of the source’s parity bits to form a new codeword to transmit to D. Various channel codes can be used at the source and relays, and the details are given below.
3.1.1 DEF with Mod-2 Parity

Instead of repeating the demodulated symbols, R can encode the demodulated bits. Our initial proposal of DEF [60] suggested a very simple encoding scheme. The source transmits uncoded bits and the relay forms parity bits based on the demodulated bits. In this case, R performs mod-2 sums (i.e., XOR) on consecutive demodulated bits to form parity bits, which are then transmitted to D. A factor graph, as described in Sec. 2.3, can be used to graphically illustrate the relationship between various bits in the received codeword. The factor graph of DEF with mod-2 parity is illustrated in Fig. 3.1. The variable nodes at the top half of the figure represent the source’s codeword and the variable nodes on the bottom represent the relay’s codeword, where $v_{s,i}$ represent the source’s information bits. Similar to the source’s codeword, $v_{r,i}$ and $v_{r,p}$ represent the relay’s information (demodulated source’s bits) and parity bits respectively, and only the relay’s parity bits are transmitted. The factor nodes connecting $v_{s,i}$ and $v_{r,i}$ represents the relationship between the transmitted source’s codeword bits and the demodulated bits at the relay. One drawback of this relaying scheme is that the amount of data relayed is fixed, and no encoding can be performed on the relayed codeword to improve the performance on the R-D link. In the next two sections, DEF with more complex channel codes is introduced. Simulation results for DEF are presented in Sec. 3.3.

3.1.2 DEF with LDGM Codes

The LDGM code was studied in depth in [61,62]. It can be implemented as a rateless code, where the parity bits are formed from the mod-2 sums of randomly chosen information bits. For DEF with LDGM [63], a LDGM encoder is used at the source and relay to encode the relayed bits. At the source node, the information bits fed into the encoder are bits that we wish to communicate to D, whereas at the relay nodes, the information bits fed into the encoder are demodulated bits of the source’s codeword. The encoding of
Figure 3.1: Factor graph of DEF with simple parities formed at the relay.

A LDGM codeword is detailed here. Let \( q \) be the number of information bits chosen to form a parity bit, and let \( \Pr(q) \) be the probability that a parity bit is formed from the XOR of \( q \) information bits. The encoding involves the following steps:

1. Choose \( q \) according to the distribution \( \Pr(q) \).
2. Randomly choose \( q \) bits from the demodulated bits, and
3. Perform mod-2 sum of the \( q \) bits to form a parity bit.

The above steps are repeated until the required number of parity bits is formed. After this is completed, the information bits are concatenated with the parity bits to form a systematic LDGM codeword. Usually the number of information bits chosen to form the parity is small compared to the blocklength, hence the name given to the code. Also, when the check degree is small, the messages passed from check nodes to variable nodes contain more information. For example, for a binary code, the LLR message passed from a check node to a variable node is dominated by the minimum of all the LLR messages received from the neighbouring variable nodes. The larger the magnitude of the LLR message, the more certain a factor node is regarding the value of a given variable. When the check node degree is large, the possibility that at least one of these variable messages is small becomes higher, and less information can gleaned from the message. Hence it
is desirable to use a small check node degree. However, it is also important to use a check node degree large enough such that protection is provided for large number of information bits through the use of parity bits. If the check node degree is too small, then there may exist a large number of information bits with no parity bits associated with them. Throughout the dissertation the number of information bits XORed to form each parity bit is arbitrarily chosen to be 11 for illustration; other values can be used as well.

A factor graph of the source’s and relay’s codewords associated with received signal at D is illustrated in Fig. 3.2. As with the previous factor graph, the variable nodes in the top half of the figure represent the source’s codeword, while the variable nodes in the bottom half represent the relay’s codeword. Here $v_{s,i}$ and $v_{s,p}$ represent the source’s information and parity bits, and $v_{r,i}$ and $v_{r,p}$ represent the relay’s information and parity bits. As illustrated in the figure, random information bits are chosen to generate parity bits through mod-2 sums. If a bit in the source’s codeword is relayed, there is a link between the source’s code bit and the corresponding relay’s code bit. The demodulated source’s codeword bits chosen to be relayed are used as information bits in the relay’s codeword. The parity bits, again, are formed from the mod-2 sum of randomly selected information bits.

LDGM codes have the advantage that the encoding procedure is extremely simple, requiring only simple hardware. In the context of relay channels, only the relay’s codeword needs to be stored, reducing the amount of storage required. One drawback of LDGM code is that, due to the fact that the variable nodes representing the parity bits only have one edge, minimal protection is provided for the parity bits, resulting in an error floor \cite{error_floor_ref}. 
3.1.3 DEF with RA Code

Instead of LDGM encoders, RA encoders can be used at the source and relays, and the use of DEF with RA was introduced by Professors Andrew Eckford\(^1\) and Raviraj Adve in [1]. Since this is used extensively in the simulations throughout this dissertation, details on this coding scheme is presented here. The author of this dissertation is responsible for generating the simulation results presented here.

Unlike with LDGM codes, where additional parity bits is generated when needed, a pre-determined number of parity bits are generated and the required code rate is obtained by puncturing the parity bits. Similar to DEF with LDGM codes, the information bits fed into the source encoder are bits that we wish to communicate to D, whereas the information bits fed into the relay encoders are demodulated bits of the source’s codeword. The encoding of a punctured systematic RA code involves the following steps:

1. Repeat the information bits \(q\) times,

\(^{1}\text{Department of Computer Science and Engineering, York University, 4700 Keele Street, Toronto, Ontario, Canada M3J 1P3.}\)
2. Interleave the repeated information bits,

3. Feed the interleaved bits $b_{in}$ into a rate-1 convolutional encoder with transfer function $1/(1 + D)$, where the output is given by

$$b_{out}(n) = \begin{cases} b_{in}(n) & \text{for } n = 1, \\ b_{in}(n) \oplus b_{out}(n-1) & \text{for } n > 1, \end{cases}$$

4. Concatenate the information bits with output from rate-1 encoder, and

5. Puncture the parity bits to obtained the codeword with desired code rate.

A factor graph for DEF with punctured systematic RA codes is illustrated in Fig. 3.3, where $\Pi$ represents random interleaving. Again, the variable nodes at the top of the figure represent the source’s codeword, while the variable nodes at the bottom of the figure represent the relay’s codeword. The grey parity bits are bits that are eliminated during the puncturing process. As illustrated in the figure, the information bits are repeated $q$ times, interleaved, and fed into the rate-1 convolutional encoder. As usual, the factor nodes between the source’s and relay’s variable nodes describe the relationship between the source’s and relay’s code bits.
As is clear from the procedure described above, the encoding is extremely simple, requiring only simple hardware. The puncturing can be performed during the encoding process, reducing the storage needed before transmission. Similar to LDGM codes, in order to guarantee good performance, the information bits must be included in the code-word when the parity bits are punctured [57]. For the rest of the thesis, the repetition parameter $q$ is set to 3, such that the code rate ranges from $1/4$ to 1.

In summary, the encoding process for DEF is fairly simple. The source transmits a codeword. The relay demodulates its received signal and uses these demodulated bits to form a new codeword. In keeping with the simplicity theme, here we suggest mod-2 parity, LDGM or RA codes at the relay.

### 3.2 Decoding

By representing DEF with channel codes as factor graphs, the SPA can be used for decoding in an iterative manner. In each iteration step, the messages from variable nodes to factor nodes and from factor nodes to variable nodes are calculated. The decoding at D can be performed in two different ways: parallel or serial decoding. In parallel decoding, the source’s and relay’s codewords are decoded in parallel at D. After each iteration of decoding the individual codewords, messages, or soft information obtained from decoding, are exchanged between the source’s and relay’s variable nodes. For serial decoding, the relay’s codeword is decoded, and the messages obtained after the maximum number of iterations is reached are passed from the relay’s variable nodes to the source’s variable nodes. Serial decoding is simpler than parallel decoding, as fewer messages are passed between the variable and factor nodes, but parallel decoding provides better performance in general. The messages exchange between the source’s and relay’s variable nodes are detailed below.
3.2.1 Relay-to-Source Messages

When serial or parallel decoding is used, messages are passed from the relay’s variable nodes to the source’s variable nodes. The messages passed to the source’s variable nodes account for the S-R channel quality through the bit-flip probability over the S-R channel, $p_{S,R_j}$, which is given by

$$p_{S,R_j} = \frac{1}{2} \text{erfc}(\sqrt{\gamma_{S,R_j}}), \quad (3.1)$$

where $\text{erfc}(\cdot)$ is the complementary error function. Discussion on how this information can be obtained by D can be found in Sec. 3.4. Let the $n$th bit of the decoded source’s codeword $\hat{c}_{S,R_j}$ be the $t_{j,n}$th bit of the relay’s codeword at $R_j$, $c_{R_j}$. Then

$$\Pr(c_S(n)|c_{R_j}(t_{j,n})) = \begin{cases} 1 - p_{S,R_j} & \text{for } c_S(n) = c_{R_j}(t_{j,n}), \\ p_{S,R_j} & \text{for } c_S(n) \neq c_{R_j}(t_{j,n}). \end{cases} \quad (3.2)$$

This equation describes the relationship between the source’s code bit and its corresponding relayed value, which may not be the same because of potential demodulation errors.

Let $M_{R_j}(t_{j,n})$ be the summed message at the relay’s variable node $c_{R_j}(t_{j,n})$, and $L_{S,j}(n)$ be the message that is passed to the $n$th bit of the source’s codeword. This is illustrated in Fig. 3.4. From (3.2) the relayed bit is different than its corresponding source’s bit with probability $p_{S,R_j}$. Hence, using Bayes’ rule

$$\Pr(c_S(n)) = (1 - p_{S,R_j})\Pr(c_{R_j}(t_{j,n}) = c_S(n)) + p_{S,R_j}\Pr(c_{R_j}(t_{j,n}) \neq c_S(n)). \quad (3.3)$$

The message passed to the $n$th bit of the source’s codeword from the $t_{j,n}$th bit of $c_{R_j}$ is given by

$$L_{S,j}(n) = \log \left[ \frac{(1 - p_{S,R_j})\Pr(c_{R_j}(t_{j,n}) = 0) + p_{S,R_j}\Pr(c_{R_j}(t_{j,n}) = 1)}{p_{S,R_j}\Pr(c_{R_j}(t_{j,n}) = 0) + (1 - p_{S,R_j})\Pr(c_{R_j}(t_{j,n}) = 1)} \right]$$

$$= \log \left[ \frac{(1 - p_{S,R_j})\frac{\Pr(c_{R_j}(t_{j,n}) = 0)}{\Pr(c_{R_j}(t_{j,n}) = 1)} + p_{S,R_j}\frac{\Pr(c_{R_j}(t_{j,n}) = 1)}{\Pr(c_{R_j}(t_{j,n}) = 0)}}{p_{S,R_j}\frac{\Pr(c_{R_j}(t_{j,n}) = 0)}{\Pr(c_{R_j}(t_{j,n}) = 1)} + (1 - p_{S,R_j})\frac{\Pr(c_{R_j}(t_{j,n}) = 1)}{\Pr(c_{R_j}(t_{j,n}) = 0)}} \right]$$

$$= \log \left[ \frac{(1 - p_{S,R_j}) \exp \{M_{R,j}(t_{j,n})/2\} + p_{S,R_j} \exp \{-M_{R,j}(t_{j,n})/2\}}{p_{S,R_j} \exp \{M_{R,j}(t_{j,n})/2\} + (1 - p_{S,R_j}) \exp \{-M_{R,j}(t_{j,n})/2\}} \right]. \quad (3.4)$$
With some simple manipulation, it can be shown that

$$L_{S,j}(n) = 2 \tanh^{-1} \left[ \tanh \left( \frac{1}{2} M_{R,j}(t_{j,n}) \right) \tanh \left( \frac{1}{2} \log \left( \frac{1 - p_{S,R_j}}{p_{S,R_j}} \right) \right) \right].$$  \tag{3.5}

Comparing this equation to (2.16), we can see if the relationship between the \(n\)th source’s bit and \(t_{j,n}\)th relay’s bit is represented by a factor graph, we can connect the two variable nodes with a factor node. An auxiliary variable node, \(u_j(n)\), which has a value of 0 and 1 with probability \(1 - p_{S,R_j}\) and \(p_{S,R_j}\) respectively, is introduced. Then \(c_{R_j}(t_{j,n}) = c_S(n) \oplus u_j(n)\).

Note that to form the message \(L_{S,j}(n)\) the destination needs to know the S-R\(_j\) channel quality. If the S-R channel quality is poor, \(p_{S,R_j}\) is close to 0.5 and (3.5) goes to 0. Hence less importance is given to the message if the S-R channel quality is poor. This reduces the effects of error propagation, where incorrect data forwarded by the relay worsens the performance of the system, and can be seen with DF when incorrect data is forwarded by the relays. With \(r\) relays are present, the messages going into the source’s variable nodes from all the relays’ variable nodes are given by

$$L_{S}(n) = \sum_{j=1}^{r} L_{S,j}(n),$$  \tag{3.6}

where \(L_{S,j}(n) = 0\) if \(R_j\) does not transmit bit \(n\) of the source’s codeword.

### 3.2.2 Source-to-Relay Messages

If parallel decoding is used, the messages passed from the source’s variable nodes to the relay’s variable nodes must be calculated. These messages are very similar to that described above. In fact, when there is only one relay, the messages can be obtained by replacing the sum message for the relay’s variable node with that for the source’s variable node in (3.5). In the case with multiple relays, the messages passed from the source’s variable nodes to the relays’ variable nodes are slightly different as messages collected from other relays are also used to calculate the messages to be passed to each relay node.
If $R_j$ is relaying bit $n$ of the source’s codeword as bit $t_{j,n}$, the bit of the relay’s codeword, the message passed from $S$ to $R_j$ is given by

$$L_{R_j}(t_{j,n}) = \log \left( \frac{1 - p_{S,R_j}}{p_{S,R_j}} \exp\{M_{S,j}(n)/2\} + p_{S,R_j} \exp\{-M_{S,j}(n)/2\} \right)$$

(3.7)

$$= 2 \tanh^{-1} \left[ \tanh \left( \frac{1}{2} M_{S,j}(n) \right) \tanh \left( \frac{1}{2} \log \left( \frac{1 - p_{S,R_j}}{p_{S,R_j}} \right) \right) \right]$$

(3.8)

where $M_{S,j}(n) = \sum_{j' \neq j} L_{S,j'}(n)$. Note that the decoding process, as described here, assumes that the destination knows which bits are relayed by the individual relays. This can be implemented by, for example, the destination sharing a seed with the relays in a pseudo-random number generator.

### 3.3 Simulation Results

In this section, we present results of simulations used to test the DEF scheme. For AWGN channels, it is assumed that the instantaneous SNR is the same for all channels, and for fading channels, it is assumed that the average SNR over all channels is the same. For simplicity, it is assumed that the transmitted energy for each symbol is the same, and the SNR is normalized such that the amount of energy used for each source’s information
bit, or $E_b/N_0$, is equal. It is assumed that the number of source’s information bits is $k_S = 2000$. In all the simulation results presented here, since only one relay is used, the subscript used to refer to different relays is removed.

### 3.3.1 DEF with Mod-2 Parity

Simulation results for DEF with the simple mod-2 parity of Sec. 3.1.1 in AWGN channels are given in Fig. 3.5, where the BER and FER for various blocklengths are shown. As mentioned earlier, the amount of data relayed is fixed, and $l_R = l_S$. The BER for various blocklengths ($l_S$) is almost identical, whereas the FER is best when $l_S$ is small. The error performance curves are quite flat, and the performance can be improved significantly with the use of slightly more sophisticated channel codes, as illustrated in the following sections.

Simulation results for DEF with simple parity in fading channels are illustrated in Fig. 3.6. Similar to the error performance in AWGN channels, the BER for various blocklengths is almost identical, and the FER is smallest when the blocklength is small. Note that even with such a simple encoding scheme, a diversity order of 2 can be achieved with the use of one relay.

### 3.3.2 DEF with LDGM Codes

Simulation results for DEF with LDGM codes in AWGN channels are given in Fig. 3.7. As suggested earlier, DEF with LDGM codes is very flexible; the amount of data relayed can be changed very easily. The performance of DEF with LDGM codes with different $\epsilon_{par}$, fraction of source’s codeword parity bits relayed, and different source code rates $R_S$ is illustrated in the plot. The source’s information bits are always relayed, i.e. $\epsilon_{info} = 1$. In the figure, $R_S$ ranges from $1/4$ to $3/4$, and $\epsilon_{par} = 0$ represents the case where none of the parity bits in the source’s codeword are relayed, while $\epsilon_{par} = 1$ represents the case where all the parity bits are relayed. In all cases, the relay code rate is fixed at $R_R = 1/2$. 
Figure 3.5: BER (dashed lines) and FER (solid lines) of DEF with mod-2 parity in AWGN channels.

Figure 3.6: BER (dashed lines) and FER (solid lines) of DEF with mod-2 parity in fading channels.
For the SNR range shown, the FER is lower when $\epsilon_{\text{par}} = 0$. One possible reason is that when the transmission power is fixed, with higher $\epsilon_{\text{par}}$, more energy is consumed by the relay, whereas with $\epsilon_{\text{par}} = 0$, the energy expended at the relay is reduced. As all the channels have the same SNR, the S-R-D link is not as reliable as the S-D link, and more energy should be allocated to the source. For $R_S \leq 1/2$, an error floor is present for the range of FER shown.

Simulation results for DEF with LDGM codes in fading channels are illustrated in Fig. 3.8. Similar to the previous plot, $R_S$ ranges from $3/4$ to $1/4$ for both $\epsilon_{\text{par}} = 0$ and $\epsilon_{\text{par}} = 1$. In the low SNR regime, $R_S = 1/4$ with $\epsilon_{\text{par}} = 0$ has the best performance, while at high SNR, $R_S = 1/2$ with $\epsilon_{\text{par}} = 1$ has the best performance. At low SNR, it is more beneficial to use a lower code rate at S and allocate more transmission energy to S, while at high SNR, it is more important to allocate more transmission energy to R. One possible explanation is that at high average SNR, the FER is dominated by occurrences of a poor S-D link, and in those cases it is especially beneficial to have path diversity. However, the difference between the two schemes with different $R_S$ and $\epsilon_{\text{par}}$ parameters is very small. In all cases, a diversity order of 2 is obtained.

### 3.3.3 DEF with RA Codes

Simulation results for DEF with punctured systematic RA codes in AWGN channels are illustrated in Fig. 3.9. Similar to previous plots, the simulation results for $R_S$ ranging from $3/4$ to $1/4$ are shown for both $\epsilon_{\text{par}} = 0$ and $\epsilon_{\text{par}} = 1$. The error floors that were visible in Fig. 3.7 are not visible here as RA codes do not suffer from high error floors. In addition, the steep decrease in FER occurs at a lower SNR compared to DEF with LDGM codes, as RA is a better code in AWGN channels. This comes at a cost of a slightly more complex encoding procedure. In all cases, for a fixed $R_S$, the relaying scheme with $\epsilon_{\text{par}} = 0$ performs better than that with $\epsilon_{\text{par}} = 1$, and the relaying scheme with $\epsilon_{\text{par}} = 0$ and $R_S = 1/4$ has the best performance.
Figure 3.7: FER of DEF with LDGM code in AWGN channels, where $\epsilon_{\text{par}} = 0$ (solid lines) or $\epsilon_{\text{par}} = 1$ (dashed lines).

Figure 3.8: FER of DEF with LDGM code in fading channels, where $\epsilon_{\text{par}} = 0$ (solid lines) or $\epsilon_{\text{par}} = 1$ (dashed lines).
Simulation results for DEF with punctured systematic RA codes in fading channels are illustrated in Fig. 3.10. Again, the simulation results for $R_S$ ranging from $3/4$ to $1/4$ are shown for both $\epsilon_{par} = 0$ and $\epsilon_{par} = 1$. At low SNR, the relaying scheme with $R_S = 1/4$ and $\epsilon_{par} = 0$ has the best performance, while at high SNR, the relaying scheme with $R_S = 1/4$ and $\epsilon_{par} = 1$ has the best performance. Similar to DEF with LDGM codes, it is more beneficial to allocate more transmission energy to S at low SNR where the relay does not provide much help in any case, while at high SNR, it is more useful to allocate more transmission energy to R. Again, a diversity order of 2 is obtained.

Simulation results of DEF with RA, AF and various DF relaying schemes are presented in Fig. 3.11 for comparison. Simulation results for no relaying are also shown for comparison. When no relaying is used, we allow S to use either a rate-1/2 or a rate-1/4 RA code. As mentioned earlier, when DF is used, R does not transmit if the decoded source’s codeword contains errors. In this case, we have presented results where S either remains quiet (S silent) in the event of a decoding failure, or takes over the responsibility of transmitting in the second phase (S sends). The received signals from both S and R are combined using MRC before being fed into the decoder. Finally, for AF, a rate-1/2 RA is used by S, and R merely amplifies the signal and transmits it in the second phase. The received signals from S and R are combined using (2.7) before being fed to the decoder. The average SNR over all channels are assumed to be equal, and the energy used per source’s information bit has been normalized to provide a fair comparison. When relays are not used, the FER has a diversity order of 1. For the SNR range shown in the figure, DF with S transmitting when decoding fails has the best performance. The gap between DF with S transmitting and AF decreases as the average SNR between the channels increases. The gap between DF with S transmitting and DEF with RA code is less than 3 dB for at high SNR. DF and AF provides better error rate in the scenario shown here, but DEF with RA code requires simpler hardware at the relays. There is a tradeoff between performance and complexity.
Figure 3.9: FER of DEF with RA code in AWGN channel, where $\epsilon_{\text{par}} = 0$ (*solid lines*) or $\epsilon_{\text{par}} = 1$ (*dashed lines*).

Figure 3.10: FER of DEF with RA code in fading channel, where $\epsilon_{\text{par}} = 0$ (*solid lines*) or $\epsilon_{\text{par}} = 1$ (*dashed lines*).
Figure 3.11: Comparison of no relaying, DF, AF and DEF with RA code in fading channels.

3.4 Discussion

As suggested in Section 3.2, the S-R channel quality must be available at D in order to perform decoding. This can be implemented by requiring R_j to estimate the S-R_j SNR using the received signal from S, and either the SNR or equivalently, the bit-flip probability \( p_{S,R_j} \), is transmitted to D. Assuming a slow-fading channel, this data can be found for each block of data received from S. The effect of inaccurate estimation of the received SNR at the relays, and the effect of erroneous transmission of the SNR or bit-flip probability from the R_j to D, is outside the scope of this thesis.

The simulation results presented in the previous section show that DEF, a simple, flexible relaying scheme, can be used to achieve cooperation with complexity-constrained relays, even though no decoding is performed at the relays, and the relays only perform the simple task of demodulation. Furthermore, in all cases shown here, a diversity order of 2 in fading channels is achieved with the use of 1 relay. This is possible even though
the burden of more complex calculations have been shifted from the relays to D. Hence this relaying scheme is most suitable for networks where relays have limited computer power and/or limited battery power, and D does not have any energy constraints, and more complex hardware is available. This result, in conjunction with the performance in AWGN channels, suggests that DEF is a suitable relaying scheme that can provide good error performance with low-complexity hardware.

In order to further increase the flexibility of DEF, limited decoding can be used in conjunction with DEF. In the case where nodes possess slightly more complex hardware, in addition to simply demodulating the received bits, relays could also perform simple decoding [65], which is denoted as limited decoding here. Limited decoding allows either a limited number of iterations of the SPA [50], or simpler decoding algorithm, such as the Gallager A algorithm [66], to be performed at the relay. The partially decoded sequence, which may still contain errors, is re-encoded with another error-correcting code and retransmitted. If limited decoding is available, the effective S-R channel is improved, as compared to the case of pure DEF. The relationship between the original S-R channel BER and the effective S-R channel BER after decoding can be obtained using density evolution. As shown in [51, 65], the BER is a non-increasing function of the number of iterations. Hence, limited decoding does not increase the effective bit-flip probability over the S-R channel. The effective bit-flip probability after decoding, instead of the bit-flip probability associated with the SNR over the S-R channel, is transmitted to D to be used in the decoding process. Note that the limited decoding suggested here is different from that suggested in [47, 48]. In our implementation, the quality of the S-R channel is not embedded in the encoded data. Here the S-R channel quality is transmitted separately from the encoded data, and is embedded in, for example, the header of a transmitted block. Also, in our implementation, the decoded data is quantized before transmission, whereas in [47, 48], the range of the transmitted signal is continuous.

Since demodulation is a form of compression of the received signal, the relaying scheme
DEF introduced here is a form of compress-and-forward (CF). It is worth mentioning that there are several papers that address implementations of CF. The simplest implementation is based on a repetition code and was suggested in [59], where the received symbols are demodulated symbol-by-symbol at R, and the demodulated symbols are forwarded to D. Maximum likelihood (ML) decoding is used at D, where the bit-flip probability over the S-R channel is used to calculate the ML probability. The BER for the repetition-code based DEF using binary frequency shift keying (BFSK) was presented and analyzed in [67]. In the paper, only channel statistics is available at the destination node, and it was shown that with the use of \( r \) relays, the diversity order is between \( r/2 \) and \( (r+1)/2 \), approximately half of the number of relays. Instead of using ML decoding, weighted combining is performed at D to simplify the decoding process in [68], and it was shown that full diversity, where the diversity order is \( r + 1 \) with \( r \) relays, can be achieved. In [69], instead of transmitting individual symbols, channel codes are used at S, and the estimated symbols at the relay are XORed with estimated symbols from the previous block to obtain encoded symbols to transmit to D, where the error performance is presented. Note that even though the relaying scheme DEF with mod-2 parity developed independently in [60] and described here is quite similar to that presented in [69], the DEF schemes are developed further to include use with LDGM and RA codes to provide flexibility that can be adjusted depending on the channel conditions, and is unavailable with the work in [69]. In [8], the authors studied the achievable rate of CF with erasure. This is an implementation of CF from [6], where binary input is used at S. Some of the quantized bits are discarded, or “erased”, in order to accommodate the required code rate in order to guarantee correct decoding of the relay’s codeword at D. The framework of DEF is similar to the relaying scheme studied in [8], where the implementation of such schemes is detailed and their associated performance is presented here.

No diversity-multiplexing tradeoff results or achievable rate analysis on DEF is shown here, since it is a fixed rate code and no optimization on the coderate is performed. Only
the implementation of this simple relaying scheme is shown. Further studies is required to gather information on the diversity-multiplexing tradeoff for binary-input CF relaying schemes, and derivations on the achievable rate of binary-input CF relaying scheme can be found in [8].
Chapter 4

Fractional Cooperation

In the previous chapter we described our first contribution in designing cooperative diversity for complexity-constrained networks. In this chapter we describe our second contribution, fractional cooperation. In most of the existing literature, it is assumed that each relay either transmits all of the source’s data, or does not transmit at all. If using fractional cooperation, a relay can relay only a fraction of the source’s codeword, instead of devoting all its resources to relaying. Fractional cooperation is designed for networks where nodes have limited resources, such as transmission bandwidth and power, e.g., WSNs with battery-operated nodes. Each relay only has limited battery power, and the amount of energy it is willing to sacrifice to assist other nodes may be limited such that only a fraction of the source’s codeword can be relayed, and multiple relays, together, transmit to improve the reliability. Another example is mesh networks wherein each node takes both roles of source and relay. While a node wants to help other nodes as much as possible, it also has its own data. In this case, it may only use part of its resources to relay, while reserving the rest for its own data. This is similar to the study of relay channels with private messages in [70], where it is assumed that S has information intended for only R, and R has its own information to transmit to D, and the achievable rates of such relay channels were found.
The concept of each relay only transmitting a fraction of the source’s codeword was also examined in [46], where a cluster of nodes divide up the source’s codeword, and each relay is only responsible for a fraction of it. The concept of fractional cooperation with multiple relays introduced here, however, is different than that from [46]. Unlike the relaying scheme in [46], fractional cooperation does not require the fractions relayed by each node not overlap. This eliminates the need for the nodes to coordinate ahead of time to decide the fraction each relay is responsible for transmitting. Since no coordination is required with fractional cooperation, it is possible that some of the bits in the source’s codeword are relayed by all the relays, while there may exist bits in the source’s codeword that are not relayed at all. Note that fractional cooperation is different from the relaying scheme using hybrid-ARQ as described in [71] as with fractional cooperation, the puncturing, or erasure, occurs before the encoding, whereas with the hybrid-ARQ relaying scheme, the puncturing occurs after the encoding.

In this chapter, we develop and analyze fractional cooperation with one and multiple relays. First we derive the conditions that must be satisfied to achieve a diversity order of two with the use of one relay. Then we present the results on diversity order and outage sets for fractional cooperation with multiple relays from [2,3]. Note that even though the concept and the flexibility of this scheme is illustrated through simulation results using DEF, the results shown here are not limited to DEF, but can also be applied to other relaying schemes such as DF and AF.

4.1 Fractional Cooperation with One Relay

For the 3-node (one-relay) network, it is assumed that the relay node has its own data to transmit, and can only spare limited transmission resources to assist the source node. As stated in Section 2.1, orthogonal channels are used to facilitate communication between nodes. In the first phase, S broadcasts its transmissions, which are received by R and
D. In the second phase, the chosen bits from the source’s codeword are transmitted by R and received by D. Let $\epsilon_{info}$ and $\epsilon_{par}$ represent the fraction of source’s information and parity bits that are relayed. Hence, if $k_S$ denotes the number of information bits in the source’s codeword $c_S$, and $l_S$ is the length of the codeword, then the number of total bits relayed are $\epsilon_{info}k_S + \epsilon_{par}(l_S - k_S)$. Let $k_R$ denote the information bits in the relay codeword $c_R$, and $l_R$ be the length of $c_R$. In addition, $m_R$ denotes the number of information bits originating from the relay. Hence $k_R = m_R + [\epsilon_{info}k_S + \epsilon_{par}(l_S - k_S)]$ must be satisfied. Finally, let $R_S = \frac{k_S}{l_S}$ and $R_R = \frac{k_R}{l_R}$ denote the code rate of the source’s and relay’s codewords respectively.

### 4.1.1 Analysis Results

In this section, the conditions that must be met in order to allow the source to achieve diversity order of 2 in the frame error rate (FER) with the help of one relay is presented. In addition, it is shown that while assisting the source, the relay cannot achieve a diversity order of 2 in the FER for its own data.

**Theorem 4.1.1** In a one-relay network, let $R_S$ be the rate of the source’s codeword, and $\epsilon_{info}$ and $\epsilon_{par}$ be the fraction of the source’s information and parity bits included in the relay codeword. Then

$$R_S \leq \frac{\epsilon_{par}}{1 - \epsilon_{info} + \epsilon_{par}} \tag{4.1}$$

must be satisfied to obtain a FER with diversity order of 2 for the source’s codeword.

**Proof:** Recall that $k_S$ is the number of source’s information bits, and $l_S - k_S$ is the number of parity bits. Then since $R_S = \frac{k_S}{l_S}$, the number of source’s bits that are included in the relay codeword is given by

$$\epsilon_{info}k_S + \epsilon_{par}(l_S - k_S) = (\epsilon_{info}R_S + \epsilon_{par} - \epsilon_{par}R_S)l_S. \tag{4.2}$$

Because only some of the source’s symbols are selected for relaying, the S-R-D channel as seen from the source’s codeword point-of-view is an erasure channel, with the erasure
probability given by
\[ p_E = 1 - [(\epsilon_{info} - \epsilon_{par}) R_S + \epsilon_{par}], \] (4.3)
which is independent of the channel conditions. Given an erasure channel with erasure probability \( p_E \), the capacity is \( 1 - p_E \) [43]. This means that the code rate, \( R_S \), must be less than or equal to \( 1 - p_E \) for successful decoding. Thus, the necessary condition for successful decoding on the relay link translates to
\[ R_S \leq 1 - p_E \]
\[ = (\epsilon_{info} - \epsilon_{par}) R_S + \epsilon_{par}, \] (4.4)
and the condition in (4.1) follows from minor manipulation of this expression.

In order to achieve diversity order of 2, the decoder must have the ability to decode the codeword in the event of an outage over the S-D channel, or equivalently, when the instantaneous received symbol SNR of the S-D channel, \( \gamma_{SD} \), is 0 (otherwise, a minimum SNR is required on the S-D link, which is, by definition, a system with diversity order 1). Hence, in order for successful decoding in the event of an outage over the S-D channel, the condition in the above equation must be satisfied. Thus, with the S-D channel present, the condition must be met in order to achieve a diversity order of 2.

Some examples where (4.4) is satisfied include \( \epsilon_{info} = 1 \) and \( \epsilon_{par} = 0 \) for any \( R_S \), and \( R_S = 1/2 \) and \( \epsilon_{info} + \epsilon_{par} = 1 \). These conditions must be met in order to achieve a diversity order of 2 with the use of 1 relay.

Let \( \epsilon' = \epsilon_{info} + \epsilon_{par} \) be a fixed value. Taking the derivative of the right-hand side of (4.1) with respect to \( \epsilon_{info} \),
\[ \frac{d}{d\epsilon_{info}} \left[ \frac{\epsilon' - \epsilon_{info}}{1 - \epsilon_{info} + (\epsilon' - \epsilon_{info})} \right] = \frac{-1 + \epsilon'}{(1 + \epsilon' - 2\epsilon_{info})^2}. \] (4.5)
If \( \epsilon' > 1 \), then the derivative is positive and the right-hand side of (4.1) is maximized when \( \epsilon_{info} \) is maximized. This means that all the source’s information bits should be relayed to allow large source code rate while still satisfying the constraint. If \( \epsilon' = 1 \), then
the right-hand side remains the same irrespective of the allocation of relay resources. If 
\( \epsilon' < 1 \), then the derivative is negative and the right-hand side of (4.1) is maximized when 
\( \epsilon_{info} \) is minimized, and all the source’s parity bits should be relayed to allow large source 
code rate while still satisfying the constraint.

**Corollary 4.1.1** The FER of the relay codeword has diversity order less than 2 for DEF 
with fractional cooperation.

**Proof:** In order for the relay to achieve a diversity order of 2, the condition in (4.4) 
must be satisfied for the relay codeword as well. Let \( \tilde{\epsilon}_{par} \) be the fraction of relay parity 
bits that are relayed, and \( \tilde{\epsilon}_{info} \) be the fraction of \( k_R \) relay information bits that are 
relayed. From the setup of the scheme, \( \tilde{\epsilon}_{par} \), is 0, since its parity bits are not relayed by 
any other nodes, and (4.4) becomes

\[
R_R \leq \tilde{\epsilon}_{info} R_R, \tag{4.6}
\]

where \( \tilde{\epsilon}_{info} \) is the fraction of relay information bits that are also transmitted by the 
source, i.e., bits from the source that are relayed by R. This means that (4.6) translates 
to \( \tilde{\epsilon}_{info} \geq 1 \). However, since \( \tilde{\epsilon}_{info} < 1 \) as long as the relay is using a part of its codeword 
to transmit its own data, diversity order of 2 cannot be achieved for the relay codeword.

This corollary, at first glance, seems obvious since no other node is forwarding in-
formation for the relay. However, as illustrated in the simulation results, the relay can 
achieve a FER with diversity order greater than 1. Since the relay’s codeword includes 
the source’s information bits, decoding the source’s codeword helps in decoding the relay 
codeword as well.

### 4.1.2 Simulation Results

This section presents results of simulations for two different scenarios. Let the distance 
between S and D, between S and R, and between R and D be \( d_{SD} \), \( d_{SR} \) and \( d_{RD} \) re-
respectively. In the first scenario R is closer to S than it is to D, where $d_{SR} = 0.4d_{SD}$ and $d_{RD} = d_{SD}$. The average received signal energy is proportional to $d^{-\alpha}$, where $d$ is the distance between the transmitter and receiver and $\alpha$ is the path loss exponent [72]. In both scenarios a path loss exponent of $\alpha = 4$ is used. The relationship between the average received symbol SNR of the various channels is given by

$$\bar{\gamma}_{SD} = (0.4)^{\frac{1}{4}} \bar{\gamma}_{SR} = \bar{\gamma}_{RD}.$$ 

In the second scenario R is closer to D than it is to S, where $d_{SR} = d_{SD}$ and $d_{RD} = 0.4d_{SD}$. The relationship between the average received symbol SNR of the various channels is given by

$$\bar{\gamma}_{SD} = \bar{\gamma}_{SR} = (0.4)^{\frac{1}{4}} \bar{\gamma}_{RD}.$$ 

For the rest of the thesis, unless stated otherwise, it is assumed that in plots of simulation results, SNR either represents the instantaneous received symbol SNR for AWGN, or the average received symbol SNR as defined in the manner stated in (2.4).

For each scenario, simulation results for two different cases are presented and compared. In both cases $\epsilon_{par} = 0$, and a rate-1/2 punctured systematic RA codes are used at S, with $l_S = 4000$. In order to provide a fair comparison $l_R = 4000$ is fixed, such that the energy consumed for transmission in both cases is the same.

The parameters $\epsilon_{info}$ and $R_R$ are varied to show two different ways to employ fractional cooperation. In the first case, the rate of the relay code $R_R = 1/2$ is fixed, and $\epsilon_{info}$, the fraction of source’s information bits, is adjusted accordingly to allow the required relay information bits, $m_R$, to be transmitted. The relationship between $R_R$ and $\epsilon_{info}$ is given by

$$k_R = R_R l_R = m_R + \epsilon_{info} k_S.$$ 

In the second case, instead of varying $\epsilon_{info}$, $\epsilon_{info} = 1$ is fixed, and $R_R$ is adjusted accordingly to allow the required $m_R$ to be transmitted. Hence, in the first case, for $m_R = 1000$ and $m_R = 500$, $\epsilon_{info} = 1/2$ and $\epsilon_{info} = 3/4$ respectively. In the second case,
for $m_R = 1000$ and $m_R = 500$, $R_R = 3/4$ and $R_R = 5/8$ respectively. The simulation results for the first scenario are shown in Fig. 4.1. The solid lines represent FER for the source’s information bits, and the dash-dot lines represent FER for the $m_R$ relay’s information bits.

When $R_R = 1/2$, the source and relay FER are almost the same, whereas the source and relay FER are vastly different when $R_R$ is varied to relay all the source’s information bits. From the plot, it is shown that the diversity order is approximately 1 for all the curves, except the ones representing the FER for the source’s codewords with $\epsilon_{info} = 1$. The increase in diversity order of the source FER comes at a cost of increasing the relay code rate, and the loss is evident from the shift of the relay FER curves. The loss, compared with its constant $R_R$ counterpart, is about 2 dB when $R_R$ is increased from $1/2$ to $3/4$ for $m_R = 1000$, and about 1 dB when it is increased to $5/8$ for $m_R = 500$. There is also a slight decrease in the source FER with $\epsilon_{info} = 1$ when $m_R$ is increased, as illustrated by the FER curve shift in the plot.

The condition necessary to obtain a diversity order of 2 is given by (4.4). In the simulation results presented above, the condition is satisfied when $\epsilon_{info} = 1$. It is illustrated in Fig. 4.1 that when the condition is satisfied, a diversity order of 2 can be observed for the source FER, and when the condition is not satisfied, the diversity order is less than 2. This agrees with Theorem 4.1.1. In addition, the diversity order of the relay FER is always less than 2, and this agrees with our analysis from Corollary 4.1.1.

The simulation results for the scenario where the relay is closer to the destination node is shown in Fig. 4.2. When $R_R = 1/2$, similar to the previous case, the FER performance for both the source and relay are better with a smaller $m_R$. This is the tradeoff between data rate and error performance. The figure also illustrates an interesting result: the FER of the relay codewords has diversity order slightly larger than 1 for all cases even though no other node forwards messages for the relay. By assisting the source node, the relay also gains an increase in diversity order when the R-D channel is good. In
addition, the FER of the source’s codewords with $\epsilon_{info} = 1$ has diversity order of 2. Taken together with the simulation results from Fig. 4.1, all these observations confirm the results from Sec. 4.1. In the plot, the FER curves of the source’s codewords for $\epsilon_{info} = 1$ and $m_R = 500$ and $m_R = 1000$ are almost identical. This probably due to the fact that the error performance is limited by the S-R channel quality such that a slight change in $R_R$ does not affect the performance.

4.2 Fractional cooperation with Multiple Relays

In some scenarios, the resources available to each relay that can be used for assisting other nodes is insufficient to guarantee a diversity order of at least 2. If several relays are available, they can cooperate and together increase the diversity order of the system. In order to simplify the implementation, the relays randomly choose a fraction of the source’s codeword to relay. No coordination among the relays is required. As mentioned earlier in this chapter, this is different from the incremental redundancy relaying scheme suggested in [46], where the fraction is pre-assigned. In fractional cooperation, because the subset of bits chosen by each relay is random, it is possible that some of the source’s symbols are not relayed, while others may be relayed by every available relay. Clearly, allowing relays to choose the source’s symbols randomly is not the optimal relay assignment in terms of performance, but fractional cooperation is extremely simple and scales well with increases in network size as no communication between nodes is required to assign the fraction transmitted by each relay.

Two different types of fractional cooperation is defined here: regular and irregular. With regular fractional cooperation, the fraction relayed by each relay is the same. Irregular fractional cooperation, on the other hand, allows the fraction relayed by different nodes to be different. Irregular fractional cooperation may be utilized when the relaying nodes have different levels of battery power, and hence need to regulate the amount of
Figure 4.1: FER of source’s (solid lines) and relay’s (dashed lines) codewords for fractional cooperation with one relay, where R is closer to S.

Figure 4.2: FER of source’s (solid lines) and relay’s (dashed lines) codewords for fractional cooperation with one relay, where R is closer to D.
resources they can provide to assist other nodes. Another scenario when irregular fractional cooperation can be used is when nodes have their own data to transmit, and the amount of resources left that can be used to assist other nodes varies. The analysis in the rest of this chapter assumes that the blocklength of the source’s codeword is infinite.

4.2.1 Diversity Order Analysis for Fractional Cooperation

Professor Andrew W. Eckford is the main contributor for the theoretical analysis of diversity order for fractional cooperation with multiple relays shown below, which were published in [2]. The author of the dissertation is responsible for the generation of simulation results and editing comments for [2]. An overview of these results are presented here to aid in the presentation of additional material developed by the dissertation author based on these results. For more details on the proofs for the results presented here, readers are referred to [2].

Assume that $r$ relays are available to assist S. Let $\epsilon_j$ be the fraction of the source’s codeword relayed by $R_j$, where it is assumed that no distinction is made between the fraction of information and parity bits relayed. This allows the following results to be applicable to both systematic and non-systematic codes.

**Lemma 4.2.1** Let $r_c$ represent the smallest non-negative integer such that if the number of relays is less than $r_c$ and $\gamma_{SD} = 0$, then a system outage always occurs. Under these assumptions, $r_c$ exists and $r_c < \infty$.

Essentially, if a set of relays choose at random a fraction $\epsilon$ of the source’s codeword to transmit, then there exists a number $r_c$ such that if the number of relays is less than $r_c$, then there is always an outage if the S-D link is absent. This result is also true if the fractions relayed by $R_j$ are different. The value of $r_c$ is dependent on $\epsilon_j$, the encoding and decoding scheme and the relaying scheme.
Lemma 4.2.2 If $r < r_c$, then by including the S-D channel, the system diversity order is 1, and if $r = r_c$, then the system diversity order is 2.

This is a direct extension of Lemma 4.2.1. A diversity order of 2 is achieved only if the source’s codeword can be decoded even when the S-D is below a SNR threshold required for correct decoding at D. Hence, in order to guarantee correct decoding in case of outage on the S-D link, the number of relays must be large enough to ensure that the amount of data transmitted through the relays is sufficient to ensure correct decoding. The number of relays required is given by $r_c$.

Theorem 4.2.1 For a relay system with $r$ relays, where $r_c \leq r < \infty$, diversity order is given by $r - r_c + 2$. That is, each additional relay adds an order of diversity, regardless of system parameters.

The diversity order can be interpreted as the number of independent paths available for the data to travel from the transmitter to the receiver. If $d$ independent paths are available, then even if $d-1$ of them are in outage, successful decoding can still be achieved. Hence diversity order can be thought of as 1 added to the maximum number of relays that can be in outage and yet successful decoding is achieved. Building on Lemma 4.2.2, Theorem 4.2.1 shows that if $r \geq r_c$ relays are available to assist S, then even if $r - r_c + 1$ of the nodes is in outage, enough data arrives at D to ensure correct decoding. Hence the diversity order is $r - r_c + 2$.

4.2.2 Outage Set Analysis for Fractional Cooperation

Professor Eckford is the main contributor for the theoretical analysis on outage sets for fractional cooperation with multiple relays shown here, which were published in [3]. The author of this dissertation is the main contributor to the derivation of the distribution of the upper bound for diversity order, $d^*$, for a given set of $\epsilon_j$, and the use of this upper
bound to derive a lower bound on \( r_c \). For more details on the proofs for the results presented here, readers are referred to [3].

Define \( b_j(n) \) as the function where \( b_j(n) = 1 \) if the \( n \)th bit of the source’s codeword is relayed by \( R_j \), and \( b_j(n) = 0 \) otherwise. Let \( b_j = [b_j(1), b_j(2), \ldots, b_j(l_s)] \) and \( B = \{b_1, b_2, \ldots, b_r\} \) be the collection of cooperation sequences.

Recall that \( \gamma_{S,R_j} \) and \( \gamma_{R_j,D} \) are the instantaneous channel SNR over the S-\( R_j \) and \( R_j \)-D links respectively. A set of relays is an outage set, \( O(\gamma_{SD}, B) \), for a given \( \gamma_{SD} \) and \( B \) if the system is in outage whenever the instantaneous SNR over the S-D link is \( \gamma_{SD} \) and \( \gamma_{S,R_j} < \eta \) or \( \gamma_{R_j,D} < \eta \) for all \( j \in O(\gamma_{SD}, B) \), for a given value \( \eta > 0 \). Let \( S(\gamma_{SD}, B) \) be the collection of all outage sets. Then

\[
\omega(\gamma_{SD}, B) = \min_{O(\gamma_{SD}, B) \in S(\gamma_{SD}, B)} |O(\gamma_{SD}, B)|
\] (4.7)

denotes the cardinality of the smallest outage set in \( S(\gamma_{SD}, B) \).

**Theorem 4.2.2** The diversity order \( d \) of an irregular fractional cooperation system is given by

\[
d = 1 + \lim_{\gamma_{SD} \to 0} \omega(\gamma_{SD}, B). \tag{4.8}
\]

First assume that the S-D link is unavailable. Then the diversity order is given by the number of relays in outage that causes a system outage, which is the cardinality of the outage set. Since the diversity order is dominated by the smallest outage set, it is given by \( \lim_{\gamma_{SD} \to 0} \omega(\gamma_{SD}, B) \). If the S-D link is available, then the diversity order is given by (4.8).

Let \( \bigvee_{j \in \mathcal{H}} b_j(n) \) represent the logical OR of the \( n \)th element of all cooperation sequences in the set \( \mathcal{H} \) and \( \mathcal{H} \subseteq \mathcal{R} \), with \( \mathcal{R} \) being the index set of the set of relays. A result of 1 indicates that the \( n \)th symbol is selected by at least one relay in the set \( \mathcal{H} \).

**Theorem 4.2.3** Suppose \( c_S \) is a codeword of a length-\( l_S \) binary code which can correct
at most a fraction of erasure $p_E$. Let $\mathcal{H} \subseteq \mathcal{R}$ be a subset of the available relays. If
\[
\sum_{n=1}^{l_S} \bigvee_{j \in \mathcal{H}} b_j(n) < l_S(1 - p_E),
\]
then the complementary subset $\bar{\mathcal{H}} = \mathcal{R} - \mathcal{H}$ is an outage set as $\gamma_{SD} \to 0$.

Assume that $\gamma_{SD} \to 0$, and all the relays in $\bar{\mathcal{H}}$ is in outage. Then signal corresponding to less than fraction $(1 - p_E)$ of the codeword is received by the destination node. Since at least fraction $(1 - p_E)$ of the codeword must be received in order to decode correctly, this causes a system outage. Hence $\bar{\mathcal{H}}$ is an outage set.

**Corollary 4.2.1** Let $\mathcal{R}_E \subseteq \mathcal{R}$ represent the largest cardinality subset satisfying (4.9). Then
\[
\lim_{\gamma_{SD} \to 0} \omega(\gamma_{SD}, \mathcal{B}) \leq |\bar{\mathcal{R}}_E|,
\]
and thus diversity order $d \leq 1 + |\bar{\mathcal{R}}_E|$.

This result is referred as the erasure channel bound. Since $\mathcal{R}_E$ satisfies (4.9), $\bar{\mathcal{R}}_E$ is an outage set. By definition, $\omega(\gamma_{SD}, \mathcal{B})$ is the cardinality of the smallest outage set, and hence $\omega(\gamma_{SD}, \mathcal{B}) \leq \bar{\mathcal{R}}_E$. Let $d^* = 1 + |\bar{\mathcal{R}}_E|$ be the upper bound on the diversity order, where $d \leq d^*$.

In some scenarios the value of $\epsilon_j$ is chosen dependent on the amount of resources each relay is willing to contribute to aiding other nodes at the moment, such that $\epsilon_j$ is a RV. Hence $\mathcal{B}$, and in turn, $\mathcal{R}_E$ and $d^*$, are random variables. In these cases, the distribution on $d^*$ in the limit as the source’s codeword length $l_S \to \infty$ can be derived. First, note that the probability of a symbol not selected by any relay in subset $\mathcal{H} \subseteq \mathcal{R}$ is given by
\[
p_N(\mathcal{H}) = \prod_{R_j \in \mathcal{H}} (1 - \epsilon_j).
\]

Recall from (4.9) that $\sum_{n=1}^{l_S} \bigvee_{j \in \mathcal{H}} b_j(n)$ is the number of bits that are transmitted by at least one relay. In the limit as $l_S \to \infty$, $\sum_{n=1}^{l_S} \bigvee_{j \in \mathcal{H}} b_j(n) = l_S(1 - p_N(\mathcal{H}))$. 
In order to find the value of \( d^* \) for a given set of \( \epsilon_j, |\mathcal{R}_E| \), the size of the largest set satisfying (4.9) must be found. Let \( \epsilon(j) \) be the ascending-ordered statistics of \( \epsilon_j \), where \( \epsilon(j) \leq \epsilon(j+1) \). Then \(|\mathcal{R}_E|\) is given by

\[
|\mathcal{R}_E| = \max \left\{ k : 1 - \prod_{j=1}^{k} (1 - \epsilon(j)) < 1 - p_E \right\} \tag{4.12}
\]

\[
= \max \left\{ k : \prod_{j=1}^{k} (1 - \epsilon(j)) > p_E \right\}. \tag{4.13}
\]

Since \(|\tilde{\mathcal{R}}_E| = r - |\mathcal{R}_E|\),

\[
|\tilde{\mathcal{R}}_E| = r - \max \left\{ k : \prod_{j=1}^{k} (1 - \epsilon(j)) > p_E \right\} \tag{4.14}
\]

\[
= \max \left\{ k : \prod_{j=1}^{r-k} (1 - \epsilon(j)) > p_E \right\}. \tag{4.15}
\]

Hence for a given realization of \( \epsilon_j \),

\[
d^* = 1 + \max \left\{ k : \prod_{j=1}^{r-k} (1 - \epsilon(j)) > p_E \right\}, \tag{4.16}
\]

and since the events \((d^* = m)\) and \((d^* = m')\) are mutually exclusive for \( m \neq m' \), the distribution is given by

\[
\Pr(d^* = m) = \begin{cases} 
\Pr(\prod_{j=1}^{r} (1 - \epsilon(j)) > p_E) & \text{for } m = 1 \\
\Pr(\prod_{j=1}^{r-m+1} (1 - \epsilon(j)) > p_E) \cap (\prod_{j=1}^{r-m+2} (1 - \epsilon(j) \leq p_E)) & \text{for } m = 2, \ldots, r, \\
\Pr(1 - \epsilon(1) \leq p_E) & \text{for } m = r + 1.
\end{cases} \tag{4.17}
\]

Let \( X_j = -\log(1 - \epsilon_j) \) and \( X(j) = -\log(1 - \epsilon(j)) \), and let \( Z_b \) be the truncated version of RV \( X \) where

\[
f_{Z_b}(z) = \begin{cases} 
\frac{f_X(z)}{F_X(b)} & \text{if } z \leq b, \\
0 & \text{otherwise}.
\end{cases}
\]

and \( f_X(x) \) and \( F_X(x) \) are the probability density function (PDF) and cumulative density function (CDF) of RV \( X \) respectively. Also, let \( Y_k = \sum_{j=1}^{k} Z_{X(k+1),j} \), where \( Z_{X(k+1),j} \) are
i.i.d. RV drawn from $f_{Z_{X_{(k+1)}}}(z)$. The distribution of $d^*$, $\Pr(d^* = m)$, can then be derived from the distribution of $\epsilon_j$ and is summarized below:

- For $m = 1$:
  \[
  \Pr(d^* = 1) = F_{\sum_{j=1}^{r} x_j}(-\log(p_E)),
  \]
  \(4.18\)

- For $m = 2, \ldots, r$:
  \[
  \Pr(d^* = m) = \int_{-\log(p_E)}^{\infty} \int_{\max(0, -\log(p_E) - x)}^{\min(-\log(p_E), (r-m+1)x)} f_{X_{(r-m+2)}}(x) f_{Y_{r-m+1}}(y) dy dx,
  \]
  \(4.19\)

- For $m = r + 1$:
  \[
  \Pr(d^* = r + 1) = (1 - F_X(-\log(p_E)))^r.
  \]
  \(4.20\)

The derivation of the distribution of $d^*$ and an illustrative example can be found in Appendix A. The analysis shown above can be used while designing a relay network, where it can be used to determine the minimum number of relays needed to provide a given performance requirement.

### 4.2.3 Simulation Results

The plots shown in this section illustrate the analysis results from Sec. 4.2.1, as well as the flexibility of fractional cooperation. In both plots, a punctured systematic RA code with $R_S = 1/2$ and $k_S = 2000$ is used, and $\bar{\gamma}_{SD} = \bar{\gamma}_{S,R_j}$. The simulation results for regular fractional cooperation are shown in Fig. 4.3. In the plot, $\epsilon = 0.125$ with $r = 10, 11, 12$, and $\epsilon = 0.1$ and $\epsilon = 0.15$ with $r = 11$ are shown to illustrate the effects of varying parameters. When $\epsilon = 0.1$, $\epsilon = 0.125$, or $\epsilon = 0.15$, 400, 500, or 600 bits of the source’s codeword is relayed by each relay respectively. For simplicity, it is assumed that the R-D channel is perfect, i.e., D can decode all the relay codewords correctly. With $\epsilon = 0.125$, the FER curves for $r = 10$, $r = 11$ and $r = 12$ have diversity order of 4, 5 and 6 respectively. Hence it can be deduced from $d = r - r_c + 2$ that $r_c = 8$. Furthermore, this agrees with the analysis of multiple relay fractional cooperation, where for $r \geq r_c$,
each additional relay increases the diversity order by 1. With $\epsilon = 0.1$ and $r = 11$, the diversity order is 3, where it can be deduced that $r_c = 10$, and with $\epsilon = 0.15$ and $r = 11$, the diversity order is 6, and $r_c = 7$.

Simulation results for irregular fractional cooperation are presented in Fig. 4.4, with $r = 11$. For a fair comparison, each curve in the figure has the same number of bits forwarded over all the relays, and the average fraction relayed, $\bar{\epsilon} = 0.125$ remains the same. Three different cases are presented: (a) each relay forwards 500 bits ($\epsilon = 0.125$); (b) the number of bits relayed by each relay ranges from 250 to 750 bits, in 50-bit increments; and (c) 5 relays forward 50 bits, 1 relay forwards 500 bits, and 5 relays forward 950 bits. The best performance is obtained when the contribution is constant for all the relays, and as the variation increases, the FER performance worsens. However, even with large variations of $\epsilon_j$ between various relaying nodes, the degradation in FER is quite small. Also, each additional relay increases the diversity order by 1 even when the contribution of a relay is small. This shows that fractional cooperation is a very robust scheme which does not require coordination between the relaying nodes.

### 4.3 Discussion

In some of the analysis results presented earlier, a distinction is made between the fraction of the source’s information and parity bits relayed with the use of the parameters $\epsilon_{info}$ and $\epsilon_{par}$. When a distinction is made between the two types of codeword bits, the results can only be used for analysis for systematic codes. These results, however, can be extended to non-systematic codes by eliminating the distinction between the source’s information and parity bits.

Even though no coordination between the relaying nodes is required to assign the fractions relayed by each node beforehand, it is necessary for the relays to inform D which section of the source’s codeword is relayed. This knowledge is required to perform
Figure 4.3: FER of DEF with RA code under regular fractional cooperation for various $r$ and $\epsilon$ values.

Figure 4.4: FER of DEF with RA code under irregular fractional cooperation for $r = 11$, where $\bar{\epsilon} = 0.125$. 
decoding at the relay. If the bits are chosen in a random fashion, instead of specifying the bits relayed, each relay can inform D the bits relayed by sharing the seed of pseudo-random number generator used with D. Using the seed, D can find the set of bits transmitted by each relay. The effect of errors that occur over the transmission of the seed value is outside the scope of this thesis.

As mentioned earlier, the value of \( r_c \) is dependent on the encoding and decoding schemes, the relaying schemes, as well as the fraction of source’s codeword relayed, \( \epsilon_j \). All these parameters must be taken into account while optimizing the system performance where fractional cooperation is used. This complicates the system design, but we believe that the flexibility brought on by the introduction of fractional cooperation is worth the extra complexity, especially if it can increase the network lifetime. In addition, we also believe the main determining factor of the value of \( r_c \) is the set of \( \epsilon_j \), and we present a method of providing an lower bound on \( r_c \) from the set of \( \epsilon_j \) below, which can aid in the system design process.

As \( r_c \) is dependent on a large number of parameters, it is difficult to find the actual value of \( r_c \) given the parameters without resorting to simulations. For example, the value of \( r_c \) is deduced from simulation results in Sec. 4.2. The analysis on outage sets presented in Sec. 4.2.2 can be used to provide an lower bound on \( r_c \) for a given set of \( \epsilon_j \). From (4.16) it can be deduced that \( d^* \geq 2 \) is a necessary condition for achieving a diversity order of 2. This condition translates to finding the smallest \( r \) such that

\[
\max \left\{ k : \prod_{j=1}^{r-k} 1 - \epsilon(j) > p_E \right\} \geq 1
\]

This means that the lower bound to \( r_c \), \( \tilde{r}_c \), is given by

\[
\tilde{r}_c = \min \left\{ r : \prod_{j=1}^{r-1} 1 - \epsilon(j) > p_E \right\}
\]

(4.21)
For regular fractional cooperation, the lower bound is given by

\[ \hat{r}_c = \min \left\{ r : r > \frac{\log(p_E)}{\log(1 - \epsilon)} + 1 \right\} \]

Equation (4.22)

\[ = \left\lceil \frac{\log(p_E)}{\log(1 - \epsilon)} + 1 \right\rceil \]

Equation (4.23)

Using density evolution, it is found that for a systematic punctured rate-1/2 RA code, \( p_E \approx 0.4 \). The lower bound to \( r_c \) for regular fractional cooperation with various values of \( \epsilon \) is tabulated in Table 4.1. Note that the values of \( r_c \) deduced from Fig. 4.3 fall above the lower bounds given in the table for all cases.

For the set of simulation results plotted in Fig. 4.4, lower bounds for \( r_c \) can be found for all cases. For case (a), \( \hat{r}_c = 8 \), as shown in Table 4.1. For case (b), \( \hat{r}_c = 10 \), and for case (c), \( \hat{r}_c = 9 \). Note that the outage set analysis is one method to find lower bounds for \( r_c \). Another method that can be used to find a lower bound on \( r_c \) is introduced in Ch. 5.

Even though the exact value of \( r_c \) is not apparent, it has been shown that fractional cooperation is a robust scheme, where no coordination between the relaying nodes is required, and the fraction of the source’s codeword transmitted by each relay can be changed easily. Even when \( \epsilon_j \) varies for different relays, each additional relay increases the diversity order by 1 for \( r \geq r_c \). Importantly, even though the performance of fractional cooperation is illustrated with the use of DEF, it is not limited to DEF, and can be used with other relaying schemes such as DF and AF. If DF is used instead of DEF, then provided correct decoding of the source’s codeword can be achieved, each relay forms a codeword based on a fraction of the source’s information bits. If AF is used, then a

<table>
<thead>
<tr>
<th>( \epsilon )</th>
<th>0.1</th>
<th>0.125</th>
<th>0.15</th>
<th>0.2</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \hat{r}_c )</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4.1: Lower bound on \( r_c \) for fractional cooperation using rate-1/2 punctured systematic RA code for various \( \epsilon \) derived from outage set analysis.
fraction of the received signal is forwarded to the destination node.

Note that the diversity order and outage set analysis done on fractional cooperation provide only asymptotic results, where it is assumed that $l_S \to \infty$. By assuming the blocklength is infinite, the probability of the fractions transmitted by different relays overlapping approaches 0, and the fraction of the source’s codeword not selected by any relay in a set $\mathcal{H}$ is given by (4.11). This assumption simplifies the analysis significantly, but is unrealistic. Since a quasi-static fading channel is used to model the channels between the nodes, where the fading coefficients are assumed to remain constant for the whole block, an infinite blocklength implies that the coherent time is infinite. There is a need to study the effects on the analysis results when the blocklength is finite.

When the blocklength is finite, however, this probability of the fractions transmitted by different relays overlapping is non-zero. The actual fraction of the source’s codeword not selected by any relay in a set $\mathcal{H}$ can be different than that stated in (4.11), and the actual outage probability will be higher than that found in the analysis. In the low SNR regime, if the blocklength is large, the outage caused by the difference between the actual and theoretical non-relayed fraction of the source’s codeword is small compared to the probability of outage from analysis, and the effects of this differences can be ignored. However, in the high SNR regime, if the blocklength is small, the difference may dominate the outage probability in actual simulations, and the effects of this difference cannot be ignored. For the blocklength and the SNR used in our simulations, the simulation results agree with our analysis, and it can be assumed that for the blocklength and the SNR regime studied the difference between the theoretical and actual outage probability is small. However, these effects need to be studied in-depth to gain an understanding of the range of SNR and blocklength where the analysis results are valid.
Chapter 5

Analysis and Design using the Union-Bhattacharyya Bound

When channel codes are used at the source, it is difficult to study the error performance of various relay channels. Most of the existing work examines the theoretical limits on transmission rate, or uses simulations to estimate error performance. In Chapters 3 and 4, error performance is obtained through (time-consuming) simulations. This makes it difficult to know whether a certain performance criterion can be met given a set of channel conditions and encoding/decoding scheme. As shown in Chapter 2, the UBB can be used to provide an upper bound on the error performance, where the structure of the code and the channel quality is taken into account. In the rest of this chapter, we use the UBB to analyze, characterize and design relay networks. We present applications of the UBB such as relay selection, upper bounding the FER in fading channels, relay channel characterization and lower bounding $r_c$ for fractional cooperation. It is the ability to analyze performance, with the implementation details accounted for, that sets this chapter apart from the literature in this area. In the next section, we apply the UBB to relay channels. We first restrict the number of relays to one for ease of presentation and then extend the results to multiple relay networks in later sections.
5.1 Union-Bhattacharyya Bound for Relay Channels

For simplicity, we first derive the UBB for one-relay networks here, where we assume that only R_j is present to relay for the source node. Let ε_{info} and ε_{par} represent the fraction of source’s information and parity symbols that are relayed by R_j. For a systematic code, \( \bar{A}_{w,h} \) represents the average number of codewords that have weight \( w \) for the information bits, and weight \( h - w \) for the parity bits. From the source’s codeword’s point of view, \( \epsilon_{info} \) of the source’s information bits are transmitted by S and relayed by the chosen relay, while \( 1 - \epsilon_{info} \) of them are only transmitted by the S. Similarly, \( \epsilon_{par} \) of the source’s parity bits are transmitted by S and the chosen relay, while \( 1 - \epsilon_{par} \) of them are only transmitted by S. In this case, applying the UBB from (2.24) here we arrive at

\[
P_f \leq \sum_{w=1}^{k} \sum_{h=1}^{n} \bar{A}_{w,h} \bar{\beta}_{info}^{w} \bar{\beta}_{parity}^{h-w},
\]  

(5.1)

where from (2.25)

\[
\bar{\beta}_{info} = (1 - \epsilon_{info}) \beta_{SD} + \epsilon_{info} \bar{\beta}_j,
\]

\[
\bar{\beta}_{parity} = (1 - \epsilon_{par}) \beta_{SD} + \epsilon_{par} \bar{\beta}_j,
\]  

(5.2)

with \( \beta_{SD} \) being the BP associated with the S-D link and \( \bar{\beta}_j \) the BP associated with the combination of the S-D link and the S-R_j-D link.

When a code bit is transmitted through \( n \) independent channels, the received signal \((y_1, y_2, \ldots, y_n)\) are independent given the transmitted bits. The BP associated with the \( n \) receive signals is given by

\[
\beta = \int \int \ldots \int \sqrt{f(y_1, y_2, \ldots, y_n|0)f(y_1, y_2, \ldots, y_n|1)}dy_1dy_2\ldots dy_n
\]

\[
= \int \int \ldots \int \prod_{k=1}^{n} \sqrt{f(y_k|0)f(y_k|1)}dy_1dy_2\ldots dy_n
\]

\[
= \prod_{k=1}^{n} \beta_{ch,k},
\]  

(5.3)
where $\beta_{ch,k}$ is the BP associated with the $k$th channel. The likelihood probability can be separated into a product form as the received signals are independent given the transmitted signal. This shows that when a code bit is transmitted over multiple channels, the BP is given by a product of the BPs associated with each channel. Hence, for the bits that are transmitted over the S-D channel and S-R$_j$-D channel, the BP is given by

$$\tilde{\beta}_j = \beta_{SD} \beta_j,$$  \hspace{1cm} (5.4)

where $\beta_j$ is the BP associated with the S-R$_j$-D link, and (5.2) becomes

$$\tilde{\beta}_{info} = \beta_{SD} (1 - \epsilon_{info} (1 - \beta_j)),$$

$$\tilde{\beta}_{parity} = \beta_{SD} (1 - \epsilon_{par} (1 - \beta_j)).$$  \hspace{1cm} (5.5)

We now present the UBB for multiple-relay channels. To simplify the derivation, the distinction made between the fraction of source’s information and parity bits is removed. Assuming that $r$ relays are used, and relay $j$ chooses to relay each of the $n$ bits of the source with probability $\epsilon_j$. Let $\mathcal{R} = \{1, 2, \ldots, r\}$ be the index set of relays, and let $\mathcal{R}_k, k = 1, \ldots, 2^r$, be all possible subsets of $\mathcal{R}$, including the empty set $\emptyset$. Then the probability of each bit of the source’s codeword relayed by the relays in set $\mathcal{R}_k$ is given by

$$\alpha_{\mathcal{R}_k} = \left( \prod_{j \in \mathcal{R}_k} \epsilon_j \right) \left( \prod_{j \in \mathcal{R} \setminus \mathcal{R}_k} (1 - \epsilon_j) \right),$$

and the associated BP is given by

$$\beta_{\mathcal{R}_k} = \prod_{j \in \mathcal{R}_k} \beta_j.$$  \hspace{1cm} (5.6)

Substituting these values into (2.25) and after summing over all $\mathcal{R}_k$ and some simple manipulation, it can be shown that for fractional cooperation

$$\tilde{\beta} = \beta_{SD} \prod_{j=1}^{r} (1 - \epsilon_j (1 - \beta_j)).$$  \hspace{1cm} (5.7)
This equation can be substituted into (2.24) to obtain the UBB for fractional cooperation

\[ P_f \leq \sum_{h=1}^{n} A_h \left( \beta_{SD} \prod_{j=1}^{r} (1 - \epsilon_j (1 - \beta_j)) \right)^h. \]  

(5.8)

### 5.2 Relay Selection for DEF

In illustrating the use of the UBB in analyzing relay networks, we begin with a simple application to relay selection. Clearly, the diversity order of the relay system can be increased by increasing the number of relays used to assist a source node. However, under the assumption that orthogonal channels are used to facilitate the communications between the various nodes, increasing the number of relays reduces the transmission rate, as the channel is further divided to allow all the nodes to transmit. Relay selection, where only the “best” relay is chosen out of a pool of available relays to assist the source, can be used to increase the diversity order without sacrificing the transmission rate. As mentioned earlier, mutual information can be used as the selection criterion, as in [29–31]. Another selection criterion is error rate, where the relay, when used, can help achieve the lowest error rate is chosen to assist the source node.

For systems where DEF is used, it is difficult to identify the best relay, as it is hard to quantify the effect of the demodulation performed at the relays. The most straightforward way to identify the best relay is to perform simulations over all possible realizations of the S-R and R-D channels and store the error rate for each of these realizations in a look-up table. However, simulations can be time-consuming, and the storage required to store the look-up table can be large as the SNR range is large. Instead of performing simulations to obtain information about the error rate at various channel realizations, density evolution, which is less time-consuming than simulations, can be used. However, density evolution is an asymptotic concept, built on the assumption of large blocklengths. Also, even though density evolution is more efficient than simulations, it is still quite time-consuming to obtain all the required data, and similar to simulations, the data needs to
be stored in a look-up table at each relay. In the following section, the calculation of \( \beta_{\text{DEF}} \), the Bhattacharyya parameter (BP) for DEF is shown, and its use as an efficient relay selection tool is presented.

### 5.2.1 Bhattacharyya Parameter for DEF

In order to simplify the presentation, the index \( j \) used to refer to the relay is dropped for the rest of this section when only one relay is chosen. Assuming that *serial decoding* is used, where the relay’s codeword is decoded before the soft information is used to assist in the decoding of the source’s codeword, the calculation of \( \beta_{\text{DEF}} \) can be separated into two steps. In the first step, the distribution of the output signal after decoding the relay’s codeword is obtained. Let the likelihood function for the R-D channel before decoding be modeled as

\[
\begin{align*}
    f(y_{R,D}(n)|c_R(n) = 0) &= g(y_{R,D}(n); 1, \sigma^2_{R,in}), \\
    f(y_{R,D}(n)|c_R(n) = 1) &= g(y_{R,D}(n); -1, \sigma^2_{R,in}),
\end{align*}
\]

where

\[
g(y; \mu, \sigma^2) = \frac{1}{2\pi\sigma^2} \exp \left\{ -\frac{(y - \mu)^2}{2\sigma^2} \right\}
\]

is the Gaussian distribution. As shown in [73], the distribution of the output signal after decoding can also be closely approximated by the Gaussian distribution if no puncturing is used at the relay’s codeword. Since an approximation is used here, when it is used the computed BP and the associated upper bound are only approximations to the actual BP and the associated upper bound respectively. However, this approximation is fairly accurate, as stated in [73] and the references within, and has been used extensively by the coding theory community. Hence we believe the impact of the approximation is quite small, and the use of UBB with this approximation in the context of the applications presented here is still valid. The parameters associated with the output signal after decoding can be found using the technique from [73]. Let the output distribution for the
R-D channel after decoding be modeled as

\[ f(y_{R,D}(n)|c_R(n) = 0) = g(y_{R,D}(n); 1, \sigma^2_{R,\text{out}}), \]
\[ f(y_{R,D}(n)|c_R(n) = 1) = g(y_{R,D}(n); -1, \sigma^2_{R,\text{out}}). \]  

(5.10)

The relationship between the variance of the channel distribution before and after decoding a rate-1/4 RA code can be found using the technique from [73], and is illustrated in Fig. 5.1, where \( \sigma^2_{R,\text{in}} = \frac{|h_{RD}|^2}{N_D} \) is the variance for the distribution in (5.9) over the R-D channel before decoding, and \( \sigma^2_{R,\text{out}} \) is the equivalent variance after decoding. The curve for \( \sigma^2_{R,\text{out}} = \sigma^2_{R,\text{in}} \) is also shown for comparison. This data is stored at the relay node, where the equivalent channel noise variance \( \sigma^2_{R,\text{out}} \) can been found through a lookup table given the instantaneous R-D channel SNR. Note that this lookup table, denoted as the \( \sigma^2_{R,\text{in}}-\sigma^2_{R,\text{out}} \) lookup table here, has to be created just once for a particular code, i.e., density evolution in real time is not required. Furthermore, the required lookup table is one-dimensional, and one expects the storage of such data can be easily implemented.

After the equivalent channel noise variance \( \sigma^2_{R,\text{out}} \) has been obtained, it is combined with the bit-flip probability \( p_{S,R} \) to find the likelihood function for the equivalent S-R-D channel after decoding the relay’s codeword

\[ f(y_{R,D}(n)|c_S(n)) = \begin{cases} 
  g(y_{R,D}(n); 1, \sigma^2_{R})\bar{p}_{S,R} + g(y_{R,D}(n); -1, \sigma^2_{R,\text{out}})p_{S,R} & \text{if } c_S(n) = 0, \\
  g(y_{R,D}(n); 1, \sigma^2_{R})p_{S,R} + g(y_{R,D}(n); -1, \sigma^2_{R,\text{out}})\bar{p}_{S,R} & \text{if } c_S(n) = 1.
\end{cases} \]  

(5.11)

where \( \bar{p}_{S,R} = 1 - p_{S,R} \). Substituting (5.11) into (2.22) we obtain

\[ \beta_{\text{DEF}} = \int_{-\infty}^{\infty} \sqrt{f(y_{R,D}(n)|c_S(n) = 0)f(y_{R,D}(n)|c_S(n) = 1)} \, dy \]  

(5.12)

\[ = \int_{-\infty}^{\infty} \left\{ [g(y; 1, \sigma^2_{R,\text{out}})\bar{p}_{S,R} + g(y; -1, \sigma^2_{R})p_{S,R}] \times [g(y; -1, \sigma^2_{R,\text{out}})\bar{p}_{S,R} + g(y; 1, \sigma^2_{R,\text{out}})p_{S,R}] \right\}^{1/2} \, dy \]  

(5.13)

\[ = \frac{1}{\sqrt{2\pi\sigma^2_{R,\text{out}}}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{y^2 + 1}{2\sigma^2_{R}} \right\} \times [4p_{S,R}\bar{p}_{S,R} \sinh^2(y/\sigma^2_{R,\text{out}}) + 1]^{1/2} \, dy. \]  

(5.14)
A closed form expression for (5.14) is not available, but the Taylor series expansion can be used to provide a close approximation to the integration. Note that as $p_{S,R}$ becomes very small, the S-R-D channel resembles an AWGN channel, with $\beta_{\text{DEF}} = \exp\left\{-\frac{1}{2\sigma_{R,\text{out}}^2}\right\}$.

Similarly, as $\sigma_{R,\text{out}}^2$ becomes very small, the S-R-D channel becomes a BSC, with $\beta_{\text{DEF}} = 2\sqrt{p_{S,R}p_{S,R}}$.

### 5.2.2 Selecting One Relay

As illustrated in (5.8), the smaller $\beta_j$ is, the smaller the upper bound on the FER is. Presuming that this bound is consistent with the true FER, the BP can be used as a metric to select the relay that gives the smallest FER. The selection criterion is therefore quite simple: the best relay is the one with the lowest BP, $\beta_j$.

A comparison of the results from density evolution on the complete factor graph and from the $\sigma_{R,\text{in}}^2 - \sigma_{R,\text{out}}^2$ lookup table together with calculation from (5.14) is shown in
Fig. 5.2. The parameters $\epsilon_{info}$ and $\epsilon_{par}$ are set to be 1 and 0 respectively, and the rate of the source’s and relay’s codewords are $R_S = 1/2$ and $R_R = 1/4$. It is assumed that only demodulation is performed at the relay. In addition, the S-D channel instantaneous SNR $\gamma_{S,D}$ is set to $-6$ dB. In the figure, the markers indicate the values of $\gamma_{SR}$ and $\gamma_{RD}$ that yield the given BER through the use of density evolution on the factor graph, and the lines show the contours outlining values of $\gamma_{SR}$ and $\gamma_{RD}$ that give the same $\beta_{DEF}$ values found using the $\sigma^2_{R,in}$-$\sigma^2_{R,out}$ lookup table and (5.14). As shown in the plot, $\beta_{DEF}$ does an excellent job of characterizing the S-R and R-D channel conditions that would yield a certain BER. This is quite remarkable, as all the channel conditions that give the same $\beta_{DEF}$ also give the same BER. The plot, therefore, illustrates that the BP can indeed be used as a simple and efficient performance measure. Note that the calculations described above place a very limited real-time computation burden on the relays.

Simulation results showing that relay selection using the BP approach provides perfor-
mance very close to optimal are presented here. In this example, $\epsilon_{info} = 1$ and $\epsilon_{par} = 0$, and pure demodulation is performed at the chosen relay. The rate of the source’s codeword is $R_S = 1/2$, with $l_S = 4000$, and the rate of the relay’s codeword is $R_R = 1/4$, with blocklength $l_R = 8000$. All the channels have the same average received SNR. The FER for choosing a single relay out of 2 or 3 available relays using the $\sigma_{R,in}^2 - \sigma_{R,out}^2$ lookup table and (5.14) is shown in Fig. 5.3, and the FER for one relay is also shown for reference. To show the importance of a good relay selection scheme, simulations results for a simpler relay selection is also provided for comparison. In this simpler scheme, the minimum SNR over the S-R and R-D links for each relays is found, which is denoted as min-SNR. The relay with the maximum min-SNR is chosen for relaying, and simulation results using this simple scheme are shown. In addition, simulation results for the case where for a given S-D channel, 2 or 3 relays are available to assist in the transmission, and the relay with the least number of bit errors is used. These results (by exhaustive search) are used as the optimal case, and provide a lower bound on the FER performance for DEF with relay selection. It is observed that selection cooperation using BP provides full diversity order despite the fact that only one relay is used. As illustrated in the plot, the FER for exhaustive search and our low complexity relay selection scheme based on the UBB are essentially the same, showing the relay selection scheme using BP as a selection metric provides excellent performance in minimizing the FER. Finally, the simple relay selection scheme, denoted as max-min, is sub-optimal. This shows the importance of a good relay selection scheme.

5.2.3 Relay Selection for Fractional Cooperation

In extending relay selection to fractional cooperation we must first take a detour into some new theoretical results. These results show that, as with DF and AF, relay selection provides “full” diversity order. As in the earlier section, the BP can then be used to select the relays.
Figure 5.3: FER for relay selection based on the BP, max-min over the S-R and R-D channel, and exhaustive search.

An overview of analysis for fractional cooperation was presented in Sec. 4.2, and these results are extended here to the case of relay selection. Let $r_a$ be the number of available relays. It is assumed that $r_b$ “best” relays are chosen out of a pool of $r_a$ relays, where in this case the relays are chosen based on their ability to improve the system error performance. Each chosen relay then selects, at random, a fraction of the source’s code bits, form a new codeword based on the chosen bits, and transmits the codeword to the destination.

Some definitions for the terms that will be used are provided here. Let $r_t(\gamma_{S,D})$ be the number of relays, such that if the number of available relays $r_a < r_t(\gamma_{S,D})$ then a system outage will occur, for that value of $\gamma_{S,D}$ and all S-R and R-D channels, independent of the value of $r_b$. When $r_a \geq r_t(\gamma_{S,D})$ relays are available to assist, but the number of relays chosen, $r_b$, is less than a threshold $\tilde{r}_c(\gamma_{S,D}, r_a)$, then an outage would occur for those values of $\gamma_{S,D}$ and $r_a$. In addition, let $r_t = \lim_{\gamma_{S,D} \to 0} r_t(\gamma_{S,D})$ and $\tilde{r}_c(r_a) = \lim_{\gamma_{S,D} \to 0} \tilde{r}_c(\gamma_{S,D}, r_a)$.
The analysis results presented below are based on the assumption that the blocklength of the source’s codeword $l_S \to \infty$.

**Lemma 5.2.1** For all $\gamma_{S,D}$, $r_t(\gamma_{S,D})$ exists and is finite. Also, for all $\gamma_{S,D}$ and $r_a > r_t(\gamma_{S,D})$, $\tilde{r}_c(\gamma_{S,D}, r_a)$ exists and is finite.

*Proof:* See Appendix B.

Lemma 5.2.1 shows that for any $\gamma_{S,D}$, a finite $r_t(\gamma_{S,D})$ can be found, such that if $r_a < r_t(\gamma_{S,D})$, a system outage occurs. In addition, for a given $\gamma_{S,D}$ and $r_a \geq r_t(\gamma_{S,D})$, a finite $\tilde{r}_c(\gamma_{S,D}, r_a)$ can be found, such that if $r_b < \tilde{r}_c(\gamma_{S,D}, r_a)$, a system outage occurs.

**Theorem 5.2.1** If $r_a < r_t$, the diversity order of the system is 1. For $r_a \geq r_t$, if $r_b < \tilde{r}_c(r_a)$, then the diversity order is 1 as well. If $r_a \geq r_t$ and $r_b \geq \tilde{r}_c(r_a)$, then the diversity order of the system is $r_a - \tilde{r}_c(r_a) + 2$.

*Proof:* See Appendix C.

Theorem 5.2.1 shows that if $r_a \geq r_t$ and $r_b \geq \tilde{r}_c(r_a)$, then a diversity order of $r_a - \tilde{r}_c(r_a) + 2$ can be observed. Only $\tilde{r}_c(r_a)$ relays is required to obtain the maximum diversity order for a pool of $r_a$ available relays; any additional relay chosen to assist only shifts the FER curve and does not provide an increase in diversity order. Note that similar to the analysis done on fractional cooperation in Ch. 4, this analysis is an asymptotic result, where it is assumed that the source’s codeword blocklength $l_S \to \infty$. Further studies are needed to characterize the effects of finite blocklength.

Similar to the one-relay system, relay selection can be used to choose the best $r_b$ relays out of a pool of $r_a$ available relays to provide full diversity. Using the same technique as earlier, $\beta_{DEF}$ values of each relay can be calculated, and the $r_b$ relays with the smallest $\beta_{DEF}$ values are chosen to assist the source.

In Fig. 5.4, simulation results for fractional cooperation with relay selection are shown. All the channels have the same average SNR. Here we set $\epsilon_j = \epsilon = 0.2$, and $r_a = 8$. Again
Figure 5.4: FER of fractional cooperation with relay selection, where $r_a = 8$ and $\epsilon_j = 0.2$.

the rate of the source’s codeword is $R_S = 1/2$, with $l_S = 4000$, and the rate of the relays’
codeword is $R_R = 1/4$. From the plot, it can be seen that if $r_b < 5$, the diversity order is
1, and when $r_b \geq 5$, the diversity order is 5. Recall that in Theorem 5.2.1, we have shown
that if $r_b < \tilde{r}_c(r_a)$, then the diversity order is 1, whereas if $r_b \geq \tilde{r}_c(r_a)$, the diversity order
is $r_a - \tilde{r}_c(r_a) + 2$. Hence it can be deduced from the plot that for $\epsilon = 0.2$ and $r_a = 8$,$$
\tilde{r}_c(r_a) = 5.$$

5.3 Outage Probability Analysis for Fractional Co-operation

Section 5.2 illustrated a system design application of the BP. In addition to relay se-
lection, the BP can also be used for analysis; here, to analyze the outage probability
of fractional cooperation. Simulations and density evolution can be used to obtain the
outage probability. As explained before, these methods are, however, complex and time-consuming. Here, an efficient method that can be used to obtain an upper bound on the FER with the use of UBB is presented. As illustrated earlier, the UBB provides upper bounds on the probability of frame error, and is far more efficient than simulations or density evolution. If outage is declared when the channel gives $P_f > P_{f,t}$, where $P_{f,t}$ is a predefined FER threshold, then using (2.24) it follows that

$$P_{out} = \Pr(P_f > P_{f,t})$$

$$\leq \Pr \left( \sum_{h=1}^{n} \bar{A}_h \left( \beta_{SD} \prod_{j=1}^{r} (1 - \epsilon_j (1 - \beta_j)) \right)^h > P_{f,t} \right)$$

$$= \Pr \left( \beta_{SD} \prod_{j=1}^{r} (1 - \epsilon_j (1 - \beta_j)) > \beta_t \right),$$

where $P_{f,t} = \sum \bar{A}_h \beta_h^t$.

To obtain the upper bound on $P_{out}$, random realizations of the S-D, S-R and R-D channels are first generated according to their distributions. The values of $\beta_{SD}$ and the $\beta_j$ associated with each of the $r$ relays are found. The frequency at which the union bound on the FER is above $P_{f,t}$ gives the upper bound on the outage probability. If simulations are used to find the FER, then for each channel realization, the simulations must be ran a number of times in order to obtain an accurate estimation of the average FER, and each simulation involves a pre-determined number of SPA iterations. Meanwhile, obtaining an upper bound using the BP only requires calculating the BP for all the links for each channel realization. Simulations are far more time-consuming and complex than checking whether $\bar{\beta}$ from (5.7) exceeds $\beta_t$. Density evolution is more efficient than simulations, but is relatively time-consuming and complex as well. Hence, using (5.16) is by far one of the most efficient methods to approximate the performance of a relay channel using fractional cooperation, indeed any cooperative scheme.

The use of the BP to obtain an upper bound on the outage probability is shown in Fig. 5.5. Similar to the previous plot, all the channels have the same SNR. Also, $l_S = 2000,$
Figure 5.5: Outage probability from simulation and upper bound on outage probability, where \( r = 4 \) and \( \epsilon_j = 0.3 \), and \( r = 6 \) and \( \epsilon_j = 0.2 \).

where \( R_S = 1/2 \) AND \( R_R = 1/2 \). Simulation results for two scenarios are shown. In the first case, \( r = 4 \) and \( \epsilon_j = 0.3 \), and in the second case, \( r = 6 \) and \( \epsilon_j = 0.2 \). Both simulation results and the upper bound obtained using (5.16) are shown. In both cases, the outage probability obtained from simulations are fairly close to their corresponding upper bounds. Again, it is worth emphasizing that obtaining this upper bound via the BP is very efficient when compared to obtaining the results through simulations.

5.4 Performance Evaluation and Characterization of Relay Channels

Recall that a relay can process the received signal in one of two ways: either it merely demodulates the received signal, or it demodulates and performs limited decoding on the received signal as suggested in Sec. 3.4. Similar to the pure demodulation case, it is
difficult to study the effect of limited decoding on the error performance without resorting to simulations. It is shown here that in addition to a performance measure, the UBB can also be used to approximate the effects of limited decoding. Also, with the use of the BP for DEF, which takes into account the effects of limited decoding, relay selection can be performed in networks where limited decoding is available.

Analysis of relay channels are complicated by their large parameter spaces. For instance, each additional relay adds at least two parameters: the SNR on the S-R link, and the SNR on the R-D link. Thus, including the direct link from source to destination, an $r$-relay system has at least a $(2r+1)$-dimensional parameter space. The analysis is further complicated by other important parameters such as the source and relay code rates and fractions relayed. Given a particular setting of all the parameters, it is not immediately clear whether a given performance criterion, such as FER, can be met. Furthermore, the large parameter space makes it difficult to characterize the entire space by simulation, density evolution, or any other method that examines the space at individual points.

In [46], the authors use the UBB to find the SNR threshold for IR where the decoding error becomes very small and to derive the diversity order of IR in fading channels. Here, the use of UBB for relay channel characterization is illustrated. As stated earlier, the UBB is not limited to DEF, but can be applied to other relay schemes as well. In Sec. 5.2, we calculated $\beta_{\text{DEF}}$, the BP for DEF. Here we illustrate that the UBB approach is more general, and can be applied to, for example, networks employing the relaying scheme AF. In the next section, the calculation of the BP for AF, $\beta_{\text{AF}}$ is shown. Then the application of UBB for characterization of relay channels is illustrated.

### 5.4.1 Limited Decoding

When limited decoding is available at the relay, it can be used to improve the error performance. Let $p_{\text{SR,in}}$ be the bit-flip probability over the S-R channel from (3.1). If only demodulation is performed at the relay, the S-R channel can be modeled as a BSC
with bit-flip probability $p_{SR,in}$. If limited decoding is performed at the relay, then the bit flip probability after relay processing is not equal to $p_{SR,in}$; let $p_{SR,out}$ represent this new bit flip probability after decoding, which can be obtained using density evolution.

As an example, the relationship between the raw (channel) bit-flip probability $p_{SR,in}$ and the equivalent bit-flip probability $p_{SR,out}$ after a given number of SPA iterations for a rate-1/2 punctured systematic RA code is illustrated in Fig. 5.6. In the figure, each curve represents a different number of decoding iterations; the results for 1 to 10, 15, 20, 30, 40 and 50 iterations are shown. As the number of iterations increases, $p_{SR,out}$ decreases for a fixed $p_{SR,in}$. The result for no iterations, where only demodulation is performed at the relay and $p_{SR,out} = p_{SR,in}$, is shown in the plot for comparison (represented by the *dash-dot* line). As illustrated in the plot, even a small number of iterations can improve the bit error rate over the S-R channel significantly. These values can be obtained at the relay either from a function approximating the relationship, or, similar to $\sigma_{R,out}^2$, from a lookup table.

The UBB can also be used to study the improvement that accompanies the introduction of limited decoding. The effect of limited decoding is taken into account while calculating the BP by substituting $p_{SR,out}$ into $p_{S,R}$ in (5.11). In addition, the $\beta_{DEF}$ that takes into account limited decoding can be used for relay selection in networks which allow limited decoding.

The performance improvement with limited decoding in AWGN at the relay is shown in Fig. 5.7. For the simulations shown, the SNR of the channels are $\gamma_{S,D} = -6$ dB, $\gamma_{S,R} = -1$ dB and $\gamma_{R,D} = -4$ dB. In the plot, the BER obtained using density evolution and the upper bound on the BER found using the BP approach are illustrated. It is shown that by performing small number of decoding iterations, e.g., by allowing the relay to perform 10 iterations of the SPA before relaying the source’s bits, the BER can be reduced by more than a factor of 100. Also, the UBB follows the BER curve closely for small BER.
Figure 5.6: Relationship between $p_{SR,in}$ and $p_{SR,out}$, the equivalent bit-flip probability after a given number of decoding iterations for the rate-1/2 punctured systematic RA code.

The FER improvement of relay selection with the use of limited decoding is illustrated in Fig. 5.8. The settings for the fading channel simulation is the same as those from Fig. 5.3, but here it is assumed that all the relays can perform 10 iterations of SPA before forming a hard decision on the relayed bits. The FER curves from Fig. 5.3 are also shown here for comparison, and it can be seen that only 10 iterations of SPA can provide a gain of as much as 3 dB at $FER = 10^{-2}$. For the simulation results shown in Fig. 5.8, the effective $p_{SR,out}$ after decoding is found using density evolution. Note that unlike DF, the decoded data is forwarded even if there are bit errors after decoding.

5.4.2 Bhattacharyya Parameter for AF

Similar to DEF, when AF is used, each relay can also relay only a fraction of the codeword. Assume that $R_j$ chooses to relay each bit with probability $\epsilon_j$. Let the $n$th bit of
the source’s codeword be relayed as the $t_{j,n}$th bit of the symbol vector transmitted by the $R_j$. Then
\[
x_{R_j}(t_{j,n}) = \frac{y_{S,R_j}(n)}{\sqrt{|h_{S,R_j}|^2 + N_{R_j}}}.
\]

At the destination node, the received signal from the relays are combined with the received signal from the source node, each of them scaled accordingly to reflect channel conditions and the noise variance.

With the use of fractional cooperation, each bit can be relayed by any number of relays ranging from 0 to $r$. Let $\hat{\mathcal{R}}_n$ be the set of relays that are relaying bit $n$ of the source’s codeword. Recall that for a Gaussian channel with noise variance $\sigma^2$ and SNR $\gamma = \frac{1}{2\sigma^2}$, the BP is given by $\beta = \exp\{-\gamma\}$. Hence the BP of bit $n$ for the AF coding
Figure 5.8: FER for relay selection based on the BP with 10 iterations of SPA.

scheme can be found by substituting in the equivalent SNR for AF from (2.8):

\[
\beta_{AF} = \exp \left\{ -\gamma_{S,D} - \sum_{j \in \hat{R}_n} \frac{\gamma_{S,R_j} \gamma_{R_j,D}}{\gamma_{S,R_j} + \gamma_{R_j,D} + 1} \right\} 
\]

(5.17)

\[
= \beta_{SD} \prod_{j \in \hat{R}_n} \beta_{AF,j},
\]

(5.18)

where

\[
\beta_{SD} = \exp\{-\gamma_{S,D}\};
\]

\[
\beta_{AF,j} = \exp\left\{ -\frac{\gamma_{S,R_j} \gamma_{R_j,D}}{\gamma_{S,R_j} + \gamma_{R_j,D} + 1} \right\}.
\]

(5.19)

In the case of AF, the SNR of the equivalent channel may be calculated exactly using (2.8). Using this in the BP expression in (5.17), and noting that this expression can be used to verify that a given performance criterion is satisfied, \(\beta_{AF}\) – or the equivalent expression in any other relaying system – may be used as a figure of merit for that system.

Due to the independence of the received signal given the transmitted bit, as illustrated in (5.3), for both AF and DEF the BP “contribution” for each relay can be isolated. For
example, the $\beta_{AF,j}$ and $\beta_{DEF,j}$ values are shown in Fig. 5.9. The contours of various $\beta_j$ values are plotted for different $\gamma_{S,R_j}$ and $\gamma_{R_j,D}$ values. In the figure, the solid lines represent values of $\beta_{AF,j}$ and the dash-dot lines represent values of $\beta_{DEF,j}$. In most cases, $\beta_{DEF,j} < \beta_{AF,j}$, except when $\gamma_{R_j,D}$ is much larger than $\gamma_{S,R_j}$, as indicated by the crossover points in the plot. The BP can be used to compare the performance the AF and DEF, and helps us estimate the conditions under which it is advantageous to use AF or DEF. If a relay is capable of both AF and DEF, then it can choose the relaying scheme that will provide the best error performance given the channel conditions. For example, if the $R_j$-D channel is much larger than the S-R$\_j$ channel, AF is more favorable; otherwise DEF should be used to obtain the optimal performance.

5.4.3 Decodable region

Another interesting application of the BP analysis is to specify the region in the $(2m+1)$-dimensional SNR space where a given error rate can be achieved. Depending on the
relaying scheme, we can then substitute (5.19) or (5.14) into $\beta_j$ of (5.7) to obtain $\bar{\beta}$, the BP for fractional cooperation. Assume that it is desirable to have the FER below a threshold $P_{f,t}$, and let $\beta_t$ be the value that satisfies $P_{f,t} = \sum_{h=1}^{n} A_h \beta_t^h$. Then it is easy to see that as long as the condition

$$\bar{\beta} = \prod_{j=1}^{r} (1 - \epsilon_j (1 - \beta_j)) \leq \beta_t$$

is satisfied, the FER is below $P_{f,t}$. The set of SNRs where the condition in (5.20) is satisfied defines the region in the $(2r + 1)$-dimensional SNR space where the FER requirement is fulfilled.

For source’s codewords with asymptotically large blocklength, if the asymptotic UBB threshold $c_0$ exists and is finite, it can be used to estimate the SNR region where the FER approaches zero [58]. From (2.26), it follows that successful decoding can be achieved if the condition

$$\beta_{SD} \prod_{j=1}^{r} (1 - \epsilon_j (1 - \beta_j)) < \exp(-c_0)$$

is satisfied.

Figures 5.10 and 5.11 illustrate the use of the BP to identify the decoding region. For the simulation results shown, a rate-1/2 punctured systematic RA code is used, and the blocklength of the source’s codeword, $l_s$ is 16000. In Fig. 5.10, only one relay is used, and the data is transmitted over AWGN channels, where the SNR is the same over all channels. Simulation results for both AF and DEF are shown, and the cases where $\epsilon_j = 1$ and $\epsilon_j = 1/2$ are illustrated. When DEF is used, it is assumed that only pure demodulation is performed at the relay, where $\epsilon_{info} = \epsilon_{par} = 1$ and $R_R = 1$, i.e., the relay demodulates and repeats the data. The SNR thresholds corresponding to the asymptotic UB threshold $c_0$ of the rate-1/2 punctured systematic RA code for AF and DEF with $\epsilon_j = 1/2$ and $\epsilon_j = 1$ are also indicated. As illustrated in the simulation results, DEF performs better than AF when the instantaneous SNR is the same over all the channels. In all the cases, the waterfall of the FER is about 0.5 dB away from the SNR threshold.
corresponding to \( c_0 \).

In Fig. 5.11, the FER for multiple relays in fading channels are shown. In this case, \( m = 6 \), \( \epsilon_j = 0.2 \) and the average SNR is the same over all channels. The probability of (5.21) not being satisfied, or equivalently, the outage probability \( P_{\text{out}} \) for both AF and DEF are shown in the plot. Clearly, \( P_{\text{out}} \) follows the FER closely. Although the outage probability only provides an upper bound on the FER, and does not provide an exact expression for the FER, it can be used to understand the diversity order of the system. As with the previous scenario, DEF performs better than AF when the average SNR over all the channels is the same.

5.5 Lower Bound on \( r_c \) for FC

In the previous sections, we illustrated the use of the UBB to analyze relay networks. Here we illustrate its use in providing a lower bound on the value of \( r_c \) when the fraction relayed by each relay \( R_j \) is \( \epsilon \).

**Theorem 5.5.1** A lower bound on \( r_c \) in a system where \( r \) relays are present, where each relay relays \( \epsilon \) of the source’s codeword, is given by

\[
\tilde{r}_c = \left\lceil \frac{\log(\beta_t)}{\log(1 - \epsilon)} \right\rceil.
\]

**Proof:** Assume that \( r \) relays are present in the system. From (5.20), for regular fractional cooperation, the condition \( \beta_{SD} \prod_{j=1}^{r} (1 - \epsilon(1 - \beta_j)) \leq \beta_t \) is a necessary condition for no outage. Hence if an outage occurs, the condition

\[
\beta_{SD} \prod_{j=1}^{r} (1 - \epsilon(1 - \beta_j)) > \beta_t
\]

is true. Recall from the definition given in Lemma 4.2.1, \( r_c \) represents the smallest number such that if the number of relays is less than \( r_c \) and \( \gamma_{S,D} = 0 \), then a system outage always occurs. Note that when \( \gamma_{S,D} = 0 \), \( \beta_{S,D} = 1 \). If \( r < r_c \), an outage occurs
Figure 5.10: FER for AF and DEF with blocklength $l_s = 16000$ and their corresponding SNR thresholds (dashed lines) derived from $c_0$ of the rate-1/2 RA code.

Figure 5.11: FER for AF and DEF with blocklength $l_s = 16000$ and $P_{out}$ of the rate-1/2 RA code with $m = 6$ and $\epsilon_j = 0.2$. 
Table 5.1: Lower bound on $r_c$ for fractional cooperation using rate-1/2 punctured systematic RA code for various $\epsilon$ derived from the UBB.

<table>
<thead>
<tr>
<th>$\epsilon$</th>
<th>0.1</th>
<th>0.125</th>
<th>0.15</th>
<th>0.2</th>
<th>0.25</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{r}_c$</td>
<td>10</td>
<td>8</td>
<td>7</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

for all channel conditions, even when all the S-R and R-D channels are perfect, where $\beta_j = 0$ for all $j$. The condition in (5.23) with $\beta_{SD} = 1$ and $\beta_j = 0$ simplifies to

$$(1 - \epsilon)^r > \beta_i.$$ (5.24)

After some simple manipulations, this condition is equivalent to

$$r > \frac{\log(\beta_i)}{\log(1 - \epsilon)}.$$ 

Let $\hat{r}_c = \left\lceil \frac{\log(\beta_i)}{\log(1 - \epsilon)} \right\rceil$. As $r < r_c$ is a sufficient condition for an outage, which is a sufficient condition for $r \geq \hat{r}_c$, $r < r_c$ is a sufficient condition for $r \geq \hat{r}_c$, and $\hat{r}_c$ is an lower bound on the value of $r_c$.

With the use of this result, the minimum number of relays required to provide a diversity order greater than one can be found. This is extremely useful as a design tool when fractional cooperation is used. A table of lower bounds on $r_c$ for the rate-1/2 punctured systematic RA code is given in Table 5.1, where the asymptotic value $c_0$ is used to obtain $\beta_i = \exp\{-c_0\} = 0.4$. Comparing Table 4.1 and Table 5.1, it can be observed that the values shown in Table 4.1 are the same as those in Table 5.1. We have hereby provided lower bounds to $r_c$ through two different methods: by evaluating the asymptotic UBB threshold $c_0$, and by evaluating $p_E$ obtained through density evolution.

### 5.6 Limitations

In this chapter we developed the use of the UBB to design and analyze relay channels. The applications provided here illustrate the many potential uses of this framework. How-
ever, this study of design and analysis is done here under two major assumptions: the relay’s codeword is not punctured, and serial decoding is performed at the destination. Both assumptions simplify the required analysis. When the codeword is not punctured, the LLR output from decoding the relay’s codeword can be closely approximated by the Gaussian distribution, allowing us to model the equivalent channel distribution after decoding with a Gaussian distribution and still obtain a fairly accurate result. If puncturing were used on the relay’s codeword, the equivalent channel distribution after decoding resembles a mixture of Gaussians. The distribution can still be approximated with a Gaussian distribution, but the approximate description is not as accurate as in the case without puncturing [73], and might lead to performance degradation in the case of relay selection, and possibly inaccurate analysis. The effect of this approximation, however, is outside the scope of this thesis, and further investigation on this topic is recommended.

Under the second assumption, the BP can be computed in two steps: first, find the effect of decoding the relay’s codeword, and second, find the effect of the imperfect S-R channel. As developed here, BP cannot be used for relay selection or analysis when parallel decoding is used at the destination node, where the source’s and relay’s codewords are decoded simultaneously with the SPA messages passed between the two encoders after each iteration. Obtaining the BP under parallel decoding is significantly more complex.

In addition, the union bound closely follows the error performance only at high SNR, and otherwise it is a fairly loose bound. Tighter bounds can be used to provide a better approximation of the performance of various relaying schemes, and examples of these bounds can be found in [58,74]. These better approximations, however, come at a price of more complex calculations. However, even with these assumptions and constraints, the BP approach is still an efficient method that can be used to observe the effect of various parameters without resorting to simulations or density evolution, which are comparatively time-consuming and complex.

Finally, similar to the analysis performed on fractional cooperation in Ch. 4, some
of the analysis results shown here are asymptotic results, where it is assumed that the blocklength of the source’s codeword \( l_S \to \infty \). This is unrealistic as an infinite blocklength in fading channel translates to infinite coherence time. Hence care must be taken while interpreting the results presented here, and the effects of finite blocklength must be investigated.
Chapter 6

Resource Allocation in Relay Networks

When fractional cooperation is applied to multiple-source, multiple-relay wireless systems, an interesting problem is raised: in a network comprising multiple source nodes and multiple relay nodes, how much of its available resources should each relay devote to different source nodes? The most accurate method is to perform an exhaustive search over all possible assignments. However, this is extremely inefficient and hence impractical to implement. In Ch. 5, an upper bound on the FER for fractional cooperation was presented, and with the use of this upper bound, the condition under which the average FER is below a given threshold can be found. This provides us with a method of easily checking whether the given parameter and channel condition allows the average FER to fall below a desired threshold, thereby guaranteeing successful communication.

In this chapter, two optimization problems on the performance of multiple-source, multiple-relay networks where fractional cooperation is used, are presented: in the case where the available resources allows all the source’s codewords to be decoded correctly, the energy consumption is minimized; on the other hand, if the resources does not allow correct decoding of all the source’s codeword, the worst FER over all source nodes is
minimized. Our approach in these problems is to optimize the energy used or the maximum FER over all the source nodes using the UBB, which has the advantage of being much easier to calculate than the true FER.

### 6.1 Optimization Problem Formulation

The system under consideration comprises $s$ source nodes and $r$ relay nodes. In this multiple-source, multiple-relay, system, it is desirable to optimize the allocation of the relay resources to obtain the best results. In this section, two optimization problems are presented. In the first problem, it is assumed that there is a minimum FER threshold that must be satisfied, for example, to avoid an outage. Assuming that the resources available is more than what is required to satisfy the FER threshold for all the source nodes, it would be optimal to minimize the amount of energy expended while still satisfying the FER constraint. In the second optimization problem, it is assumed that resources available do not allow the FER constraint to be satisfied for all the source nodes. Instead of optimizing the energy consumption, it is now desirable to ensure the relay resources are distributed in a manner such that the maximum FER over all source’s codewords is minimized. The formulation of the two problems is presented below.

For simplicity, it is assumed that the length of the source’s codewords is the same and is denoted as $l_S$. Let $\bar{\epsilon}_j$ be the maximum contribution $R_j$ is willing to provide. In addition, it is assumed that no encoding is available at the relay. A variation on these problems, where it is assumed that the relays can change the code rate of the relay’s codeword, is presented in a later section.

#### 6.1.1 Optimizing Energy Consumption

In this optimization problem, the goal is to minimize the total energy consumption while still achieving the threshold $P_{f,t}$ for all the source nodes. Let $\epsilon_{j,i}$ be the fraction of $c_{S_i}$
Chapter 6. Resource Allocation in Relay Networks

relayed by $R_j$. Given the threshold $\beta_t$, where $P_{f,t} = \sum \bar{A}_h \beta_t^h$, the optimization problem can be formulated as follows

$$\min_{\epsilon_{j,i}} \sum_{i=1}^{s} \sum_{j=1}^{r} \epsilon_{j,i}$$  \hspace{1cm} (6.1)$$

subject to

$$\beta_{S_i,D} \prod_{j=1}^{r} (1 - \epsilon_{j,i}(1 - \beta_{j,i})) \leq \beta_t \hspace{1cm} i = 1, \ldots, s$$

$$\sum_{i=1}^{s} \epsilon_{j,i} \leq \bar{\epsilon}_j \hspace{1cm} j = 1, \ldots, r$$

$$\epsilon_{j,i} \geq 0 \hspace{1cm} \forall i, j.$$  

where $\beta_{S_i,D} = \exp\{\gamma_{S_i,D}\}$ is the BP over the $S_i$-D channel, and $\beta_{j,i}$ is the BP over the $S_i$-$R_j$-D link. Here the first set of constraints are denoted as source constraints, and the second set of constraints are denoted as relay constraints, as they are imposed by the source and relay nodes respectively. Note that since the set of points that satisfy the source constraints is non-convex, this is a non-convex optimization problem.

### 6.1.2 Optimizing Error Rate

In this optimization problem, the goal is to minimize the maximum error rate over all source nodes, while not exceeding the amount of resources available. The optimization problem is formulated as follows

$$\min_{\epsilon_{j,i}} \tilde{\beta}_t$$  \hspace{1cm} (6.2)$$

subject to

$$\beta_{S_i,D} \prod_{j=1}^{r} (1 - \epsilon_{j,i}(1 - \beta_{j,i})) \leq \tilde{\beta}_t \hspace{1cm} i = 1, \ldots, s$$

$$\sum_{i=1}^{s} \epsilon_{j,i} \leq \bar{\epsilon}_j \hspace{1cm} j = 1, \ldots, r$$

$$\epsilon_{j,i} \geq 0 \hspace{1cm} \forall i, j.$$  

With the use of the dummy variable $\tilde{\beta}_t$, we ensure that the relay assistance is distributed so as to minimize the maximum error rate in order to provide fairness of resource allo-
6.2 Algorithm for Distributed Optimization

As mentioned earlier, both the energy and error rate optimization problems are non-convex, and there are no known algorithms guaranteed to converge to the global optimum. In addition, if formulated as a single global optimization problem, one node, such as the destination node, must collect all the information related to the allocation of resources in order to distribute them in the optimal manner. To address these issues, in this section an implementation of a distributed optimization algorithm is provided. The implementation presented below can be used to solve the energy optimization problem, where the resources available allows the source constraints to be satisfied.

The distributed optimization problem is designed in the same spirit as belief propagation in factor graphs. In large, sparsely populated wireless networks, each source node only requests help from the few relay nodes that are closest to it, and each relay node only receives requests for assistance from the few source nodes that are closest to it. The optimization problem can be broken down into multiple localized optimization problems, where the complexity of each is significantly reduced. We allow source and relay nodes to make a judgment on the best way to distribute the available resources based on its knowledge of the local network. However, similar to belief propagation, optimal results cannot be guaranteed in graphs where cycles exist.

Recall that the energy optimization problem consists of two sets of constraints: the source and relay constraints. In the distributed algorithm, this optimization problem is broken down into two subproblems. In the first subproblem, each relay divides up its resources in an optimal manner to reserve a fraction to assist each source that requests its help. This is done to decouple the relay constraints such that the optimization
problem can be solved in a distributed manner. In the second subproblem, each source node attempts to find a solution to minimize the total energy consumed given its source constraint and the resources reserved by the relays. These two subproblems are solved iteratively until a solution is found (or until it is realized that no solution is available).

### 6.2.1 Definitions and Border Points

Up to this point, it is unclear what method should be used to divide each relay’s resources such that the appropriate amount is reserved for each source node. One way to divide up the resources is presented here and justification for such a division is provided.

Before providing a description of the division of relay resources and the distributed optimization algorithm, definitions for some of the terms used are provided. Let \( \bar{\epsilon}_{j,i} \) be the resources \( R_j \) has reserved for use by \( S_i \), where \( \sum_i \bar{\epsilon}_{j,i} \leq \bar{\epsilon}_j \). Basically, \( R_j \) pre-divides its resources \( \bar{\epsilon}_j \) into \( \bar{\epsilon}_{j,i} \) for use by each source node \( S_i \). As suggested earlier, this allows the optimization problem in (6.1) to be divided into smaller subproblems that can be solved at each source node.

During optimization, each source node \( S_i \) needs to calculate the maximum and minimum required resources for relay \( j \), denoted as \( M_i(j) \) and \( m_i(j) \) respectively. The value \( M_i(j) \) represents the fraction of \( R_j \)'s codeword that \( S_i \) wants to have assigned to assist in the relaying of the codeword originating from \( S_i \) in order to use the minimal energy. In other words, \( S_i \) calculates the resources it requires from each relay to satisfy its source constraint while consuming minimal energy, where \( M_i(j) \) is the resource it wants from \( R_j \). Equivalently, \( \mathbf{M}_i = [M_i(1), M_i(2), \ldots, M_i(r)] \) is the solution to the following revised
optimization problem at $S_i$:

$$\min_{\epsilon_{j,i}} \sum_{j=1}^{r} \epsilon_{j,i} \quad (6.3)$$

subject to

$$\beta_{SD,i} \prod_{j=1}^{r} (1 - \epsilon_{j,i}(1 - \beta_{j,i})) \leq \beta_t$$

$$\epsilon_{j,i} \leq \bar{\epsilon}_j \quad j = 1, \ldots, r$$

$$\epsilon_{j,i} \geq 0. \quad j = 1, \ldots, r.$$  

The minimum required resources, $m_i(j)$, is the minimum fraction of $R_j$'s codeword needed to relay the codeword of $S_i$ in order to satisfy its source constraint, given that all the available constraints from other resources have been exhausted:

$$m_i(j) = \begin{cases} 
0 & \text{if } \beta_{SD,i} \prod_{j' \neq j} (1 - (1 - \beta_{j',i})\bar{\epsilon}_{j'}) \geq \beta_t, \\
\frac{1}{1 - \beta_{j,i}} \left(1 - \frac{\beta_t}{\beta_{SD,i}} \right) & \text{otherwise.}
\end{cases} \quad (6.4)$$

As shown in the above equation, if its source constraint can be satisfied even if $\bar{\epsilon}_j = 0$ with the given values of $\bar{\epsilon}_{j'}$, for $j' \neq j$, then $m_i(j) = 0$.

Let

$$\hat{\epsilon}_{j,i} = \frac{1}{1 - \beta_{j,i}} \left(1 - \frac{\beta_t}{\beta_{SD,i}} \right) \quad (6.5)$$

be the upper bound of the resources $S_i$ requests of $R_j$ to satisfy its source constraint. For each relay $R_j$ the vector $q_{j,k} = [q_{j,k}(1), q_{j,k}(2), \ldots, q_{j,k}(s)]$, where $q_{j,k}(i)$ corresponds to the resources allocated by $R_j$ to $S_i$, is a border point if each source node belongs to exactly one of the three index sets: $W_{M,j}$, $W_{m,j}$ and $W'_j$. The source node index are sorted into the corresponding set depending on the condition it satisfies:

- If $q_{j,k}(i) = M_i(j)$, then $i \in W_{M,j}$,
- If $q_{j,k}(i) = m_i(j)$, then $i \in W_{m,j}$,
- If $q_{j,k}(i) \in (m_i(j), M_i(j))$, then $i \in W'_j$, where $|W'_j| \leq 1$, i.e. $W'_j$ can have at most one element.
and let the set of all border points at $R_j$ be denoted as $S_{BP,j}$. Assume that the globally optimal solution is

$$
\epsilon^* = [\epsilon^*_{1,1}, \epsilon^*_{1,2}, \ldots, \epsilon^*_{1,s}, \epsilon^*_{2,1}, \ldots, \epsilon^*_{2,s}, \ldots, \epsilon^*_{r,s}].
$$

Next, separate $\epsilon^*$ into vectors $\epsilon^*_j$, which is associated with $R_j$, where

$$
\epsilon^*_j = [\epsilon^*_j, \ldots, \epsilon^*_j,s] \quad \forall j.
$$

Then it can be shown that $\epsilon^*_j \in S_{BP,j}$, i.e., the optimal solution is a border point. The proof is provided in Appendix D.

By establishing the fact that $\epsilon^*_j \in S_{BP,j}$, we can restrict the search area for the optimal solution by allowing all the relays to only allow border points to be assigned as reserved resources, $\bar{\epsilon}_{j,i}$, in the local optimization problem in (6.3). Through this, we guarantee that the globally optimal point is included in the search space used in this distributed optimization algorithm.

### 6.2.2 Distributed Algorithm

Clearly, while solving the local optimization problem in (6.3), assigning the resources of one relay to source node $S_i$ reduces the amount of resources it can allocate to other source nodes, $S_{i'}$, for $i' \neq i$. Since the optimization problem has been decoupled by having relays reserve resources for each source node, it is extremely difficult to keep track of such changes. In the distributed optimization algorithm the calculation is simplified by first making the assumption that the available resources from other relays does not change while solving the local optimization problem in (6.3). Assume that in the $n$th step of the iterative optimization algorithm, the resources available at $R_j$ is $\bar{\epsilon}_{j,(n)}$, and $\bar{\epsilon}_{j,(1)}$ is the original available resources at $R_j$. The steps involved in $n$th iteration of are described below:

1. Each source node $S_i$ finds the maximum required resources $M_i$, which is the solution
to the optimization problem in (6.3) where $\bar{\epsilon}_j = \bar{\epsilon}_{j,(n)}$. It then finds the minimum required resources $m_i$ using (6.4). The values $M_i(j)$ and $m_i(j)$ are passed to $R_j$.

2. After receiving the $M_i(j)$ and $m_i(j)$ from each source node, $R_j$ finds every border point, and the $k$th border point is denoted as $q_{j,k}$, where $q_{j,k}(i)$ is the component associated with $S_i$. The set of values $q_{j,k}(i)$ corresponding to $i$th component of $q_{j,k}$ are then sent to the respective source nodes.

3. For each $q_{j,k}(i)$ received from $R_j$, $S_i$ solves the optimization problem in (6.3) by substituting $\bar{\epsilon}_j = q_{j,k}(i)$. As stated earlier, in order to simplify the optimization process, the values $\bar{\epsilon}_{j'}$ for $j' \neq j$ remain unchanged. If a solution exists and is given by $[\epsilon_{1,i}^+, \ldots, \epsilon_{J,i}^+]$ then set $E_{j,k,i} = \sum_{j'=1}^s \epsilon_{j',i}^+$. If no solution can be found, then set $E_{j,k,i} = \infty$.

4. The calculated minimum energy associated with each border point from each source node are collected by $R_j$ to obtain the summed value $E_{j,k} = \sum_i E_{j,k,i}$. Let

$$k_j^* = \arg\min_k E_{j,k}$$

be the index of the border point with minimum summed energy over all border points, and $q_{j,k_j^*}$ be the associated border point vector. Let $S_{M,j}$ be the index set of source nodes where $q_{j,k_j^*}(i) = M_i(j)$. Then for all $i$ in $S_{M,j}$, $M_i(j)$ of relay $j'$’s codeword is allocated to relay $S_i$’s codeword. The available constraints are updated via

$$\bar{\epsilon}_{j,(n+1)} = \bar{\epsilon}_{j,(n)} - \sum_{i \in S_{M,j}} M_i(j). \quad (6.6)$$

5. Check to see whether all the source node constraints have been satisfied, or all the relays’ resources have been exhausted. If either condition is true, then this process is terminated. Otherwise, repeat the process using the updated constraints $\bar{\epsilon}_{j,(n+1)}$. 

This algorithm used to minimize the energy consumed to satisfy a given set of constraints can also be used for error rate optimization. An initial value of $\tilde{\beta}_t = \beta_t$ is set, and the algorithm is used to check whether the given constraints can be satisfied. If so, then $\tilde{\beta}_t$ is increased; otherwise $\tilde{\beta}_t$ is decreased. The algorithm is then run iteratively with the adjusted $\tilde{\beta}_t$ value until the difference between the minimum $\tilde{\beta}_t$ value where the condition can be satisfied and the maximum $\tilde{\beta}_t$ value where the condition cannot be satisfied is within a pre-defined value $\delta > 0$.

6.3 Simulation Results

In this section, simulation results are presented to show that by optimizing the distribution of relay resources, the error performance of the system can be improved while minimizing the energy consumption. In all the simulations presented here, a punctured systematic rate-1/2 RA code is used for encode the source’s symbols, where $l_S = 4000$ and $R_R = 1$. In Fig. 6.1, the FER in an AWGN channel for a two-source two-relay system is shown, where the SNR between the nodes are given by

$$
\gamma_{S_1,D} = -4 \text{ dB}, \quad \gamma_{S_2,D} = -2 \text{ dB}, \quad \gamma_{R_1,D} = 0 \text{ dB}, \quad \gamma_{R_2,D} = 0 \text{ dB},
$$

$$
\gamma_{S_1,R_1} = -2 \text{ dB}, \quad \gamma_{S_1,R_2} = -2 \text{ dB}, \quad \gamma_{S_2,R_1} = -1 \text{ dB}, \quad \gamma_{S_2,R_2} = -1 \text{ dB}.
$$

In this case, the resources available are not adequate to ensure correct encoding at both source nodes. All the resources are exhausted to minimize the maximum FER over both source nodes, as indicated in (6.2). In the plot, the contours for the maximum FER over both source nodes for different $\epsilon_{1,1}$ and $\epsilon_{2,1}$ values are shown, where $\bar{\epsilon}_1 = \bar{\epsilon}_2 = 0.8$. As all the resources are exhausted in this scenario, $\epsilon_{1,2} = \bar{\epsilon}_1 - \epsilon_{1,1}$ and $\epsilon_{2,2} = \bar{\epsilon}_2 - \epsilon_{2,1}$. In the plot, the marker indicates the optimal values of $\epsilon_{1,1}$ and $\epsilon_{2,1}$ obtained from the optimization problem (6.2), where $\epsilon_{1,1} = \epsilon_{2,1} = 0.63$, and the FER is 0.2933 for both source nodes. As shown in the plot, the minimization of the maximum FER is achieved, where resources are distributed such that both source nodes have the same FER, even though $S_2$ has better
Figure 6.1: Contours of maximum FER for a two-source two-relay network with AWGN.

links to the relays and the destination node than \( S_1 \). This optimization is therefore extremely useful when the link quality is very different for the source nodes and fair resource allocation is desired.

In the next set of simulation results, the minimization of energy consumption while meeting the FER constraints in fading channels is illustrated. Similar to the previous plot, this system consists of two source and two relay nodes. In addition, \( \bar{\epsilon}_1 = \bar{\epsilon}_2 = 0.8 \). In Fig. 6.2, the maximum FER over the two source nodes are shown. The average SNR for the source-destination links are varied, while the average SNR over the rest of the links are

\[
\bar{\gamma}_{S_1,R_1} = -2 \text{ dB} \,, \quad \bar{\gamma}_{S_2,R_1} = -1 \text{ dB} \,, \quad \bar{\gamma}_{R_1,D} = 0 \text{ dB} \,, \\
\bar{\gamma}_{S_1,R_2} = -2 \text{ dB} \,, \quad \bar{\gamma}_{S_2,R_2} = -1 \text{ dB} \,, \quad \bar{\gamma}_{R_2,D} = 0 \text{ dB} \,.
\]

In the plot, the solid lines represent contours the maximum average FER for the case
Figure 6.2: Contours of maximum FER for two-source two-relay network with fading channels using optimization (solid) and equal allocation (dash-dot).

where optimization is performed, and the dash-dot lines represent contours for the maximum average FER for the case where the resources are distributed evenly among the two source nodes. For all the cases, optimization improves the maximum average FER over both source nodes compared to the case when equal allocation is used.

In Fig. 6.3, the value of $\sum_{i,j} \epsilon_{j,i}$ corresponding to Fig. 6.2 is shown. From these plots, it can be observed that even though the energy used with optimization is less than that if the resources were distributed evenly without optimization, the maximum FER is smaller compared to the case where no optimization is performed as well, where $\sum_{i,j} \epsilon_{j,i} = \bar{\epsilon}_1 + \bar{\epsilon}_2 = 1.6$.

Simulation results are shown in Table 6.1 to illustrate the performance of the distri-
Figure 6.3: Plot of $\sum_{i,j} \epsilon_{j,i}$ under optimization for a two-source two-relay network in fading channels.

The distribution algorithm introduced in Sec. 6.2. In the 3-source 3-relay network,

$$\bar{\gamma}_{S_i,D} = -2 \text{ dB} \quad \text{for} \quad i = 1, 2, 3,$$
$$\bar{\gamma}_{S_i,R_j} = 1 \text{ dB} \quad \text{for} \quad i \neq j, \quad \bar{\gamma}_{S_i,R_j} = 2 \text{ dB} \quad \text{for} \quad i = j,$$
$$\bar{\gamma}_{R_j,D} = 0 \text{ dB} \quad \text{for} \quad j = 1, 3, \quad \bar{\gamma}_{R_j,D} = 1 \text{ dB} \quad \text{for} \quad j = 2.$$

and $\bar{\epsilon}_j = 0.8$ for $j = 1, 2, 3$. In addition to the distributed algorithm, simulation results obtained using sequential quadratic programming (SQP) [75] is also illustrated. As mentioned earlier, since the optimization problem is a non-convex optimization problem, different initial points used in the conventional optimization schemes may generate different solutions. Hence random points are generated and used as initial points for optimization while using SQP; SQP20 uses 20 randomly generated initial points to run the optimization algorithm, and the best result is chosen to run simulations on the rate-1/2 RA code. Using simulations for 500 randomly generating channel coefficients for the given setting, it was found that the optimal solution found from 20 random points (SQP20) coincides with that found from 200 random points (SQP200) about 85% of the
time. The distribution of the number of random points used to arrive at the same optimal point as that from SQP200 is shown in Fig. 6.4. We believe that SQP20 can be used to approximate the global optimum, without requiring the number of optimizations to be exceedingly large. Various parameters such as FER, maximum FER over all source nodes, energy consumed to assist each source node, and total energy consumed by the relays are shown in the table. As expected, the maximum FER over all source nodes is smallest with SQP20, where the total energy consumed is smallest as well. Even though the maximum FER and energy consumed is larger with our distributed optimization algorithm, the difference is quite small for this 3-source 3-relay network.

Recall that when the amount of resources available cannot support the desired $\beta_t$, the maximum average FER over all source nodes must be minimized given the relay resources. A binary search, where the upper or lower bound on the search region is reduced by half every iteration, can be used to find the value of $\tilde{\beta}_t$ such that the source
### 6.4 Discussion

In this chapter we used the BP approach to develop an optimization formulation for resource allocation in the context of fractional cooperation. Here a distributed algorithm was developed that proves effective in reducing energy consumption. As stated earlier, this algorithm does not guarantee the global optimum. However, the fact that it is a distributed algorithm makes it an excellent method to obtain a solution in large networks such as WSNs, as it scales well with the network size. As illustrated in the simulation re-
results, any optimization on the energy usage and error performance allows the network to improve its performance, even if the optimal solution cannot be guaranteed. The conditions under which the global optimal can be obtained, as well as the rate of convergence, should be investigated in future studies.

Instead of using repetition codes, where the relay merely repeats the bits chosen to relay, DEF allows the relays to form codewords based on the chosen bits and transmit to the destination node, with examples including LDGM and RA codes as presented in Ch. 3. In this case, the cost of transmitting one bit is not equal for all the nodes anymore, and is scaled by the reciprocal of the rate of the channel code used. If a rate $R_{S_i,R_j}$ code is used for the bits relayed by $R_j$ for $S_i$, the optimization becomes

$$
\min \sum_{i=1}^{l} \sum_{j=1}^{J} \frac{\epsilon_{j,i}}{R_{S_i,R_j}} \quad (6.7)
$$

subject to

$$
\beta_{SD,i} \prod_{j=1}^{r} (1 - \epsilon_{j,i}(1 - \beta_{j,i})) \leq \tilde{\beta}_t \quad i = 1, \ldots, s
$$

$$
\sum_{i=1}^{s} \frac{\epsilon_{j,i}}{R_{S_i,R_j}} \leq \bar{\epsilon}_j \quad j = 1, \ldots, r
$$

$$
\epsilon_{j,i} \geq 0 \quad \forall i, j.
$$

where in this case, $\beta_{j,i}$ is a function of $R_{S_i,R_j}$.

As the dependence of $\beta_{j,i}$ on $R_{S_i,R_j}$ varies for different source-relay and relay-destination link quality, it is difficult to characterize the relationship between $\beta_{j,i}$ and $R_{S_i,R_j}$. To simplify the optimization, the problem can be solved by separating it into two subproblems. In the first subproblem, the optimal values of $\epsilon_{j,i}$ are obtained from solving (6.7) for a fixed rate $R_{S_i,R_j}$ (and hence fixed $\beta_{j,i}$) for each pair of source and relay nodes. The optimal values of $\epsilon_{j,i}$ are denoted as $\epsilon^*_{j,i}$, and the value used in the optimization problem is denoted as $R^*_{S_i,R_j}$. In the second subproblem, optimal values of $\epsilon_{j,i}$ and $R_{S_i,R_j}$ are solved for each source-relay pair, given the channel qualities over the S-R and R-D link. This is
done by solving the following optimization problem for each $\epsilon_{j,i}$

$$\min \quad (1 - \epsilon_{j,i}(1 - \beta_{j,i}))$$

subject to

$$\frac{\epsilon_{j,i}}{R_{S_i,R_j}} \leq \frac{\epsilon^*_{j,i}}{R^*_{S_i,R_j}}$$

such that the fraction used by each relay for each source is fixed. In some cases, this optimization subproblem can be difficult to solve. An example is when DEF is used, where the relationship between $R_{S_i,R_j}$ and $\beta_{j,i}$ cannot be derived easily. In these cases, instead of allowing the range of $R_{S_i,R_j}$ to be continuous, the restriction of only allowing finite number of possible value for $R_{S_i,R_j}$ can be imposed. Then $\beta_{j,i}$ can be calculated for each possible value of $R_{S_i,R_j}$, and the optimal value is selected. After obtaining the optimal $R_{S_i,R_j}$ for each source-relay pair, the value is substituted into the first subproblem and the process is repeated until the change in $\sum_{i,j} \epsilon_{j,i}/R_{S_i,R_j}$ is less than $\delta$, where $\delta > 0$.

As both of these optimization subproblems provide solutions that are non-increasing, a local minimum can be found.

As suggested in Ch. 4, a node can be a source and a relay node simultaneously. Even though a distinction is made between source and relay nodes in the above derivations, this distinction can be removed when applied to networks where a node can act as both a source and relay. In this case, the resources available to a node will be divided between its role as a source and as a relay. This extension, however, is outside the scope of this dissertation and further studies are required for investigation.
Chapter 7

Conclusions

7.1 Summary

In this dissertation we developed system concepts and related analysis tools for networks with relatively simple nodes. Specifically, we relax two assumptions that are prevalent in current literature on relay channels. The first assumption is that the relaying nodes possess hardware that allow them to perform complex tasks such as decoding the source’s codeword and/or storing large amount of data, which is required if DF or AF is used respectively. The relaying scheme DEF is introduced, where relaying nodes only perform the simple task of demodulating the received bits, and because each transmitted symbol is only represented by one bit, a large amount of storage is not needed to store the data after demodulation and before transmission to D. Even though only simple operations are performed at the relays, a diversity order of two can be obtained with one relay. In order to improve the performance, encoding can be performed on the demodulated codes. Also, if the hardware allows it, limited decoding can be performed at the relay to improve the effective S-R channel. The flexibility of DEF, where the complexity of the scheme can be changed depending on the resources available, makes it a robust relaying scheme.
The second assumption is that relaying nodes must either relay the complete source’s codeword or not relay at all. The concept of fractional cooperation, where each relay can transmit only part of a source’s codeword is proposed and discussed in detail here. The condition which must be satisfied to achieve a diversity order of two with one relay is presented. When multiple relays are used to assist S, each relay can relay only part of the source’s codeword, and together provide assistance to S. It has been shown in [2] that as long as the number of relays present is larger than $r_c$, each additional relay increases the diversity order by one. Outage set analysis on fractional cooperation is also presented. The parameter $r_c$ associated with fractional cooperation depends on various parameters such as the encoding and decoding schemes, the relaying schemes, and the fraction of the source’s codeword relayed. Using the analysis results on outage sets, a lower bound on the value of $r_c$ can be found.

In addition to relaxing the two assumptions frequently used in relay channels, this dissertation developed the UBB to design and analyze relay channels. The UBB allows one to analyze the network accounting for the specifics of this implementation. One of the applications of UBB is relay selection. While relay selection for DF and AF has well-defined criteria, choosing the optimal relay for DEF can be difficult due to the demodulation at the relay. The BP can be used as a metric to choose the relay so as to minimize the error rate. Not only can BP be used to select one relay, it can also be used to select multiple relays when fractional cooperation is applied to the relay channel. In addition to relay selection, the UBB can be used to provide an upper bound of the performance of fractional cooperation in fading channels. It can also be used as an efficient method to characterize the relay channels and to assess the performance of the relaying scheme with the resources available. Lastly, it was shown that the UBB can also be used to derive an lower bound on the value of $r_c$.

Finally, this dissertation formulated and solved certain optimization problems for resource allocation with fractional cooperation. Two optimization problems were pre-
presented: (a) minimize energy consumption while ensuring the average FER is below a pre-defined threshold for all source nodes, and (b) minimize the maximum FER over all source nodes while ensuring that only the available resources are used. A sub-optimal distributed algorithm is also introduced to solve the optimization problems. By optimizing the resource allocation, system performance such as energy consumption and maximum FER can be improved.

7.2 Suggestions for Future Research

Below are some suggested research topics that extend the work presented here.

- Asymptotic analysis results on fractional cooperation are presented in Ch. 4, where it is assumed that $l_S \to \infty$. Since in practical systems, the blocklength of the source’s codewords and the fading channel coherence time are finite, fractional cooperation with finite blocklength should be analyzed in future work.

- Lower bounds for $r_c$ can be found through either the outage set analysis on fractional cooperation, or through the UBB analysis. It is unclear that for a given $\epsilon_j$, a set encoding and decoding scheme, and a set the relaying scheme, what the value of $r_c$ is. The determination of $r_c$ with given set of parameters should be examined.

- As mentioned in Ch. 5, if the relay’s codeword is punctured, the output distribution of the R-D channel after decoding, $f(y_{R,D}(n)|c_R(n))$, can still be approximated with a Gaussian distribution, but the approximate description is not as accurate as in the case without puncturing [73]. The effect of this approximation on relay selection and performance analysis should be investigated.

- The application of UBB to relay channels presented in Ch. 5 is suitable when serial decoding is performed at D. Design and analysis of relay channels when parallel decoding is performed at D requires further study.
• As suggested in Ch. 2, the concepts developed in this dissertation can be extended to modulation schemes other than BPSK. If a modulation scheme other than BPSK is used, then the associated bit-flip probability will be dependent on the mapping between the encoded bits and the modulated symbols. In addition, the UBB must be modified to account for symbol, instead of bit, errors over the S-R channel. Further studies are needed to analyze the performance of such systems.

• A simple distributed algorithm that can be used for solving the optimization problems presented in Ch. 6 was shown. However, the algorithm does not guarantee the global optimal solution. Further investigation is needed to study under what conditions the algorithm would provide the global optimum, and the number of iterations needed to arrive at a local minimum solution.

• In [76], the concept of minimum path difference (MPD) was introduced. MPD, which describes the Euclidean distance between the received codeword and the decoded codeword found through the Viterbi algorithm [49], is a metric that can be used to improve the threshold-based selection DF. It was shown that MPD is a better metric used to estimate the number of errors in a block compared to the SNR. Instead of discarding the received block of data when the CRC is not satisfied, MPD can be calculated for sub-blocks of each data block, where each sub-block is relayed if its associated MPD is below a given threshold, and discarded if its associated MPD is above the threshold.

When sub-blocks are relayed or discarded depending on their MPD values, the system resembles one where fractional cooperation is applied and the DF relaying scheme is employed. Analysis of the performance fractional cooperation can be extended to such systems, where the fraction relayed by each relay is a random variable that is dependent on the instantaneous S-R channel.
Appendix A

Derivation of Distribution of $d^*$

Let RV $X_j$ and $X_{(j)}$ be functions of $\epsilon_j$ and $\epsilon_{(j)}$ respectively, where $\epsilon_{(j)}$ is the ascending-ordered statistics of $\epsilon_j$:

$$X_j = -\log(1 - \epsilon_j),$$

$$X_{(j)} = -\log(1 - \epsilon_{(j)}).$$

Then with a change of variables (4.17) becomes

$$\Pr(d^* = m) = \begin{cases} 
\Pr(\sum_{i=1}^{r} X_{(j)} < -\log(p_E)) & m = 1, \\
\Pr((\sum_{i=1}^{r-m+1} X_{(j)} < -\log(p_E)) \cap (\sum_{i=1}^{r-m+2} X_{(j)} \geq -\log(p_E))) & m = 2, \ldots, r, \\
\Pr(X_{(1)} \geq -\log(p_E)) & m = r + 1. 
\end{cases} \quad (A.1)$$

For $m = r + 1$, since $X_{(1)} \leq X_{(j)}$ for all $i$,

$$\Pr(X_{(1)} \geq -\log(p_E)) = \prod_{j=1}^{r} \Pr(X_j \geq -\log(p_E))$$

and

$$\Pr(d^* = m) = (1 - F_X(-\log(p_E)))^r. \quad (A.2)$$

To find $\Pr(d^* = m)$ for $m = 1, \ldots, r$, we first need to find the distribution of $Y_n = \sum_{j=1}^{n} X_{(j)}$. Since $X_{(j)} \leq X_{(j+1)}$, define the new RV $Z_b$ as a truncated version of RV $X$.
Appendix A. Derivation of Distribution of $d^*$

with

$$f_{Z_b}(z) = \begin{cases} \frac{f_X(z)}{F_X(b)} & \text{if } z \leq b, \\ 0 & \text{otherwise.} \end{cases}$$  \hspace{1cm} (A.3)$$

where $b$ is a given parameter. Then given that $X_{(n+1)} = x_{n+1}$

$$Y_n = \sum_{j=1}^{n} Z_{x_{n+1},j},$$  \hspace{1cm} (A.4)

where $Z_{b,j}$ are i.i.d. RV drawn from the distribution $f_{Z_b}(z)$.

Since $\epsilon_{(j)} \in [0,1]$, $X_{(j)}$, and hence $Z_{X_{(j)}}$, are non-negative RV. Let $\Phi_{Z_b}(s)$ be the characteristic function of RV $Z_b$, then if $\Phi_{Z_b}(s)$ exists [77],

$$\Phi_{Y_n}(s) = [\Phi_{Z_{x_{n+1}}}(s)]^n, \hspace{1cm} (A.5)$$

and

$$f_{Y_n}(y) = \mathcal{L}^{-1}(\Phi_{Y_n}(s)), \hspace{1cm} (A.6)$$

where $\mathcal{L}^{-1}(\cdot)$ is the inverse Laplace transform function. Hence

$$\Pr(d^* = m) = \begin{cases} \Pr(Y_r < -\log(p_E)) & \text{for } m = 1, \\ \Pr((Y_{r-m+1} < -\log(p_E)) \cap (Y_{r-m+2} \geq -\log(p_E))) & \text{for } m = 2, \ldots, r. \end{cases}$$  \hspace{1cm} (A.7)

For $m = 1$, let $Y_r = \lim_{b \to \infty} \sum_{j=1}^{r} Z_{b,j} = \sum_{j=1}^{r} X_j$. Then

$$\Pr(d^* = 1) = F_{Y_r}(-\log(p_E)) = F_{\sum_{j=1}^{r} X_j}(-\log(p_E)). \hspace{1cm} (A.8)$$

Obtaining $\Pr(d^* = m)$ for $m = 2, \ldots, r$ involves a few extra steps. Note that the event

$$(Y_{r-m+1} < -\log(p_E)) \cap (Y_{r-m+2} \geq -\log(p_E))$$

is equivalent to the event

$$\sum_{c=0}^{-\log(p_E)} (Y_{r-m+1} = c) \cap (X_{r-m+2} \geq -\log(p_E) - c).$$
The above condition places constraints on the possible values of $X_{(r-m+2)}$. Substituting $c = Y_{r-m+1} = \sum_{j=1}^{r-m+1} Z_{(r-m+2)}$ into the right-hand side of the equation, for a given realization of $X_{(r-m+2)}$, $x_{r-m+2}$,

$$x_{r-m+2} + \sum_{j=1}^{r-m+1} Z_{x_{r-m+2}} \geq -\log(p_E)$$

According to its definition, $Z_{x_{(r-m+2)}} \leq x_{r-m+2}$, and

$$(r-m+2)x_{r-m+2} \geq -\log(p_E)$$

$$x_{r-m+2} \geq \frac{-\log(p_E)}{r-m+2}$$  \hspace{1cm} (A.9)

The following conditions must be satisfied for the region of integration for $y_{r-m+1}$

1. $y_{r-m+1} \geq 0$ and $y_{r-m+1} \geq -\log(p_E) - x_{r-m+2}$,

2. $y_{r-m+1} < -\log(p_E)$ and $y_{r-m+1} \leq (r-m+1)x_{r-m+2}$.

Combining all the conditions, for $m = 2, \ldots, r$ we have

$$\Pr(d^* = m) = \int_{\frac{-\log(p_E)}{r-m+2}}^{\infty} \int_{\max(0, -\log(p_E) - x_{r-m+2})}^{\min(-\log(p_E), (r-m+1)x_{r-m+2})} f_{X_{(r-m+2)}}(x_{r-m+2})
$$

$$f_{Y_{r-m+1}}(y_{r-m+1})dy_{r-m+1}dx_{r-m+2}. \hspace{1cm} (A.10)$$

Recall that $X_{(r-m+2)}$ is the $(r-m+2)$th order statistics with PDF

$$f_{X_{(r-m+2)}}(x) = \frac{r!}{(r-m+1)!1!(m-2)!} (F_X(x))^{r-m+1} f_X(x)(1-F_X(x))^{m-2}. \hspace{1cm} (A.11)$$

Substituting into (A.10)

$$\Pr(d^* = m) = \frac{r!}{(r-m+1)!1!(m-2)!} \int_{\frac{-\log(p_E)}{r-m+2}}^{\infty} \int_{\max(0, -\log(p_E) - x_{r-m+2})}^{\min(-\log(p_E), (r-m+1)x_{r-m+2})} (F_X(x_{r-m+2}))^{r-m+1}
$$

$$f_X(x_{r-m+2})(1-F_X(x_{r-m+2}))^{m-2} f_{Y_{r-m+1}}(y_{r-m+1})dy_{r-m+1}dx_{r-m+2}. \hspace{1cm} (A.12)$$

An example is presented below to illustrate the results. Let $\epsilon$ be a uniformly distributed RV, with PDF

$$f_{\epsilon}(\epsilon) = \begin{cases} 
1 & \text{if } \epsilon \in [0, 1], \\
0 & \text{otherwise},
\end{cases}$$
Appendix A. Derivation of Distribution of $d^*$

and it can be shown that $X = -\log(1 - \epsilon)$ is an exponentially distributed with PDF

$$f_X(x) = \begin{cases} e^{-x} & \text{if } x \geq 0, \\ 0 & \text{if } x < 0, \end{cases}$$

Then $Z_b$, the truncation version of $X$ has PDF

$$f_{Z_b}(z) = \begin{cases} \frac{e^{-z}}{1-e^{-z}} & \text{if } z \in [0, b], \\ 0 & \text{otherwise}, \end{cases} \quad (A.13)$$

and characteristic function

$$\Phi_{Z_b}(s) = \begin{cases} \frac{1-e^{-(1+s)b}}{(1-e^{-y})(1+s)} & \text{if } b \text{ is finite}, \\ \frac{1}{1+s} & \text{otherwise}. \end{cases} \quad (A.14)$$

Assume that $X_{(r-m+2)} = x_{r-m+2}$, it can be shown that:

$$f_{Y_{r-m+1}}(y) = \mathcal{L}^{-1} \left[ \Phi_{Z_{x_{r-m+2}}}(s) \right] \quad (A.15)$$

$$= \begin{cases} e^{-y} \frac{y^{r-m}}{(r-m)!} u(y) & \text{if } m = 1, \\ \frac{e^{-y}}{(1-e^{-x_{r-m+2}})^{r-m+1} (r-m)!} \sum_{i=0}^{r-m+1} (-1)^i (i - ix_{r-m+2})^{r-m} u(y - ix_{r-m+2}) & \text{if } m = 2, \ldots, r, \end{cases} \quad (A.16)$$

where $u(t - c)$ is the unit step function with $u(t - c) = 1$ for $t \geq c$ and 0 otherwise.

Incidentally, for $m = 1$, $Y_r$ is an Erlang RV, and $\Pr(d^* = 1)$ is given by

$$\Pr(d^* = 1) = F_{Y_r}(-\log(p_E)) = 1 - p_E \sum_{k=0}^{r-1} \frac{(-\log(p_E))^k}{k!}. \quad (A.18)$$

To find $\Pr(d^* = m)$ for $m = 2, \ldots, r$, we first note that

$$f_{X_{(r-m+2)}}(x_{r-m+2}) = \frac{r!}{(r - m + 1)!(m-2)!} e^{-(m-1)x_{r-m+2}} (1 - e^{-x_{r-m+2}})^{r-m+1}. \quad (A.19)$$
Next, we substitute (A.17) and (A.19) into (A.12)

\[
\Pr(d^* = m) = \frac{r!}{(r - m + 1)!(m - 2)!} \int_{-\log(p_E)}^{-\log(p_E)} \int_{\max(0, -\log(p_E) - x)}^{\min(-\log(p_E), (r-m+1)x)} e^{-(m-1)x} (1 - e^{-x})^{r-m+1} \left( \frac{e^{-y}}{(1 - e^{-x})^{r-m+1}} \right) \frac{1}{(r-m)!} \sum_{i=0}^{r-m+1} \left( \begin{array}{c} r - m + 1 \\ i \end{array} \right) (-1)^i (y - ix)^{r-m} u(y - ix) dy dx \]

\[
= \frac{r!}{(r - m + 1)!(m - 2)!(r - m)!} \sum_{i=0}^{r-m+1} \left( \begin{array}{c} r - m + 1 \\ i \end{array} \right) (-1)^i \int_{-\log(p_E)}^{-\log(p_E)} \int_{\max(0, -\log(p_E) - x, ix)}^{\min(-\log(p_E), (r-m+1)x)} e^{-y} (y - ix)^{r-m} dy dx. \tag{A.20}
\]

After some manipulation, it can be shown that

\[
\Pr(d^* = m) = \frac{r!}{(r - m + 1)!(m - 2)!(r - m)!} \left\{ \sum_{i=0}^{r-m} \left( \begin{array}{c} r - m + 1 \\ i \end{array} \right) (-1)^i \int_{-\log(p_E)}^{-\log(p_E)} e^{-(m-1)x} \int_{x}^{\min(-\log(p_E), (r-m+1)x)} e^{-y} (y - ix)^{r-m} dy dx \right. \\
+ \sum_{i=0}^{r-m} \left( \begin{array}{c} r - m + 1 \\ i \end{array} \right) (-1)^i \int_{-\log(p_E)}^{-\log(p_E)} e^{-(m-1)x} \int_{-\log(p_E) - x}^{\max(0, -\log(p_E))} e^{-y} (y - ix)^{r-m} dy dx \\
+ \sum_{i=0}^{r-m} \left( \begin{array}{c} r - m + 1 \\ i \end{array} \right) (-1)^i \int_{-\log(p_E)}^{-\log(p_E)} e^{-(m-1)x} \int_{ix}^{\min(-\log(p_E), (r-m+1)x)} e^{-y} (y - ix)^{r-m} dy dx \right\}. \tag{A.22}
\]
Appendix B

Proof of Lemma 5.2.1

Proof: This proof follows closely the steps used to prove Lemma 5.2.1 in [2]. For simplicity, it is assumed that the fraction of the source codeword relayed are the same for all relays, where $\epsilon_j = \epsilon$ and $\gamma_{S,R} = 0$. In addition, if $\gamma_{S,R_j}$ or $\gamma_{R_j,D}$ falls below the threshold $\eta$ then $R_j$ is not used for relaying. Let $p_a = \Pr(\gamma_{S,R_j} > \eta \cap \gamma_{R_j,D} > \eta)$. Given that the $r_b$ best relays are chosen out of a pool of $r_a$ relays, then the probability of a bit not chosen by any of these relays is

$$p_{nr} = \sum_{j=0}^{r_b} {r_a \choose j} (1 - p_a)^r_j p^j_a (1 - \epsilon)^j + (1 - \epsilon)^r_b \sum_{j=r_b+1}^{r_a} {r_a \choose j} (1 - p_a)^r_a - j p^j_a$$  \hspace{1cm} (B.1a)

$$= \sum_{j=0}^{r_a} {r_a \choose j} (1 - p_a)^r_a - j p^j_a (1 - \epsilon)^j$$

$$+ \sum_{j=r_b+1}^{r_a} {r_a \choose j} (1 - p_a)^r_a - j p^j_a [(1 - \epsilon)^r_b - (1 - \epsilon)^j]$$  \hspace{1cm} (B.1b)

$$= [1 - \epsilon p_a]^r_a + \sum_{j=r_b+1}^{r_a} {r_a \choose j} (1 - p_a)^r_a - j p^j_a [(1 - \epsilon)^r_b - (1 - \epsilon)^j].$$  \hspace{1cm} (B.1c)

Since the second term in (B.1c) is always positive, given a fixed $r_a$, $p_{nr}$ decreases as $r_b$ increases, and the smallest $p_{nr}$ is reached when $r_b = r_a$.

For an uncoded transmission over the S-R-D link, a bit error occurs when either of the follow two events occur: (a) the bit is not relayed, or (b) the bit is relayed but an
error occurs over either the S-R or R-D link. The FER can be upper bounded by

\[ P_f \leq 1 - (1 - p_{nr} - (1 - p_{nr})2\zeta (1 - \zeta)^{l_S}, \]  

(B.2)

where \( \zeta \) represents the upper bound on the probability of a bit error over a link with SNR \( \gamma > \eta \) and is given by

\[ \zeta = \frac{1}{2}\operatorname{erfc}(\sqrt{\eta}), \]

and \( l_S \) is the length of the source’s codeword. Let \( r_b = r_a \), and the second term in (B.1c) vanishes. The upper bound on the FER can be decreased by decreasing \( \zeta \) through increasing \( \eta \), which in turn reduces \( p_a \). Even though increasing \( \eta \) reduces \( p_a \), \( r_a \) can be increased such that \( p_{nr} \to 0 \) such that any frame error rate criterion for outage probability can be satisfied, for some value of \( r_a \) (denoted as \( r_t \)). Now assume that \( r_a \geq r_t \), then there exists a \( r_b \leq r_a \) large enough such that the FER criterion can be satisfied, where the value of \( r_b \) satisfying the criterion is denoted as \( \tilde{r}_c(r_a) \).

It can be shown that the result holds if we relax the assumptions stated at the beginning of the proof. As illustrated in the proof in [2], it can be shown that for the cases where \( \gamma_{S,D} > 0 \), the results stated above holds. Now consider the case where the fractions relayed by each relay are not necessarily the same. First, we let \( \epsilon_{\min} = \min_j \epsilon_j \).

Consider two different systems, where in System A, relay \( i \) relays fraction \( \epsilon_j \) of the source codeword, and in System B, all relays relay \( \epsilon_{\min} \) of the source codeword. For System B, it has been shown that a corresponding \( r_t(\gamma_{S,D}) \) and \( \tilde{r}_c(\gamma_{S,D}, r_a) \) can be found. Note that the \( p_{nr} \) associated with System A is smaller by the \( p_{nr} \) associated with System B for the same values of \( r_a \) and \( r_b \). Hence the \( r_t(\gamma_{S,D}) \) and \( \tilde{r}_c(\gamma_{S,D}, r_a) \) associated with System A are upper-bounded by those associated with System B, which are finite. This shows that the corresponding \( r_t(\gamma_{S,D}) \) and \( \tilde{r}_c(\gamma_{S,D}, r_a) \) also exist when the fraction relayed by each relay are not identical.
Appendix C

Proof of Theorem 5.2.1

Proof: This proof, again, follows closely the steps that were used to prove Theorem 5.2.1 in [2]. For simplicity, we have set the average SNR over all channels to $\bar{\gamma}$.

From the definition of $r_t$, if $r_a < r_t$, then the system is in outage if the S-D link is not present. Hence with the S-D link, the system has diversity order of 1. Also, from the definition of $\tilde{r}_c(r_a)$, if $r_a \geq r_t$, but $r_b < \tilde{r}_c(r_a)$, then the system has diversity order of 1 as well. For $r_a \geq r_t$ and $r_b \geq \tilde{r}_c(r_a)$, we will show that a diversity order of $r_a - \tilde{r}_c(r_a) + 2$ can be achieved. Let $\gamma_{SD,0} = \max\{\gamma_{S,D} : \tilde{r}_c(\gamma_{S,D}, r_a) = \tilde{r}_c(r_a)\}$ be the SNR required on the SD channel such that for any $\gamma_{S,D} < \gamma_{SD,0}$, $\tilde{r}_c(\gamma_{S,D}, r_a) = \tilde{r}_c(r_a)$. For each number of relays $r_b \geq \tilde{r}_c(r_a)$, there exists $\gamma_{suf,r} > 0$ such that if $\gamma_{S,D} < \gamma_{SD,0}$ and at least $r_b - \tilde{r}_c(r_a) + 1$ of the $r_b$ chosen relays have $\gamma_{S,R_j} < \gamma_{suf,r}$ or $\gamma_{R_j,D} < \gamma_{suf,r}$, then an outage occurs. Note that if at least one of the $r_b$ chosen relays are in outage, i.e., at least one of the $r_b$ chosen relays have $\gamma_{S,R_j} < \gamma_{suf,r}$ or $\gamma_{R_j,D} < \gamma_{suf,r}$, that implies that all of the $r_a - r_b$ relays that are not chosen must be in outage as well, since the chosen $r_b$ relays are the best out of the $r_a$ available relays. Let

$$p_{o,r} := \Pr(\gamma_{S,R_j} < \gamma_{suf,r} \cup \gamma_{R_j,D} < \gamma_{suf,r}).$$

(C.1)

As shown in [2], $\Pr(\gamma_{S,D} < \gamma_{SD,0}) = \Theta(\bar{\gamma}^{-1})$, $p_{o,r} = \Theta(\bar{\gamma}^{-1})$ and $1 - p_{o,r} = \Theta(1)$. A lower
Appendix C. Proof of Theorem 5.2.1

Bound on the probability of outage $P_{\text{out}}$ is

$$P_{\text{out}} \geq \Pr(\gamma_{S,D} < \gamma_{SD,0}) p_{o,r}^{r_a-r_b} \sum_{j=r_b-\tilde{r}_c(r_a)+1}^{r_b} \binom{r_b}{j} p_{o,r}^j (1-p_{o,r})^{r_b-j}$$  \hspace{1cm} (C.2a)

$$= \Theta(\bar{\gamma}^{-1}) \Theta(\bar{\gamma}^{-(r_a-r_b)}) \Theta(\bar{\gamma}^{-(r_b-\tilde{r}_c(r_a)+1)})$$  \hspace{1cm} (C.2b)

$$= \Theta(\bar{\gamma}^{-(r_a-\tilde{r}_c(r_a)+2)})$$  \hspace{1cm} (C.2c)

Similarly, let $\gamma_{\text{nec}}$ be the maximum $\gamma_{S,R_j}$ or $\gamma_{R_j,D}$ for any relay $j$, such that if the system in outage, then $\gamma_{S,D} \leq \gamma_{SD,0}$ and $\gamma_{S,R_j} < \gamma_{\text{nec}}$ or $\gamma_{R_j,D} < \gamma_{\text{nec}}$ for at least $r_b - \tilde{r}_c(r_a) + 1$ of the $r_b$ chosen relays. The analysis of the necessary condition is similar to (C.2a), with $\leq$ in place of $\geq$, and $\gamma_{\text{nec}}$ in place of $\gamma_{\text{suf}}$ in (C.1). This shows that for $r_b \geq \tilde{r}_c(r_a)$, the diversity order of the system is $r_a - \tilde{r}_c(r_a) + 2$. \hfill \blacksquare
Appendix D

Proof of Border Points

Assume that the globally optimal solution for (6.1) is

\[ \epsilon^* = [\epsilon_{1,1}^*, \epsilon_{1,2}^*, \ldots, \epsilon_{1,s}^*, \epsilon_{2,1}^*, \ldots, \epsilon_{r,s}^*]. \]

Here we show that if \( \epsilon^* \) is separated into vectors \( \epsilon_j^* \), which is associated with \( R_j \), where

\[ \epsilon_j^* = [\epsilon_{j,1}^*, \ldots, \epsilon_{j,s}^*] \]

then \( \epsilon_j^* \in S_{BP,j} \).

We will first illustrate with the simple two-source two-relay network. To simplify the analysis, two variables are introduced:

\[ a_{j,i} = 1 - \beta_{j,i}, \]
\[ \beta_i = \beta_t / \beta_{S_i,D}. \]

Then substituting the \( a_{j,i} \) into (6.5) we obtain

\[ \hat{\epsilon}_{j,i} = \frac{1}{a_{j,i}} (1 - \beta_i) \quad \text{for } i, j = 1, 2, \]  

(D.1)

where \( \hat{\epsilon}_{j,i} \) is an upper bound on the resources \( S_i \) requests of \( R_j \) to satisfy its constraints. Recall that the source constraint that needs to be satisfied at \( S_1 \) is

\[ (1 - a_{1,1} \epsilon_{1,1})(1 - a_{2,1} \epsilon_{2,1}) \leq \beta_1. \]  

(D.2)
If $\beta_1 \geq 1$, then this constraint is satisfied without any relaying and can be ignored. Assume that $\beta_1 < 1$, and let $\bar{\epsilon}_{1,1}$ and $\bar{\epsilon}_{2,1}$ be the resources $R_1$ and $R_2$ have reserved for $S_1$ respectively, where $\epsilon_{1,1} \leq \bar{\epsilon}_{1,1}$ and $\epsilon_{2,1} \leq \bar{\epsilon}_{2,1}$. An equation describing the set of points $(\epsilon_{1,1}, \epsilon_{2,1})$ where (D.2) is satisfied with equality is

$$
\epsilon_{2,1} = \frac{1}{a_{2,1}} \left( 1 - \frac{\beta_1}{1 - a_{1,1} \epsilon_{1,1}} \right),
$$

and all such points are plotted in Fig. D.1, with the curve referred as the source constraint contour. The resources $R_1$ and $R_2$ has reserved for $S_i$, $\bar{\epsilon}_{1,1}$ and $\bar{\epsilon}_{2,1}$, are also shown in the figure, where $\bar{\epsilon}_{1,1} < \bar{\epsilon}_{1,1}$ and $\bar{\epsilon}_{2,1} < \bar{\epsilon}_{2,1}$. For a given point on the source constraint contour, the total energy used by the relays to assist $S_1$ is given by

$$
E_1 = \epsilon_{1,1} + \frac{1}{a_{2,1}} \left( 1 - \frac{\beta_1}{1 - a_{1,1} \epsilon_{1,1}} \right).
$$

Differentiating with respect to $\epsilon_{1,1}$ we arrive at

$$
\frac{dE_1}{d\epsilon_{1,1}} = 1 + \frac{a_{1,1}}{a_{2,1} (1 - a_{1,1} \epsilon_{1,1})^2},
$$

$$
\frac{d^2 E_1}{d\epsilon_{1,1}^2} = \frac{a_{1,1}^2}{a_{2,1} (1 - a_{1,1} \epsilon_{1,1})^3}.
$$

This shows that $E_1$ is a concave function of $\epsilon_{1,1}$, and the minimum values can only be one of the two values:

- $\epsilon_{1,1} = \min\{\epsilon_{1,1}, \hat{\epsilon}_{1,1}\}$, corresponding to $P_2$ in Fig. D.1, where

$$
\epsilon_{2,1} = \begin{cases} 
\frac{1}{a_{2,1}} \left( 1 - \frac{\beta_1}{1 - a_{1,1} \epsilon_{1,1}} \right) & \text{if } \epsilon_{1,1} = \bar{\epsilon}_{1,1}, \\
0 & \text{if } \epsilon_{1,1} = \hat{\epsilon}_{1,1}.
\end{cases}
$$

- $\epsilon_{2,1} = \min\{\bar{\epsilon}_{2,1}, \hat{\epsilon}_{2,1}\}$, corresponding to $P_1$ in Fig. D.1, where

$$
\epsilon_{1,1} = \begin{cases} 
\frac{1}{a_{1,1}} \left( 1 - \frac{\beta_1}{1 - a_{2,1} \epsilon_{2,1}} \right) & \text{if } \epsilon_{2,1} = \bar{\epsilon}_{2,1}, \\
0 & \text{if } \epsilon_{2,1} = \hat{\epsilon}_{2,1}.
\end{cases}
$$
This shows that at the optimal point on its source constraint contour a source node exhausts its reserved resources, or has obtained its requested resources. In Fig. D.1, $P_2$ is on a lower energy contour than $P_1$. Hence $P_2$ is the optimal point for $S_1$ for the given reserved resources $\bar{\epsilon}_{1,1}$ and $\bar{\epsilon}_{2,1}$.

The study of trivial cases where $\beta_1 \geq 1$ or $\beta_2 \geq 1$ is true is omitted here as no reservation of resources for source nodes is required. Under the assumption $\beta_1 < 1$ and $\beta_2 < 1$, the set of points satisfying the sources constraints for both $S_1$ and $S_2$ are plotted in Fig. D.2, where the plot corresponding to $S_2$ has been rotated. In the plot, the dashed lines refer to division of resources reserved for each source node: If $\bar{\epsilon}_{1,1}$ is reserved for $S_1$, then $\bar{\epsilon}_{1,2} = \bar{\epsilon}_1 - \bar{\epsilon}_{1,1}$ is reserved for $S_2$ by $R_1$. Similarly, $\bar{\epsilon}_{2,1}$ and $\bar{\epsilon}_{2,2} = \bar{\epsilon}_2 - \bar{\epsilon}_{2,1}$ are resources reserved by $R_2$ for $S_1$ and $S_2$ respectively. From the previous argument, given the resources reserved by each relay for each source node, the optimal point for $S_1$ must be one of the two points indicated by the circles, $P_1$ and $P_2$, and the optimal point for $S_2$ must be one of the two points indicated by the crosses, $P_3$ and $P_4$. The point on each source constraint contour is determined to be optimal if it lies on the lowest energy
Figure D.2: Optimal points for two-source two-relay system with reserved resources $\bar{\epsilon}_{1,1}, \bar{\epsilon}_{1,2}, \bar{\epsilon}_{2,1}$ and $\bar{\epsilon}_{2,2}$.

There are four possible combinations:

- **$P_2$ and $P_3$:** In this case, the total energy is given by

  \[
  E = E_1 + E_2 \\
  = \bar{\epsilon}_{1,1} + \frac{1}{a_{2,1}} \left(1 - \frac{\beta_1}{1 - a_{1,1} \bar{\epsilon}_{1,1}}\right) + \bar{\epsilon}_{2,2} + \frac{1}{a_{1,2}} \left(1 - \frac{\beta_2}{1 - a_{2,2} \bar{\epsilon}_{2,2}}\right). 
  \]

  \[
  \text{(D.9)} \\
  \text{(D.10)}
  \]

  The total energy can be reduced by increasing $\bar{\epsilon}_{2,2}$ and $\bar{\epsilon}_{1,1}$. Since increasing $\bar{\epsilon}_{1,1}$ (and hence decreasing $\bar{\epsilon}_{1,2}$) pushes the corresponding $P_4$ to a higher energy contour, and increasing $\bar{\epsilon}_{2,2}$ (and hence decreasing $\bar{\epsilon}_{2,1}$) pushes the corresponding $P_1$ to a higher energy contour, $P_2$ and $P_3$ remain the optimal points for minimizing the total energy consumed. The minimized energy is found when $\bar{\epsilon}_{1,1} = \min(\bar{\epsilon}_1, \bar{\epsilon}_{1,1}) = \tilde{\epsilon}_{1,1}$ and $\bar{\epsilon}_{2,2} = \min(\bar{\epsilon}_2, \bar{\epsilon}_{2,2}) = \tilde{\epsilon}_{2,2}$. This is a border point.
• $P_1$ and $P_4$: In this case, the total energy is given by

$$E = \tilde{\epsilon}_{2,1} + \frac{1}{a_{1,1}} \left(1 - \frac{\beta_1}{1 - a_{1,1} \tilde{\epsilon}_{2,1}}\right) + \bar{\epsilon}_{1,2} + \frac{1}{a_{2,2}} \left(1 - \frac{\beta_2}{1 - a_{1,2} \bar{\epsilon}_{1,2}}\right). \quad \text{(D.11)}$$

The total energy can be reduced by increasing $\tilde{\epsilon}_{2,1}$ and $\bar{\epsilon}_{1,2}$. Since increasing $\tilde{\epsilon}_{2,1}$ pushes $P_3$ to a higher energy contour, and increasing $\bar{\epsilon}_{1,2}$ pushes $P_2$ to a higher energy contour, $P_1$ and $P_4$ will remain the optimal points for minimizing the total energy consumed. The minimized energy is found when $\tilde{\epsilon}_{2,1} = \min(\tilde{\epsilon}_2, \bar{\epsilon}_{2,1}) = \tilde{\epsilon}_{2,1}$ and $\bar{\epsilon}_{1,2} = \min(\bar{\epsilon}_1, \bar{\epsilon}_{1,2}) = \bar{\epsilon}_{1,2}$. This is a border point.

• $P_2$ and $P_4$: In this case, the total energy is given by

$$E = \bar{\epsilon}_{1,1} + \frac{1}{a_{2,1}} \left(1 - \frac{\beta_1}{1 - a_{1,1} \bar{\epsilon}_{1,1}}\right) + (\bar{\epsilon}_1 - \bar{\epsilon}_{1,1}) + \frac{1}{a_{2,2}} \left(1 - \frac{\beta_2}{1 - a_{1,2}(\bar{\epsilon}_1 - \bar{\epsilon}_{1,1})}\right)$$

$$= \bar{\epsilon}_1 + \frac{1}{a_{2,1}} \left(1 - \frac{\beta_1}{1 - a_{1,1} \bar{\epsilon}_{1,1}}\right) + \frac{1}{a_{2,2}} \left(1 - \frac{\beta_2}{1 - a_{1,2}(\bar{\epsilon}_1 - \bar{\epsilon}_{1,1})}\right). \quad \text{(D.13)}$$

Differentiating with respect to $\bar{\epsilon}_{1,1}$ we have

$$\frac{dE}{d\bar{\epsilon}_{1,1}} = -\frac{a_{1,1}}{a_{2,1} (1 - a_{1,1} \bar{\epsilon}_{1,1})^2} + \frac{a_{1,2}}{a_{2,2} [1 - a_{1,2}(\bar{\epsilon}_1 - \bar{\epsilon}_{1,1})]^2}, \quad \text{(D.14)}$$

$$\frac{d^2E}{d\bar{\epsilon}_{1,1}^2} = -\frac{a_{1,1}^2}{a_{2,1} (1 - a_{1,1} \bar{\epsilon}_{1,1})^3} + \frac{a_{1,2}^2}{a_{2,2} [1 - a_{1,2}(\bar{\epsilon}_1 - \bar{\epsilon}_{1,1})]^3}. \quad \text{(D.15)}$$

This shows that $E$ is a concave function of $\bar{\epsilon}_{1,1}$. For a given $\bar{\epsilon}_{1,1}$ and $\bar{\epsilon}_{1,2}$,

- if $\frac{dE}{d\bar{\epsilon}_{1,1}} < 0$, the total energy consumed can be increased by increasing $\bar{\epsilon}_{1,1}$ and decreasing $\bar{\epsilon}_{1,2}$, where $E_1$ decreases while $E_2$ increases. This continues until

  * $\bar{\epsilon}_{1,1} = \min(\bar{\epsilon}_1, \bar{\epsilon}_{1,1})$, where $\bar{\epsilon}_{1,1}$ has either reached its limit ($\bar{\epsilon}_{1,1} = \bar{\epsilon}_1$), or increasing $\bar{\epsilon}_{1,1}$ does not decrease the total energy further ($\bar{\epsilon}_{1,1} = \bar{\epsilon}_{1,1}$). The optimal point has be reached. This is one of the border points.

  * $P_4$ has been pushed to a higher energy contour such that $P_4$ is on a higher energy contour than $P_3$; $P_3$ has replaced $P_4$ as the new optimal point on the constraint curve for $S_2$. As shown earlier, if $P_2$ and $P_3$ are the optimal points, then it leads to a border point.
Appendix D. Proof of Border Points

- if \( \frac{dE}{d\bar{\epsilon}_{1,1}} > 0 \), the total energy consumed can be increased by decreasing \( \bar{\epsilon}_{1,1} \) and increasing \( \bar{\epsilon}_{1,2} \), where \( E_2 \) decreases while \( E_1 \) increases. This continues until
  * \( \bar{\epsilon}_{1,2} = \min(\bar{\epsilon}_1, \bar{\epsilon}_{1,2}) \). The optimal point has been reached. This is one of the border points.

* \( P_2 \) has been pushed to a higher energy contour such that \( P_2 \) is on a higher energy contour than \( P_1 \); \( P_1 \) has replaced \( P_2 \) as the new optimal point on the constraint curve for \( S_1 \). As shown earlier, if \( P_1 \) and \( P_4 \) are the optimal points, then it leads to a border point.

**P_1** and **P_3**: In this case, the total energy is given by

\[
E = \bar{\epsilon}_{2,1} + \frac{1}{a_{1,1}} \left( 1 - \frac{\beta_1}{1 - a_{2,1}\bar{\epsilon}_{2,1}} \right) + (\bar{\epsilon}_2 - \bar{\epsilon}_{2,1}) + \frac{1}{a_{1,2}} \left( 1 - \frac{\beta_2}{1 - a_{2,2}(\bar{\epsilon}_2 - \bar{\epsilon}_{2,1})} \right) \\
= \bar{\epsilon}_2 + \frac{1}{a_{1,1}} \left( 1 - \frac{\beta_1}{1 - a_{2,1}\bar{\epsilon}_{2,1}} \right) + \frac{1}{a_{1,2}} \left( 1 - \frac{\beta_2}{1 - a_{2,2}(\bar{\epsilon}_2 - \bar{\epsilon}_{2,1})} \right) .
\]  
(D.16)

(D.17)

Using similar analysis as above, for a given \( \bar{\epsilon}_{2,1} \) and \( \bar{\epsilon}_{2,2} \),

- If \( \frac{dE}{d\bar{\epsilon}_{2,1}} < 0 \), increase \( \bar{\epsilon}_{2,1} \) until
  * \( \bar{\epsilon}_{2,1} = \min(\bar{\epsilon}_2, \bar{\epsilon}_{2,1}) \). This is one of the border points.

* \( P_3 \) has been pushed to a higher energy contour such that \( P_4 \) has replaced \( P_3 \) as the optimal point on the constraint curve for \( S_2 \). As stated above, \( P_1 \) and \( P_4 \) being the optimal point on their respective constraint curve leads to a border point.

- If \( \frac{dE}{d\bar{\epsilon}_{2,2}} > 0 \), decrease \( \bar{\epsilon}_{2,2} \) until
  * \( \bar{\epsilon}_{2,2} = \min(\bar{\epsilon}_2, \bar{\epsilon}_{2,2}) \). This is one of the border points.

* \( P_1 \) has been pushed to a higher energy contour such that \( P_2 \) has replaced \( P_1 \) as the optimal point on the constraint curve for \( S_1 \). As stated above, \( P_2 \) and \( P_3 \) being the optimal point on their respective constraint curve leads to a border point.
Appendix D. Proof of Border Points

This shows that for a two-source two-relay network, if a solution exists, the optimal solution belongs to the set of border points.

Next we show that if this is also true for a network with s source and r relay nodes. First, assume $\epsilon_j \notin S_{BP,j}$ for at least one $j$, where $|W_j'| > 1$. Without loss of generality, let $j = 1$ and $\{1, 2\} \in W_1'$. From the earlier explanation, the total energy consumed can be further reduced by either increasing or decreasing $\bar{\epsilon}_{1,1}$ and changing $\bar{\epsilon}_{1,2}$ in a complementary manner, while keeping $\bar{\epsilon}_{1,1} + \bar{\epsilon}_{1,2}$ constant.

- If the energy can be reduced by increasing $\bar{\epsilon}_{1,1}$, then it is increased.

  - If $\bar{\epsilon}_{1,1} + \bar{\epsilon}_{1,2} \geq M_1(1) + m_2(1)$, the new reserved resources from $R_1$ are $\bar{\epsilon}'_{1,1} = M_1(1)$ and $\bar{\epsilon}'_{1,2} = \bar{\epsilon}_{1,1} + \bar{\epsilon}_{1,2} - M_1(1)$. This moves 1 from $W_1'$ to $W_{M,1}$, reducing $|W_1'|$ by 1. If $|W_1'| > 1$ is still true, then this process is repeated between $S_2$ and $S_{i'}$ where $i' \in W_1'$, until $|W_1'| \leq 1$; otherwise this is a border point.

  - If $\bar{\epsilon}_{1,1} + \bar{\epsilon}_{1,2} < M_1(1) + m_2(1)$, the new constraints are

    \[
    \bar{\epsilon}'_{1,1} = \bar{\epsilon}_{1,1} + \bar{\epsilon}_{1,2} - m_2(1), \\
    \bar{\epsilon}'_{1,2} = m_2(1).
    \]

    This moves 2 from $W_1'$ to $W_{m,1}$, reducing $|W_1'|$ by 1. If $|W_1'| > 1$, then this process is repeated between $S_1$ and $S_{i'}$ where $i' \in W_1'$, until $|W_1'| \leq 1$; otherwise this is a border point.

- If the energy can be reduced by increasing $\bar{\epsilon}_{1,2}$, then it is increased.

  - If $\bar{\epsilon}_{1,1} + \bar{\epsilon}_{1,2} \geq M_2(1) + m_1(1)$, the new reserved resources from $R_2$ are $\bar{\epsilon}'_{1,2} = M_2(1)$ and $\bar{\epsilon}'_{1,1} = \bar{\epsilon}_{1,1} + \bar{\epsilon}_{1,2} - M_2(1)$. This moves 2 from $W_1'$ to $W_{M,1}$, reducing $|W_1'|$ by 1. If $|W_1'| > 1$ is still true, then this process is repeated between $S_1$ and $S_{i'}$ where $i' \in W_1'$; otherwise this is a border point.
Appendix D. Proof of Border Points

- If $\bar{\varepsilon}_{1,1} + \bar{\varepsilon}_{1,2} < M_2(1) + m_1(1)$, the new constraints are

$$
\bar{\varepsilon}'_{1,2} = \bar{\varepsilon}_{1,1} + \bar{\varepsilon}_{1,2} - m_1(1),
$$
$$
\bar{\varepsilon}'_{1,1} = m_1(1).
$$

This moves 1 from $W'_1$ to $W_{m,1}$, reducing $|W'_1|$ by 1. If $|W'_1| > 1$ is still true, then this process is repeated between $S_2$ and $S_i'$ where $i' \in W'_1$; otherwise this is a border point.

As illustrated above, resource allocation resulting in a lower total energy consumption can be found by shifting the solution to one where $\varepsilon^*_j \in S_{BP,j}$. This contradicts with the original premise that $\varepsilon^*_j \not\in S_{BP,j}$ is the globally optimal solution. Hence the globally optimal solution $\varepsilon^*_j \in S_{BP,j}$ for all $j$. 

Bibliography


