On Asynchronous Communication Systems:

Capacity Bounds and Relaying Schemes

by

Mojtaba Rahmati

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved March 2013 by the
Graduate Supervisory Committee:

Tolga M. Duman, Chair
Cihan Tepedelenlioglu
Martin Reisslein
Junshan Zhang

ARIZONA STATE UNIVERSITY

May 2013

ABSTRACT

Practical communication systems are subject to errors due to imperfect time alignment among the communicating nodes. Timing errors can occur in different forms depending on the underlying communication scenario. This doctoral study considers two different classes of asynchronous systems; point-to-point (P2P) communication systems with synchronization errors, and asynchronous cooperative systems. In particular, the focus is on an information theoretic analysis for P2P systems with synchronization errors and developing new signaling solutions for several asynchronous cooperative communication systems.

The first part of the dissertation presents several bounds on the capacity of the P2P systems with synchronization errors. First, binary insertion and deletion channels are considered where lower bounds on the mutual information between the input and output sequences are computed for independent uniformly distributed (i.u.d.) inputs. Then, a channel suffering from both synchronization errors and additive noise is considered as a serial concatenation of a synchronization error-only channel and an additive noise channel. It is proved that the capacity of the original channel is lower bounded in terms of the synchronization error-only channel capacity and the parameters of both channels. On a different front, to better characterize the deletion channel capacity, the capacity of three independent deletion channels with different deletion probabilities are related through an inequality resulting in the tightest upper bound on the deletion channel capacity for deletion probabilities larger than 0.65. Furthermore, the first non-trivial upper bound on the $2K$-ary input deletion channel capacity is provided by relating the $2K$-ary input deletion channel capacity with the binary deletion channel capacity through an inequality.

The second part of the dissertation develops two new relaying schemes to alleviate asynchronism issues in cooperative communications. The first one is a single

carrier (SC)-based scheme providing a spectrally efficient Alamouti code structure at the receiver under flat fading channel conditions by reducing the overhead needed to overcome the asynchronism and obtain spatial diversity. The second one is an orthogonal frequency division multiplexing (OFDM)-based approach useful for asynchronous cooperative systems experiencing excessive relative delays among the relays under frequency-selective channel conditions to achieve a delay diversity structure at the receiver and extract spatial diversity.

To Afsaneh

and

my dear family

for their love and support.

ACKNOWLEDGEMENTS

First and foremost, I would like to give special thanks to my advisor, Dr. Tolga M. Duman, for his tremendous support, insightful guidance, invaluable assistance, and endless patience over the past four years. I would like to thank my committee members, Dr. Junshan Zhang, Dr. Cihan Tepedelenlioglu and Dr. Martin Reisslein, for their comments and patience. I would like to express my gratituede and thanks to my parents and my brother for their constant support. Lastly, and most importantly, I would like to thank my best friend, Afsaneh Nassery, for her support during the last four years.

TABLE OF CONTENTS

CHAPTER Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Time synchronization is a basic challenge in design and implementation of digital communication systems whether they are point-to-point (P2P) or cooperative systems. In fact, achieving a perfect time synchronization is not possible in many communication systems and synchronization issues are unavoidable. The synchronization problems may degrade the performance of communication systems in different ways depending on the underlying communication scenario. In this dissertation, we focus on two different cases: P2P communications with synchronization errors and asynchronous cooperative communications. More precisely, our focus is on information theoretic analysis for P2P channels suffering from synchronization errors and physical layer solutions for cooperative communication systems suffering from asynchronism among the cooperating nodes.

Different types of impairments may lead to synchronization errors in P2P communication systems. For example, imperfect knowledge of the receiver about the transmitter clock rate [7] or time-varying transmission rate [8] result in insertion and deletion errors at the receiver which degrade the system performance. Hard disk systems also suffer from synchronization errors in which varying rotation speed of the platter results in a varying bit rate, therefore, insertion and deletion errors become unavoidable [7]. Insertion errors may also occur in digital audio tape (DAT) systems due to the stretching of the tape [7]. Furthermore, bit-patterned media (BPM) recording [9] is known as a new technology that suffers from synchronization errors. The main difficulty in this new technology is the write synchronization issue that is not a problem in the conventional recording systems. In BPM recording systems, there are arrays of magnetic cells over the recording medium where each cell is magnetized by the write head to represent a bit, therefore, the imperfect alignment

between the write head and the position of the magnetic cells result in synchronization errors.

A useful channel model for P2P systems with synchronization errors assumes that the number of received symbols may be more or less than the number of transmitted symbols. In other words, channels with synchronization errors can be well modeled using bit drop outs and/or bit insertions as well as random errors. By channels with synchronization errors, mathematically, we refer to the binary memoryless channels with synchronization errors as described by Dobrushin in [10] where every transmitted bit is independently replaced with a random number of symbols (possibly the empty string, i.e., a deletion event is also allowed), and the transmitter and receiver have no information about the position and/or the pattern of the synchronization errors. As proved in the same paper, for such channels, information stability holds and Shannon capacity exists. Digital communication systems under perfect synchronization assumption are well modeled and widely studied in the literature from different aspects including computation of channel capacity and development of coding principles. However, information theoretic analysis performed under the perfect synchronization assumption does not easily extend to channels with synchronization errors. That is, due to the memory introduced into the received sequence by the synchronization errors, an information theoretic study of these channels proves to be very challenging.

In the existing literature, several specific instances of the Dobrushin's model are more widely studied. For instance, by a proper selection of the stochastic channel transition matrix, one obtains the independent and identically distributed (i.i.d.) deletion channel which represents one of the simplest models allowing for bit drop-outs. In an i.i.d. deletion channel, the transmitted symbols are either received correctly and in the right order or deleted from the transmitted sequence altogether

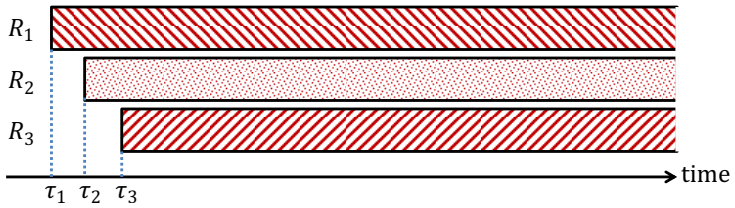

Figure 1.1: The received signal at the destination of an asynchronous relay system with three relays.

with a certain probability d independent of each other. Neither the receiver nor the transmitter knows the positions of the deleted symbols. Despite the simplicity of the model, the capacity for this channel is still unknown, and only a few upper and lower bounds are available [5, 11–13]. Other special cases of the general model by Dobrushin are the Gallager model allowing for insertion, deletion and substitution errors in which every transmitted bit is either deleted with probability of d, replaced with two random bits with probability of i, flipped with probability of f or received correctly with probability of $1 - d - i - f$. There are also some capacity upper and lower bounds for the Gallager's insertion/deletion channel model in the literature, e.g., [14].

Another communication scenario in which timing issues may degrade the system performance is the case of cooperative communication systems. In cooperative communications, a group of nodes, known as relays, help a source node deliver its data to the destination [15]. Due to the distributed nature of the wireless cooperative communication systems, achieving a perfect time alignment among signals received from different nodes is not always possible. For instance, Fig. 1.1 shows the received signal at the destination of an asynchronous cooperative system with three relays which is a superposition of three unaligned signals, where τ_i represents the beginning of the arrival of the signal transmitted by the i-th relay (R_i), indicating a relative delay of $\tau_i - \tau_j$ between the signals received from the i-th and j-th relays ($i, j \in \{1, 2, 3\}$).

As a specific scenario in which relative time delays among the relay nodes can be significant, we focus on asynchronous underwater acoustic (UWA) cooperative communication systems. In a UWA cooperative communication system, the time differences among signals received from geographically separated nodes can be excessive due to the low speed of sound in water. For instance, if the relative distance between two nodes with respect to another is 500 m, then their transmissions experience a relative delay of 333 ms. Considering, for instance, that in an orthogonal frequency division multiplexing (OFDM)-UWA cooperative communication scheme with 512 sub-carriers over a total bandwidth of 8 kHz, the OFDM block duration is only 64 ms, the excessive delay of 333 ms becomes very problematic. Furthermore, UWA channels are highly time varying due to the large Doppler spreads and Doppler shift effects (or, Doppler scaling effects) [16].

The existing signaling solutions for asynchronous terrestrial radio cooperative communications rely on quasi-static fading assumptions with limited delays among signals received from different relays at the destination, e.g., see [17] and references therein. Therefore, conventional physical layer solutions designed specifically for terrestrial radio cooperative communications cannot be directly applied for the cooperative UWA scenarios which are asynchronous with large relative delays among the nodes under highly time-varying frequency selective channel conditions. For instance, in systems employing OFDM, e.g., [3, 18], the existing solutions are effective when the maximum length of the relative delays among signals received from various nodes are less than the length of an OFDM block which is not a practical assumption for the case of UWA communications. A trivial generalization of existing OFDM-based results to compensate for large relative delays may be to increase the OFDM block lengths. The main drawback is that inter carrier interference (ICI) is increased due to time variations of the UWA channels. Another trivial solution is to increase the length

of the cyclic prefix (CP). This is not an efficient solution either, since it dramatically decreases the spectral efficiency of the system.

1.1 Outline of Dissertation

In Chapter 2, we first give the general model for memoryless channels with synchronization errors, by illustrating several specific models such as binary input insertion/deletion/substitution channel, $2K$-ary input deletion channel, binary input symmetric q-ary output channels (BSQC) with synchronization errors and binary input additive white Gaussian noise (BI-AWGN) channels with synchronization errors, and provide a review of existing capacity bounds on channels with synchronization errors. Then, we present the system model for asynchronous cooperative communication systems and review some existing works on asynchronous terrestrial radio cooperative communication systems.

In Chapter 3, we consider achievable rates over binary input insertion and deletion channels for small values of insertion and deletion probabilities by computing bounds on the mutual information rate of the insertion and deletion channels for independent uniformly distributed (i.u.d.) input sequences. We consider three different deletion channel models: the usual i.i.d. deletion channel, i.i.d. deletion-substitution channel and i.i.d. deletion channel with additive white Gaussian noise (AWGN). For the insertion channel case we assume that the transmitted bits are replaced with two bits with a certain probability independently of any other insertion events. We consider two specific cases: Gallager's model where the pair of bits are random and uniform over the four possibilities, and the sticky channel where transmitted bits are simply duplicated.

In Chapter 4, we consider binary input symmetric output channels with synchronization errors which also suffer from other type of impairments such as substitutions, erasures, additive white Gaussian noise (AWGN) etc. We present several lower

bounds on the capacity of such a channel, where we show that if the channel with synchronization errors can be decomposed into a cascade of two channels where only the first one suffers from synchronization errors and the second one is a memoryless channel, a lower bound on the capacity of the original channel in terms of the synchronization error-only channel capacity and the parameters of the original channel can be derived. We illustrate that, with our approach, it is possible to derive tighter bounds compared to the currently available bounds in the literature for certain classes of channels, e.g., deletion-substitution channels and deletion-AWGN channels for a range of signal to noise ratio (SNR) values.

In Chapter 5, we first prove a simple result that the parallel concatenation of two different independent deletion channels with deletion probabilities d_1 and d_2, in which every input bit is either transmitted over the first channel with probability of λ or over the second one with probability of $1-\lambda$, is nothing but another deletion channel with deletion probability of $d = \lambda d_1 + (1-\lambda)d_2$. We then provide an upper bound on the concatenated deletion channel capacity $C(d)$ in terms of the weighted average of $C(d_1)$, $C(d_2)$ and the parameters of the three channels. An interesting consequence of this bound is that $C(\lambda d_1 + (1-\lambda)) \leq \lambda C(d_1)$ which enables us to provide an improved upper bound on the capacity of the i.i.d. deletion channels, i.e., $C(d) \leq 0.4143(1-d)$ for $d \geq 0.65$. Using the same approach we are also able to improve upon existing upper bounds on the capacity of the deletion-substitution channel.

In Chapter 6, we derive the first non-trivial upper bound on the non-binary deletion channel capacity and reduce the gap between the existing achievable rates and upper bounds. To derive the new upper bounds we first prove an inequality between the capacity of a $2K$-ary deletion channel with deletion probability d, denoted by $C_{2K}(d)$, and the capacity of the binary deletion channel with the same deletion

probability, $C_2(d)$, that is, $C_{2K}(d) \leq C_2(d) + (1-d)\log(K)$. Then by employing the existing upper bounds on the capacity of the binary deletion channel, we obtain upper bounds on the capacity of the $2K$-ary deletion channel. We illustrate via examples the use of the new bounds and discuss their asymptotic behavior as $d \to 0$.

In Chapters 7 and 8, we turn our attention to a physical layer study of cooperative communication systems which suffer from time asynchronism among the signals received from different relays at the destination node. Our main motivation to consider asynchronous cooperative communication systems is their application in cooperative UWA communications in which due to the low speed of the sound in water, relative delays among relay nodes become extremely large.

In Chapter 7, a relay communication system with two amplify and forward (AF) relays under fading channel conditions is considered in which the signals received from the relay nodes are not necessarily time aligned and both relays share the same time and frequency bands to communicate with the destination. We propose a new time-reversal (TR)-space time block coding (STBC) scheme for the considered cooperative system for which in comparison with the existing STBC solutions with the same data block lengths, a smaller time guard needs to be added to guarantee robustness against asynchronism among the cooperating nodes. Assuming full channel state information (CSI) at the receiver, we obtain the optimal maximum likelihood (ML) detector structure for the proposed scheme. We also propose a sub-optimal lower complexity detector. Through numerical examples, we verify that the new scheme extracts full spatial diversity out of the system.

In Chapter 8, we develop a new OFDM based scheme to combat the asynchronism problem in cooperative UWA communication systems without adding a long CP (in the order of the long relative delays) at the transmitter. More precisely, we show that by adding a much more manageable (short) CP at the source, utilizing

full-duplex AF relaying at the relays, and appropriate CP removal at the destination, we can obtain a delay diversity structure out of the received signal at the destination node for effective processing and exploitation of the spatial diversity. We provide a pairwise error probability (PEP) analysis of the system for both time invariant and block fading channel scenarios and show that the system achieves full spatial diversity. By full CSI at the destination node, a Viterbi decoder can be employed to recover the original data which is observed through an equivalent delay diversity scheme. Through numerical examples, we evaluate the performance of the proposed scheme for time-varying multipath channels with Rayleigh fading channel taps, modeling UWA channels. We compare our results with those of the existing schemes and find that while for time invariant channels, the performance is similar, for time varying cases (typical in UWA communications) the proposed scheme is significantly superior.

Finally, we provide a summary of the accomplishments and conclusions in Chapter 9.

Chapter 2

Preliminaries

In this chapter, we first focus on memoryless P2P channels with synchronization errors. We review the general model for such channels [10], give some specific channel models which are used in the dissertation and review existing results on the capacity of synchronization error channels by focusing on insertion/deletion channel models. Then, we turn our attention to the asynchronous cooperative communication systems by providing the general system model and reviewing existing results on asynchronous cooperative radio terrestrial communication systems under both flat fading and frequency selective channel conditions. Throughout the dissertation, by channels with synchronization errors, we mean P2P channels with synchronization errors.

The chapter is organized as follows. In Section 2.1, we first present the channel models for P2P memoryless channels with synchronization errors focusing on the specific models considered in this thesis, then review the existing capacity upper and lower bounds for the considered channel models. In Section 2.2, we focus on asynchronous cooperative communication systems where we first review the general asynchronous cooperative system model, then provide a review on existing signaling solution for asynchronous cooperative systems. Finally, we summarize the chapter in Section 2.3.

2.1 Memoryless Channels with Synchronization Errors

2.1.1 Channel Models

A general memoryless channel with synchronization errors [10] is defined via a stochastic matrix $\{p(y_i|x_i), y_i \in \mathcal{Y}, x_i \in \mathcal{X}\}$ where $\mathcal{X}$ is the input alphabet (e.g., for a binary input channel $\mathcal{X} = \{0, 1\}$), and $\mathcal{Y}$ is the set of output symbols which may contain the null strings, $0 \leq p(y_i|x_i) \leq 1$, and $\sum_{y_i \in \mathcal{Y}} p(y_i|x) = 1$. In the other words, based on

Dobrushin's model [10], in a memoryless channel with synchronization errors, every transmitted bit is independently replaced with a random number of symbols such that deletion of a symbol is also probable, and the transmitter and receiver have no information about the position and pattern of the synchronization errors. Different specific models on channels with synchronization errors are considered in the literature, including insertion/deletion channels, e.g., the Gallager insertion/deletion channel [4], the sticky channel [19] and the segmented insertion/deletion channel [20]. As another example, the model in [21] considers timing errors modeled as a discrete-valued Markov process.

The most commonly used version of the general model by Dobrushin is the Gallager model [4] allowing for insertions, deletions and substitution errors in binary input channels in which every transmitted bit is either deleted with probability of d, replaced with two random bits with probability of i or flipped with probability of f or received correctly with probability of $1 - d - i - f$. The Gallager model can also be considered as the cascade of an insertion/deletion channel with the same insertion and deletion probabilities and a binary symmetric channel (BSC) with cross error probability of s where $s = \dfrac{f}{1-d-i}$. Substituting $i = s = 0$ in the Gallager model results into the binary deletion-only channel model in which only bit drop-out are possible. The binary deletion channel model can be easily generalized to the non-binary input cases as considered in [22] to model information transmission over a finite buffer channel in which every transmitted symbol may be independently deleted with probability of d.

In a sticky channel [19], only duplication errors can occur, so an insertion channel is obtained in which each symbol is independently duplicated with probability i. The number of duplications can also be a random variable by following a certain distribution function on the set of positive integers. An example of the sticky channel

is typing on a keyboard where instead of pressing a key only once, it is held for a long time. In [19], two special cases for the number of duplications in a binary input channel are studied in detail, namely elementary i.i.d. duplication channel where the number of duplications is not random and it is exactly one, and geometric i.i.d. duplication channel where the number of duplications follows a geometric distribution. The received sequence from a sticky channel preserves the block structure of the transmitted sequence, i.e., in transmitting each contiguous block of zeros (ones) a contiguous block of zeros (ones) is received which makes the study of the channel easier than the general insertion/deletion case. For example, considering 11101001101000 as the input sequence of the sticky channel, 1111110010001100011000 could be a possible output sequence. If we interpret the input and output sequences to 31122112 and 62132323 as the sequences of block lengths, respectively, then we see that the length of the block length sequences are the same. This property is exploited for calculating upper and lower bounds bounds on the capacity of the sticky channel [19].

Segmented insertion/deletion channel is introduced in [20] as another special case of the insertion/deletion channels to model synchronization errors arising from small differences between transmitter and receiver clock frequencies. Consecutively transmitted bits over the segmented deletion or insertion channel are (implicitly) partitioned into segments of b bits and each segment is either received correctly without any synchronization errors with probability of $1 - d$, or is corrupted by a fixed number of deletion or insertion errors with probability of d. As an instance, consider segments consisting of 4 bits and at most one deletion error within each segment, then in transmitting 110010111010, deletions of the first, sixth and eleventh bits are probable but deletions of both the first and second bits are not possible.

As another possible model, we can consider a binary input symmetric q-ary output channel (BSQC) with synchronization errors. As depicted in Fig. 2.1, such a

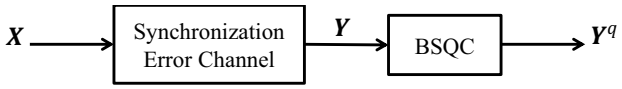

Figure 2.1: Binary input symmetric q-ary output channel with synchronization errors.

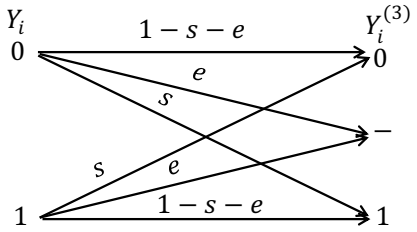

Figure 2.2: Input-output relation in the substitution/erasure channel ($P(Y_i^{(3)}|Y_i)$ for all $1 \leq i \leq |\boldsymbol{y}|$).

channel can be expressed as a concatenation of two independent channels in which the first one is a channel with only synchronization errors with input sequence $\boldsymbol{X}$ and output sequence $\boldsymbol{Y}$, and the second one is a BSQC with input sequence $\boldsymbol{Y}$ and output sequence $\boldsymbol{Y}^{(q)}$, and by a symmetric channel we refer to the definition given in [23, p. 94]. That is, a channel is symmetric if by dividing the columns of the transition matrix into sub-matrices, in each sub-matrix, each row is a permutation of any other row and each column is a permutation of any other column. For example, a channel with independent substitution, erasure and synchronization errors (sub/ers/synch channel) can be considered as a concatenation of a channel with only synchronization errors with input sequence $\boldsymbol{X}$ and output sequence $\boldsymbol{Y}$ and a substitution/erasure channel (binary input ternary output channel) with input sequence $\boldsymbol{Y}$ and output sequence $\boldsymbol{Y}^{(3)}$. In a substitution/erasure channel, each bit is independently flipped with probability s or erased with probability e, as depicted in Fig. 2.2. Another specific example is a binary input symmetric quaternary output channel with synchronization errors which can be decomposed into two independent channels such that the first one is a memoryless synchronization error channel and the second one is a memoryless binary input symmetric quaternary output channel shown in Fig. 2.3.

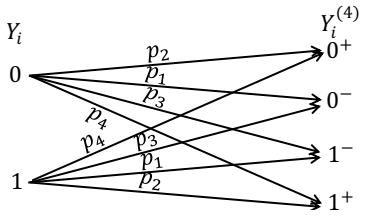

Figure 2.3: Input-output relation in the binary input quaternary output channel ($P(Y_i^{(4)}|Y_i)$ for all $1 \leq i \leq |\boldsymbol{y}|$).

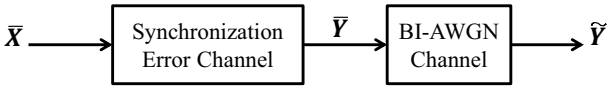

Figure 2.4: AWGN channel with synchronization errors.

We consider the case of a binary synchronization error channel in the presence of AWGN as well, in which the bits are transmitted using binary phase shift keying (BPSK) and the received signal contains AWGN in addition to synchronization errors. As illustrated in Fig. 2.4, this channel can be considered as the cascade of two independent channels where the first channel is a synchronization error channel and the second one is a binary input AWGN (BI-AWGN) channel. We use $\overline{\boldsymbol{X}}$ to denote the input sequence to the first channel which is a BPSK modulated version of the binary input sequence $\boldsymbol{X}$, i.e., the i-th symbol satisfies $\overline{X}_i = 1 - 2X_i$, and $\overline{\boldsymbol{Y}}$ to denote the output sequence of the first channel which input to the second one. $\widetilde{\boldsymbol{Y}}$ is the output sequence of the second channel that is the noisy version of $\overline{\boldsymbol{Y}}$, i.e.,

$$\widetilde{Y}_i = \overline{Y}_i + Z_i,$$

where Z_i's are i.i.d. Gaussian random variables with zero mean and a variance of σ^2, and $\widetilde{Y}_i$ and $\overline{Y}_i$ are the i^{th} received and transmitted bits of the second channel, respectively.

Note that the binary input symmetric q-ary output channel can be considered as a q-level quantized output version of the BI-AWGN channel (with a symmetric

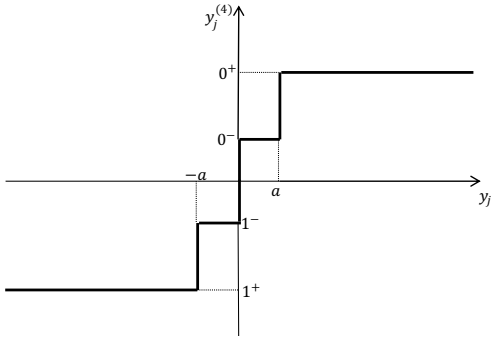

Figure 2.5: Four level quantization.

Table 2.1: Transition probabilities of the quantized AWGN channel.

	$P(Y_j^{(4)}\|\overline{Y}_j)$			
Y_j	$Y_j^{(4)}=0^+$	$Y_j^{(4)}=0^-$	$Y_j^{(4)}=1^-$	$Y_j^{(4)}=1^+$
0	$p_2=1-Q(\frac{1-a}{\sigma})$	$p_1=Q(\frac{1-a}{\sigma})-Q(\frac{1}{\sigma})$	$p_3=Q(\frac{1}{\sigma})-Q(\frac{1+a}{\sigma})$	$p_4=Q(\frac{1+a}{\sigma})$
1	p_4	p_3	p_1	p_2

quantizer). For example the case of 4-level quantizer is illustrated in Fig. 2.5, i.e.,

$$Y_j^{(4)}=\begin{cases} 0^+ & if \quad \widetilde{Y}_j \geq a \\ 0^- & if \quad 0 \leq \widetilde{Y}_j < a \\ 1^- & if \quad -a \leq \widetilde{Y}_j < 0 \\ 1^+ & if \quad \widetilde{Y}_j < -a \end{cases}. \tag{2.1}$$

Transition probabilities of the binary input symmetric quaternary output channel, resulting from a 4-level quantization based on (2.1) are reported in Table. 2.1 (where $Q(.)$ denotes the right tail probability of the standard normal distribution).

2.1.2 Bounds on the Capacity of Insertion and Deletion Channels

Dobrushin [10] proved under very general conditions that for a discrete memoryless channel with synchronization errors, Shannon's theorem on transmission rates applies and the information and transmission capacities are equal. The proof hinges on showing that information stability holds for the insertion/deletion channels and, as a result, capacity per bit of an i.i.d. insertion/deletion channel can be obtained by

$\lim_{N\to\infty} \max_{P(\boldsymbol{X})} \frac{1}{N} I(\boldsymbol{X};\boldsymbol{Y})$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the transmitted and received sequences, respectively, and N is the length of the transmitted sequence. On the other hand, there is no single-letter or finite-letter formulation which may be amenable for the capacity computation, and no results are available providing the exact value of the limit. In [24], authors extend Dobrushin's result on discrete memoryless channels with synchronization errors to the case of continuous output memoryless channels with synchronization errors.

Our main focus in this dissertation is on the insertion/deletion channels as the most common specific model for memoryless channels with synchronization errors, therefore, in the following, we review existing upper and lower bounds on the capacity of such channels.

Gallager [4] considered the use of convolutional codes over channels with synchronization errors, and derived an expression which represents an achievable rate for channels with insertion, deletion and substitution errors (whose model is specified earlier). The approach is to consider transmission of i.u.d. binary information sequences by convolutional coding and modulo-2 addition of a pseudo-random binary sequence (which could be considered as a watermark used for synchronization purposes), and computation of a rate that guarantees success by sequential decoding. The achievable rate, or the capacity lower bound, is given by the expression

$$C \geq 1 + d\log d + i\log i + c\log c + f\log(f), \tag{2.2}$$

where C is the channel capacity, $c = (1-d-i)(1-s)$ is the probability of correct reception, and $f = (1-d-i)s$ is the probability that a flipped version of the transmitted bit is received. The logarithm is taken base 2 resulting in transmission rates in bits/channel use. Substituting $i = 0$ in (2.2) gives a lower bound on the

capacity of the deletion-substitution channel C_{ds}, for $d \leq 0.5$, as

$$C_{ds} \geq 1 - H_b(d) - (1-d)H_b(s), \tag{2.3}$$

where $H_b(d) = -d\log(d) - (1-d)\log(1-d)$ is the binary entropy function. It is interesting to note that for $i = s = 0$ ($d = s = 0$) and $d \leq 0.5$ ($i \leq 0.5$), a lower bound on the capacity of the deletion-only channel (insertion-only channel) with deletion (insertion) probability of d (i), is equal to the capacity of a binary symmetric channel with a substitution error probability of d (i).

In an early work, Vvedenskaya and Dobrushin [25] has employed Monte Carlo methods to compute an achievable rate for the binary deletion channel by generating input codewords according to an m-th order Markov chain ($0 \leq m \leq 2$). Furthermore, in [22, 26], authors argue that, since the deletion channel has memory, optimal codebooks for use over deletion channels should have memory. Therefore, in [5, 26, 27], achievable rates are computed by using a random codebook of rate R with $2^{n \cdot R}$ codewords of length n, while each codeword is generated independently according to a symmetric first-order Markov process. At the receiver side, different decoding algorithms are proposed, e.g., in [26], if the number of codewords in the codebook that contain the received sequence as a subsequence is only one, the transmission is successful, otherwise an error is declared. The proposed decoding algorithms result in an upper bound for the incorrect decoding probability. Finally, the maximum value of R that results in success as $n \to \infty$ is an achievable rate, hence a lower bound on the transmission capacity of the deletion channel. The lower bound (2.2), for $i = s = 0$, is also proved in [26] using a different approach compared to the one taken by Gallager [4], i.e., by using codewords with i.u.d. elements.

In [12], a lower bound on the capacity of the binary deletion channel is obtained directly by lower bounding the information capacity $\lim_{N\to\infty} \frac{1}{N} \max_{P(\boldsymbol{X})} I(\boldsymbol{X};\boldsymbol{Y})$. In [12],

input sequences are considered as alternating blocks of zeros and ones (runs), where the length of the runs L are i.i.d. random variables following a particular distribution over positive integers with a finite expectation and finite entropy ($E(L), H(L) < \infty$ where $E(\cdot)$ and $H(\cdot)$ denote the expected value and entropy, respectively).

Another interesting lower bound on the binary deletion channel capacity is presented in [11] as $C_d \geq \frac{1-d}{9}$ for all d. In [11], the binary i.i.d. deletion channel is considered as a Poisson-repeat channel in which each bit is independently replaced by a Poisson number copies of itself with mean l. Then, by showing that if L_1 is a lower bound on the capacity of the Poisson-repeat channel then $L_1(1-d)$ is a lower bound on the deletion channel capacity, the lower bound $C_d \geq \frac{1-d}{9}$ is obtained. Obviously, for low values of d, this is a loose lower bound, e.g., for $d = 0$ the capacity of the deletion channel is equal to 1 while the derived lower bound is equal to $\frac{1}{9}$. However, as $d \to 1$, the lower bound becomes interesting.

There are also a few results on the capacity of the sticky channel in the literature [5, 12, 19]. In [5, 12], the authors derive lower bounds by using the same approach employed for the deletion channel. Whereas in [19], several upper and lower bounds are obtained by resorting to the Blahut-Arimoto algorithm (BAA) [28, 29] in an appropriate manner.

In [6, 30], Monte Carlo methods are used for computing lower bounds on the capacity of the insertion/deletion channels based on reduced-state techniques. In [30], the input process is assumed to be a stationary Markov process and lower bounds on the capacity of the deletion and insertion channels are obtained via Monte Carlo simulations considering both the first and second-order Markov processes as input. In [6], information rates for i.u.d. input sequences are computed for several channel models using a similar Monte Carlo approach where in addition to the insertions/deletions, effects of intersymbol interference (ISI) and AWGN are also investigated.

As a trivial upper bound on the deletion channel capacity, one can consider the erasure channel capacity, i.e., $C_d \leq 1 - d$. This is because, a genie-aided deletion channel with deletion probability d in which the positions of the deleted bits are revealed to the decoder is nothing but an erasure channel with erasure probability d. There are several papers deriving non-trivial upper bounds on the capacity of the insertion/deletion channels as well. Fertonani and Duman in [13] present several upper bounds on the capacity of the i.i.d. deletion channel by providing the decoder (and possibly the encoder) with some genie-aided information about the deletion process resulting in auxiliary channels whose capacities are certainly upper bounds on the capacity of the i.i.d. deletion channel. By providing the decoder with appropriate side information, a memoryless channel is obtained in such a way that BAA can be used for evaluating the capacity of the auxiliary channels (or, at least for computing a provable upper bound on their capacities). They also prove that by subtracting some value from the derived upper bounds, lower bounds on the capacity can be derived. The intuition is that the subtracted information is more than extra information added by revealing certain aspects of the deletion process. A nontrivial upper bound on the deletion channel capacity is also obtained in [31] where a different genie-aided decoder is considered. Furthermore, Fertonani and Duman in [14] extend their work [13] to compute several upper and lower bounds on the capacity of channels with insertion, deletion and substitution errors as well.

In two recent papers [32, 33], asymptotic capacity expressions for the binary deletion channel for small deletion probabilities are developed. It is proved in [33] that $C_d \leq 1-(1-O(d))H_b(d)$ (where $O(.)$ represents the standard Landau (big-O) notation) which clearly shows that for small deletion probabilities, $1 - H_b(d)$ is a tight lower bound on the deletion channel capacity. In [32], an expansion of the capacity for small deletion probabilities is computed with several dominant terms in an explicit form.

The idea of capacity expansion for $d \to 0$ employed in [32] is extended to the sticky and segmented deletion channel cases in [34] and [35], respectively. In [34], authors provide the sticky channel capacity expansion with several dominant terms valid for both $d \to 0$ and $d \to 1$. In [35], the capacity expansion is provided for the segmented deletion channel as $\frac{d}{b} \to 0$ where b and d denote the segment length and the probability of one bit loss from each segment, respectively.

Although binary deletion channels have received significant attention over the years, and many upper and lower bounds on their capacity have been derived, such studies for the non-binary case are largely missing. In the following we give a review of the state of the art.

Non-trivial lower bounds on the capacity of the non-binary deletion channels are provided in [22] where two different bounds are derived. More precisely, the achievable rates of the $2K$-ary input deletion channel are computed for i.i.d. and Markovian codebooks by considering a simple decoder which decides in favor of a sequence if the received sequence is the subsequence of only one transmitted sequence. The derived achievable rates for a $2K$-ary input deletion channel are given by

$$C_{2K} \geq \log\left(\frac{2K}{2K-1}\right) + (1-d)\log(2K-1) - H_b(d), \tag{2.4}$$

by considering i.i.d. codebooks, and

$$C_{2K} \geq \sup_{\gamma>0,\ 0<p<1} \left[-(1-d)\log\left((1-q)A + qB\right) - \gamma\log(e)\right] \tag{2.5}$$

by considering Markovian codebooks, with $q = \frac{1}{2K}\left(1 + \frac{(1-d)(2K-1)(2Kp-1)}{2K-1-d(2Kp-1)}\right)$, $A = \frac{e^{-\gamma}(1-p)}{(2K-1)(1-e^{-\gamma}(1-\frac{1-p}{2K-1}))}$ and $B = e^{-\gamma}\left((1-p)A + p\right)$.

Non-binary input alphabet channels with synchronization errors are also considered in [36] where the capacity of memoryless synchronization error channels in the presence of noise and the capacity of channels with week synchronization errors

(i.e., the transmitter and receiver are partly synchronized) have been studied. The main focus of the work in [36] is on asymptotic behavior of the channel capacity for large values of K.

2.2 Asynchronous Cooperative Communications

In this section, we turn our attention to cooperative communication systems suffering from time asynchronism among the relay nodes which may degrade the system performance. We first present the cooperative communication system model under consideration and then review some existing solutions applicable to radio terrestrial asynchronous cooperative communication systems.

2.2.1 Cooperative Communications

Multi-input multi-output (MIMO) communication systems have attracted a great deal of attention in recent years [37–39] as a promising solution to improve the performance of wireless communications systems by providing spatial diversity. Different signaling solutions are employed to extract the maximum possible enhancement in comparison with the single input single output (SISO) communication systems out of the MIMO communication systems, e.g., space-time coding (STC) [40] scheme which was first introduced by Alamouti in [41] for a two transmit antenna set up. However, due to the limited size and cost constraints, using more than one transmit and/or receive antennas may not always be possible. Furthermore, geographically separated nodes may form virtual transmit elements, known as relays, to provide spatial diversity in a distributed manner known as cooperative communication [15]. Due to the distributed nature of the cooperative communication systems, the signals traveling through various paths may experience different delays. Therefore, the signals received from different nodes may not be in perfect time alignment, and as a result the signaling solutions originally designed for MIMO communication and coopera-

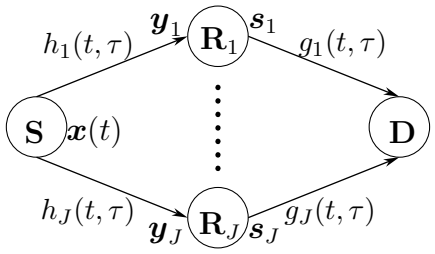

Figure 2.6: Relay channel with J relays.

tive communication systems under perfect synchronization assumptions may not be directly applicable in asynchronous cooperative communication systems.

In addition to the asynchronism issues, depending on the application and communication medium, different channel conditions may be experienced by the cooperating nodes, e.g., multipath or single tap channels, time varying or quasi-static channels. In this dissertation, we focus on two different scenarios, time varying frequency selective and quasi-static flat fading channel conditions. In the following, we present the cooperative communication system model for time varying frequency selective channel conditions by the understanding that the quasi-static flat fading channel is a special case. A general cooperative relay system with J relay nodes (R_j and $j \in \{1, \cdots, J\}$) helping a source node S deliver its data to a destination node D is shown in Fig. 2.6 in which all the channels from the source to the relays, i.e., $h_i(t,\tau)$ ($j \in \{1, \cdots, J\}$), and from the relays to the destination, i.e., $g_j(t,\tau)$ ($j \in \{1, \cdots, J\}$), are time varying frequency selective fading channels. The signal $\boldsymbol{x}(t)$ is broadcast by the source and received as $\boldsymbol{y}_j(t)$ at the j-th relay. We assume that there is no direct link between the source and the destination, however, one can model the system such that the signal broadcast by the source node is received at the destination node as well. Furthermore, each relay forwards the signal $\boldsymbol{s}_j(t)$ to the destination, i.e., performs amplify and forward relaying. We assume that the signal corresponding to the j-th relay is received D_j seconds later than the signal corresponding to the first relay,

i.e., $D_1 = 0$. Therefore, the received signal at the destination $\boldsymbol{y}(t)$ can be written as

$$\boldsymbol{y}(t) = \sum_{j=1}^{J} \int_{-\infty}^{\infty} g_j(t,\tau) s(t-\tau-D_j) d\tau. \tag{2.6}$$

We adopt the general asynchronous cooperative system model given in this section in Chapters 7 and 8, and propose new relaying schemes to combat asynchronism issues at the destination.

Different cooperation protocols and relaying techniques have been considered in the literature for terrestrial radio communication systems. Time division multiple access (TDMA) cooperation protocols have been first discussed in [15, 42] in which the relays are assumed to be half duplex. In half-duplex relaying, transmission from the source to the destination is done in two phases. In the first phase, the source is active by broadcasting its signal while the relay nodes are silent. In the second phase, the relay nodes are active where each relay transmits its interpretation of the signal received during the first phase to the destination based on the specific signaling approach. One may increase the spectral efficiency of the cooperative system by employing full-duplex relay nodes, e.g., [43, 44]. In full-duplex relaying schemes, no time division is performed through the transmission and the transmission and reception by each relay is done simultaneously.

A relay can employ different techniques to help the source node deliver its signal to the destination. Most common techniques considered in the literature are AF and decode and forward (DF) relaying protocols [44]. A DF relay first decodes the signal received from the source, then encodes and forwards its decoded sequence to the destination. In AF relaying, the signal received from the source is simply amplified and forwarded to the destination without any decoding/encoding procedures at the relays while different amplifying protocols can be used. In general, for both AF and DF relaying protocols, different power allocation strategies can be employed.

2.2.2 Review of Existing Signaling Solutions for Asynchronous Cooperative Communication Systems

To combat the asynchronism issues in the cooperative communication systems and obtain the maximum achievable spatial diversity (and possibly multipath diversity in frequency selective channels), different signaling solutions are provided in the literature, e.g., OFDM-based distributed space-time block coding (DSTBC) transmission schemes in [1, 3, 18, 45–47] and single carrier (SC)-DSTBC schemes in [2, 47–53]. In the following we review the existing signaling solutions under two different channel conditions, quasi-static flat fading and frequency selective channel conditions. Note that any flat fading channel (frequency non-selective channel) is in fact a frequency selective channel (multipath channel) with only one channel tap, therefore, the signaling solutions under frequency selective channel conditions are applicable for flat fading channel conditions as well.

2.2.2.1 Flat Fading Channel Conditions

Here, we provide a literature review on signaling solutions for asynchronous cooperative communication systems under quasi-static flat fading channel conditions, i.e., all the communicating channels under consideration are one-tap channels which are fixed during a specific transmission block.

In [47], an asynchronous cooperative communication system with DF relaying is considered. It is first argued that under flat fading channel conditions, the relative delays among the signals received from different relays make the effective channel at the destination frequency selective. Then, time reversal space-time coding (TR-STC) and space-time (ST)-OFDM schemes appropriate for frequency selective channels are employed to combat the asynchronism issues. For TR-STC scheme, time-reversal and complex conjugation operations and for the ST-OFDM scheme, OFDM symbol gen-

Table 2.2: Relaying strategy of the scheme proposed in [1].

	First OFDM symbol	Second OFDM Symbol
R_1	$\boldsymbol{Y}_{1,1}$	$\zeta_F(\boldsymbol{Y}_{1,2})$
R_2	$-\boldsymbol{Y}^*_{2,2}$	$\zeta_F\left(\boldsymbol{Y}^*_{2,1}\right)$

eration and complex conjugation operations are all implemented at the relay nodes. The main focus of the work is on the case that the relative delay among the relays are less than the transmitted symbols. However, a discussion on generalization to larger relative delays is also provided for which a time guard greater than the maximum possible delay needs to be inserted among every two transmitted blocks, either SC-based or OFDM-based. The main drawback of the scheme for large relative delay is the reduction in the spectral efficiency.

In [1], the ST-OFDM scheme proposed in [47] for DF relay nodes is generalized to asynchronous cooperative systems with two AF relays to achieve an Alamouti coding structure at the destination. To reduce the relay node computational complexity, the OFDM symbol generation process is implemented at the source node and the generated OFDM blocks are separated by adding a CP longer than the maximum possible delay between the relays. At the relay nodes, only time-reversal and complex conjugation operations are implemented as shown in Table 2.2 where two consecutive transmitted OFDM blocks are considered and $\boldsymbol{Y}_{i,j}$ represents the signal received at the i-th relay corresponding to the j-th transmitted OFDM block. Furthermore, $\zeta_F(.)$ denotes the time reversal operation over the OFDM block, i.e., $\zeta_F([Z_0, Z_1, \ldots, Z_{N-1}]) = [Z_0, Z_{N-1}, \ldots, Z_1]$, and $(.)^*$ denotes the complex conjugation operation. We would like to define another time-reversal operation as $\zeta(.)$ on SC-based transmitted blocks as $\zeta([z_0, \ldots, z_{N-1}]) = [z_{N-1}, \ldots, z_0]$ which is used in Chapter 7. The scheme is further extended to the arbitrary number of relay nodes in [45].

There are several works in the literature that develop ST code designs achieving spatial diversity in asynchronous cooperative systems, e.g., [49, 50, 53]. In [49], the ST coding design approach developed in [54] for MIMO communication systems is generalized to asynchronous cooperative systems assuming that the possible relative delays among the relays are multiples of the symbol duration T. The structure of the optimal detector is derived, and a sub-optimal detector with a reduced complexity is also presented. The focus is first on the transmission of the binary phase shift keying (BPSK) modulated symbols and it is shown that the sufficient condition to achieve full spatial diversity is to have a shift full rank (SFR) binary matrix, i.e., the matrix is full rank for all possible row shifts. Then, the scheme is generalized to higher modulation levels following the approach given in [55]. Furthermore, it is shown through simulations in [49] that with perfect channel state information and perfect knowledge about the relative delays at the destination, the asynchronous cooperative system performance is improved in comparison with the perfectly synchronized cooperative system. In [53], the code structure developed in [49] is extended to the case of fractional relative delays among the relays where the relative delays are considered as multiples of $\frac{p}{q}T$ with $\frac{p}{q}$ being a rational number.

In [50], the threaded algebraic space-time codes originally developed for perfectly synchronized MIMO communication systems in [56] are extended to the case of asynchronous cooperative systems with reasonably complex near-optimal lattice decoders for arbitrary number of relay nodes. The developed code design structure gives flexibility to obtain different transmission rates and can easily employ different signaling constellations.

2.2.2.2 Frequency Selective Channel Conditions

Flat fading channel assumption is not valid for certain wireless communication systems, e.g., UWA communication systems [16]. In the following, we review several

Table 2.3: Transmission scheme of the relays proposed in [2].

	Frame 0	Frame 1	$\cdots$	Frame $p_{i,j}$	$\cdots$	Frame P
Relay 1	$\boldsymbol{s}_1$	-$\zeta(\boldsymbol{s}_2^*)$	$\cdots$		$\cdots$	
Relay 2	$\boldsymbol{s}_2$	$\zeta(\boldsymbol{s}_1^*)$	$\cdots$		$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Relay i	$\boldsymbol{s}_i$		$\cdots$	$-\zeta(\boldsymbol{s}_j^*)$	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Relay j	$\boldsymbol{s}_j$		$\cdots$	$\zeta(\boldsymbol{s}_i^*)$	$\cdots$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Relay J	$\boldsymbol{s}_J$		$\cdots$		$\cdots$	$\zeta(\boldsymbol{s}_{J-1}^*)$

existing works on asynchronous cooperative communication systems under frequency selective channel conditions.

In [2], a DSTBC transmission approach is proposed which achieves both full spatial and full multipath diversities. The main drawback is that by increasing the number of relay nodes, the spectral efficiency is dramatically reduced. In a cooperative communication system with J relays each symbol block of length $J(N+1)$ is divided into J symbol sub-blocks of length $N+1$ as $\boldsymbol{s}_j = \{s_j(0), \cdots, s_j(N)\}$ for $j = \{1, \cdots, J\}$ and is cooperatively transmitted in $P+1 = \binom{J}{2} + 1$ time frames. The transmission scheme of the relays is reported in Table 2.3, where in the frame $p_{i,j} = \frac{(i-1)(2J-i)}{2} + j - i$ only nodes i and j $(i < j)$ are active while the other nodes are silent. Furthermore, by linearly combining the received frames at the destination, the transmitted symbol sub-block is obtained as

$$y_j(n) = \boldsymbol{g}^T \boldsymbol{s}_j(n - d_j) + w_j(n), \tag{2.7}$$

where d_j denotes the delay associated with the j-th relay (assuming $D_j = d_j T$ with T denoting the symbol time duration), $\boldsymbol{s}_j(n) = [s_j(n), s_j(n-1), \cdots, s_j(n-2L)]^T$,

$\boldsymbol{g}^T = \sum_{j=1}^{J} [g_j^*(L), \cdots, g_j^*(0)]\mathcal{G}_j$, $w_j(n)$ are i.i.d. zero mean Gaussian noise samples and

$$\mathcal{G}_j = \begin{bmatrix} g_j(0) & \cdots & \cdots & g_j(L) & 0 & \cdots & \cdots & 0 \\ 0 & g_j(0) & \cdots & \cdots & g_j(L) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & g_j(0) & \cdots & \cdots & g_j(L) & 0 \\ 0 & \cdots & \cdots & 0 & g_j(0) & \cdots & \cdots & g_j(L) \end{bmatrix}. \tag{2.8}$$

In fact $\boldsymbol{g}$ represents a maximal ratio combiner at the destination, therefore, the scheme proposed in [2] achieves the diversity order of $J(L+1)$. Furthermore, for the system with two relay nodes, the scheme gives an Alamouti coding structure at the receiver.

In [51], authors propose a distributed space-time trellis coding (DSTTC) scheme which under certain conditions can achieve both full cooperative and multipath diversities in asynchronous cooperative communication systems. It is assumed that in each time frame only the relays which correctly decoded the transmitted data sequence will participate in communication and forward convolutionally encoded data sequences to the destination. At the destination, Viterbi decoding is employed as the optimal ML decoding algorithm. Furthermore, it is shown that a generator matrix can achieve full spatial and multipath diversities if and only if the generator matrix has a full rank in the binary field for arbitrary relative delays among the relays. The authors only provide the exact code construction for the asynchronous cooperative network with at most three relays, and use computer search for cases with more than three relays, e.g., they give a code for $J = 4$ and $L = 2$.

The authors in [51] use the same idea as in [52] to develop distributed linear convolutional space-time coding (DLCSTC) schemes. Instead of Viterbi decoding, i.e., the optimal ML decoding, they utilize several sub-optimal equalizers and show that zero forcing (ZF), minimum mean square error (MMSE) and MMSE decision

feedback equalizer (MMSE-DFE) achieve full diversity for frequency selective channels. A construction scheme for arbitrary values of J and L is provided along with a few specific codes for different values of J and L.

The power of OFDM transmission schemes to combat the multipath effects under frequency selective channels is widely studied in the literature, e.g., see [57] and references therein. In some recent literature, e.g., [3, 18, 46], OFDM-based schemes are employed to overcome both asynchronism among the relay nodes and the multipath effects over frequency selective channels.

In [18], a space-frequency coding approach is proposed which achieves both full spatial and full multipath diversities. In a given time frame, it is assumed that J relays correctly decode the transmitted data sequence of length N_s and cooperate in transmitting the data sequence to the destination while there is no direct link between the source and destination. First, the data sequence of length N_s is encoded by an $N_s \times J$ code matrix, then the j-th relay takes an N point inverse fast Fourier transform (IFFT) of the j-th column of the matrix code and forwards the resulting block along with its CP which forms the desired OFDM block. To combat both asynchronism among the relays and inter-frame interference, the length of the CP has to be at least $D_{max} + \max_{j,l}\{\tau_j(l)\}$ with D_{max} denoting the maximum possible relative delay among the relays and $\tau_j(l)$ denoting the delay of the l-th path from the j-th relay to the destination. At the destination, the CP is removed and an N point FFT of the remaining noisy block is computed. Since the relays are not perfectly synchronized, the CP removal can be only aligned with one of the relays and it would be unaligned for the others. Due to the properties of the OFDM technique the only effect of this misalignment is a phase shift in the time domain version of the received signal that can be covered in the channel coefficients. In addition, the authors provide high rate space-frequency code construction techniques to achieve full diversity.

The OFDM-based scheme designed under flat fading channel assumption in [1] is extended to the frequency selective channel conditions in [3] in which OFDM modulation is implemented at the source node by adding a CP to resolve both multipath and asynchronism issues and relays only perform time reversal and complex conjugation operations. The main difference with the scheme in [1] is in the CP removal for which after conventional CP removal, an appropriate cyclic shifted version of the received signal is fed to the FFT block.

A STC cooperative system using OFDM transmission is proposed in [46], where a complete OFDM frame structure aimed at timing and frequency synchronization and channel estimation at the receiver side is proposed. A cooperative system with only one relay and two phase transmission is considered where in the first phase (silent phase) only the source transmits its signals and in the second phase (cooperation phase) if the relay decodes the received signal during the silent phase correctly, then both the source and relay transmit, otherwise only source transmits in the cooperation phase as well. In the cooperation phase, the source node transmits half of the space-time coded version of the original transmitted signal (during the silent phase) and relay (if participating in cooperation) transmits the other half. By choosing the length of the CP greater than the maximum value of the delay between the source and the relay, asynchronism between the source and the relay is taken care of.

2.3 Chapter Summary

In this chapter, we gave the general system models and provided brief reviews of the related existing literature for both P2P systems with synchronization errors and asynchronous cooperative communication systems. We presented the general channel models for memoryless P2P systems with synchronization errors and gave some specific models which are considered in the remainder of the thesis. We reviewed existing

results on upper and lower bounds on the capacity of the insertion/deletion channels as the specific models considered in the rest of the thesis. We then turned our attention to asynchronous cooperative communication systems, and presented the general cooperative communication systems model. We also provided a review on existing signaling solutions to combat the asynchronism issues in cooperative communication systems.

Chapter 3

Analytical Lower Bounds on the Capacity of Insertion and Deletion Channels

In this chapter, we develop several analytical lower bounds on the capacity of binary insertion and deletion channels by considering i.u.d. inputs and computing lower bounds on the mutual information between the input and output sequences. For the deletion channel, we consider three different models: usual i.i.d. deletion channel, i.i.d. deletion-substitution channel and i.i.d. deletion channel with AWGN. The latter two models are introduced to incorporate effects of the channel noise along with the synchronization errors. For the insertion channel case we assume that the transmitted bits are replaced with two bits with a certain probability independently of any other insertion events. We consider two specific cases: Gallager's model where the pair of bits are random and uniform over the four possibilities, and the sticky channel where transmitted bits are duplicated. The general approach taken is similar in all cases, however the specific computations differ. Furthermore, the approach yields a useful lower bound on the capacity for a wide range of deletion probabilities for the deletion channels, while it provides a beneficial bound only for very low insertion probabilities for the insertion models adopted. We emphasize the importance of these results by noting that 1) our results are the first analytical bounds on the capacity of deletion-AWGN channels, 2) the results developed are the best available analytical lower bounds on the deletion-substitution case, 3) for the deletion only channel, our results competes well with the best available lower bounds for small deletion probabilities and they explicitly obtain the first order terms in the recently proved capacity expansions, 4) for both sticky and Gallager insertion channel models, the new lower bounds improve the existing results for small insertion probabilities.

We first give an introduction on insertion and deletion channels and state the main contributions of this chapter in Section 3.1. In Section 3.2, we introduce our

general approach for lower bounding the mutual information of the input and output sequences for insertion/deletion channels. In Section 3.3, we apply the introduced approach to the i.i.d. deletion, deletion-substitution and deletion-AWGN channels and present analytical lower bounds on their capacities, and compare the resulting expressions with earlier results. In Section 3.4, we provide lower bounds on the capacity of the sticky and random insertion channels and comment on our results with respect to the existing literature. In Section 3.5, we compute the lower bounds for a number of insertion/deletion channels, and finally, we provide a summary of the chapter in Section 3.6.

3.1 Introduction

In modeling digital communication systems, we often assume that the transmitter and the receiver are completely synchronized; however, achieving a perfect time-alignment between the transmitter and receiver clocks is not possible in all communication systems and synchronization errors are unavoidable. A useful model for synchronization errors assumes that the number of received bits may be less or more than the number of transmitted bits. In other words, insertion/deletion channels may be used as appropriate models for communication channels that suffer from synchronization errors. Due to the memory introduced by the synchronization errors, an information theoretic study of these channels proves to be very challenging. For instance, even for seemingly simple models such as an i.i.d. deletion channel, an exact calculation of the capacity is not possible and only upper/lower bounds (which are often loose) are available.

In this chapter, we compute analytical lower bounds on the capacity of the i.i.d. deletion channel: without substitution errors, with substitution errors and in the presence of additive white Gaussian noise (AWGN), and random insertion channel with random insertions or duplications, by lower bounding the mutual information

rate between the transmitted and received sequences for i.u.d. inputs. We particularly focus on the small insertion/deletion probabilities with the premise that such small values are more practical from an application point of view. Specific models adopted are as follows. For random insertion and deletion channels, we use the general insertion/deletion channel model proposed in [4], where every bit is independently deleted with probability d or replaced with two randomly chosen bits with probability i, while neither the transmitter nor the receiver have any information about the positions of deletions and insertions, and undeleted bits are flipped with probability s and bits are received in the correct order. By a deletion-only channel we refer to an insertion/deletion channel with $i = s = 0$; by a deletion-substitution channel we refer to an insertion/deletion channel with $i = 0$; by a deletion-AWGN channel we refer to a deletion channel in which undeleted bits are received in the presence of AWGN, that can be modeled by a combination of a deletion-only channel with an AWGN channel such that every bit first goes through a deletion-only channel and then through an AWGN channel. By a sticky channel we refer to a binary insertion channel that only duplication errors are possible such that every bit is duplicated with probability of i [19]; and, by a random insertion channel we refer to an insertion/deletion channel with $d = s = 0$.

We note that our idea is somewhat similar to the idea of directly lower bounding the information capacity instead of lower bounding the transmission capacity as employed in [12]. However, there are fundamental differences in the main methodology as will become apparent later. For instance, our approach provides a procedure that can easily be employed for many different channel models with synchronization errors as such we are able to consider deletion-substitution, deletion-AWGN and random insertion channels, in addition to the deletion-only and sticky channel formulations as studied in [12]. Other differences include adopting a finite-length

transmission which is proved to yield a lower bound on the capacity after subtracting some appropriate term, and the complexity in computing numerically the final expression is much lower in many versions of our results.

Finally, we emphasize that the new approach and the obtained results in the existing literature are improved in several different aspects. In particular, the contributions of the chapter include

- development of a new approach for deriving achievable information rates for insertion/deletion channels that can be applied to the other channel models and possibly other input distributions as well,
- the first analytical lower bound on the capacity of the deletion-AWGN channel,
- tighter analytical lower bounds on the capacity of the deletion-substitution channel for all values of deletion and substitution probabilities compared to the existing analytical results,
- tighter analytical lower bounds on the capacity of the random insertion and sticky channels for small values of insertion probabilities compared to the existing lower bounds,
- very simple lower bounds on the capacity of insertion/deletion channels.

Regarding the final point, we note that our results for the i.i.d. deletion channel are in agreement with the asymptotic results of [32, 33] in the sense of capturing the dominant capacity expansion terms. Our results, however, are provable lower bounds on the capacity, while the existing asymptotic results are not amenable for numerical calculation (as they contain big-O terms).

3.2 Main Approach

We rely on lower bounding the information capacity of insertion/deletion channels directly as justified by [10], where it is shown that, for a memoryless channel with synchronization errors, the Shannon's theorem on transmission rates applies and the information and transmission capacities are equal, and thus every lower bound on the information capacity of an insertion/deletion channel is a lower bound on the transmission capacity of the channel. Our approach is different than most existing work on finding lower bounds on the capacity of the insertion/deletion channels where typically the transmission capacity is lower bounded using a certain codebook and particular decoding algorithms. The idea we employ is similar to the work in [12] which also considers the information capacity $\lim_{N\to\infty} \frac{1}{N} \max_{P(\boldsymbol{X})} I(\boldsymbol{X};\boldsymbol{Y})$ and directly lower bounds it using a particular input distribution to arrive at an achievable rate result.

Our primary focus is on the small deletion and insertion probabilities. As also noted in [32], for such probabilities it is natural to consider binary i.u.d. input distribution. This is justified by noting that when $d = i = 0$, i.e., for a binary symmetric channel, the capacity is achieved with independent and symmetric binary inputs, and hence we expect that for small deletion/insertion probabilities, binary i.u.d. inputs are not far from the optimal input distribution.

Our methodology is to consider a finite length transmission of i.u.d. bits over the insertion/deletion channel, and to compute (more precisely, lower bound) the mutual information between the input and the resulting output sequences. As proved in [13] for a deletion-only channel, such a finite length transmission in fact results in an upper bound on the mutual information supported by the insertion/deletion channels; however, as also shown in [13], if a suitable term is subtracted from the mutual information, a provable lower bound on the achievable rate, hence the channel

capacity, results. The following theorem provides this result in a slightly generalized form compared to [13].

Theorem 1. *For binary input random insertion and i.i.d. deletion channels, for any input distribution and any $n > 0$, the channel capacity C can be lower bounded by*

$$C \geq \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y}) - \frac{1}{n} H(\boldsymbol{T}), \tag{3.1}$$

where $H(\boldsymbol{T}) = -\sum_{j=0}^{n} \left[\binom{n}{j} p^j (1-p)^{n-j} \log \left(\binom{n}{j} p^j (1-p)^{n-j} \right) \right]$ with the understanding that $p = d$ for the deletion channel case and $p = i$ in the insertion channel case, and n is the length of the input sequence $\boldsymbol{X}$.

Proof. This is a slight generalization of a result in [13] which shows that Eqn. (3.1) is valid for the i.i.d. deletion channel. It is easy to see that [13], for any random process $\boldsymbol{T}^N$, and for any input distribution $P(\boldsymbol{X}^N)$, we have

$$C \geq \lim_{N\to\infty} \frac{1}{N} I(\boldsymbol{X}^N; \boldsymbol{Y}^N, \boldsymbol{T}^N) - \lim_{N\to\infty} \frac{1}{N} H(\boldsymbol{T}^N), \tag{3.2}$$

where C is the capacity of the channel and N is the length of the input sequence $\boldsymbol{X}^N$. We assume that the input bits in both insertion and deletion channels are divided into Q blocks of length n ($\boldsymbol{X}^N = \{\boldsymbol{X}_j\}_{j=1}^{Q}$). We define the random process $\boldsymbol{T}^N$ in the following manner. For an random insertion channel, $\boldsymbol{T}^{N,i}$ is formed as the sequence $\boldsymbol{T}^{N,i} = \{T_j^i\}_{j=1}^{Q}$ which denotes the number of insertions that occur in each block of length n transmission. For a deletion channel, $\boldsymbol{T}^{N,d} = \{T_j^d\}_{j=1}^{Q}$ represents the number of deletions occurring in transmission of each block. Since insertions (deletions) for different blocks are independent, the random variables $T_j = T_j^{N,i}$ ($T_j^{N,d}$) $j = 1, ..., Q$ are i.i.d., and transmission of different blocks are independent. Therefore, we can rewrite Eqn. (3.2) as

$$\begin{aligned} C &\geq \frac{1}{n} I(\boldsymbol{X}_j; \boldsymbol{Y}_j) - \frac{1}{n} H(\boldsymbol{T}_j) \\ &= \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y}) - \frac{1}{n} H(\boldsymbol{T}). \end{aligned} \tag{3.3}$$

Noting that the random variable denoting the number of deletions or insertions as a result of n bit transmission is binomial with parameters n and d (or, i) the proof follows. □

Several comments on the specific calculations involved are in order. Theorem 1 shows that for any input distribution and any transmission length, Eqn. (3.1) results in a lower bound on the capacity of the deletion channel (or insertion channel). Therefore, employing any lower bound on the mutual information rate $\frac{1}{n}I(\boldsymbol{X};\boldsymbol{Y})$ in Eqn. (3.1) also results in a lower bound on the capacity of the insertion/deletion channel. Due to the fact that obtaining the exact value of the mutual information rate for any n is infeasible, we first derive a lower bound on the mutual information rate for i.u.d. input sequences and then employ it in Eqn. (3.1). Based on the formulation of the mutual information, obviously

$$I(\boldsymbol{X};\boldsymbol{Y}) = H(\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}), \tag{3.4}$$

thus by calculating the exact value of the output entropy or lower bounding it and obtaining the exact value of the conditional output entropy or upper bounding it, the mutual information is lower bounded. For deletion and random insertion channels, we are able to obtain the exact value of output sequence probability distribution when i.u.d. input sequences are used, hence the exact value of the output entropy (the differential output entropy for the deletion-AWGN channel) is available. However, for the sticky channel obtaining the exact probability of all output sequences resulting from i.u.d. input sequences of length n is infeasible and we are only able to obtain the exact probability of output sequences with at most 2 insertions. Therefore, we are able to obtain a lower bound on the output entropy (by also deriving a manageable lower bound on the remaining terms of the entropy expression). For small insertion probabilities, by focusing on the outputs with at most two insertions and by choosing

a reasonable n, we do not lose too much information, because the probability of more than two insertions is much lower than the dominant events.

In deriving the conditional output entropy for the deletion-only, deletion-substitution and random insertion channels and also deriving the conditional differential entropy of the output sequence for the deletion-AWGN channel, we cannot obtain the exact probability of all the possible output sequences conditioned on a given input sequence. For deletion channels, we compute the probability of all possible deletion patterns for a given input sequence, and treat the resulting sequences as if they are all distinct to find a provable upper bound on the conditional entropy term. Clearly, we are losing some tightness, as different deletion patterns may result in the same sequence at the channel output. For the random insertion channel, we calculate the conditional probability of the output sequences resulting from at most one insertion, and derive an upper bound on the part of the conditional output entropy expression that results from the output sequences with multiple insertions. For the sticky channel, we are able to compute the conditional probabilities of all possible output sequences for all possible inputs and as a result obtain the exact value of the conditional output entropy.

3.2.1 Notation

We denote a finite binary sequence of length n with K runs by $(b; n_1, n_2, ..., n_K)$, where $b \in \{0, 1\}$ denotes the first run type and $\sum_{k=1}^{K} n_k = n$. For example, the sequence 001111011000 can be represented as (0;2,4,1,2,3). We use four different ways to denote different sequences; $\boldsymbol{x}(b; n^x; K^x)$ represents every sequence belonging to the set of sequences of length n^x with K^x runs and by the first run of type b, $\boldsymbol{x}(b; n^x; K^x; l)$ represents a sequence $\boldsymbol{x}(b; n^x; K^x)$ which has l runs of length one ($l = \sum_{k=1}^{K^x} \delta(n_k^x - 1)$ where $\delta(.)$ denotes the Kronecker delta function), $\boldsymbol{x}(n^x)$ represents every sequence of length n^x, and $\boldsymbol{x}$ shows every possible sequence. The set of all input sequences

is shown by $\mathcal{X}$, and the set of output sequences of the deletion, sticky and random insertion channels are shown by $\mathcal{Y}^d$, $\mathcal{Y}^s$ and $\mathcal{Y}^i$, respectively. $\mathcal{Y}^d_{-a}$, $\mathcal{Y}^s_{+b}$ and $\mathcal{Y}^i_{+c}$ denote the set of output sequences resulting from a deletions, b duplications and c random insertions, respectively, and $\mathcal{Y}^d(\boldsymbol{x}-a)$, $\mathcal{Y}^s(\boldsymbol{x}+b)$ and $\mathcal{Y}^i(\boldsymbol{x}+c)$ denote the set of output sequences resulting from a deletions from, b duplications into, and c random insertions into, the input sequence $\boldsymbol{x}$, respectively. We denote the deletion pattern of length J in a sequence of length n with K runs by $D(n;K;J) = (j_1, j_2, ..., j_K)$, where j_k denotes the number of deletions in the k-th run and $\sum_{k=1}^{K} j_k = J$, and also we denote the duplication pattern of length L in a sequence of length n with K runs by $I(n;K;L) = (l_1, l_2, ..., l_K)$, where l_k denotes the number of duplications in the k-th run and $\sum_{k=1}^{K} l_k = L$. The outputs resulting from a given deletion pattern $D(n;K;J) = (j_1, j_2, ..., j_K)$ and a given duplication pattern $I(n;K;L) = (l_1, l_2, ..., l_K)$, are denoted by $D(n;K;J) * \boldsymbol{x}(n;K) = (n_1 - j_1, n_2 - j_2, ..., n_K - j_K)$ and $I(n;K;L) * \boldsymbol{x}(n;K) = (n_1 + l_1, n_2 + l_2, ..., n_K + l_K)$, respectively. The sets $\mathcal{D}^n_K(J)$ and $\mathcal{I}^n_K(L)$ represent the set of all deletion patterns of length J and the duplication patterns of length L into a sequence of length n and with K runs, respectively.

3.3 Lower Bounds on the Capacity of Deletion Channels

As mentioned earlier, we consider three different variations of the binary deletion channel: i.i.d. deletion-only channel, i.i.d. deletion and substitution channel (deletion-substitution channel), and i.i.d. deletion channel in the presence of AWGN (deletion-AWGN channel). The results utilize the idea and approach of the previous section. For ease of exposition, we start with the deletion-only channel even though the other two are generalizations of this model, and the results boil down to the one on the deletion-only channel if $s = 0$ (or $\sigma^2 = 0$ for the deletion-AWGN channel). We then consider the extensions to the latter two channel models.

3.3.1 I.I.D. Deletion Channel

A binary input i.i.d. deletion channel is a channel in which each bit is independently deleted with a probability d and the receiver and the transmitter do not have any information about the position of deletions, and non-deleted bits are received correctly and in the correct order.

Lemma 1. *For any $n > 0$, the capacity of an i.i.d. deletion channel C_d, with a deletion probability of d is lower bounded by*

$$C_d \geq 1 - d - H_b(d) + \frac{1}{n}\sum_{j=1}^{n} W_j(n)\binom{n}{j}d^j(1-d)^{n-j}, \tag{3.5}$$

where

$$W_j(n) = \frac{1}{\binom{n}{j}}\sum_{l=1}^{n-1} 2^{-l-1}(n-l+3)\sum_{j'=1}^{j}\binom{l}{j'}\binom{n-l}{j-j'}\log\binom{l}{j'} + 2^{-n+1}\log\binom{n}{j}. \tag{3.6}$$

□

Before proving the lemma, we would like to make a few comments. First of all, the new lower bound is tighter than the one proved in [4] (Eqn. (2.2) with $i = s = 0$) which is the simplest analytical lower bound on the capacity of the deletion channel. The amount of improvement in (3.5) over the one in (2.2) is $\frac{1}{n}\sum_{j=1}^{n} W_j(n)\binom{n}{j}d^j(1-d)^{n-j} - d$, which is guaranteed to be positive.

In [32], it is shown that

$$C_d = 1 + d\log(d) - A_1 d + O(d^{1.4}), \tag{3.7}$$

where $A_1 = \log(2e) - \sum_{l=1}^{\infty} 2^{-l-1} l\log(l)$, and $O(p^\alpha)$ represents the standard Landau (big O) notation. A similar result in [33] is provided, that is

$$C_d \leq 1 - (1 - O(d))H_b(d), \tag{3.8}$$

which shows that $1 - H_b(d)$ is a tight lower bound for small deletion probabilities. If we consider the new capacity lower bound in (3.5), and represent $(1-d)\log(1-d)$ by its Taylor series expansion, we can readily write

$$C_d \geq 1 + d\log(d) - (\log(2e) - W_1(n))d + d^2 f(n,d), \tag{3.9}$$

where $f(n,d)$ is a polynomial function. On the other hand for $W_1(n)$, if we let n go to infinity, we have

$$\begin{aligned} \lim_{n\to\infty} W_1(n) &= \lim_{n\to\infty} \left[\frac{1}{n} \sum_{l=1}^{n-1} 2^{-l-1}(n-l+3)l\log(l) + \frac{\log(n)}{2^{n-1}} \right] \\ &= \sum_{l=1}^{\infty} 2^{-l-1} l \log(l). \end{aligned} \tag{3.10}$$

Therefore, we observe that the derived lower bound on the capacity captures the first order term of the capacity expansion (3.7). This is an important result as the the capacity expansions in [32, 33] are asymptotic and do not lend themselves for a numerical calculation of the transmission rates for any non-zero value of the deletion probability.

A final comment is on a simplified form of the lower bound for small values of deletion probability. By invoking the inequalities $(1-d)^m \geq [1 - md + \binom{m}{2}d^2 - \binom{m}{3}d^3]$ and $(1-d)^m \geq 1 - md$, and ignoring some positive terms ($d^j(1-d)^{n-j}$ for $j \geq 3$), we can write

$$\begin{aligned} C_d \geq\ & 1 - H_b(d) + d(W_1(n) - 1) + d^2 \frac{n-1}{2}(W_2(n) - 2W_1(n)) \\ & + d^3 \binom{n-1}{2}(W_1(n) - W_2(n)) - d^4 \binom{n-1}{3} W_1(n). \end{aligned} \tag{3.11}$$

We now turn our attention to the proof of Lemma 1. We need the following two propositions. The first one gives the exact value of the output entropy for an i.i.d. deletion channel with i.u.d. inputs of length n. The second one provides an upper bound on the conditional output entropy for i.u.d. inputs of length n as obtaining the exact value of the mutual information rate seems infeasible.

Proposition 1. *For an i.i.d. deletion channel with i.u.d. input sequences of length n, the output entropy is given by*

$$H(\boldsymbol{Y}) = n(1-d) + H(\boldsymbol{T}), \tag{3.12}$$

where $\boldsymbol{Y}$ denotes the output sequence and $H(\boldsymbol{T})$ is as defined in Eqn. (3.1)*.*

Proof. By using the fact that with i.u.d. input sequences, all the elements of the set of outputs with j deletions $\mathcal{Y}_{-j}^{d}$ are identically distributed, we have

$$P(\boldsymbol{y}(n-j)) = \frac{1}{2^{n-j}}\binom{n}{j}d^j(1-d)^{n-j}, \tag{3.13}$$

where $\binom{n}{j}d^j(1-d)^{n-j}$ is the probability of exactly j deletions occurring in n use of the channel. Therefore, we obtain

$$\begin{aligned} H(\boldsymbol{Y}) &= \sum_{\boldsymbol{y}} -P(\boldsymbol{y})\log(P(\boldsymbol{y})) \\ &= \sum_{j=0}^{n}\binom{n}{j}d^j(1-d)^{n-j}\log\left(\frac{2^{n-j}}{\binom{n}{j}d^j(1-d)^{n-j}}\right) \\ &= n(1-d) + H(\boldsymbol{T}). \end{aligned} \tag{3.14}$$

□

Proposition 2. *For an i.i.d. deletion channel, with i.u.d. input sequences of length n, the entropy of the output sequence $\boldsymbol{Y}$ conditioned on the given input $\boldsymbol{X}$, is upper bounded by*

$$H(\boldsymbol{Y}|\boldsymbol{X}) \leq nH_b(d) - \sum_{j=1}^{n} W_j(n)\binom{n}{j}d^j(1-d)^{n-j}, \tag{3.15}$$

where $W_j(n)$ is given in Eqn. (3.5)*.*

Proof. To obtain the conditional output entropy, we need to compute the probability of all possible output sequences resulting from every possible input sequence $\boldsymbol{x}$ ($P(\boldsymbol{Y}|\boldsymbol{x})$). For a given $\boldsymbol{x} = (b; n_1, n_2, ..., n_k)$ and for a specific deletion pattern

$D(n;K;j) = (j_1, ..., j_K)$, such that j_k denotes the number of deletions in the k-th run, we can write

$$P\left(D(n;K;j) = (j_1, ..., j_K)\middle|\boldsymbol{x}(b; n_1, ..., n_K)\right) = \binom{n_1}{j_1}...\binom{n_K}{j_K} d^j (1-d)^{n-j}. \quad (3.16)$$

However, there is a difficulty as two different possible deletion patterns, $D(n;K;j) = (j_1, ..., j_K)$ and $D'(n;K;j) = (j'_1, ..., j'_K)$, may convert a given input sequence $\boldsymbol{x}(n;K)$ into the same output sequence, i.e., $D(n;K;j) * x(n;K) = D'(n;K;j) * x(n,K)$. This occurs when successive runs are completely deleted, for example, in transmitting $(1;2,1,2,3,2) = 1101100011$, if the second, third and fourth runs are completely deleted, by deleting one bit from the first run, $(1,1,2,3,0) * (1;2,1,2,3,2) = (1;1,0,0,0,2) = 111$, or the last run, $(0,1,2,3,1) * (1;2,1,2,3,2) = (1;2,0,0,0,1) = 111$, the same output sequences are obtained. This difficulty can be addressed using

$$\sum_t -p_t \left(\log \sum_t p_t\right) \leq \sum_t -p_t \log(p_t), \quad (3.17)$$

which is trivially valid for any set of probabilities $(p_1, ..., p_t, ...)$. Applying this result in Eqn. (3.16), an upper on the conditional output entropy is obtained. Hence, for $\boldsymbol{x}(b;n;K^x) = (b; n_1^x, ..., n_{K^x}^x)$, we can write (for more details see Appendix A)

$$H\left(\boldsymbol{Y}\middle|\boldsymbol{x}(b;n;K^x)\right) \leq nH_b(d) - \sum_{j=0}^{n} d^j(1-d)^{n-j} \sum_{k=1}^{K^x} \sum_{j_k=0}^{j} \binom{n_k^x}{j_k}\binom{n-n_k^x}{j-j_k} \log\binom{n_k^x}{j_k}. \quad (3.18)$$

Therefore, by considering i.u.d. input sequences, we have

$$\begin{aligned} H(\boldsymbol{Y}|\boldsymbol{X}) &= \sum_{\boldsymbol{x}\in\mathcal{X}} \frac{1}{2^n} H(\boldsymbol{Y}|\boldsymbol{x}) \\ &\leq nH_b(d) - \sum_{j=0}^{n} \frac{d^j(1-d)^{n-j}}{2^n} \sum_{\boldsymbol{x}\in\mathcal{X}} \sum_{k=1}^{K^x} \sum_{j_k=0}^{j} \binom{n_k^x}{j_k}\binom{n-n_k^x}{j-j_k} \log\binom{n_k^x}{j_k}. \end{aligned} \quad (3.19)$$

On the other hand, we can write

$$\sum_{\boldsymbol{x}\in\mathcal{X}}\sum_{k=1}^{K^x}\sum_{j_k=0}^{j}\binom{n_k^x}{j_k}\binom{n-n_k^x}{j-j_k}\log\binom{n_k^x}{j_k}=\sum_{j'=0}^{j}\sum_{l=1}^{n}\#R(l,n)\binom{l}{j'}\binom{n-l}{j-j'}\log\binom{l}{j'}, \tag{3.20}$$

where $\#R(l,n)$ is the number of runs of length of l among all 2^n input sequences with a length of n. It is obvious that $\#R(n,n)=2$, and for $1\leq l\leq n-1$, we have

$$\#R(l,n)=2\sum_{K=2}^{n-l+1}\binom{n-l-1}{K-2}K=2^{n-l-1}(n-l+3). \tag{3.21}$$

Finally, by substituting Eqns. (3.20) and (3.21) in Eqn. (3.19), Eqn. (3.15) results, completing the proof. □

We can now complete the proof of the main lemma of the section.

Proof of Lemma 1: In Theorem 1, we showed that for any input distribution and any transmission length, Eqn. (3.1) results in a lower bound on the capacity of the i.i.d. deletion channel. On the other hand, any lower bound on the information rate can also be used to derive lower bound on the capacity. Due to the definition of the mutual information, Eqn. (3.4), by obtaining the exact value of the output entropy (Proposition 1) and upper bounding the conditional output entropy (Proposition 2) the mutual information is lower bounded. Finally, by substituting Eqns. (3.12) and (3.15) into Eqn. (3.1), Lemma 1 is proved. □

3.3.2 Deletion-Substitution Channel

In this section, we consider a binary deletion channel with substitution errors in which each bit is independently deleted with probability d, and transmitted bits are independently flipped with probability s. The receiver and the transmitter do not have any information about the position of deletions or the transmission errors. As shown in Fig. 3.1, this channel can be considered as a cascade of an i.i.d. deletion channel with a deletion probability d and output sequence $\boldsymbol{Y}$, and a binary symmetric

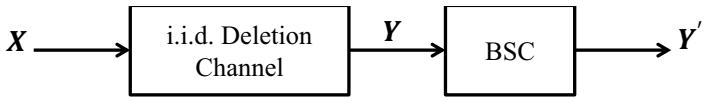

Figure 3.1: Deletion-substitution channel.

channel (BSC) with a cross-over error probability s and output sequence $\boldsymbol{Y}'$. For such a channel model the following lemma is a lower bound on the capacity.

Lemma 2. *For any $n > 0$, the capacity of the i.i.d. deletion-substitution channel C_{ds}, with a substitution probability s and a deletion probability d, is lower bounded by*

$$C_{ds} \geq 1 - d - H_b(d) + \frac{1}{n}\sum_{j=1}^{n} W_j(n)\binom{n}{j}d^j(1-d)^{n-j} - (1-d)H_b(s), \qquad (3.22)$$

where $W_j(n)$ is as given in Eqn. (3.5). □

Before proving the lemma, we would like to emphasize that the only existing analytical lower bound on the capacity of deletion-substitution channels is derived in [4] (Eqn. (2.3)). In comparing the lower bound in Eqn. (2.3) with the lower bound in Eqn. (3.22), we observe that the new lower bound improves the previous one by $\frac{1}{n}\sum_{j=1}^{n} W_j(n)\binom{n}{j}d^j(1-d)^{n-j} - d$, which is guaranteed to be positive.

In comparing the lower bounds in Eqns. (3.5) and (3.22), we observe that the lower bound on the capacity of the i.i.d. deletion channel and the one for the i.i.d. deletion-substitution channel differs by $(1-d)H_b(s)$. This gap between the lower bounds is intuitive, since due to the data processing inequality, the capacity of the deletion-substitution channel is a lower bound on the capacity of the deletion-only channel. Also, for a BSC the capacity is reduced compared to the noiseless case by $H_b(s)$ per received bits; for the deletion-substitution channel case, the number of received bits is $(1-d)n$, hence this difference is reasonable.

We need the following two propositions in the proof of Lemma 2. In Proposition 3, we obtain the exact value of the output entropy in the deletion-substitution

channel with i.u.d. input sequences, while Proposition 4 gives an upper bound on the conditional output entropy with i.u.d. bits transmitted through the deletion-substitution channel.

Proposition 3. *For an i.i.d. deletion-substitution channel with i.u.d. input sequences of length n, we have*

$$H(\boldsymbol{Y}') = n(1-d) + H(\boldsymbol{T}), \tag{3.23}$$

where $\boldsymbol{Y}'$ denotes the output sequence of the deletion-substitution channel and $H(\boldsymbol{T})$ is as defined in Eqn. (3.1).

Proof. By using the facts that all the elements of the set $\mathcal{Y}^d_{-j}$ are identically distributed, which are inputs into the BSC channel, and a fixed length i.u.d. input sequence into a BSC result in i.u.d. output sequences, all elements of the set $\mathcal{Y}'^d_{-j}$ are also identically distributed. Hence,

$$P(\boldsymbol{y}'(n-j)) = \frac{1}{2^{n-j}} \binom{n}{j} d^j (1-d)^{n-j}, \tag{3.24}$$

and clearly, the output entropy of the deletion-substitution channel is equal to the output entropy of the deletion channel $H(\boldsymbol{Y}') = H(\boldsymbol{Y})$, where $H(\boldsymbol{Y})$ is given in Eqn. (3.12), completing the proof of the proposition. □

Proposition 4. *For a deletion-substitution channel with i.u.d. input sequences, the entropy of the output sequence $\boldsymbol{Y}'$ conditioned on the input $\boldsymbol{X}$ of length n bits, is upper bounded by*

$$H(\boldsymbol{Y}'|\boldsymbol{X}) \leq nH_b(d) - \sum_{j=1}^{n} W_j(n) \binom{n}{j} d^j (1-d)^{n-j} + n(1-d)H_b(s), \tag{3.25}$$

where $W_j(n)$ is given in Eqn. (3.5). □

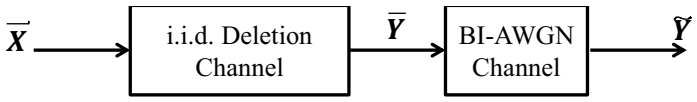

Figure 3.2: Deletion-AWGN channel.

The main idea in the proof of Proposition 4 is same as the proof of Proposition 2 whose details are provided in Appendix B. We can now complete the proof of the main lemma of the section.

Proof of Lemma 2: By substituting the exact value of the output entropy, Eqn. (3.23), and an upper bound on the conditional output entropy, Eqn. (3.25), into Eqn. (3.1), Lemma 2 is proved. □

3.3.3 Deletion-AWGN Channel

In this section, a binary deletion channel in the presence of AWGN is considered, where the bits are transmitted using binary phase shift keying (BPSK) and the received signal contains AWGN in addition to the deletion errors. As illustrated in Fig. 3.2, this channel can be considered as a coscade of two independent channels where the first channel is an i.i.d. deletion channel and the second one is a binary input AWGN (BI-AWGN) channel. We use $\bar{\boldsymbol{X}}$ to denote the input sequence to the first channel which is a BPSK modulated version of the binary input sequence $\boldsymbol{X}$, i.e., $\bar{x}_i = 1 - 2x_i$ and $\bar{\boldsymbol{Y}}$ to denote the output sequence of the first channel input to the second one. $\widetilde{\boldsymbol{Y}}$ is the output sequence of the second channel that is the noisy version of $\bar{\boldsymbol{Y}}$, i.e.,

$$\widetilde{y}_i = \bar{y}_i + z_i, \tag{3.26}$$

where z_i's are i.i.d. Gaussian random variables with zero mean and a variance of σ^2, and $\widetilde{y}_i^d$ and $\bar{y}_i$ are the i^{th} received and transmitted bits of the second channel, respectively. Therefore, for the probability density function of the i^{th} channel output,

we have

$$\begin{aligned} f_{\tilde{y}_i}(\eta) &= f_{\tilde{y}_i}(\eta|\bar{y}_i=1)P(\bar{y}_i=1) + f_{\tilde{y}_i}(\eta|\bar{y}_i=-1)P(\bar{y}_i=-1) \\ &= \frac{1}{\sqrt{2\pi}\sigma}\left[P(\bar{y}_i=1)e^{-\frac{(\eta-1)^2}{2\sigma^2}} + P(\bar{y}_i=-1)e^{-\frac{(\eta+1)^2}{2\sigma^2}}\right]. \end{aligned} \tag{3.27}$$

In the following lemma, an achievable rate is provided over this channel.

Lemma 3. *For any $n > 0$, the capacity of the deletion-AWGN channel with a deletion probability of d and a noise variance of σ^2 is lower bounded by*

$$C_{d,AWGN} \geq 1 - d - H_b(d) + \frac{1}{n}\sum_{j=1}^{n} W_j(n)\binom{n}{j}d^j(1-d)^{n-j} - (1-d)E\left[\log(1+e^{\frac{-2\boldsymbol{z}}{\sigma^2}})\right], \tag{3.28}$$

where $W_j(n)$ is as given in Eqn. (3.5)*, $E[.]$ is statistical expectation and $\boldsymbol{z}\sim\mathcal{N}(0,\sigma^2)$.*□

Before giving the proof of the above lemma, we provide several comments about the result. First, the desired lower bound in Eqn. (3.28) is the only analytical lower bound on the capacity of the deletion-AWGN channel. In the current literature, there are only simulation based lower bounds, e.g., [6] which employs Monte-Carlo simulation techniques. Furthermore, the procedure employed in [6] is only useful for deriving lower bounds for small values of deletion probability, e.g., $d \leq 0.1$, while the lower bound in Eqn. (3.28) is useful for a much wider range.

For $d = 0$, the lower bound in Eqn. (3.28) is equal to $1 - E\left[\log(1+e^{\frac{-2\boldsymbol{z}}{\sigma^2}})\right]$ which is the capacity of the BI-AWGN channel. Similar to the deletion-substitution channel case, the bound on the capacity of the deletion-AWGN channel differs from the one on the capacity of the deletion channel by some amount $(1-d)E\left[\log(1+e^{\frac{-2\boldsymbol{z}}{\sigma^2}})\right]$. Also, by substituting $\sigma^2 = 0$, we obtain the lower for the deletion-only channel. Finally, we note that the term in Eqn. (3.28) which contains $E\left[\log(1+e^{\frac{-2\boldsymbol{z}}{\sigma^2}})\right]$ can be easily computed by numerical integration with an arbitrarily accuracy (it involves only an one-dimensional integral).

We need the following two propositions in the proof of Lemma 3. In the following proposition, the exact value of the differential output entropy in the deletion-AWGN channel with i.u.d. input bits is calculated.

Proposition 5. *For an i.i.d. deletion-AWGN channel with i.u.d. input sequences of length n, we have*

$$h(\widetilde{\boldsymbol{Y}}) = n(1-d)\left(\log(2\sigma\sqrt{2\pi e}) - E\left[\log(1+e^{-\frac{2\boldsymbol{z}}{\sigma^2}})\right]\right) + H(\boldsymbol{T}), \tag{3.29}$$

where $h(.)$ denotes the differential entropy function, $\widetilde{\boldsymbol{Y}}$ denotes the output of the deletion-AWGN channel, $\boldsymbol{z} \sim \mathcal{N}(0,\sigma^2)$, and $H(\boldsymbol{T})$ is as defined in Eqn. (3.1).

Proof. For the differential entropy of the output sequence, we can write

$$\begin{aligned} h(\widetilde{\boldsymbol{Y}}) &= h(\widetilde{\boldsymbol{Y}}) + H(\boldsymbol{T}|\widetilde{\boldsymbol{Y}}) \\ &= h(\widetilde{\boldsymbol{Y}}, \boldsymbol{T}) \\ &= h(\widetilde{\boldsymbol{Y}}|\boldsymbol{T}) + H(\boldsymbol{T}), \end{aligned} \tag{3.30}$$

where the first equality results by using the fact that by knowing the received sequence, the number of deletions is known and $\boldsymbol{T}$ is determined, i.e., $H(\boldsymbol{T}|\widetilde{\boldsymbol{Y}}) = 0$, and the last equality is obtained by using a different expansion of $h(\widetilde{\boldsymbol{Y}}, \boldsymbol{T})$. On the other hand, we can write

$$\begin{aligned} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{T}) &= \sum_{j=0}^{n} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{T}=j)P(\boldsymbol{T}=j) \\ &= \sum_{j=0}^{n} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{T}=j)\binom{n}{j}d^j(1-d)^{n-j}. \end{aligned} \tag{3.31}$$

We know that all the elements of the set $\bar{\mathcal{Y}}^d_{-j}$ are i.i.d., the same as $\mathcal{Y}^d_{-j}$, and as given in Eqn. (3.13), $P(\bar{\boldsymbol{y}}(n-j)) = P(\bar{\boldsymbol{y}}, \boldsymbol{T}=j) = \frac{1}{2^{n-j}}\binom{n}{j}d^j(1-d)^{n-j}$. So we can write

$$\begin{aligned} P(\bar{\boldsymbol{y}}|\boldsymbol{T}=j) &= \frac{P(\bar{\boldsymbol{y}}, \boldsymbol{T}=j)}{P(\boldsymbol{T}=j)} \\ &= \frac{1}{2^{n-j}}, \end{aligned} \tag{3.32}$$

and as a result $P(\bar{y}_i = 1|\boldsymbol{T} = j) = P(\bar{y}_i = -1|\boldsymbol{T} = j) = \frac{1}{2}$ (for $1 \leq i \leq n - j$). By employing this result in Eqn. (3.27), we have

$$f_{\widetilde{y}_i}(\eta) = \frac{1}{2\sqrt{2\pi}\sigma}\left[e^{-\frac{(\eta-1)^2}{2\sigma^2}} + e^{-\frac{(\eta+1)^2}{2\sigma^2}}\right]. \tag{3.33}$$

where $f_{\widetilde{y}_i}(\eta)$ denotes the probability density function (PDF) of the continuous random variable $\widetilde{y}_i$. Noting also that the deletions happen independently, $\widetilde{y}_i$'s are i.i.d. and we can write

$$\begin{aligned} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{T} = j) &= (n-j)h(\widetilde{y}_i) \\ &= (n-j)\int_{-\infty}^{\infty} -f_{\widetilde{y}_i}(\eta)\log\left(f_{\widetilde{y}_i}(\eta)\right)d\eta \\ &= (n-j)\left(\log(2\sigma\sqrt{2\pi e}) - E\left[\log(1+e^{-\frac{2\boldsymbol{z}}{\sigma^2}})\right]\right), \end{aligned} \tag{3.34}$$

where $\boldsymbol{z} \sim \mathcal{N}(0, \sigma^2)$. By substituting Eqn. (3.34) into Eqn. (3.31), we obtain

$$\begin{aligned} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{T}) &= \sum_{j=0}^{n}(n-j)\binom{n}{j}d^j(1-d)^{n-j}\left(\log(2\sigma\sqrt{2\pi e}) - E\left[\log(1+e^{-\frac{2\boldsymbol{z}}{\sigma^2}})\right]\right) \\ &= n(1-d)\left(\log(2\sigma\sqrt{2\pi e}) - E\left[\log(1+e^{-\frac{2\boldsymbol{z}}{\sigma^2}})\right]\right), \end{aligned} \tag{3.35}$$

and by using Eqns. (3.35) and (3.30), Eqn. (3.29) is obtained. □

In the following proposition, we derive an upper bound on the differential entropy of the output conditioned on the input for deletion-AWGN channel.

Proposition 6. *For a deletion-AWGN channel with i.u.d. input bits, the differential entropy of the output sequence $\widetilde{\boldsymbol{Y}}$ conditioned on the input $\boldsymbol{X}$ of length n, is upper bounded by*

$$h(\widetilde{\boldsymbol{Y}}|\boldsymbol{X}) \leq nH_b(d) - \sum_{j=1}^{n} W_j(n)\binom{n}{j}d^j(1-d)^{n-j} + n(1-d)\log(2\sigma\sqrt{2\pi e}), \tag{3.36}$$

where $W_j(n)$ is given in Eqn. (3.5).

Proof. For the conditional differential entropy of the output sequence given the length n input $\boldsymbol{X}$, we can write

$$\begin{aligned} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{X}) &= h(\widetilde{\boldsymbol{Y}}|\boldsymbol{X}) + H(\boldsymbol{T}|\widetilde{\boldsymbol{Y}},\boldsymbol{X}) \\ &= H(\boldsymbol{T}) + h(\widetilde{\boldsymbol{Y}}|\boldsymbol{T},\boldsymbol{X}), \end{aligned} \tag{3.37}$$

where in the first equality we used the fact that by knowing $\boldsymbol{X}$ and $\widetilde{\boldsymbol{Y}}$, the number of deletions is known, i.e., $H(\boldsymbol{T}|\widetilde{\boldsymbol{Y}},\boldsymbol{X}) = 0$. The second equality is obtained by using a different expansion of $h(\widetilde{\boldsymbol{Y}},\boldsymbol{T}|\boldsymbol{X})$ and also using the fact that the deletion process is independent of the input $\boldsymbol{X}$, i.e., $H(\boldsymbol{T}|\boldsymbol{X}) = H(\boldsymbol{T})$. On the other hand, we also have

$$\begin{aligned} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{T},\boldsymbol{X}) &= \sum_{j=0}^{n} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{X},\boldsymbol{T}=j)P(\boldsymbol{T}=j) \\ &= \sum_{j=0}^{n} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{X},\boldsymbol{T}=j)\binom{n}{j}d^j(1-d)^{n-j}. \end{aligned} \tag{3.38}$$

To obtain $h(\widetilde{\boldsymbol{Y}}|\boldsymbol{X},\boldsymbol{T}=j)$, we need to compute $f_{\widetilde{\boldsymbol{y}}|\boldsymbol{x},j}(\eta)$ for any given input sequence $\boldsymbol{x} = (b; n_1, n_2, ..., n_K)$ and different values of j. As in the proofs of Propositions 2 and 4, if we consider the outputs of the deletion channel resulting from different deletion patterns of length j from a given $\boldsymbol{x}$, as if they are distinct and also use the result in Eqn. (3.17), an upper bound on the differential output entropy conditioned on the input sequence $\boldsymbol{X}$ results. We relegate the details of this computation and completion of the proof of the proposition to Appendix C. □

We can now state the proof of the main lemma of the section.

Proof of Lemma 3: By substituting the exact value of the differential output entropy in Eqn. (3.29), and the upper bound on the differential output entropy conditioned on the input in Eqn. (3.36), in Eqn. (3.4) a lower bound on the mutual information rate of a deletion-AWGN channel is obtained, hence the lemma is proved. □

3.4 Lower Bounds on the Capacity of Insertion Channels

We now turn our attention to the insertion channels and derive lower bounds on the capacity of the sticky and random insertion channels by employing the approach proposed in Section 3.2. We start with the sticky channel model for which the derivations have more similarity with those in Section 3.3 (for the i.i.d. deletion channel). We then consider the case of random insertions in Section 3.4.2.

3.4.1 Sticky Channel

Another useful model for channels suffering from synchronization errors is sticky channel in which only duplication errors are allowed and every bit is independently duplicated with a probability i. The number of duplications can also be modeled as a random variable by following a particular distribution on the set of positive integers. In [19], two special cases for the number of duplications are studied in detail, namely elementary i.i.d. duplication channel where the number of duplications is exactly one, and geometric i.i.d. duplication channel where the number of duplications follows a geometric distribution. In the following, by a sticky channel we refer to an elementary i.i.d. duplication channel. The following is a lower bound on the capacity of such channels.

Lemma 4. *For any $n > 0$, the capacity of the sticky channel C_s with a duplication probability of i is lower bounded by*

$$\begin{aligned} C_s \;\geq\; & (1-i)^n - H_b(i) + i(1-i)^{n-1}\Big(\log(n) - S_1(n) + n\Big) \\ & + \frac{n-1}{2} i^2 (1-i)^{n-2}\left(n+1+\log\binom{n}{2} - S_2(n)\right) \\ & + \frac{1}{n}\sum_{j=1}^{n} W_j(n)\binom{n}{j} i^j (1-i)^{n-j}, \end{aligned} \tag{3.39}$$

where

$$S_1(n) = \frac{1}{2^{n-1}} \sum_{k=1}^{n} \binom{n-1}{k-1} \log(k),$$

$$S_2(n) = \sum_{l=0}^{n-2} \sum_{K=l+1}^{\lfloor \frac{n+l}{2} \rfloor + 1} \frac{\binom{n+1-K}{K-l-1}\binom{K}{l}}{2^{n-1} n(n-1)} A(n,K,l) \log\left(A(n,K,l)\right),$$

$A(n,K,l) = (n-K)^2 + n + K - 2 - 2l$ *and* $W_j(n)$ *is as given in Eqn.* (3.5). □

To best of our knowledge, no simple expressions exists in the literature bounding the capacity of the sticky channel. There are several analytical lower bounds with time consuming numerical calculations [5, 12], and lower bounds employing the BAA [19] in the literature.

The result in (3.39), can be further lower bounded by a simpler expression that only contains some powers of i, that is,

$$\begin{aligned} C_s \geq & 1 - H_b(i) + i\Big(\log(n) - S_1(n) + W_1(n) \Big) \\ & + i^2 \frac{n-1}{2} \Big(2S_1(n) + \log\frac{n-1}{n} - S_2(n) - 2W_1(n) + W_2(n) \Big) \\ & + i^3 \binom{n-1}{2} \Big(W_1(n) + S_2(n) - S_1(n) - W_2(n) - \log(n-1) - \frac{n}{3} \Big) \\ & - i^4 \binom{n-1}{3} \Big(\log(n) + n - S_1(n) + W_1(n) \Big), \end{aligned} \tag{3.40}$$

where the inequalities $(1-p)^m \geq [1 - mp + \binom{m}{2}p^2 - \binom{m}{3}p^3]$ and $(1-p)^m \geq 1 - mp$ are employed. Finally we note that the lower bound in Eqn. (3.39) is only useful for small values of duplication probability.

We utilize the same approach as in the case of a deletion channel by considering i.u.d. input sequences. However, contrary to the deletion channel proofs, we can obtain the conditional output entropy precisely but we cannot obtain the exact value of the output entropy, hence we only lower bound it. Before giving the proof of

Lemma 4, we present two propositions needed in the proof. In the following proposition, the output entropy of the sticky channel with i.u.d. input sequences is lower bounded.

Proposition 7. *For a sticky channel with i.u.d. input sequences of length n, we have*

$$\begin{aligned} H(\boldsymbol{Y}) \;\geq\; & n(1-i)^n + ni(1-i)^{n-1}\left(\log(n) - S_1(n) + n\right) \\ & + \binom{n}{2} i^2 (1-i)^{n-2}\left(n+1+\log\binom{n}{2} - S_2(n)\right) + H(\boldsymbol{T}), \end{aligned} \tag{3.41}$$

where $\boldsymbol{Y}$ denotes the output sequence, $S_1(n)$ and $S_2(n)$ are as given in Eqn. (3.39)*, and $H(\boldsymbol{T})$ is as given in Eqn.* (3.1)*.*

Proof. To compute $H(\boldsymbol{Y})$, we need to determine the probabilities of all possible output sequences, which is clearly infeasible for large values of n. For the output entropy, we have

$$H(\boldsymbol{Y}) = \sum_{m=n}^{2n} \sum_{K=1}^{n} \sum_{\boldsymbol{y}(0;m;K)} 2P\left(\boldsymbol{y}(0;m;K)\right) \log\left(\frac{1}{P(\boldsymbol{y}(0;m;K))}\right), \tag{3.42}$$

hence, we first obtain the exact probability distribution of the output sequences resulting from at most two duplications, see Appendix D for details, and calculate

$$\sum_{m=n}^{n+2} \sum_{K=1}^{n} \sum_{\boldsymbol{y}(0;m;K)} 2P\left(\boldsymbol{y}(0;m;K)\right) \log\left(\frac{1}{P(\boldsymbol{y}(0;m;K))}\right),$$

precisely; then we derive a lower bound by considering the contribution to the output entropy of sequences with more than two duplications. We can write

$$\begin{aligned} H(\boldsymbol{Y}) \;\geq\; & \sum_{m=n}^{n+2} \sum_{K=1}^{n} \sum_{\boldsymbol{y}(0;m;K)} 2P\left(\boldsymbol{y}(0;m;K)\right) \log\left(\frac{1}{P(\boldsymbol{y}(0;m;K))}\right) \\ & + \sum_{m=n+3}^{2n} P(|\boldsymbol{y}|=m) \log\left(\frac{1}{P(|\boldsymbol{y}|=m)}\right) \\ =\; & \sum_{m=n}^{n+2} \sum_{K=1}^{n} \sum_{\boldsymbol{y}(0;m;K)} 2P\left(\boldsymbol{y}(0;m;K)\right) \log\left(\frac{1}{P(\boldsymbol{y}(0;m;K))}\right) \\ & + \sum_{j=3}^{n} \binom{n}{j} i^j (1-i)^{n-j} \log\left(\frac{1}{\binom{n}{j} i^j (1-i)^{n-j}}\right), \end{aligned} \tag{3.43}$$

where in deriving the inequality, we used the concavity of the function $-\sum_t p_t \log(p_t)$ ($p_t > 0$ and $\sum_t p_t \leq 1$) and the last equality is obtained noting that with probability $\binom{n}{j} i^j (1-i)^{n-j}$, the length of the output sequence is $n+j$, i.e., $P(|\boldsymbol{y}| = n+j) = \binom{n}{j} i^j (1-i)^{n-j}$. We can then rewrite (3.43) as

$$\begin{aligned}
H(\boldsymbol{Y}) \;\geq\; & 2^n \left(P\left(\boldsymbol{y}(n)\right) \log\left(\frac{1}{P\left(\boldsymbol{y}(n)\right)} \right) \right) \\
& + \sum_{K=1}^{n} \left[2 \binom{n}{k-1} P(\boldsymbol{y}(0; n+1; K)) \log\left(\frac{1}{P(\boldsymbol{y}(0; n+1; K))} \right) \right] \\
& + \sum_{l=0}^{n-2} \sum_{K=l+1}^{\lfloor \frac{n+l}{2} \rfloor + 1} \left[2 \binom{n+1-K}{K-l-1} \binom{K}{l} P(\boldsymbol{y}(0; n+2; K; l)) \log\left(\frac{1}{P(\boldsymbol{y}(0; n+2; K; l))} \right) \right] \\
& + n(1-i)^n \log(1-i) + ni(1-i)^{n-1} \log\left(ni(1-i)^{n-1} \right) \\
& + \binom{n}{2} i^2 (1-i)^{n-2} \log\left(\binom{n}{2} i^2 (1-i)^{n-2} \right) + H(\boldsymbol{T}),
\end{aligned} \tag{3.44}$$

where we used the facts that there are 2^n output sequences of length n ($|\mathcal{Y}_0| = 2^n$); $\binom{n}{K-1}$ output sequences beginning by 0, with length $n+1$ and K runs ($|\boldsymbol{y}(0; n+1; K)| = \binom{n}{K-1}$); $\binom{n+1-K}{K-l-1}\binom{K}{l}$ output sequences beginning by 0, with length $n+2$, K runs and l runs with length 1 $\left(|\boldsymbol{y}(0; n+2; K; l)| = \binom{n+1-K}{K-l-1}\binom{K}{l}\right)$; and $P(\boldsymbol{y}(0; n+j; K)) = P(\boldsymbol{y}(1; n+j; K))$. For the output sequences resulting from at most two duplications, we have (for more details see Appendix D)

$$P(\boldsymbol{y}) = \begin{cases} \frac{(1-i)^n}{2^n} & |\boldsymbol{y}| = n \\ \frac{(n+1-K) i (1-i)^{n-1}}{2^n} & \boldsymbol{y} = \boldsymbol{y}(b; n+1; K) \\ \frac{i^2 (1-i)^{n-2}}{2^{n+1}} \left((n-K)^2 + n + K - 2 - 2l \right) & \boldsymbol{y} = \boldsymbol{y}(b; n+2; K; l) \end{cases}, \tag{3.45}$$

so by substituting Eqn. (3.45) into Eqn. (3.44), Eqn. (3.41) is obtained. □

The conditional output entropy of the sticky channel with i.u.d. inputs is calculated in the following proposition.

Proposition 8. *For a sticky channel with the output sequence denoted by $\boldsymbol{Y}$, using i.u.d. input sequences and for any $n > 0$, we have*

$$H(\boldsymbol{Y}|\boldsymbol{X}) = nH_b(i) - \sum_{j=1}^{n} W_j(n) \binom{n}{j} i^j (1-i)^{n-j}, \tag{3.46}$$

where $W_j(n)$ is given in Eqn. (3.5)*.*

Proof. To obtain $H(\boldsymbol{Y}|\boldsymbol{X})$, we first need to obtain $H(\boldsymbol{Y}|\boldsymbol{x})$ for different (fixed) input sequences $\boldsymbol{x}$. Contrary to the deletion channel, in a sticky channel and for a given $\boldsymbol{x}$, distinct $I(n; K; j)$'s, which determine the number of duplications in different runs, result in different output sequences, i.e., $I \neq I'$ guarantees that $I * \boldsymbol{x} \neq I' * \boldsymbol{x}$. On the other hand, for $\sum_{k=1}^{K} j_k = j$, $0 \leq j_k \leq \min(j, n_k)$ and $0 \leq j \leq n$, we have

$$\begin{aligned} P\Big(\boldsymbol{y} = (b; n_1 + j_1, ..., n_K + j_K)\Big|\boldsymbol{x}(b; n, K)\Big) &= P\left(I(j_1, j_2, ..., j_K)\right) \\ &= \binom{n_1}{j_1} ... \binom{n_K}{j_K} i^j (1-i)^{n-j}. \end{aligned} \tag{3.47}$$

Clearly, $P\Big(\boldsymbol{y} = (b; n_1 + j_1, ..., n_K + j_K)\Big|\boldsymbol{x}(b; n, K)\Big)$ in Eqn. (3.47) is similar to $P\left(D(j_1, j_2, ..., j_K)\right)$ in Eqn. (3.16) which is used in deriving the upper bound on the conditional output entropy of the deletion channel. But Eqn. (3.47) is the exact output probability conditioned on a given input for the sticky channel which gives the exact conditional output entropy. The rest of the proof follows identical steps used in deriving the conditional output entropy of the deletion channel, provided in the proof of Proposition 2. □

Proof of Lemma 4: By utilizing the results of Propositions 7 and 8 in Theorem 1, Lemma 4 is proved. □

3.4.2 Random Insertion Channel

In this section, we consider the Gallager model [4] for insertion channels in which every transmitted bit is independently replaced by two random bits with probability

of i while neither the receiver nor the transmitter have information about the position of the insertions. The following lemma provides the main results of this section.

Lemma 5. *For any $n>0$, the random insertion channel capacity is bounded by*

$$\begin{aligned} C_i \;\geq\; & (1-i)^n - H_b(i) + \left(S_3(n) - \frac{3n+1}{4n} + n\right) i(1-i)^{n-1} + i^{n-1}(1-i)\log(n) \\ & + \frac{1}{n}\left(1-(1-i)^n - ni(1-i)^{n-1} - i^n - ni^{n-1}(1-i)\right)\log\binom{n}{2}, \end{aligned} \tag{3.48}$$

where $S_3(n) = \dfrac{1}{4n}\displaystyle\sum_{l=1}^{n-1} 2^{-l}\left[(n+1-l)(l+2)\log(l+2) + 2(l+1)\log(l+1)\right] + \dfrac{\log(n)}{2^{n+1}}$. □

To the best of our knowledge, the only analytical lower bound on the capacity of the random insertion channel is derived in [4] (Eqn. (2.2) for $d = s = 0$). Our result improves upon this result for small values of insertion probabilities as will be apparent with numerical examples.

Similar to the deletion and sticky channel cases, we can write a simpler lower bound as

$$\begin{aligned} C_i \geq\; & 1 - H_b(i) + \left(S_3(n) - \frac{3n+1}{4n}\right) i - \frac{n-1}{2}\left(2S_3(n) - \frac{3n+1}{2n} + n - \log\binom{n}{2}\right) i^2 \\ & - \binom{n-1}{2}\left(\log\binom{n}{2} - S_3(n) - \frac{2n}{3} + \frac{3n+1}{4n}\right) i^3 - \binom{n-1}{3}\left(S_3(n) + n - \frac{3n+1}{4n}\right) i^4. \end{aligned}$$

To prove the above lemma, we need the following two propositions. The output entropy of the random insertion channel with i.u.d. input sequences is calculated in the first one.

Proposition 9. *For a random insertion channel with i.u.d. input sequences of length n, we have*

$$H(\boldsymbol{Y}) = n(1+i) + H(\boldsymbol{T}). \tag{3.49}$$

where $\boldsymbol{Y}$ denotes the output sequence and $H(\boldsymbol{T})$ is as defined in Eqn. (3.1).

Proof. Similar to the proof of Proposition 1, we use the fact that

$$P(\boldsymbol{y}(n+j)) = \frac{1}{2^{n+j}}\binom{n}{j} i^j (1-i)^{n-j}. \tag{3.50}$$

Therefore, by employing Eqn. (3.50) in computing the output entropy, we obtain

$$H(\boldsymbol{Y}) = -\sum_{j=0}^{n} \binom{n}{j} i^j (1-i)^{n-j} \log\left(\frac{1}{2^{n+j}}\binom{n}{j} i^j (1-i)^{n-j}\right) = n(1+i) + H(\boldsymbol{T}),$$

which concludes the proof. □

In the following proposition, we present an upper bound on the conditional output entropy of the random insertion channel with i.u.d. input sequences for a given input of length n.

Proposition 10. *For a random insertion channel with input and output sequences denoted by $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively, with i.u.d. input sequences of length n, we have*

$$H(\boldsymbol{Y}|\boldsymbol{X}) \leq n(1+i) - n(1-i)^n + nH_b(i) - n\left(S_3(n) - \frac{3n+1}{4n} + n\right) i(1-i)^{n-1}$$
$$- \left(1-(1-i)^n - ni(1-i)^{n-1} - i^n - ni^{n-1}(1-i)\right) \log\binom{n}{2} - ni^{n-1}(1-i)\log(n). \tag{3.51}$$

Proof. For the conditional output sequence distribution for a given input sequence, we can write

$$p(\boldsymbol{y}|\boldsymbol{x}(b;n;K)) = \begin{cases} (1-i)^n & \boldsymbol{y} = \boldsymbol{x}(b;n;K) \\ \frac{n_1+1}{4} i(1-i)^{n-1} & \boldsymbol{y} = (b; n_1+1, ..., n_K) \\ \frac{n_K+1}{4} i(1-i)^{n-1} & \boldsymbol{y} = (b; n_1, ..., n_K+1) \\ \frac{n_k+2}{4} i(1-i)^{n-1} & \boldsymbol{y} = (b; n_1, ..., n_k+1, ..., n_K)(1<k<K) \\ \frac{1}{4} i(1-i)^{n-1} & \boldsymbol{y} = (b; n_1, ..., n'_{k,1}, 2, n'_{k,2}, ..., n_K) \\ \frac{2}{4} i(1-i)^{n-1} & \boldsymbol{y} = (b; n_1, ..., n''_{k,1}, 1, n''_{k,2}, ..., n_K) \\ \frac{1}{4} i(1-i)^{n-1} & \boldsymbol{y} = (\bar{b}; 1, n_1, ..., n_k, ..., n_K) \\ \frac{1}{4} i(1-i)^{n-1} & \boldsymbol{y} = (b; n_1, ..., n_k, ..., n_K, 1) \\ \epsilon_{y,x} & |\boldsymbol{y}| \geq n+2 \end{cases}, \tag{3.52}$$

where $n'_{k,1}+n'_{k,2}=n_k-1$ ($n'_{k,1}$, $n'_{k,2}\geq 0$), $n''_{k,1}+n''_{k,2}=n_k$ ($n''_{k,1}$, $n''_{k,2}\geq 1$), and $\epsilon_{y,x}$ represents $p(\boldsymbol{y}|\boldsymbol{x}(b;n;K))$ for given $\boldsymbol{y}$ with $|\boldsymbol{y}|\geq 2$. Hence, we obtain

$$\begin{aligned}H(\boldsymbol{Y}|\boldsymbol{x}(b;n;K^x)) &= -(1-i)^n\log(1-i)^n \\ &\quad -i(1-i)^{n-1}\Big(n\log(i(1-i)^{n-1})-1.5n-0.5K^x\Big) \\ &\quad -\frac{1}{4}i(1-i)^{n-1}\Bigg((n_1^x+1)\log(n_1^x+1)+(n_{K^x}^x+1)\log(n_{K^x}^x+1) \\ &\quad +\sum_{k=2}^{K^x-1}(n_k^x+2)\log(n_k^x+2)\Bigg)+H_\epsilon(\boldsymbol{x}),\end{aligned} \tag{3.53}$$

where $H_\epsilon(\boldsymbol{x})$ is the term related to the outputs resulting from more than one insertion. Therefore, by considering i.u.d. input sequences, we have

$$\begin{aligned}H(\boldsymbol{Y}|\boldsymbol{X}) =& -(1-i)^n\log(1-i)^n - ni(1-i)^{n-1}\left(\log(i(1-i)^{n-1})-\frac{7n+1}{4n}+S_3(n)\right) \\ &+H_\epsilon(\boldsymbol{X}),\end{aligned} \tag{3.54}$$

where $H_\epsilon(\boldsymbol{X})=\sum_{\boldsymbol{x}\in\mathcal{X}}\frac{H_\epsilon(\boldsymbol{x})}{2^n}$ and

$$\begin{aligned}S_3(n)=\frac{1}{2^{n+2}n}\sum_{x,K^x\neq 1}\Bigg[&(n_1^x+1)\log(n_1^x+1)+(n_{K^x}^x+1)\log(n_{K^x}^x+1) \\ &+\sum_{k=2}^{K^x-1}(n_k^x+2)\log(n_k^x+2)\Bigg]+\frac{\log(n)}{2^{n+1}},\end{aligned}$$

which can be written as

$$\begin{aligned}S_3(n) =&\frac{1}{2^{n+2}n}\Bigg[2\sum_{x,K^x\neq 1}\left[(n_1^x+1)\log(n_1^x+1)-(n_1^x+2)\log(n_1^x+2)\right] \\ &+\sum_{\boldsymbol{x}}\sum_{k=1}^{K^x}(n_k^x+2)\log(n_k^x+2)\Bigg]+\frac{\log(n)}{2^{n+1}} \\ =&\frac{1}{4n}\sum_{l=1}^{n-1}2^{-l}\left[(n+1-l)(l+2)\log(l+2)+2(l+1)\log(l+1)\right]+\frac{\log(n)}{2^{n+1}}.\end{aligned} \tag{3.55}$$

Here we have used the same approach used in the proof of Proposition 4, and considered the fact that there are 2^{n-l} sequences of length n with $n_1=l$ or $n_K=l$.

If we assume that all the possible outputs resulting from k insertions ($k \geq 2$) for a given $\boldsymbol{x}$ are equiprobable, since

$$-\sum_{j=1}^{J} p_j \log(p_j) \leq -\left(\sum_{j=1}^{J} p_j\right) \log\left(\frac{\sum_{j=1}^{J} p_j}{J}\right), \tag{3.56}$$

we can upper bound $H_\epsilon(\boldsymbol{x})$. That is,

$$\begin{aligned} H_\epsilon(\boldsymbol{x}) &= \sum_{k=2}^{n} \sum_{\boldsymbol{y}\in(\boldsymbol{x},k)} -Q(\boldsymbol{y}|\boldsymbol{x}) \log\Big(Q(\boldsymbol{y}|\boldsymbol{x})\Big) \leq \sum_{k=2}^{n} -\epsilon_k \log\left(\frac{\epsilon_k}{|(\boldsymbol{x},k)|}\right) \\ &\leq \sum_{k=2}^{n} -\epsilon_k \log\left(\frac{\epsilon_k}{2^{n+k}}\right), \end{aligned} \tag{3.57}$$

where $\epsilon_k = \sum_{\boldsymbol{y}\in(\boldsymbol{x},k)} Q(\boldsymbol{y}|\boldsymbol{x}) = \binom{n}{k} i^k (1-i)^{n-k}$ is the probability of k insertions for transmission of n bits, and the last inequality results by using the fact that $|\mathcal{Y}^i(\boldsymbol{x}+k)| \leq 2^{n+k}$, where $|\mathcal{Y}^i(\boldsymbol{x}+k)|$ denotes the number of output sequences resulting from k insertions into a given input sequence $\boldsymbol{x}$. After some algebra, we arrive at

$$\begin{aligned} H_\epsilon(\boldsymbol{X}) \leq & n(1+i) + nH_b(i) - n(1-i)^n - (n+1)ni(1-i)^{n-1} \\ & + (1-i)^n \log(1-i)^n + ni(1-i)^{n-1} \log i(1-i)^{n-1} \\ & - ni^{n-1}(1-i)\log(n) - (1 - i^n - (1-i)^n - ni(1-i)^{n-1} \\ & - ni^{n-1}(1-i)) \log\binom{n}{2}. \end{aligned} \tag{3.58}$$

Finally, by substituting Eqn. (3.58) into Eqn. (3.54), the upper bound (3.51) is obtained. □

Proof of Lemma 5: By substituting the exact value of the output entropy (Eqn. (3.49)) and the upper bound on the conditional output entropy (Eqn. (3.51)) of the random insertion channel with i.u.d. input sequences into Eqn. (3.4), a lower bound on the achievable information rate is obtained, hence the lemma is proved. □

3.5 Numerical Examples

We now present several examples of the lower bounds on the insertion/deletion channel capacity for different values of n and compare them with the existing ones in the literature.

3.5.1 I.I.D. Deletion Channel

Here, we numerically evaluate the lower bounds derived on the capacity of the i.i.d. deletion channel and compare them with existing results in the literature. The lower bounds on the capacity of the i.i.d. deletion channel in Eqns. (3.5) and (3.11) are functions of n, so for different values of n, different lower bounds result. For example, for $n = 10000$, $n = 1000$, $n = 100$ and $n = 10$, the Eqn. (3.11) evaluates to

$$C_d \geq 1 - H_b(d) + 0.2884d - 1.4355d^2 - 6.4384 \times 10^7 d^3 - 2.146 \times 10^{11} d^4, \quad (3.59)$$

$$C_d \geq 1 - H_b(d) + 0.2868d - 1.4334d^2 - 6.4003 \times 10^5 d^3 - 2.1318 \times 10^8 d^4, \quad (3.60)$$

$$C_d \geq 1 - H_b(d) + 0.2711d - 1.4121d^2 - 6027.7d^3 - 1.9937 \times 10^5 d^4, \quad (3.61)$$

and

$$C_d \geq 1 - H_b(d) + 0.1094d - 1.2027d^2 - 30.3160d^3 - 93.1881d^4, \quad (3.62)$$

respectively. We observe that by increasing n, these lower bounds effectively capture the first order term expression of the capacity expansion.

In Table 3.1 the effect of changing n is numerically evaluated where the lower bounds in Eqns. (3.5) and (3.11), for $n = 100$ and $n = 1000$, along with several existing results are presented which show that the new lower bounds on the capacity are in the range of the tightest existing lower bounds. On the other hand, they do not improve the best known results. In comparing the lower bounds in Eqns. (3.5) and (3.11), we observe that for small values of deletion probability ($nd < 0.1$), the

Table 3.1: Lower bounds on the capacity of the deletion channel[1].

d	1−LB from [4]	1−LB from [5]	1−LB (3.5) $n = 1000$	1−LB (3.60)	1−LB (3.5) $n = 100$	1−LB (3.61)
10^{-5}	1.81×10^{-4}	1.7763×10^{-4}	1.7765×10^{-4}	1.7770×10^{-4}	1.7781242×10^{-4}	1.7781243×10^{-4}
10^{-4}	1.4739×10^{-3}	1.4442×10^{-3}	1.4444×10^{-3}	1.5399×10^{-3}	1.44593×10^{-3}	1.44595×10^{-3}
10^{-3}	0.01141	0.011120	0.011122	0.2902	0.11139	0.11145

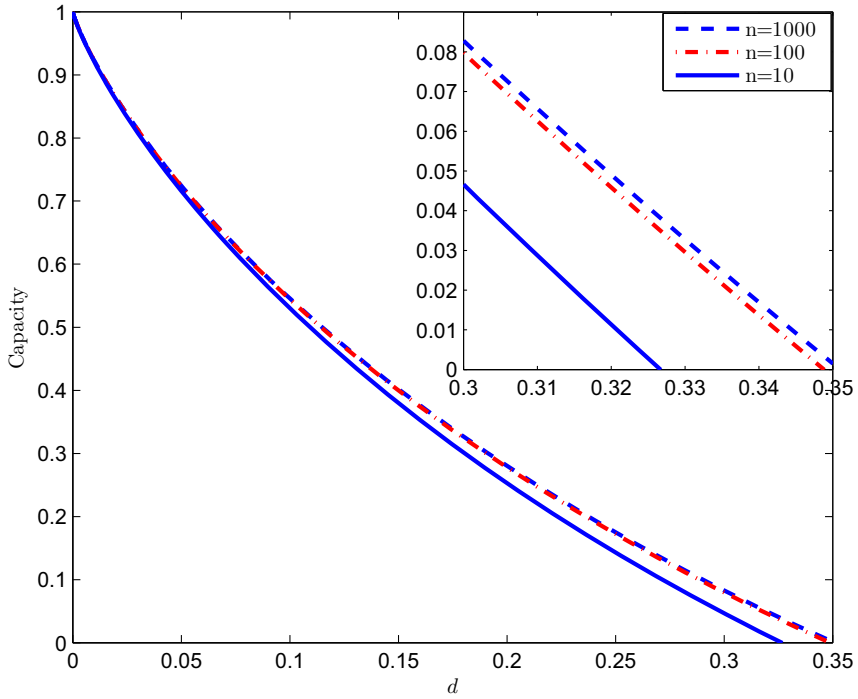

Figure 3.3: Lower bounds on the deletion channel capacity resulting from different values of block length n.

simplified lower bound is very close to the original lower bound while it is a simple expression lower bound.

In Fig. 3.3, we compare lower bounds on the capacity of the deletion channel resulting from different values of n. We observe that by increasing n from 10 to 100, a tighter lower bound is obtained but from 100 to 1000, the resulting improvement is not significant. In Fig. 3.4, the lower bound in Eqn. (3.5) for $n = 1000$ is compared with the lower bounds in [4] and [5]. We observe that the new lower bound in Eqn. (3.5) is

[1]Note that in Tables 3.1-3.4, we depict ″1−lower bound″ if the lower bound is close to 1.

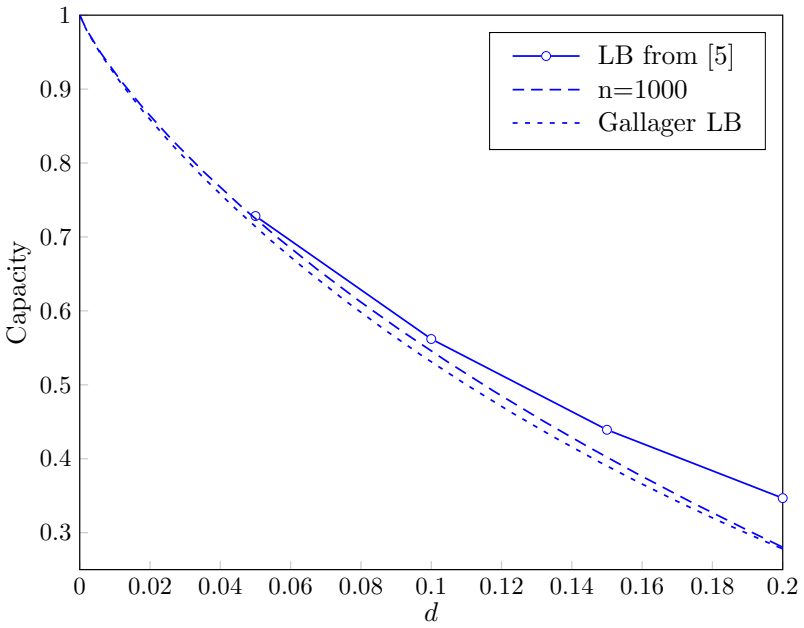

Figure 3.4: Comparison of the lower bound (3.5) for $n = 1000$ with lower bounds presented in [4] and [5].

tighter than one in [4] and for small values of d, as we observe in Table 3.1, it is very close to the lower bound of [5], which is the tightest lower bound for small values of d available.

3.5.2 Deletion-Substitution Channel

In Table 3.2, we compare the lower bound (3.22) for $n = 100$ and $n = 1000$ with the one in [4]. We observe that the bound improves the result of [4] for the entire range of d and s, and also as we expected, by increasing n from 100 to 1000, a tighter lower bound for all values of d and s is obtained.

3.5.3 Deletion-AWGN Channel

We now compare the derived analytical lower bound on the capacity of the deletion-AWGN channel with the simulation based bound of [6] which is the achievable information rate of the deletion-AWGN channel for i.u.d. input sequences obtained by Monte-Carlo simulations. As we observe in Fig. 3.5, the lower bound (3.28) is very

Table 3.2: Lower bounds on the capacity of the deletion-substitution channel.

d	s	1−LB (2.3)	1−LB (3.22) $n = 1000$	1−LB (3.22) $n = 100$
10^{-5}	10^{-5}	3.6104×10^{-4}	3.5817×10^{-4}	3.5834×10^{-4}
10^{-5}	10^{-4}	1.6535×10^{-3}	1.6506×10^{-3}	1.6508×10^{-3}
10^{-5}	10^{-3}	1.15881×10^{-2}	1.15853×10^{-2}	1.15854×10^{-2}
10^{-4}	10^{-5}	1.6535×10^{-3}	1.6248×10^{-3}	1.6264×10^{-3}
10^{-4}	10^{-4}	2.9459×10^{-3}	2.9172×10^{-3}	2.9188×10^{-3}
10^{-4}	10^{-3}	1.2879×10^{-2}	1.2850×10^{-2}	1.2852×10^{-2}
10^{-3}	10^{-5}	1.1588×10^{-2}	1.1302×10^{-2}	1.1319×10^{-2}
10^{-3}	10^{-4}	1.2879×10^{-2}	1.2593×10^{-2}	1.261×10^{-2}
10^{-3}	10^{-3}	2.2804×10^{-2}	2.2518×10^{-2}	2.2535×10^{-2}

d	s	LB (2.3)	LB (3.22) $n = 1000$	LB (3.22) $n = 100$
0.01	0.01	0.8392	0.8419	0.8418
0.01	0.03	0.7268	0.7373	0.7293
0.01	0.10	0.4549	0.4576	0.4575
0.05	0.01	0.6368	0.6476	0.6469
0.05	0.03	0.5289	0.5397	0.5390
0.05	0.10	0.2681	0.2789	0.2781
0.10	0.01	0.4583	0.4729	0.4716
0.10	0.03	0.3561	0.3707	0.3693
0.10	0.10	0.1089	0.1236	0.1222

close to the simulation results of [6] for small values of deletion probability but it does not improve them. This is not unexpected, because we further lower bounded the achievable information rate for i.u.d. input sequences while in [6], the achievable information rate for i.u.d. input sequences is obtained by Monte-Carlo simulations without any further lower bounding. The new bound is analytical and very easy to compute while the result in [6] requires lengthly simulations. Furthermore, the procedure employed in [6] is only useful for deriving lower bounds for small values of deletion probability, e.g., $d \leq 0.1$, while the lower bound (3.28) holds for a much wider range of deletion probabilities.

3.5.4 Sticky Channel

Similar to the deletion channel cases, the lower bounds on the capacity of the sticky channel in Eqns. (3.39) and (3.40) are functions of n, therefore, for different values of n, different lower bounds result. For example, for $n = 10$, $n = 100$ and $n = 1000$, the Eqn. 3.40 evaluates to

$$C_s \geq 1 - H_b(i) + 2.0329i - 1.7784i^2 - 178.9562i^3 - 1.0108 \times 10^3 i^4, \tag{3.63}$$

$$C_s \geq 1 - H_b(i) + 2.2639i - 2.1157i^2 - 1.7247 \times 10^5 i^3 - 1.6040 \times 10^7 i^4, \tag{3.64}$$

and

$$C_s \geq 1 - H_b(i) + 2.2861i - 2.1529i^2 - 8.3722 \times 10^7 i^3 - 8.3047e \times 10^{10} i^4, \tag{3.65}$$

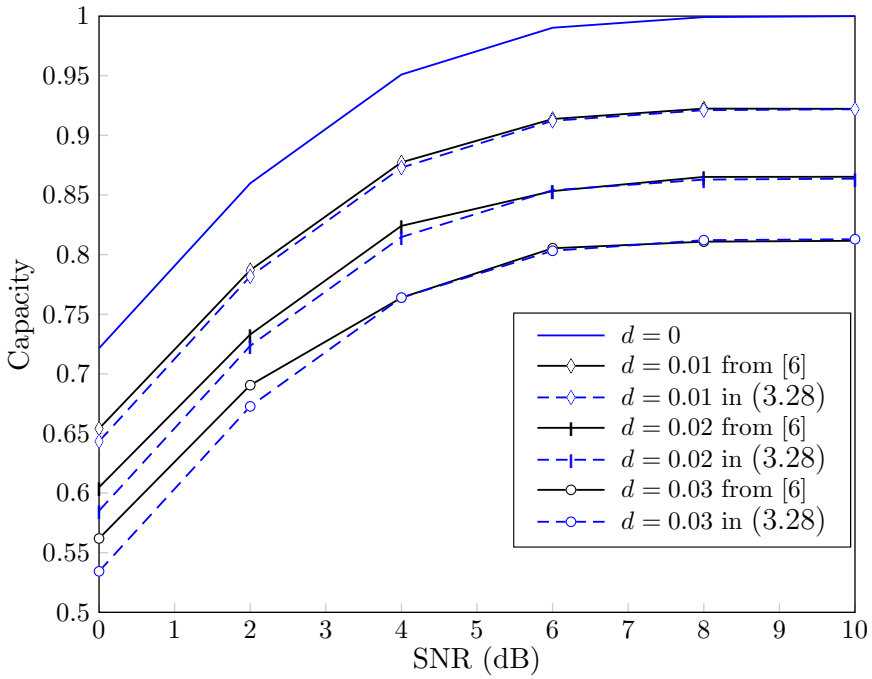

Figure 3.5: Comparison between the lower bound (3.28) for $n = 1000$ with the lower bound in [6] versus SNR for different deletion probabilities.

Table 3.3: Lower bounds on the capacity of the sticky channel.

i	1−LB from [19]	1−LB from [5]	1−LB (3.39) $n = 1000$	1−LB (3.65)	1−LB (3.64)	1−LB (3.63)
10^{-6}	2.9999×10^{-5}	1.9085×10^{-5}	1.9079×10^{-5}	1.9089×10^{-5}	1.9110×10^{-5}	1.94×10^{-5}
10^{-5}	1.6863×10^{-4}	1.5770×10^{-4}	1.5695×10^{-4}	1.5783×10^{-4}	1.5787×10^{-4}	1.6099×10^{-4}
10^{-4}	1.26×10^{-3}	1.2442×10^{-3}	1.3199×10^{-3}	1.43×10^{-3}	1.2499×10^{-3}	1.2699×10^{-3}
10^{-3}	9.09×10^{-3}	9.129×10^{-3}	86.311×10^{-3}	3.425×10^{-3}	9.342×10^{-3}	9.381×10^{-3}

respectively. In Table 3.3, we compare the lower bounds (3.39) and (3.40), with lower bounds presented in [5, 19] for small values of i. We observe that for small values of the duplication probability our lower bound improves the existing ones.

In Fig. 3.6, we compare different lower bounds for different values of n. We observe that for different ranges of duplication probability different values of n result in the tightest lower bound. This is because we derived a lower bound on the part of the output entropy resulting from outputs with more than two duplications which is tight for $np \ll 1$ and by increasing np becomes a loose lower bound.

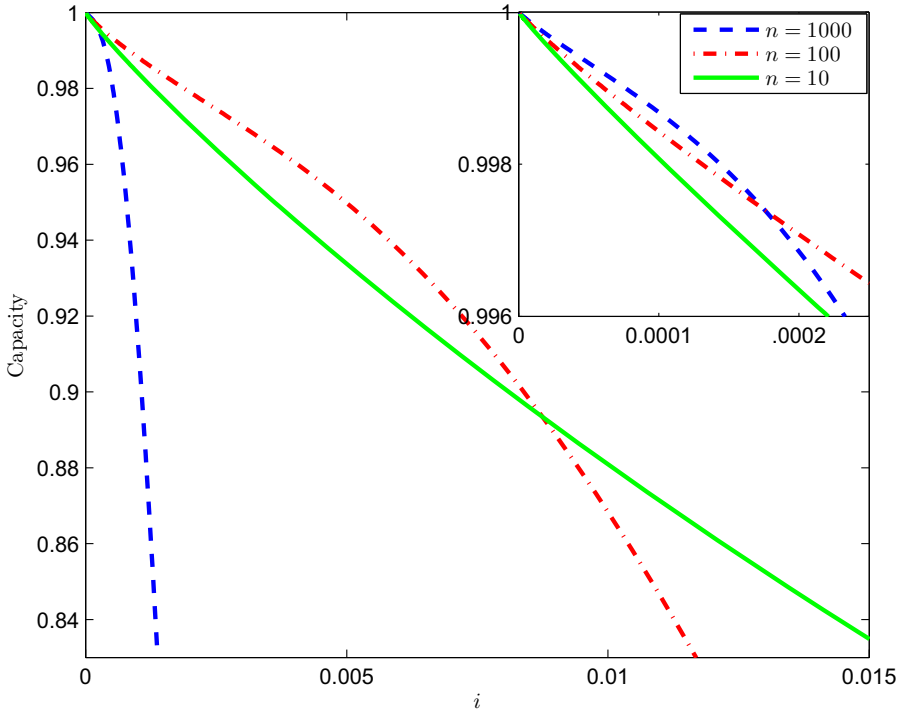

Figure 3.6: Comparison of the lower bounds on the capacity of the sticky channel resulting from different values of block length n.

3.5.5 Random Insertion Channel

We now numerically evaluate the lower bounds derived on the capacity of the random insertion channel. Similar to the previous cases, different values of n result in different lower bounds. In Table 3.4 and Fig. 3.7, we compare the lower bound in Eqn. (3.48) with the Gallager lower bound $(1 - H_b(i))$, where the reported values are obtained for the optimal value of n.

We observe that for larger i, smaller values of n give the tightest lower bounds. This is not unexpected since in upper bounding $H(\boldsymbol{Y}|\boldsymbol{X})$, we computed the exact value of $p(\boldsymbol{y}|\boldsymbol{x})$ for at most one insertion, i.e., $|\boldsymbol{y}| = |\boldsymbol{x}|$ or $|\boldsymbol{y}| = |\boldsymbol{x}| + 1$, and upper bounded the part of the conditional entropy resulting form more than one insertion. Therefore, for a fixed i by increasing n, the probability of having more than one

Table 3.4: Lower bounds on the capacity of the random insertion channel.

i	1−LB from [4]	1−LB (3.48)	optimal value of n
10^{-6}	2.14×10^{-5}	2.007×10^{-5}	121
10^{-5}	1.81×10^{-4}	1.68×10^{-4}	57
10^{-4}	1.47×10^{-3}	1.35×10^{-3}	27
10^{-3}	1.14×10^{-2}	1.02×10^{-2}	13
10^{-2}	8.07×10^{-1}	7.14×10^{-2}	7

i	LB from [4]	LB (3.48)	optimal value of n
0.03	0.8056	0.8276	5
0.05	0.7136	0.7442	5
0.10	0.5310	0.5702	4
0.15	0.3901	0.4230	4
0.20	0.2781	0.2962	3
0.23	0.2220	0.2283	3
0.25	0.1887	0.1853	3

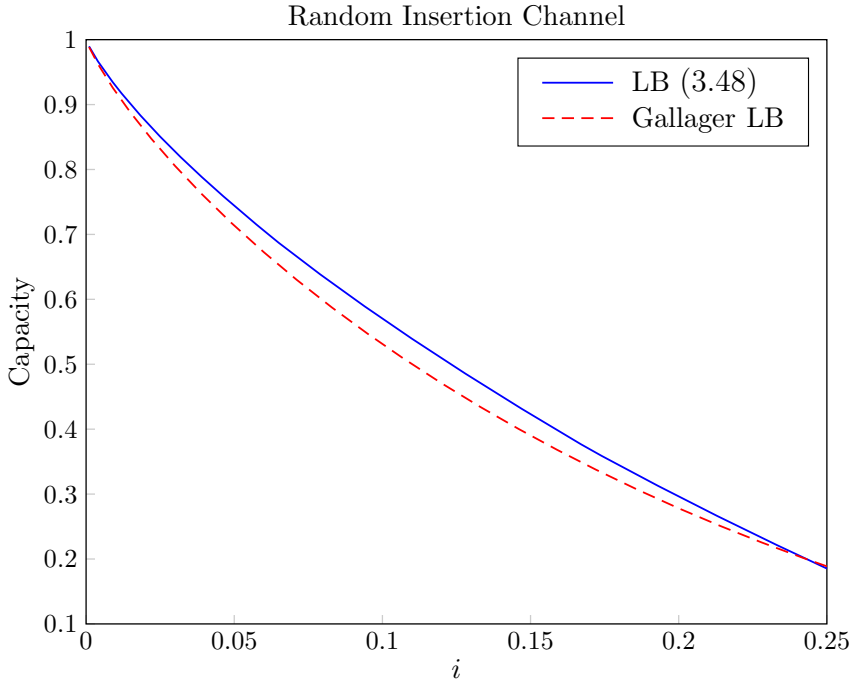

Figure 3.7: Comparison of the lower bound (3.48) with lower bound presented in [4].

insertion increases and as a result the upper bound becomes loose. We also observe that the lower bound (3.48) improves upon the Gallager's lower bound [4] for $i < 0.25$, e.g., for $i = 0.1$, we achieve an improvement of 0.0392 bits/channel use.

3.6 Chapter Summary

We have presented several analytical lower bounds on the capacity of the insertion/deletion channels by lower bounding the mutual information rate for i.u.d. input sequences. We have derived the first analytical lower bound on the capacity of the

deletion-AWGN channel which for small values of deletion probability is very close to the existing simulation based lower bounds. The lower bound presented on the capacity of the deletion-substitution channel improves the existing analytical lower bound for all values of deletion and substitution probabilities. For random insertion and sticky channels, the presented lower bounds improve the existing ones for small values of insertion probability. The lower bound on the capacity of the i.i.d. deletion channel, for small values of deletion probability, is very close to the tightest available lower bounds, and is in agreement with the first order expansion of the channel capacity for $d \to 0$, while our result is a strict lower bound for all range of d. Our approach for lower bounding the mutual information rate for small probabilities of synchronization errors is general an may be applicable to other channel models and other input distributions as well.

Chapter 4

Achievable Rates for Noisy Channels with Synchronization Errors

Several lower bounds on the capacity of binary input symmetric output channels with synchronization errors in addition to substitution and erasure errors and in the presence of white Gaussian noise (WGN) are derived in this chapter. More precisely, we show that if the channel with synchronization errors can be decomposed into a cascade of two channels where only the first one suffers from synchronization errors, a lower bound on the capacity of the original channel related to the capacity of the one with only synchronization errors can be derived. We derive lower bounds on the mutual information rate between the transmitted and received sequences for input sequence distribution which achieves the capacity of the channel with only synchronization errors. To derive the lower bounds, we do not need to know the exact capacity achieving input distributions, in fact, we only upper bound the performance degradation of the system due to the effect of the substitution and erasure errors or AWGN. The main advantage of the presented lower bounds is that we can employ any lower bound derived on the capacity of the channels with synchronization errors in lower bounding the capacity of the noisy channels with synchronization errors. The result on channel with synchronization errors in addition to erasure errors is the first result on lower bounding the capacity of such a channel. On the other hand, the derived lower bound on the capacity of channels with synchronization errors in the presence of WGN improves the existing lower bounds on the capacity of the AWGN channel with deletion errors for high SNR values.

The chapter is organized as follows. We start with an introduction in Section 4.1. In Section 4.2, we give two lemmas and one proposition which will be useful in the proof of the result on binary input symmetric q-ary output channels (BSQC) with synchronization errors. In Section 4.3, we derive lower bounds on the

capacity of the BSQC channels with synchronization errors where we first focus on sub/ers/synch channel (binary input symmetric ternary output channel) and binary input symmetric quaternary output channel, then give the results for arbitrary values of q. We directly lower bound the capacity of the AWGN/synch channel, again with respect to the capacity of the synchronization error channel, in Section 4.4. We give some numerical examples of the results and compare them with the existing results in Section 4.5. We summarize the result of the chapter in Section 4.6.

4.1 Introduction

In different communication systems, depending on the transmitting medium and the system design, different limiting factors degrade the performance of the system. Imperfect alignment between transmitter and receiver clocks in a digital communication system can be one of the limiting factors of the system performance which can be modeled by synchronization errors. On the other hand, additive noises are unavoidable in digital communication systems including systems with synchronization errors. The main object of this chapter is to include additive noises in analyzing the digital communication systems with synchronization errors to obtain more realistic results.

Here, we focus on finding achievable rates for the channels which can be considered as concatenation of two independent channels where the first one is a binary channel with only synchronization errors and the second one is either a memoryless binary input symmetric q-ary output channel (BSQC) or an AWGN channel. For instance, the first channel can be a binary insertion/deletion channel and the second one can be a binary symmetric channel (BSC) or a substitution/erasure channel (ternary output channel $q = 3$). We first consider the ternary ($q = 3$) and quaternary ($q = 4$) output cases, respectively, then generalize the results for arbitrary values of q. We obtain achievable rates of the concatenated channel with respect to the capacity of the synchronization error channel by lower bounding the information rate of the

concatenated channel for input distributions which achieve the capacity of the synchronization error channel. In fact, we derive the lower bounds without knowing the exact capacity achieving input distribution of the synchronization error channel. We only derive an upper bound on the system performance degeneration by considering the effect the second channel. Therefore, we can employ every lower bounds derived on the capacity of the synchronization error channel in lower bounding the capacity of the concatenated channel, where the results can be employed on any model on memoryless channels with synchronization errors.

Dobrushin [10] proves that Shannon's theorem holds in a memoryless channel with synchronization errors where he shows information stability holds for memoryless channels with synchronization errors such that we can write $\lim_{N\to\infty} \max_{P(\boldsymbol{X})} \frac{1}{N} I(\boldsymbol{X};\boldsymbol{Y})$, where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the transmitted and received sequences, respectively, and N is the length of the transmitted sequence. Therefore, the information and transmission capacities of the memoryless channels with synchronization errors are equal and we can employ any lower bound on the information capacity as a lower bound on the transmission capacity of a channel with synchronization errors.

4.1.1 Example of a Synchronization Error Channel Decomposition into Two Independent Channels

The procedure used in this chapter can be employed for any channel which can be decomposed into two independent channels such that the first one is a memoryless synchronization error channel and the second one is a symmetric memoryless channel with no effect on the length of the input sequence. Therefore, if we can also decompose a synchronization error channel into two channels with described properties, we can derive lower bounds on the capacity of the synchronization error channel by employing this approach. The advantage of this decomposition is in decomposing the original synchronization error channel into a well characterized synchronization error

Table 4.1: Transition probabilities of the hypothetical synchronization error channel.

	$P(Y_j\|X_j)$					
X_j	$Y_j=0$	$Y_j=1$	$Y_j=00$	$Y_j=01$	$Y_j=10$	$Y_j=11$
0	$\left(\frac{1}{2}-(\alpha+\beta)\right)\left(1+\sqrt{\frac{\alpha-\beta}{\alpha+\beta}}\right)$	$\left(\frac{1}{2}-(\alpha+\beta)\right)\left(1-\sqrt{\frac{\alpha-\beta}{\alpha+\beta}}\right)$	α	β	β	α
1	$\left(\frac{1}{2}-(\alpha+\beta)\right)\left(1-\sqrt{\frac{\alpha-\beta}{\alpha+\beta}}\right)$	$\left(\frac{1}{2}-(\alpha+\beta)\right)\left(1+\sqrt{\frac{\alpha-\beta}{\alpha+\beta}}\right)$	α	β	β	α

Table 4.2: Transition probabilities of two independent channels giving rise to the synchronization error channel given in Table 4.1.

	$P(Z_j\|X_j)$			
X_j	$Z_j=0$	$Z_j=1$	$Z_j=00$	$Z_j=11$
0	$1-2(\alpha+\beta)$	0	$\alpha+\beta$	$\alpha+\beta$
1	0	$1-2(\alpha+\beta)$	$\alpha+\beta$	$\alpha+\beta$

	$P(Y_j\|Z_j)$	
Z_j	$Y_j=0$	$Y_j=1$
0	$0.5+0.5\sqrt{\frac{\alpha-\beta}{\alpha+\beta}}$	$0.5-0.5\sqrt{\frac{\alpha-\beta}{\alpha+\beta}}$
1	$0.5-0.5\sqrt{\frac{\alpha-\beta}{\alpha+\beta}}$	$0.5+0.5\sqrt{\frac{\alpha-\beta}{\alpha+\beta}}$

channel and a memoryless channel such that lower bounding the capacity of the new synchronization error channel could be simpler than lower bounding the capacity of the original synchronization error channel. In the following, we provide an example of a hypothetical synchronization error channel that can be decomposed into another hypothetical synchronization channel and a memoryless BSC.

In Table 4.1, the transition probabilities of a hypothetical synchronization error channel are given. It can be shown that this channel can be decomposed into two independent channels given in Table 4.2 such that the first one is a synchronization error channel and the second one is a BSC channel.

4.2 Entropy Bounds for Binary Input q-ary Output Channels with Synchronization Errors

In the following two lemmas, we provide a lower bound on the output entropy and an upper bound on the conditional output entropy of the binary input q-ary output channel in terms of the the corresponding output entropies of the synchronization error channel, respectively. We then give a proposition that will be useful in the proof of the result on BSQC channels with synchronization errors (note that the following

two lemmas hold for any binary input q-ary output channels with synchronization errors regardless of any symmetry).

Lemma 6. *In any binary input q-ary output channel with synchronization errors and for all non-negative integer values of q, we have*

$$H(\boldsymbol{Y}^{(q)}) \geq H(\boldsymbol{Y}) - E_{\boldsymbol{M}}\left\{\log\left(\sum_{\boldsymbol{y}^{(q)}}\sum_{\boldsymbol{y},p(\boldsymbol{y})\neq 0} p(\boldsymbol{y}^{(q)}|\boldsymbol{y},\boldsymbol{M})p(\boldsymbol{y}^{(q)}|\boldsymbol{M})\right)\right\}, \quad (4.1)$$

where $\boldsymbol{M}$ is the random variable denoting the length of the received sequence, $\boldsymbol{Y}$ denotes the output sequence of the synchronization error channel and the input sequence of the binary input q-ary output channel, and $\boldsymbol{Y}^{(q)}$ denotes the output sequence of the binary input q-ary output channel.

Proof. By using two different expansions of $H(\boldsymbol{Y}^{(q)}, \boldsymbol{M})$, we have

$$\begin{aligned} H(\boldsymbol{Y}^{(q)}, \boldsymbol{M}) &= H(\boldsymbol{Y}^{(q)}) + H(\boldsymbol{M}|\boldsymbol{Y}^{(q)}) \\ &= H(\boldsymbol{Y}^{(q)}|\boldsymbol{M}) + H(\boldsymbol{M}). \end{aligned} \quad (4.2)$$

Hence, we can write

$$H(\boldsymbol{Y}^{(q)}) = H(\boldsymbol{Y}^{(q)}|\boldsymbol{M}) + H(\boldsymbol{M}), \quad (4.3)$$

where we used the fact that by knowing $\boldsymbol{Y}^{(q)}$, random variable $\boldsymbol{M}$ is also known, i.e., $H(\boldsymbol{M}|\boldsymbol{Y}^{(q)}) = 0$. By using the same approach for $H(\boldsymbol{Y})$, we have

$$H(\boldsymbol{Y}) = H(\boldsymbol{Y}|\boldsymbol{M}) + H(\boldsymbol{M}). \quad (4.4)$$

Finally, we can write

$$\begin{aligned} H(\boldsymbol{Y}^{(q)}) - H(\boldsymbol{Y}) &= H(\boldsymbol{Y}^{(q)}|\boldsymbol{M}) - H(\boldsymbol{Y}|\boldsymbol{M}) \\ &= \sum_{m} p(m)\left[H(\boldsymbol{Y}^{(q)}|\boldsymbol{M}=m) - H(\boldsymbol{Y}|\boldsymbol{M}=m)\right], \end{aligned} \quad (4.5)$$

where $p(m) = P(\boldsymbol{M} = m)$. On the other hand, due to the definition of the entropy, we can write

$$\begin{aligned} & H(\boldsymbol{Y}^{(q)}|\boldsymbol{M}=m) - H(\boldsymbol{Y}|\boldsymbol{M}=m) \\ & \quad = E_{\boldsymbol{Y}^{(q)}}\{-\log(p(\boldsymbol{Y}^{(q)}))|\boldsymbol{M}=m\} - E_{\boldsymbol{Y}}\{-\log(p(\boldsymbol{Y}))|\boldsymbol{M}=m\} \\ & \quad = E_{(\boldsymbol{Y},\boldsymbol{Y}^{(q)})}\left\{-\log\left(\frac{p(\boldsymbol{Y}^{(q)})}{p(\boldsymbol{Y})}\right)\middle|\boldsymbol{M}=m\right\} \\ & \quad = -\sum_{\boldsymbol{y}^{(q)}}\sum_{\boldsymbol{y},p(\boldsymbol{y})\neq 0} p(\boldsymbol{y}^{(q)}|\boldsymbol{y},\boldsymbol{M}=m)p(\boldsymbol{y}|\boldsymbol{M}=m)\log\left(\frac{p(\boldsymbol{y}^{(q)}|\boldsymbol{M}=m)}{p(\boldsymbol{y}|\boldsymbol{M}=m)}\right), \end{aligned}$$

where $E_{\boldsymbol{Z}}\{.\}$ denotes the expected value with respect to the random variable $\boldsymbol{Z}$. Now due to the fact that $-\log(x)$ is a convex function of x, we apply Jensen's inequality to write

$$\begin{aligned} & H(\boldsymbol{Y}^{(q)}|\boldsymbol{M}=m) - H(\boldsymbol{Y}|\boldsymbol{M}=m) \\ & \quad \geq -\log\left(\sum_{\boldsymbol{y}^{(q)}}\sum_{\boldsymbol{y},p(\boldsymbol{y})\neq 0} p(\boldsymbol{y}^{(q)}|\boldsymbol{y},\boldsymbol{M}=m)p(\boldsymbol{y}|\boldsymbol{M}=m)\frac{p(\boldsymbol{y}^{(q)}|\boldsymbol{M}=m)}{p(\boldsymbol{y}|\boldsymbol{M}=m)}\right) \\ & \quad = -\log\left(\sum_{\boldsymbol{y}^{(q)}}\sum_{\boldsymbol{y},p(\boldsymbol{y})\neq 0} p(\boldsymbol{y}^{(q)}|\boldsymbol{y},\boldsymbol{M}=m)p(\boldsymbol{y}^{(q)}|\boldsymbol{M}=m)\right). \end{aligned} \tag{4.6}$$

By substituting this result into (4.5), the proof follows. □

Lemma 7. *In any binary input q-ary output channel with synchronization errors and for any input distribution, we have*

$$H(\boldsymbol{Y}^{(q)}|\boldsymbol{X}) \leq H(\boldsymbol{Y}|\boldsymbol{X}) + E\{\boldsymbol{M}\}H(Y_j^{(q)}|Y_j), \tag{4.7}$$

where Y_j denotes the j-th output bit of the synchronization error channel and j-th input bit of the binary input q-ary output channel and $Y_j^{(q)}$ denotes the output symbol of the binary input q-ary output channel corresponding to the input bit Y_j.

Proof. For the conditional output entropy, we can write

$$\begin{aligned} H(\boldsymbol{Y}^{(q)}, \boldsymbol{Y}|\boldsymbol{X}) &= H(\boldsymbol{Y}^{(q)}|\boldsymbol{X}) + H(\boldsymbol{Y}|\boldsymbol{Y}^{(q)}, \boldsymbol{X}) \\ &= H(\boldsymbol{Y}|\boldsymbol{X}) + H(\boldsymbol{Y}^{(q)}|\boldsymbol{Y}, \boldsymbol{X}) \\ &= H(\boldsymbol{Y}|\boldsymbol{X}) + H(\boldsymbol{Y}^{(q)}|\boldsymbol{Y}), \end{aligned} \tag{4.8}$$

where the last equality follows since $\boldsymbol{X} \to \boldsymbol{Y} \to \boldsymbol{Y}^{(q)}$ form a Markov chain. Therefore,

$$\begin{aligned} H(\boldsymbol{Y}^{(q)}|\boldsymbol{X}) &= H(\boldsymbol{Y}|\boldsymbol{X}) + H(\boldsymbol{Y}^{(q)}|\boldsymbol{Y}) - H(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{Y}^{(q)}) \\ &\leq H(\boldsymbol{Y}|\boldsymbol{X}) + H(\boldsymbol{Y}^{(q)}|\boldsymbol{Y}). \end{aligned} \tag{4.9}$$

On the other hand, by using the fact that by knowing $\boldsymbol{Y}$, $\boldsymbol{M}$ is also known, we have

$$H(\boldsymbol{Y}^{(q)}|\boldsymbol{Y}) = H(\boldsymbol{Y}^{(q)}|\boldsymbol{M}, \boldsymbol{Y}). \tag{4.10}$$

Furthermore, since the second channel is memoryless, we obtain

$$\begin{aligned} H(\boldsymbol{Y}^{(q)}|\boldsymbol{Y}, \boldsymbol{M}) &= \sum_m p(m) H(\boldsymbol{Y}^{(q)}|\boldsymbol{Y}, \boldsymbol{M} = m) \\ &= \sum_m p(m) m H(Y_j^{(q)}|Y_j) \\ &= E_{\boldsymbol{M}}\{M\} H(Y_j^{(q)}|Y_j), \end{aligned} \tag{4.11}$$

which concludes the proof. □

By combining the results of Lemmas 6 and 7, we obtain

$$\begin{aligned} I(\boldsymbol{X}; \boldsymbol{Y}^q) \geq\ & I(\boldsymbol{X}; \boldsymbol{Y}) - E_{\boldsymbol{M}} \left\{ \log \left(\sum_{\boldsymbol{y}^{(q)}} \sum_{\boldsymbol{y}, p(\boldsymbol{y}) \neq 0} p(\boldsymbol{y}^{(q)}|\boldsymbol{y}, \boldsymbol{M}) p(\boldsymbol{y}^{(q)}|\boldsymbol{M}) \right) \right\} \\ & - E\{\boldsymbol{M}\} H(Y_j^{(q)}|Y_j), \end{aligned} \tag{4.12}$$

which gives a lower bound on the mutual information between the transmitted and received sequences of the concatenated channel $I(\boldsymbol{X}; \boldsymbol{Y}^q)$ in terms of the mutual information between the transmitted and received sequences of the synchronization error channel $I(\boldsymbol{X}; \boldsymbol{Y})$.

Proposition 11. *For any $\boldsymbol{X}$, $\boldsymbol{Y}$ and $\boldsymbol{Y}^{(q)}$ forming a Markov chain $\boldsymbol{X} \to \boldsymbol{Y} \to \boldsymbol{Y}^{(q)}$, if*

$$I(\boldsymbol{X}; \boldsymbol{Y}^{(q)}) \geq I(\boldsymbol{X}; \boldsymbol{Y}) + A,$$

where A is a constant, then the capacity of the channels $\boldsymbol{X} \to \boldsymbol{Y}^{(q)}$ ($C_{\boldsymbol{X} \to \boldsymbol{Y}^{(q)}}$) and $\boldsymbol{X} \to \boldsymbol{Y}$ ($C_{\boldsymbol{X} \to \boldsymbol{Y}}$) satisfy

$$C_{\boldsymbol{X} \to \boldsymbol{Y}^{(q)}} \geq C_{\boldsymbol{X} \to \boldsymbol{Y}} + A. \tag{4.13}$$

Proof. Using the input distribution which achieves the capacity of the channel $\boldsymbol{X} \to \boldsymbol{Y}$, $P(\boldsymbol{X})$, we can write

$$\begin{aligned} \lim_{n \to \infty} \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y}^{(q)}(\boldsymbol{X})) &\geq \lim_{n \to \infty} \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y}(\boldsymbol{X})) + A \\ &= C_{\boldsymbol{X} \to \boldsymbol{Y}} + A. \end{aligned} \tag{4.14}$$

Hence, for the capacity of the channel $\boldsymbol{X} \to \boldsymbol{Y}^{(q)}$, we have

$$\begin{aligned} C_{\boldsymbol{X} \to \boldsymbol{Y}^{(q)}} &= \lim_{n \to \infty} \frac{1}{n} \max_{P(\boldsymbol{X})} I(\boldsymbol{X}; \boldsymbol{Y}^{(q)}) \\ &\geq \lim_{n \to \infty} \frac{1}{n} I(\boldsymbol{X}; \boldsymbol{Y}^{(q)}(X)) \\ &\geq C_{\boldsymbol{X} \to \boldsymbol{Y}} + A, \end{aligned} \tag{4.15}$$

which concludes the proof. □

Due to the result in (4.12) and the result of Proposition 11, the capacity of the concatenated channel can be lower bounded in terms of the capacity of the synchronization error channel and the parameters of the second (memoryless) channel.

4.3 Achievable Rates over Binary Input Symmetric q-ary Output Channels with Synchronization Errors

In this section, we focus on the BSQC channels with synchronization errors (as introduced in Section 2.1.1) and provide lower bounds on the capacity of the channel.

We first develop the results for sub/ers/synch channel and binary input symmetric quaternary output channel, respectively. Then give the results for general (odd and even) q, respectively.

4.3.1 Substitution/Erasure Channels with Synchronization Errors

The following theorem gives a lower bound on the capacity of the sub/ers/synch channel with respect to the capacity of the synchronization error channel. In a sub/ers channel, every transmitted bit is either flipped with probability of s, or erased with probability of e or received correctly with probability of $1-s-e$ independent of each other.

Theorem 2. *The capacity of the sub/ers/synch channel C_{ses} can be lower bounded by*

$$C_{ses} \geq C_s - r\left[H(s,e,1-s-e) + \log\left((1-e)^2 + 2e^2\right)\right], \tag{4.16}$$

where C_s denotes the capacity of the synchronization error channel, $r = \lim_{n\to\infty} \frac{E\{\boldsymbol{M}\}}{n}$, n and m denote the length of the transmitted and received sequences, respectively.

Before giving the proof of Theorem 2, we consider some special cases of this result. Since we have considered the general synchronization error channel model of Dobrushin [10], the lower bound (4.16) holds for many different models on channels with synchronization errors. A popular model for channels with synchronization errors is the Gallager's ins/del model[1] in which every transmitted bit is either deleted with probability of d or replaced with two random bits with probability of i or received correctly with probability of $1-d-i$ independent of each other while neither the transmitter nor the receiver have any information about the insertion and/or deletion errors. If we employ the Gallager's model in deriving the lower bounds, for the

[1] In fact, Gallager's model in general refers to a channel with insertion, deletion and substitution errors, but with Gallager's ins/del model we refer to the case with $s = 0$ (i.e., substitution error probability being zero).

parameter r, we have

$$\begin{aligned} r &= \lim_{n\to\infty} \frac{E\{\boldsymbol{M}\}}{n} \\ &= \lim_{n\to\infty} \frac{1}{n} n E\{|s_j|\} \\ &= 1 - d + i, \end{aligned} \tag{4.17}$$

where $|s_j|$ denotes the length of the output sequence in one use of the ins/del channel, and the equality results since the channel is memoryless. By utilizing the result of (4.17) in (4.16), we obtain the following two corollaries.

Corollary 1. *The capacity of the sub/ers/ins/del channel C_{seid} is lower bounded by*

$$C_{seid} \geq C_{id} - (1 - d + i)\left[H(s, e, 1 - s - e) + \log\left((1 + e)^2 + 2e^2\right)\right], \tag{4.18}$$

where C_{id} denotes the capacity of an insertion/deletion channel with parameters d and i.

Taking $e = 0$ in this channel model gives the ins/del/sub channel, hence we have the following corollary.

Corollary 2. *The capacity of the ins/del/sub channel C_{ids} can be lower bounded by*

$$C_{ids} \geq C_{id} - (1 - d + i)H_b(s), \tag{4.19}$$

To prove Theorem 2, we need the following two lemmas. In the first one we give a lower bound on the output entropy of the sub/ers/synch channel related to the output entropy of the insertion/deletion channel, while in the second one we give an upper bound on the conditional output entropy of the sub/ers/synch channel, related to the conditional output entropy of the insertion/deletion channel.

Lemma 8. *For a sub/ers/synch channel, for any input distribution, we have*

$$H(\boldsymbol{Y}^{(3)}) \geq H(\boldsymbol{Y}) - E\{\boldsymbol{M}\} \log\left((1 - e)^2 + 2e^2\right), \tag{4.20}$$

where $\boldsymbol{Y}$ denotes the output sequence of the synchronization error channel and input sequence of the substitution/erasure channel, and $\boldsymbol{Y}^{(3)}$ denotes the output sequence of the substitution/erasure channel.

Proof. Using the result of Lemma 6, we only need to obtain an upper bound on

$$\sum_{\boldsymbol{y}^{(3)}} \sum_{\boldsymbol{y}, p(\boldsymbol{y})\neq 0} p(\boldsymbol{y}^{(3)}|\boldsymbol{y}, \boldsymbol{M}=m) p(\boldsymbol{y}^{(3)}|\boldsymbol{M}=m)$$

for all values of m. On the other hand for $p(\boldsymbol{y}^{(3)}|\boldsymbol{y}, \boldsymbol{M}=m)$, we have

$$\begin{aligned} p(\boldsymbol{y}^{(3)}|\boldsymbol{y}, \boldsymbol{M}=m) &= \prod_{i=1}^{m} p(Y_i^{(3)}|Y_i) \\ &= e^{j_1} s^{j_2} (1-s-e)^{m-j_1-j_2}, \end{aligned} \tag{4.21}$$

where j_1 denotes the number of transitions $0 \to -$ or $1 \to -$ and j_2 denotes the number of transitions $0 \to 1$ or $1 \to 0$. E.g., $p(011-|0000) = p(0|0)p(1|0)p(1|0)p(-|0) = es^2(1-e-s)$. On the other hand, for a fixed output sequence $\boldsymbol{y}^{(3)}$ of length m with j_1 erased symbols "$-$", there are $2^{j_1}\binom{m-j_1}{j_2}$ possibilities among all m-tuples such that $d(\boldsymbol{y}^{(3)})_e = j_1$, i.e., the number of erased symbols in $\boldsymbol{y}^{(3)}$, and $d(\boldsymbol{y}, \boldsymbol{y}^{(3)})_s = j_2$, i.e., the number of positions in $\boldsymbol{y}$ and $\boldsymbol{y}^{(3)}$ in which $Y_j^{(3)}$'s are the flipped versions of Y_j, therefore we can write

$$\begin{aligned} \sum_{\boldsymbol{y}, p(\boldsymbol{y})\neq 0} p(\boldsymbol{y}^{(3)}|\boldsymbol{y}, \boldsymbol{M}=m) &\leq \sum_{j_2=0}^{m-j_1} 2^{j_1} \binom{m-j_1}{j_2} e^{j_1} s^{j_2} (1-s-e)^{m-j_1-j_2} \\ &= 2^{j_1} e^{j_1} (1-e)^{m-j_1}. \end{aligned} \tag{4.22}$$

Note that in deriving the inequality in (4.6), the summation is taken over the values of $\boldsymbol{y}$ with $p(\boldsymbol{y}) \neq 0$. However, in (4.22) the summation is taken over all possible values of $\boldsymbol{y}$ of length m (over all m-tuples), i.e., $p(\boldsymbol{y}) = 0$ or $p(\boldsymbol{y}) \neq 0$, which results in the lower bound in (4.22). Furthermore, by using the fact that the probability of having

j_1 erasures in a sequence of length m is equal to $\binom{m}{j_1}e^{j_1}(1-e)^{m-j_1}$, we obtain

$$\begin{aligned}
&\sum_{\boldsymbol{y}^{(3)}} p(\boldsymbol{y}^{(3)}|\boldsymbol{M}=m) \sum_{\boldsymbol{y},p(\boldsymbol{y})\neq 0} p(\boldsymbol{y}^{(3)}|\boldsymbol{y},\boldsymbol{M}=m) \\
&\quad \leq \sum_{\boldsymbol{y}^{(3)}} P(d(\boldsymbol{y}^{(3)})_e = j_1|\boldsymbol{M}=m) 2^{j_1} e^{j_1}(1-e)^{m-j_1} \\
&\quad = \sum_{j_1=0}^{m} \binom{m}{j_1} e^{j_1}(1-e)^{m-j_1}(2e)^{j_1}(1-e)^{m-j_1} = \left((1-e)^2+2e^2\right)^m . \qquad (4.23)
\end{aligned}$$

By substituting this result into (4.1), we arrive at

$$H(\boldsymbol{Y}^{(3)}) - H(\boldsymbol{Y}) \;\geq\; -E\{\boldsymbol{M}\}\log\left((1+e)^2+2e^2\right), \qquad (4.24)$$

concluding the proof. □

It is also worth noting that any capacity achieving input distribution over a discrete memoryless channel results in strictly positive output probabilities for possible output sequences of the channel ([23, p. 95]). Therefore, for special synchronization error channel models in which for any possible length of the output sequence m, all the m-tuple output sequences are probable, e.g., i.i.d. deletion channel or i.i.d. random insertion channel, capacity achieving input distributions ($p(\boldsymbol{x})$) would result in strictly positive output probability distributions for all m-tuple output sequences, i.e., $p(\boldsymbol{y}^q) > 0$ for all $\boldsymbol{y}^q$ of length m and all possible m. Hence, the bounds in (4.22) and (4.23) can be thought as equalities for these cases.

Lemma 9. *In any sub/ers/synch channel and for any input distribution, we have*

$$H(\boldsymbol{Y}^{(3)}|\boldsymbol{X}) \leq H(\boldsymbol{Y}|\boldsymbol{X}) + E\{\boldsymbol{M}\}H(e,s,1-e-s). \qquad (4.25)$$

Proof. Due to the result of Lemma 7 and the fact that in a substitution/erasure channel, regardless of the distribution of $\boldsymbol{Y}_j$, we can write

$$H(Y_j^{(3)}|Y_j) = H(e,s,1-e-s), \qquad (4.26)$$

hence the proof follows. □

We can now complete the proof of the main theorem.

Proof of Theorem 2: By substituting the results of Lemmas 8 and 9 into the definition of mutual information, for the same input distribution given to both synchronization error and sub/ers/synch channels, we obtain

$$I(\boldsymbol{X};\boldsymbol{Y}^{(3)}) \geq I(\boldsymbol{X};\boldsymbol{Y}) - E\{\boldsymbol{M}\}\left[H(s,e,1-s-e) + \log\left((1+e)^2+2e^2\right)\right]. \quad (4.27)$$

By using the result of Proposition 11, the proof is completed. ■

4.3.2 Binary Input Symmetric Quaternary Output Channels with Synchronization Errors

In this subsection, we consider a binary input symmetric quaternary output channel with synchronization errors as described in Section 2.1.1.

Theorem 3. *The capacity of the binary input symmetric quaternary output channel with synchronization errors C_{sq} can be lower bounded by*

$$C_{sq} \geq C_s - r\left[H(p_1,p_2,p_3,p_4) + \log\left((p_1+p_3)^2+(p_2+p_4)^2\right)\right], \quad (4.28)$$

where C_s denotes the capacity of the synchronization error only channel, and r is as defined in (4.16).

Note that, the presented lower bound is true for all memoryless synchronization error channel models. Therefore, similar to the sub/ers/synch channel we can specialize the results to the Gallager insertion/deletion channel as given in the following corollary.

Corollary 3. *The capacity of binary input symmetric quaternary output channel with insertion/deletion errors (following Gallager's model) C_{qid} is lower bounded by*

$$C_{qid} \geq C_{id} - (1-d+i)\left[H(p_1,p_2,p_3,p_4) + \log\left((p_1+p_3)^2+(p_2+p_4)^2\right)\right]. \quad (4.29)$$

To prove Theorem 3, we need the two lemmas below where the first one gives a lower bound on the output entropy of the binary input quaternary output channel with synchronization errors related to the output entropy of the synchronization error channel, and the second one gives an upper bound on the conditional output entropy of the binary input quaternary output channel with synchronization errors, related to the conditional output entropy of the synchronization error channel.

Lemma 10. *In any binary input quaternary output channel with synchronization errors and for any input distribution, we have*

$$H(\boldsymbol{Y}^{(4)}) \geq H(\boldsymbol{Y}) - E\left\{\boldsymbol{M}\right\} \log\left((p_1+p_3)^2 + (p_2+p_4)^2\right), \tag{4.30}$$

where $\boldsymbol{Y}$ *denotes the output sequence of the synchronization error channel and input sequence of the binary input quaternary output channel, and* $\boldsymbol{Y}^{(4)}$ *denotes the output sequence of the binary input quaternary output channel corresponding to the input sequence* $\boldsymbol{Y}$*.*

Proof. Similar to the proof of Lemma 8, we use the result of Lemma 6 by taking the summation over all possible sequences of length m, i.e., regardless of $p(\boldsymbol{y}) = 0$ or $p(\boldsymbol{y}) \neq 0$, which results into a looser lower bound. On the other hand, for $p(\boldsymbol{y}^{(4)}|\boldsymbol{y}, \boldsymbol{M} = m)$, we have

$$\begin{aligned} p(\boldsymbol{y}^{(4)}|\boldsymbol{y}, \boldsymbol{M} = m) &= \prod_{i=1}^{m} p(Y_i^{(4)}|Y_i) \\ &= p_1^{j_1} p_2^{j_2} p_3^{j_3} p_4^{m-j_1-j_2-j_3}, \end{aligned} \tag{4.31}$$

where j_1 denotes the number of transitions $0 \to 0^-$ or $1 \to 1^-$, j_2 denotes the number of transitions $0 \to 0^+$ or $1 \to 1^+$, and j_3 denotes the number of transitions $0 \to 1^-$ or $1 \to 0^-$. E.g., $p(0^-1^+0^+1^-|0000) = p(0^-|0)p(1^+|0)p(0^+|0)p(1^-|0) = p_1p_2p_3p_4$. Furthermore, for a fixed output sequence $\boldsymbol{y}^{(4)}$ of length m with j 0^- symbols, k 0^+

symbols, l 1^- symbols and $m-j-k-l$ 1^+ symbols, there are $\binom{j}{i_1}\binom{k}{i_2}\binom{l}{i_3}\binom{m-j-k-l}{i_4}$ possibilities among all m-tuples (for $\boldsymbol{y}$) such that $d(\boldsymbol{y},\boldsymbol{y}^{(4)})_{0\to 0^-}=i_1$, $d(\boldsymbol{y},\boldsymbol{y}^{(4)})_{0\to 0^+}=i_2$, $d(\boldsymbol{y},\boldsymbol{y}^{(4)})_{0\to 1^-}=i_3$ and $d(\boldsymbol{y},\boldsymbol{y}^{(4)})_{0\to 1^+}=i_4$. By defining $m^-(\boldsymbol{y}^{(4)})=\#\{t\le m|y_t^{(4)}\in\{0^-,1^-\}\}$, i.e., the number of the times $y_t^{(4)}=0^-$ or $y_t^{(4)}=1^-$, and $m^+(\boldsymbol{y}^{(4)})=\#\{t\le m|y_t^{(4)}\in\{0^+,1^+\}\}$, i.e., the number of the times $y_t^{(4)}=0^+$ or $y_t^{(4)}=1^+$, we can write

$$\begin{aligned}
&\sum_{\boldsymbol{y},p(\boldsymbol{y})\neq 0} p(\boldsymbol{y}^{(4)}|\boldsymbol{y},\boldsymbol{M}=m)\\
&\le \sum_{i_1=0}^{j}\binom{j}{i_1}p_1^{i_1}p_3^{j-i_1}\sum_{i_2=0}^{k}\binom{k}{i_2}p_2^{i_2}p_4^{k-i_2}\sum_{i_3=0}^{l}\binom{l}{i_3}p_3^{i_3}p_1^{l-i_3}\sum_{i_4=0}^{m-j-k-l}\binom{m-j-k-l}{i_4}p_4^{i_4}p_2^{m-j-k-l-i_4}\\
&=(p_1+p_3)^{j+l}(p_2+p_4)^{m-j-l}\\
&=(p_1+p_3)^{m^-(\boldsymbol{y}^{(4)})}(p_2+p_4)^{m^+(\boldsymbol{y}^{(4)})}. \qquad (4.32)
\end{aligned}$$

By taking the summation over all possible output sequences of length m, and using the fact that the probability of having the output $\boldsymbol{y}^{(4)}$ with length m containing m^- 0^- or 1^- is $\binom{m}{m^-}(p_1+p_3)^{m^-}(p_2+p_4)^{m-m^-}$, we obtain

$$\begin{aligned}
&\sum_{\boldsymbol{y}^{(4)}} p(\boldsymbol{y}^{(4)}|\boldsymbol{M}=m)\sum_{\boldsymbol{y}} p(\boldsymbol{y}^{(4)}|\boldsymbol{y},\boldsymbol{M}=m)\\
&\qquad=\sum_{\boldsymbol{y}^{(4)}} p(\boldsymbol{y}^{(4)}|\boldsymbol{M}=m)(p_1+p_3)^{m^-(\boldsymbol{y}^{(4)})}(p_2+p_4)^{m^+(\boldsymbol{y}^{(4)})}\\
&\qquad=\sum_{m^-=0}^{m}\binom{m}{m^-}(p_1+p_3)^{m^-}(p_2+p_4)^{m-m^-}(p_1+p_3)^{m^-}(p_2+p_4)^{m-m^-}\\
&\qquad=\left((p_1+p_3)^2+(p_2+p_4)^2\right)^m, \qquad (4.33)
\end{aligned}$$

By substituting the result of (4.33) into the result of Lemma 6, we obtain

$$H(\mathbf{Y}^{(4)})\ge H(\mathbf{Y})-E_{\boldsymbol{M}}\{\boldsymbol{M}\}\log\left((p_1+p_3)^2+(p_2+p_4)^2\right), \qquad (4.34)$$

which concludes the proof. □

Table 4.3: Transition probabilities for a binary input 5-ary output channel.

	$P(Y_j^{(q)}\|\bar{Y}_j)$				
Y_j	$Y_j^{(q)}=-2$	$Y_j^{(q)}=-1$	$Y_j^{(q)}=0$	$Y_j^{(q)}=1$	$Y_j^{(q)}=2$
-1	p_2	p_1	p_0	p_{-1}	p_{-2}
1	p_{-2}	p_{-1}	p_0	p_1	p_2

Lemma 11. *For a binary input quaternary output channel with synchronization errors, for any input distribution, we have*

$$H(\boldsymbol{Y}^{(4)}|\boldsymbol{X}) \leq H(\boldsymbol{Y}|\boldsymbol{X}) + E_{\boldsymbol{M}}\{\boldsymbol{M}\}\, H(p_1,p_2,p_3,p_4). \tag{4.35}$$

Proof. Substituting the straightforward result $H(Y_j^{(4)}|Y_j) = H(p_1,p_2,p_3,p_4)$ in the result of Lemma 7 concludes the proof. □

We can now complete the proof of Theorem 3.

Proof of Theorem 3: Using the results of Lemmas 10 and 11, we obtain

$$I(\boldsymbol{X};\boldsymbol{Y}^{(4)}) \geq I(\boldsymbol{X};\boldsymbol{Y}) - nr\left[H(p_1,p_2,p_3,p_4) + \log\left((p_1+p_3)^2+(p_2+p_4)^2\right)\right]. \tag{4.36}$$

Hence, due the result in Proposition 11, the proof is complete. ■

4.3.3 Binary Input Symmetric q-ary Output Channel with Synchronization Errors (Odd q Case)

In this subsection, we consider a binary input symmetric q-ary output channel with synchronization errors for an arbitrary odd value of q, where we represent the transition probability values $P\left(Y_j^{(q)}=k|\bar{Y}_j=b\right)$ for different values of $b\in\{-1,1\}$ and $k=\{-\frac{q-1}{2},\cdots,-1,0,1,\cdots,\frac{q-1}{2}\}$ by $P\left(Y_j^{(q)}=k|\bar{Y}_j=b\right)=p_{k\times b}$. For instance, Table 4.3 shows transition probabilities for a binary input 5-ary output channel.

The main result on the BSQC channel with synchronization errors with odd q is a generalized version of the result in Theorem 2.

Table 4.4: Transition probabilities for a binary input symmetric 6-ary output channel.

	$P(Y_j^{(q)}\|\bar{Y}_j)$					
Y_j	$Y_j^{(q)}=-3$	$Y_j^{(q)}=-2$	$Y_j^{(q)}=-1$	$Y_j^{(q)}=1$	$Y_j^{(q)}=2$	$Y_j^{(q)}=3$
-1	p_3	p_2	p_1	p_{-1}	p_{-2}	p_{-3}
1	p_{-3}	p_{-2}	p_{-1}	p_1	p_2	p_3

Theorem 4. *The capacity of the BSQC channel with synchronization errors C_{Qs} for an odd q can be lower bounded by*

$$C_{Qs} \geq C_s - r\left(H(p_{-\frac{q-1}{2}},\cdots,p_{\frac{q-1}{2}}) + \log\left(2p_0^2 + \sum_{k=1}^{\frac{q-1}{2}}(p_k+p_{-k})^2\right)\right), \tag{4.37}$$

where C_s denotes the capacity of the binary input synchronization error channel.

Proof. The proof of the theorem is given in Appendix E. □

4.3.4 Binary Input Symmetric q-ary Output Channel with Synchronization Errors (Even q Case)

We now consider the generalization of the result of Theorem 3 for even q. For the transition probabilities of the binary input q-ary output channel with $b \in \{-1,1\}$ and $k = \{-\frac{q}{2},\cdots,-1,1,\cdots,\frac{q}{2}\}$, we define $P\left(Y_j^{(q)}=k|\bar{Y}_j=b\right)=p_{k\times b}$. For instance, Table 4.4 shows transition probabilities for a binary input 6-ary output channel.

The main result on the BSQC channel with synchronization errors for any q is given in the following theorem.

Theorem 5. *Capacity of the BSQC channel with synchronization errors C_{Qs}, for any even q can be lower bounded by*

$$C_{Qs} \geq C_s - r\left[H(p_{-\frac{q}{2}},\cdots,p_{-1},p_1,\cdots,p_{\frac{q}{2}}) + \log\left(\sum_{k=1}^{\frac{q}{2}}(p_k+p_{-k})^2\right)\right], \tag{4.38}$$

where C_s denotes the capacity of the binary input synchronization error channel.

Proof. The proof of Theorem 5 is given in Appendix F. □

4.4 Achievable Rates over BI-AWGN Channels with Synchronization Errors

In this section, a binary synchronization error channel in the presence of AWGN is considered as defined in Section 2.1.1

Before giving the main results on AWGN channels with synchronization errors, we would like to make some comments on the information stability of such a channel.

4.4.1 Information Stability of Memoryless Discrete Input Continuous Output Channels with Synchronization Errors

It is shown in [58] that the Shannon's theorem holds in any information stable channel. In [10], the information stability of the memoryless discrete input discrete output channels with synchronization errors is proved which shows that the Shannon's theorem holds in such a channel. It can be observed that the proofs used in [10] can be also generalized to the continuous output case as discussed in this section.

To prove the information stability, it is sufficient to prove the existence of the limit

$$C = \lim_{N\to\infty} \frac{1}{N} C_N = \lim_{N\to\infty} \frac{1}{N} \max_{P(\boldsymbol{X})} I(\boldsymbol{X}; \widetilde{\boldsymbol{Y}}), \tag{4.39}$$

which is the information capacity of the channel, and the existence of an information stable sequence of two random variables $\left(\boldsymbol{X}, \widetilde{\boldsymbol{Y}}\right)$, which achieves the capacity of the channel.

The only difference between the channel considered here with the channel considered by Dobrushin in [10], is that in the continuous output case the output symbols belong to an infinite set. This difference does not have any effect on the steps of proofs however. The existence of the limit in [10, Section IV] is proved based on the memoryless property of the channel which also holds in the continuous output case.

In the case of the existence of an information stable sequence achieving the capacity ([10, Section V]), there is no need to condition on discrete output symbol values, and all the reasoning hold for the continuous output case as well. The key point in the proof is that the channel is stationary which also holds for the continuous output case, such that the same genie-aided channel as the one considered for the discrete output channel can also be considered for the continuous output channel. The genie-aided channel is obtained by inserting markers through the transmission after transmitting each block of length k, where the entire length of transmission is $K = gk + l$ $(l < k)$.

The other point in the proof is the number of possibilities in converting the output of the original channel $\widetilde{\boldsymbol{Y}}$ into the output of the genie-aided channel $\widetilde{\boldsymbol{Y}}'$, i.e., $|f^{-1}(\widetilde{Y})|$ where $\widetilde{Y} = f(\widetilde{Y}')$. Since for the continuous output case we still have $\lim_{g\to\infty} \dfrac{\max_{\widetilde{Y}} |f^{-1}(\widetilde{Y})|}{g} \to 0$, the proof holds.

Since, both capacity convergence and existence of an information stable sequence which achieves the capacity remain valid in the continuous output case as well, we can conclude that the memoryless discrete input continuous output channels with synchronization errors are also information stable and, as a result, the Shannon's theorem applies in such a channel considered in this section as well.

4.4.2 Capacity Lower bounds for AWGN Channels with Synchronization Errors

Here, we present two results on the capacity of an AWGN/synch channel. Both results are generalizations of results for discrete output cases when the number of quantization levels goes to infinity and the quantization level goes to zero. The first result is obtained by employing a uniform quantizer while in deriving the second result a non-uniform quantizer is employed. The second result is the main result in this section since provides a tighter lower bound compared to the first result.

In the following theorem, we present the first result on lower bounding the capacity of the AWGN/synch channel.

Theorem 6. *The capacity of the AWGN/synch channel C_{As} can be lower bounded by*

$$C_{As} \geq C_s - r \log\left(\sqrt{\frac{e}{2}}(1+e^{-\frac{1}{\sigma^2}})\right), \tag{4.40}$$

where C_s denotes the capacity of the synchronization error channel, and signal to noise ratio (SNR) of the BI-AWGN channel is equal to $\frac{1}{2\sigma^2}$.

Here, we give an outline of the proof and defer the details of the proof to Appendix G. To prove Theorem 6, we consider a quantized version of the output symbols where a uniform quantizer is used with the number of quantization levels going to infinity. I.e., by quantizing the output symbols into $2M$-levels and considering uniform quantization levels of Δ, for p_m ($m = \{-M, \cdots, -1, 1, \cdots, M\}$) which denotes the probability that the continuous output symbol, $\widetilde{Y}_j$, being quantized to the bm-th quantization level ($b \in \{-1, 1\}$) conditioned on $\bar{X}_j = b$ being transmitted, we obtain

$$p_m = \begin{cases} Q(\frac{1-m\Delta}{\sigma}) - Q(\frac{1-(m-1)\Delta}{\sigma}) & , \quad m > 0 \\ Q(\frac{1+(|m|-1)\Delta}{\sigma}) - Q(\frac{1+|m|\Delta}{\sigma}) & , \quad m < 0 \end{cases}, \tag{4.41}$$

where $Q(.)$ is the tail probability of the standard normal distribution. By substituting (4.41) in the result of Theorem 5, we can write

$$C_{As} \geq C_s - r \lim_{M\to\infty, \Delta\to 0} \left[H(p_{-M}, \cdots, p_{-1}, p_1, \cdots, p_M) + \log\left(\sum_{m=1}^{M} (p_m + p_{-m})^2\right)\right]. \tag{4.42}$$

Finally, by using the fact that when $M \to \infty$ and $\Delta \to 0$, we have $p_m = f(1-m\Delta)\Delta$, whit $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$, after some algebra (given in Appendix G), Theorem 6 is proved.

We obtain this result as a straightforward generalization of the discrete output channel results by employing a symmetric uniform quantizer, but the result may not

be tight. For instance, for $\sigma = 0$, i.e., the noiseless scenario, the result does not match with the trivial result which is $C_{As} = C_s$ for $\sigma = 0$. We expect that if we apply an appropriate non-uniform quantizer on the output symbols of the AWGN/synch channel, we can achieve a tighter lower bound on its capacity (which also agrees with the trivial result for $C_{As} = C_s$ for $\sigma = 0$). By using this idea, we present our main result on the capacity of an AWGN/synch channel in the following theorem by using a symmetric non-uniform quantizer.

Theorem 7. *Let C_s denote the capacity of the synchronization error channel, then for the capacity of the AWGN/synch channel C_{As}, we obtain*

$$C_{As} \geq C_s - r\left[\log(e)\left(\frac{2}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}} - \frac{2}{\sigma^2}Q\left(\frac{1}{\sigma}\right)\right) + \log\left(1 + Q\left(\frac{1}{\sigma}\right) + e^{\frac{4}{\sigma^2}}Q\left(\frac{3}{\sigma}\right)\right)\right]. \tag{4.43}$$

Proof. To prove the theorem, we first define an appropriate symmetric non-uniform quantizer with $2M$ quantization levels. Then, by letting M go to infinity and employing the result of Theorem 5, we complete the proof.

In general, by utilizing any symmetric quantizer with $2M$ quantization levels on the output symbols $\widetilde{Y}_j$, for the transition probabilities of the resulting binary input symmetric $2M$-ary output channel, we have

$$p_m = P(Y^{(2M)} = bm | \bar{X}_j = b) = \begin{cases} P(t_{m-1} < \widetilde{Y}_j < t_m) & , \quad 0 < m \leq M \\ P(-t_m < \widetilde{Y}_j < -t_{m-1}) & , \quad -M \leq m < 0 \end{cases},$$

where $t_{-m} = -t_m$, $t_0 = 0$ and $t_{m-1} < t_m$ for $m = \{1, \cdots, M\}$. We choose the quantization step sizes, i.e., $\Delta_m = t_m - t_{m-1}$ for $m = \{1, \cdots, M\}$, to satisfy $p_1 = p_2 = \cdots = p_M$. Note that due to symmetry of the quantizer $\Delta_{-m} = \Delta_m$ (as illustrated in Fig. 4.1). On the other hand, by defining $P = Q(\frac{1}{\sigma})$, we have $\sum_{m=1}^{M} p_{-m} = P$ and $\sum_{m=1}^{M} p_m = 1 - P$ which results in $p_m = \frac{1-P}{M}$ for $m = \{1, \cdots, M\}$.

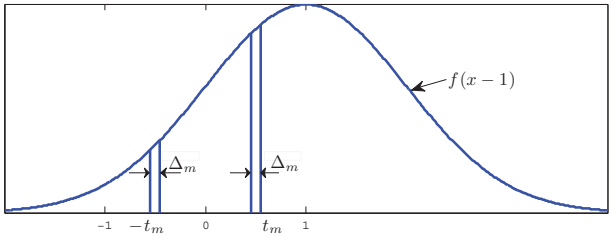

Figure 4.1: Symmetric non-uniform quantizer step sizes.

Using the result of Theorem 5, to derive a lower bound on the capacity of the channel with $2M$-level quantized outputs, we need to obtain $H(p_{-M}, \cdots, p_M) + \log\left(\sum_{m=1}^{M}(p_m + p_{-m})^2\right)$. In the following, we first compute the exact values of $H_M = H(p_{-M}, \cdots, p_{-1}, p_1, \cdots, p_M) - \log(M)$ and $\log\left(\sum_{m=1}^{M}(p_m + p_{-m})^2\right) + \log(M)$. For H_M, we have

$$\begin{aligned} H_M &= -\sum_{m=1}^{M} p_m \log(p_m) - \sum_{m=1}^{M} p_{-m} \log(p_{-m}) - \log(M) \\ &= -(1-P)\log(1-P) - \sum_{m=1}^{M} p_{-m} \log(M p_{-m}). \end{aligned} \tag{4.44}$$

To calculate $-\sum_{m=1}^{M} p_{-m} \log(M p_{-m})$, we first derive a relation between p_m and p_{-m} by using the fact that $\Delta_m = \Delta_{-m}$. For large M and $m = \{1, \cdots, M\}$, we have $p_m \cong f(1 - t_m)\Delta_m$ and $p_{-m} \cong f(1 + t_m)\Delta_m$, where $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$. Furthermore, since $p_m = \frac{1-P}{M}$ for $m = \{1, \cdots, M\}$ and $\frac{f(1+t_m)}{f(1-t_m)} \cong e^{-\frac{2t_m}{\sigma^2}}$, we can write

$$\begin{aligned} p_{-m} &\cong \frac{f(1+t_m)}{f(1-t_m)} p_m \\ &= \frac{1-P}{M} e^{-\frac{2t_m}{\sigma^2}}, \end{aligned} \tag{4.45}$$

with the understanding that the approximation becomes exact as $M \to \infty$. By using this result and the fact that $\sum_{m=1}^{M} p_{-m} = P$, we obtain

$$\begin{aligned} \lim_{M\to\infty} -\sum_{m=1}^{M} p_{-m} \log(M p_{-m}) &= \lim_{M\to\infty} -\sum_{m=1}^{M} p_{-m} \log(1-P) - \lim_{M\to\infty} \sum_{m=1}^{M} p_{-m} \log\left(e^{-\frac{2t_m}{\sigma^2}}\right) \\ &= -P\log(1-P) - \lim_{M\to\infty} \sum_{m=1}^{M} p_{-m} \log\left(e^{-\frac{2t_m}{\sigma^2}}\right). \end{aligned} \tag{4.46}$$

Furthermore, for $\lim_{M\to\infty} -\sum_{m=1}^{M} p_{-m}\log\left(e^{-\frac{2t_m}{\sigma^2}}\right)$, we can write

$$\begin{aligned}\lim_{M\to\infty} -\sum_{m=1}^{M} p_{-m}\log\left(e^{-\frac{2t_m}{\sigma^2}}\right) &= \lim_{M\to\infty}\log(e)\sum_{m=1}^{M} f(1+t_m)\Delta_m\frac{2t_m}{\sigma^2}\\ &= \log(e)\int_0^{\infty} f(1+t)\frac{2t}{\sigma^2}dt\\ &= \log(e)\frac{2}{\sigma^2}\int_0^{\infty}\frac{t}{\sqrt{2\pi}\sigma}e^{-\frac{(t+1)^2}{2\sigma^2}}dt\\ &= \log(e)\left(\frac{2}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}}-\frac{2}{\sigma^2}P\right). \end{aligned} \tag{4.47}$$

By substituting (4.47) and (4.46) into (4.44), we obtain

$$\lim_{M\to\infty} H_M = -\log(1-P)+\log(e)\left(\frac{2}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2\sigma^2}}-\frac{2}{\sigma^2}P\right). \tag{4.48}$$

At this point, we only need to obtain the exact value of $\sum_{m=1}^{M}(p_m+p_{-m})^2$, where we have

$$\begin{aligned}\sum_{m=1}^{M} M(p_m+p_{-m})^2 &= \sum_{m=1}^{M} M(p_m^2+2p_mp_{-m}+p_{-m}^2)\\ &= (1-P)^2+2P(1-P)+\sum_{m=1}^{M} Mp_{-m}^2. \end{aligned} \tag{4.49}$$

Furthermore, if we let M go to infinity, for $\sum_{m=1}^{M} Mp_{-m}^2$, we can write

$$\begin{aligned}\lim_{M\to\infty}\sum_{m=1}^{M} Mp_{-m}^2 &= \lim_{M\to\infty}\sum_{m=1}^{M} Mf(1+t_m)\Delta_m\frac{f(1+t_m)}{f(1-t_m)}p_m\\ &= \lim_{M\to\infty}(1-P)\sum_{m=1}^{M}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t_m+1)^2}{2\sigma^2}}e^{-\frac{2t_m}{\sigma^2}}\Delta_m\\ &= (1-P)\int_0^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t+1)^2}{2\sigma^2}}e^{-\frac{2t}{\sigma^2}}dt\\ &= (1-P)\int_0^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t+3)^2-8}{2\sigma^2}}dt\\ &= (1-P)e^{\frac{4}{\sigma^2}}Q\left(\frac{3}{\sigma}\right). \end{aligned} \tag{4.50}$$

Using the results of (4.50) and (4.48), we obtain

$$\begin{aligned}\lim_{M\to\infty}\left(H(p_{-M},\cdots,p_{-1},p_1,\cdots,p_M)+\log\left(\sum_{m=1}^{M}(p_m+p_{-m})^2\right)\right)&\\ =\lim_{M\to\infty}\left(H_M+\log\left(\sum_{m=1}^{M}M(p_m+p_{-m})^2\right)\right)&\\ =\log(e)\left(\frac{2e^{-\frac{1}{2\sigma^2}}}{\sqrt{2\pi}\sigma}-\frac{2}{\sigma^2}P\right)+\log\left(1+P+e^{\frac{4}{\sigma^2}}Q\left(\frac{3}{\sigma}\right)\right).&\end{aligned} \tag{4.51}$$

Finally, by substituting this result into (4.38), the proof follows. □

By employing a symmetric non-uniform quantizer, we achieve a tighter lower bound on the capacity of the AWGN/synch channel compared to the lower bound in Theorem 6. The result is also in agreement with the trivial result $C_{As} = C_s$ ($\sigma = 0$). A primary advantage of the derived lower bound in (4.43) is that we can use any lower bound on the capacity of the synchronization error only channel to lower bound the capacity of the AWGN/synch channel.

4.5 Numerical Examples

In this section, we give several numerical examples of the lower bounds on the capacity of the ins/del/sub and del/AWGN channel and compare them with the existing ones in the literature. As we are aware of there are no result on lower bounding the capacity of the ins/del/sub/ers and ins/del/AWGN channels to compare them with our results.

4.5.1 Insertion/Deletion/Substitution Channel

In Table 4.5, we compare the lower bound (4.19) with existing lower bounds in [4, 14]. We employ the lower bound derived in [12] as the lower bound on the capacity of the deletion channel and the lower bound in [14] as the lower bound on the capacity of the ins/del channel in lower bound (4.19). Note that the Gallager's model in [4] by

Table 4.5: Comparison between the lower bound derived on the capacity of the ins/del/sub channel with existing lower and upper bounds.

d	i	s	LB from [4]	LB (4.19)	LB from [14]	UB from [14]
0.001	0.00	0.001	0.9772	0.9775	0.9773	0.9856
0.001	0.00	0.010	0.9079	0.9082	0.9081	0.9163
0.001	0.00	0.100	0.5201	0.5204	0.5210	0.5292
0.010	0.00	0.001	0.9079	0.9107	0.9091	0.9586
0.010	0.00	0.010	0.839	0.842	0.842	0.886
0.010	0.00	0.100	0.454	0.458	0.466	0.510
0.100	0.00	0.001	0.5207	0.5514	0.5346	0.7300
0.100	0.00	0.010	0.458	0.489	0.492	0.644
0.100	0.00	0.100	0.108	0.140	0.211	0.363
0.100	0.10	0.001	0.0689	0.1678	0.1761	0.4504
0.100	0.10	0.010	0.013	0.0984	0.139	0.438

parameters d, i and p_c can be considered as concatenation of an ins/del channel with parameters d and i, and a BSC channel with cross error probability of s such that $p_c = (1-s)(1-d-i)$. The advantage of the lower bound (4.19) is in using the tightest lower bound on the capacity of the ins/del channel in lower bounding the capacity of the overall channel, i.e., the information rate of the overall channel is lower bounded for the input distribution which resulted in the tightest lower bound on the capacity of the ins/del channel. We observe that for $i = 0$, a fixed d and small values of s, the lower bound (4.19) improves the lower bound given in [14]. This is not unexpected, because for small values of s the input distribution achieving the capacity of the deletion channel is not far from the optimal input distribution of the del/sub channel. We also observe that the lower bound (4.19) outperforms the lower bound given in [4], but for the case $i \neq 0$ does not improve the lower bound given in [14], since as the lower bound on the capacity of ins/del channel we used the result in [14] and lower bounded further to achieve lower bound on the capacity of the overall channel.

4.5.2 Insertion/Deletion/AWGN Channel

Here, we give some numerical examples of the lower bound (4.43) on the capacity of the ins/del/AWGN channel and compare them with the existing results. As we are aware of, no upper or lower bounds are derived on the capacity of the ins/del/AWGN channel. There are only a few results on the capacity of the deletion/AWGN channel, e.g., the simulation based bound of [6] which is the achievable information rate of the deletion/AWGN channel for i.u.d. input sequences obtained by Monte-Carlo simulations and the analytical result given in Chapter 3 which is a lower bound on the information rate for i.u.d. input sequences.

As we observe in Fig. 4.2, the lower bound (4.43) is far away from the simulation results of [6] for small SNR values and small deletion probability values. This is not unexpected, because in [6], the achievable information rate for i.u.d. input sequences is obtained by Monte-Carlo simulations which requires lengthly simulations. Furthermore, the procedure employed in [6] is only useful for deriving lower bounds for small values of deletion probability, e.g., $d \leq 0.1$, while the lower bound (4.28) holds for the entire range of deletion probabilities by employing any lower bound on the capacity of the deletion channel in lower bounding the capacity of the deletion/AWGN channel. We also observe that, since in deriving the lower bound (4.43) on the capacity of the deletion/AWGN channel, we employ the tightest lower bound presented on the capacity of the deletion channel, for large values of SNR, the lower bound (4.43) improves the lower bound given in [6].

4.6 Chapter Summary

In this chapter, we presented several lower bounds on the capacity of binary input symmetric output channels with synchronization errors in addition to substitutions, erasures or AWGN. We showed that the capacity of any channel with synchronization

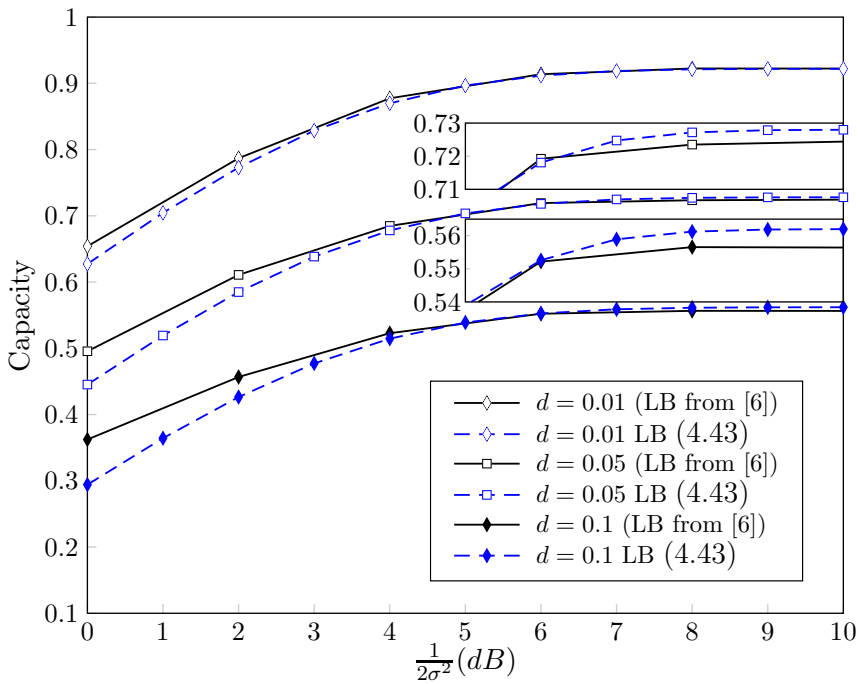

Figure 4.2: Comparison between the lower bound (4.43) with the lower bound in [6] versus SNR for different deletion probabilities.

errors which can be considered as a cascade of two channels (where only the first one suffers from synchronization errors and the second one is a memoryless channel) can be lower bounded in terms of the capacity of the first channel and the parameters of the second channel. We considered two classes of channels: binary input symmetric q-ary output channels (e.g., for $q = 3$ a binary input channel with substitutions and erasures) with synchronization errors and BI-AWGN channels with synchronization errors. We gave the first lower bound on the capacity of substitution/erasure channel with synchronization errors and the first analytical result on the capacity of BI-AWGN channel with synchronization errors. We also demonstrated that the lower bounds developed on the capacity of the del/AWGN channel for small σ^2 values and the del/sub channel for small values of s improve the existing results.

Chapter 5

Improvement of the Deletion Channel Capacity Upper Bound

Memoryless channels with deletion errors as defined by a stochastic channel matrix allowing for bit drop outs are considered in which transmitted bits are either independently deleted with probability d or unchanged with probability $1-d$. Such channels are information stable, hence their Shannon capacity exists. However, computation of the channel capacity is formidable, and only some upper and lower bounds on the capacity exist (see Chapter 2 as a review on the existing upper and lower bounds on the deletion channel capacity). In this chapter, we first define a new channel as a parallel concatenation of two independent deletion channels with deletion probabilities d_1 and d_2 such that any input bit is either transmitted the first deletion channel with probability λ or the second one with probability $1-\lambda$. Then by showing that the new defined channel is in fact another deletion channel with deletion probability $d = \lambda d_1 + (1-\lambda)d_2$, we are able to provide an upper bound on the concatenated deletion channel capacity $C(d)$ in terms of the weighted average of $C(d_1)$, $C(d_2)$ and the parameters of the three channels. An interesting consequence of this bound is that $C(\lambda d_1 + (1-\lambda)) \leq \lambda C(d_1)$ which enables us to provide an improved upper bound on the capacity of the i.i.d. deletion channels, i.e., $C(d) \leq 0.4143(1-d)$ for $d \geq 0.65$. This generalizes the asymptotic result by Dalai [59] as it remains valid for all $d \geq 0.65$. Using the same approach we are also able to improve upon existing upper bounds on the capacity of the deletion/substitution channel.

The chapter is organized as follows. In Section 5.1, we give a brief introduction on memoryless channels with synchronization errors and an outline of our results. In Section 5.2, we prove the main result of the chapter which relates the capacity of the three different deletion channels through an inequality. In Section 5.3, we generalize the result to the case of deletion/substitution channels and the parallel

concatenation of more than two channels. In Section 5.4, we present tighter upper bounds on the capacity of the deletion and deletion/substitution channels based on previously known best upper bounds, and comment on the limit of the capacity as the deletion probability approaches unity. We conclude the chapter in Section 5.5.

5.1 Introduction

Channels with synchronization errors can be well modeled using bit drop outs and/or bit insertions as well as random errors. There are many different models adopted in the literature to describe these errors. Among them, a relatively general model is employed by Dobrushin [10] where memoryless channels with synchronization errors are described by a channel matrix allowing for the channel outputs to be of different lengths for different uses of the channel. As proved in the same paper, for such channels, information stability holds and Shannon capacity exists. However, the determination of the capacity remains elusive as the mutual information term to be maximized does not admit a single letter or finite letter form.

In this chapter, we prove that the capacity of an i.i.d. deletion channel with deletion probability of d as an arithmetic mean of two different deletion probabilities d_1 and d_2, i.e., $d = \lambda d_1 + (1-\lambda)d_2$ for $\lambda \in [0, 1]$, can be upper bounded in terms of the capacity and the parameters of the two newly considered deletion channels. The proof relies on the simple observation that the deletion channel with deletion probability d can be considered as the parallel concatenation of two independent deletion channels with deletion probabilities d_1 and d_2 where each bit is either transmitted over the first channel with probability λ or the second channel with probability $1 - \lambda$.

Thanks to the presented inequality relation among the deletion channels capacity, we are able to improve upon the existing upper bounds on the capacity of the deletion channel for $d \geq 0.65$ [13]. The improvement is the result of the fact that the currently known best upper bounds are not convex for some range of deletion

probabilities. More precisely, our result allows us to convexify the existing deletion channel capacity upper bound for $d \geq 0.65$, leading to a significant improvement of the upper bound. In other words, we are able to prove that for $0 \leq \lambda \leq 1$, $C(\lambda d + 1 - \lambda) \leq \lambda C(d)$, resulting in $C(d) \leq 0.4143(1-d)$ for $d \geq 0.65$ which is tighter than the result in [13]. The same result for the asymptotic scenario $d \to 1$ was also obtained in [59] using a different approach; however our result is valid for $d \geq 0.65$ hence more general. We also note that the best known limiting lower bound (as $d \to 1$) is $0.1185(1-d)$ [11]. We also demonstrate that a similar improvement is possible for the case of deletion/substitution channels. As an example, we can prove that for $s = 0.03$, an improved capacity upper bound is obtained for $d \geq 0.6$ over the best existing result given in [14].

5.2 Main Theorem

In this section, we provide the main result of the chapter on the capacity of the deletion channel and its proof. Furthermore, we present a simple proof for the special case with $d_2 = 0$, i.e., $C(\lambda d_1 + 1 - \lambda) \leq \lambda C(d_1)$.

The theorem below states our basic result whose proof hinges on a simple observation.

Theorem 8. *Let $C(d)$ denotes the capacity of the i.i.d. deletion channel with deletion probability d, $\lambda \in [0,1]$ and $d = \lambda d_1 + (1-\lambda)d_2$, then we have*

$$\begin{aligned} C(d) \;\leq\; & \lambda C(d_1) + (1-\lambda)C(d_2) + (1-d)\log(1-d) \\ & -\lambda(1-d_1)\log(\lambda(1-d_1)) - (1-\lambda)(1-d_2)\log((1-\lambda)(1-d_2)). \end{aligned} \tag{5.1}$$

Proof. Let us consider two different deletion channels, $\mathcal{C}_1$ and $\mathcal{C}_2$, with deletion probabilities d_1 and d_2, input sequences of bits $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$, and output sequences of bits $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, respectively. Denote their Shannon capacities by $C(d_1)$ and $C(d_2)$, respectively. Given a specific $\lambda \in (0,1)$, define a new binary input channel $\mathcal{C}'$ (shown

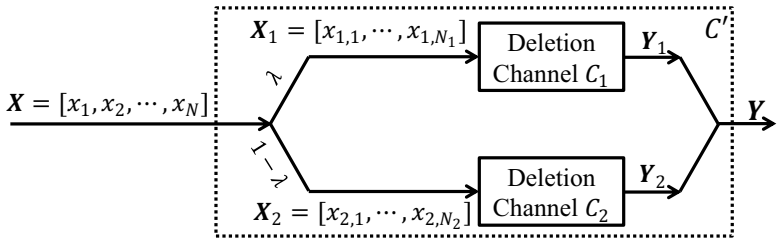

Figure 5.1: Channel Model $\mathcal{C}'$.

in Fig. 5.1) with input sequence of bits $\boldsymbol{X}$ and output sequence of bits $\boldsymbol{Y}$ as follows: each channel input symbol is transmitted through $\mathcal{C}_1$ with probability λ, and through $\mathcal{C}_2$ with probability $1-\lambda$, independently of each other. Neither the transmitter nor the receiver knows the specific realization of the "individual channel selection events," i.e., they do not know which specific subchannel a symbol is transmitted through, and which specific subchannel each output symbol is received from. The following two lemmas demonstrate that 1) the new channel is a new i.i.d. deletion channel with deletion probability $d = \lambda d_1 + (1-\lambda)d_2$, 2) if appropriate side information be provided for the transmitter and the receiver then the capacity of the genie-aided channel is upper bounded by $\lambda C(d_1) + (1-\lambda)C(d_2) + (1-d)\log(1-d) - \lambda(1-d_1)\log(\lambda(1-d_1)) - (1-\lambda)(1-d_2)\log((1-\lambda)(1-d_2))$. Combining these two results, the proof of the theorem follows easily by noting that the capacity of the new channel $\mathcal{C}'$ cannot decrease with side information. □

The following two lemmas are employed in the proof of the theorem.

Lemma 12. *$\mathcal{C}'$ as defined in the proof of the theorem above is nothing but a deletion channel with deletion probability $d = \lambda d_1 + (1-\lambda)d_2$.*

Proof. For each use of the channel $\mathcal{C}'$, for any input symbol $x \in \mathcal{X}$ and channel output $y \in \mathcal{Y}$, the transition probability is given by $P\{\mathcal{C}_1 \text{ is used}\}d_1 + P\{\mathcal{C}_2 \text{ is used}\}d_2 = \lambda d_1 + (1-\lambda)d_2$. Noting that the subchannels are memoryless and the channel selection

events are independent of each other, this transition matrix precisely defines a deletion channel with deletion probability $d = \lambda d_1 + (1-\lambda)d_2$. □

Lemma 13. *The capacity of the channel $\mathcal{C}'$ as defined in the proof of the theorem above is upper bounded by*

$$\begin{aligned} C' \leq & \lambda C(d_1) + (1-\lambda)C(d_2) + (1-d)\log(1-d) - \lambda(1-d_1)\log(\lambda(1-d_1)) \\ & - (1-\lambda)(1-d_2)\log((1-\lambda)(1-d_2)). \end{aligned}$$

Proof. We first define a new genie-aided channel which is obtained by providing the transmitter and the receiver of the channel $\mathcal{C}'$ with appropriate side information, then derive an upper bound on the capacity of the genie-aided channel which is also an upper bound on the capacity of the channel $\mathcal{C}'$. More precisely, we provide the transmitter with side information on which channel is being used for each transmitted symbol ($\boldsymbol{X} = \boldsymbol{X}_1\boldsymbol{X}_2$), and the receiver with side information on which channel the received symbol comes from ($\boldsymbol{Y} = \boldsymbol{Y}_1\boldsymbol{Y}_2$), and reveal the side information on the fragmentation information, i.e., random process $\boldsymbol{F}_y$, to the receiver such that by knowing $\boldsymbol{F}_y$, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$, one can retrieve $\boldsymbol{Y}$. $\boldsymbol{F}_y$ is defined as an M-tuple $\boldsymbol{F}_y = (f_y[1], \cdots, f_y[M])$, where M denotes the length of the received sequence $\boldsymbol{Y}$, i.e., $M = |Y|$, and $f_y[i] \in \{1, 2\}$ denotes the index of the channel the i-th received bit is coming from. We also define $\boldsymbol{F}_x$ which determines the fragmentation process from the random process $\boldsymbol{X}$ to $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ as an N-tuple $\boldsymbol{F}_x = (f_x[1], \cdots, f_x[N])$, where $f_x[i] \in \{1, 2\}$ denotes the index of the channel the i-th bits is going through.

Since $\boldsymbol{X} \to (\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x) \to (\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{F}_y) \to \boldsymbol{Y}$ form a Markov chain, we can write

$$\begin{aligned} I(\boldsymbol{X}; \boldsymbol{Y}) &\leq I(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{F}_y) \\ &= I_1 + I_2 + I_3, \end{aligned} \tag{5.2}$$

where $I_1 = I(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{Y}_1)$, $I_2 = I(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{Y}_2|\boldsymbol{Y}_1)$ and $I_3 = I(\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{F}_y|\boldsymbol{Y}_1, \boldsymbol{Y}_2)$. For I_1, we have

$$\begin{aligned} I_1 &= I(\boldsymbol{X}_1; \boldsymbol{Y}_1) + I(\boldsymbol{X}_2, \boldsymbol{F}_x; \boldsymbol{Y}_1|\boldsymbol{X}_1) \\ &= I(\boldsymbol{X}_1; \boldsymbol{Y}_1), \end{aligned} \tag{5.3}$$

where we used the fact that $P(\boldsymbol{Y}_1|\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x) = P(\boldsymbol{Y}_1|\boldsymbol{X}_1)$, i.e., $\boldsymbol{Y}_1$ is independent of $\boldsymbol{X}_2$ and $\boldsymbol{F}_x$ conditioned on $\boldsymbol{X}_1$. Furthermore, by using the facts that $P(\boldsymbol{Y}_2|\boldsymbol{X}_2, \boldsymbol{Y}_1) = P(\boldsymbol{Y}_2|\boldsymbol{X}_2)$ and $P(\boldsymbol{Y}_2|\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x, \boldsymbol{Y}_1) = P(\boldsymbol{Y}_2|\boldsymbol{X}_2)$, we obtain

$$\begin{aligned} I_2 &= I(\boldsymbol{X}_2; \boldsymbol{Y}_2|\boldsymbol{Y}_1) + I(\boldsymbol{X}_1, \boldsymbol{F}_x; \boldsymbol{Y}_2|\boldsymbol{Y}_1, \boldsymbol{X}_2) \\ &= H(\boldsymbol{Y}_2|\boldsymbol{Y}_1) - H(\boldsymbol{Y}_2|\boldsymbol{X}_2) \\ &\leq I(\boldsymbol{X}_2; \boldsymbol{Y}_2). \end{aligned} \tag{5.4}$$

We are not able to derive the exact value of I_3, therefore we derive an upper bound on I_3 which results in an upper bound on $I(\boldsymbol{X}, \boldsymbol{Y})$. For I_3, if we define $N_i = |X_i|$ and $M_i = |Y_i|$ as the length of the transmitted and received sequences form the i-th channel, respectively, then we can write

$$\begin{aligned} I_3 &= H(\boldsymbol{F}_y|\boldsymbol{Y}_1, \boldsymbol{Y}_2) - H(\boldsymbol{F}_y|\boldsymbol{Y}_1, \boldsymbol{Y}_2, \boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{F}_x) \\ &\leq H(\boldsymbol{F}_y|\boldsymbol{Y}_1, \boldsymbol{Y}_2) \\ &= H(\boldsymbol{F}_y|\boldsymbol{M}_1, \boldsymbol{M}_2). \end{aligned} \tag{5.5}$$

For fixed M_1 and M_2, there are $\binom{M_1+M_2}{M_2}$ possibilities for $\boldsymbol{F}_y = (f_y[0], \cdots, f_y[M_1])$. Therefore, we obtain

$$\begin{aligned} H(\boldsymbol{F}_y|\boldsymbol{M}_1 = M_1, \boldsymbol{M}_2 = M_2) &\leq \log\left(\binom{M_1 + M_2}{M_2}\right) \\ &\leq (M_1+M_2)\log(M_1+M_2) - M_1\log(M_1) - M_2\log(M_2), \end{aligned} \tag{5.6}$$

where we have used the inequality $\log \binom{n}{k} \leq nH_b(\frac{k}{n})$ provided in [60, p. 353]. Due to the fact that $(x+a)\log(x+a) - x\log(x)$ is a concave function of x for $a > 0$, and $E\{\boldsymbol{M}_1|\boldsymbol{M}_2 = M_2\} = (N - M_2)\frac{\lambda(1-d_1)}{\lambda+(1-\lambda)d_2}$ (see Appendix H), by applying Jensen's inequality, we can write

$$\begin{aligned}
I_3 &\leq E_{\boldsymbol{M}_1,\boldsymbol{M}_2}\{H(\boldsymbol{F}_y|\boldsymbol{M}_1,\boldsymbol{M}_2)\} \\
&\leq E_{\boldsymbol{M}_2}\Big\{(E\{\boldsymbol{M}_1|\boldsymbol{M}_2\}+\boldsymbol{M}_2)\log(E\{\boldsymbol{M}_1|\boldsymbol{M}_2\}+\boldsymbol{M}_2) \\
&\qquad -E\{\boldsymbol{M}_1|\boldsymbol{M}_2\}\log(E\{\boldsymbol{M}_1|\boldsymbol{M}_2\}) - \boldsymbol{M}_2\log(\boldsymbol{M}_2)\Big\} \\
&= E\Bigg\{\left(\frac{\lambda(N-\boldsymbol{M}_2)(1-d_1)}{\lambda+(1-\lambda)d_2}+\boldsymbol{M}_2\right)\log\left(\frac{\lambda(N-\boldsymbol{M}_2)(1-d_1)}{\lambda+(1-\lambda)d_2}+\boldsymbol{M}_2\right) \\
&\quad -\frac{\lambda(N-\boldsymbol{M}_2)(1-d_1)}{\lambda+(1-\lambda)d_2}\log\left(\frac{\lambda(N-\boldsymbol{M}_2)(1-d_1)}{\lambda+(1-\lambda)d_2}\right)-\boldsymbol{M}_2\log(\boldsymbol{M}_2)\Bigg\}. \quad (5.7)
\end{aligned}$$

Furthermore since $(a(b-x)+x)\log(a(b-x)+x) - a(b-x)\log(a(b-x)) - x\log(x)$ is a concave function of x for $a > 0$ and $0 < x \leq b$, and $E\{\boldsymbol{M}_2\} = N(1-\lambda)(1-d_2)$ (see Appendix H), by applying Jensen's inequality, we obtain

$$\begin{aligned}
I_3 &\leq N(\lambda(1-d_1)+(1-\lambda)(1-d_2))\log(N(\lambda(1-d_1)+(1-\lambda)(1-d_2))) \\
&\quad -N\lambda(1-d_1)\log(N\lambda(1-d_1)) - N(1-\lambda)(1-d_2)\log(N(1-\lambda)(1-d_2)) \\
&= N(\lambda(1-d_1)+(1-\lambda)(1-d_2))\log(\lambda(1-d_1)+(1-\lambda)(1-d_2)) \\
&\quad -N\lambda(1-d_1)\log(\lambda(1-d_1)) - N(1-\lambda)(1-d_2)\log((1-\lambda)(1-d_2)). \quad (5.8)
\end{aligned}$$

On the other hand, for $I(\boldsymbol{X}_i;\boldsymbol{Y}_i)$ $(i \in \{1,2\})$, we can write

$$\begin{aligned}
I(\boldsymbol{X}_i;\boldsymbol{Y}_i) &= I(\boldsymbol{X}_i;\boldsymbol{Y}_i,\boldsymbol{N}_i) - I(\boldsymbol{X}_i;\boldsymbol{N}_i|\boldsymbol{Y}_i) \\
&= I(\boldsymbol{X}_i;\boldsymbol{Y}_i|\boldsymbol{N}_i) + I(\boldsymbol{X}_i;\boldsymbol{N}_i) - I(\boldsymbol{X}_i;\boldsymbol{N}_i|\boldsymbol{Y}_i) \\
&\leq I(\boldsymbol{X}_i;\boldsymbol{Y}_i|\boldsymbol{N}_i) + H(\boldsymbol{N}_i) \\
&\leq I(\boldsymbol{X}_i;\boldsymbol{Y}_i|\boldsymbol{N}_i) + \log(N+1) \\
&= \sum_{N_i=0}^{N} P(\boldsymbol{N}_i = N_i) I(\boldsymbol{X}_i;\boldsymbol{Y}_i|\boldsymbol{N}_i = N_i) + \log(N+1), \quad (5.9)
\end{aligned}$$

where in deriving the first inequality we have used the facts that $H(\boldsymbol{N}_i|\boldsymbol{X}_i) = 0$ and $I(\boldsymbol{X}_i; \boldsymbol{N}_i|\boldsymbol{Y}_i) \geq 0$, and in deriving the second equality the fact that

$$H(\boldsymbol{N}_i) = -\sum_{n=0}^{N} \binom{N}{n} \lambda^n (1-\lambda)^{N-n} \log\left(\binom{N}{n} \lambda^n (1-\lambda)^{N-n}\right) \leq \log(N+1). \quad (5.10)$$

Furthermore, as it is shown in [13], for a finite length transmission over the deletion channel, the mutual information rate between the transmitted and received sequences can be upper bounded in terms of the capacity of the channel after adding some appropriate term, which can be spelled out as [13, Eqn. (39)]

$$I(\boldsymbol{X}_i; \boldsymbol{Y}_i|\boldsymbol{N}_i = N_i) \leq N_i C(d_i) + H(\boldsymbol{D}_i|\boldsymbol{N}_i = N_i), \quad (5.11)$$

where $\boldsymbol{D}_i$ denotes the number of deletion through the transmission of N_i bits over the i-th channel and

$$H(\boldsymbol{D}_i|\boldsymbol{N}_i = N_i) = -\sum_{n=0}^{N_i} \binom{N_i}{n} d_i^n (1-d_i)^{N_i-n} \log\left(\binom{N_i}{n} d_i^n (1-d_i)^{N_i-n}\right) \leq \log(N_i+1).$$

Substituting (5.11) into (5.9), we have

$$\begin{aligned} I(\boldsymbol{X}_i; \boldsymbol{Y}_i) &\leq \sum_{N_i=0}^{N} P(\boldsymbol{N}_i = N_i) \left(N_i C(d_i) + \log(N_i+1)\right) + \log(N+1) \\ &\leq \lambda_i N C(d_i) + \log(\lambda_i N + 1) + \log(N+1), \end{aligned} \quad (5.12)$$

where the last inequality results since $\log(x)$ is a concave function of x, and $\lambda_1 = \lambda$ and $\lambda_2 = 1-\lambda$. Finally, by substituting (5.12), (5.8), (5.4) and (5.3) in (5.2), we obtain

$$\begin{aligned} I(\boldsymbol{X}; \boldsymbol{Y}) \leq\ & N\lambda C(d_1) + \log(\lambda N + 1) + N(1-\lambda) C(d_2) + \log((1-\lambda)N + 1) \\ & + 2\log(N+1) + N(1-d)\log(1-d) - N\lambda(1-d_1)\log(\lambda(1-d_1)) \\ & - N(1-\lambda)(1-d_2)\log((1-\lambda)(1-d_2)). \end{aligned}$$

By dividing both sides of the above inequality by N, letting N go to infinity, and noting that the inequality is valid for any input distribution $P(\boldsymbol{X})$, the proof follows.

□

Note that for the special case of $\mathcal{C}_2$ being a pure deletion channel, i.e., $d_2 = 1$, the presented upper bound (5.15) results into $C(\lambda d_1 + 1 - \lambda) \leq \lambda C(d_1)$. One can observe that to prove the relation $C(\lambda d_1 + 1 - \lambda) \leq \lambda C(d_1)$, there is no need for the entire proof given in Lemma 13. More precisely, when $\mathcal{C}_2$ is a pure deletion channel, $\boldsymbol{X} \to \boldsymbol{X}_1 \to \boldsymbol{Y}_1 \to \boldsymbol{Y}$ form a Markov chain ($\boldsymbol{Y} = \boldsymbol{Y}_1$), therefore we can write

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &\leq I(\boldsymbol{X}_1;\boldsymbol{Y}_1) \\ &\leq \lambda N C(d_1) + \log(\lambda_1 N + 1) + \log(N+1), \end{aligned} \tag{5.13}$$

where the last inequality holds due to (5.12). Furthermore, by dividing both sides of the above inequality by N, letting N go to infinity, and the fact that the inequality is valid for any input distribution $P(\boldsymbol{X})$, we arrive at $C(\lambda d_1 + 1 - \lambda) \leq \lambda C(d_1)$.

Another observation from the result $C(\lambda d_1 + (1-\lambda)) \leq \lambda C(d_1)$ is that by series concatenation of two independent deletion channels with deletion probabilities d_1 and $1-\lambda$, we also arrive at a deletion channel with deletion probability of $d = \lambda d_1 + 1 - \lambda$. Therefore we can say that the capacity of the series concatenation of two independent deletion channels can be upper bounded in terms of the capacity of one of them and the parameters of the other.

5.3 Some Generalizations and Implications

5.3.1 Generalization to the Case of Deletion/Substitution Channel

In a deletion/substitution channel (special case of the Gallager channel model without any insertions) with parameters (d,f), any transmitted bit is either deleted with probability of d or flipped with probability of f or received correctly with probability of $1 - d - f$, where neither the transmitter nor the receiver have any information about the position of the deleted and flipped bits. It is easy to show that the result of Theorem 8 can also be generalized to the deletion/substitution channel as given in the following corollary.

Corollary 4. *Let $C(d, f)$ denotes the capacity of the deletion/substitution channel with deletion probability d and flip probability f, $\lambda \in [0, 1]$, $d = \lambda d_1 + (1 - \lambda)d_2$ and $f = \lambda f_1 + (1 - \lambda)f_2$, then we have*

$$C(d, f) \leq \lambda C(d_1, f_1) + (1 - \lambda)C(d_2, f_2) + (1 - d)\log(1 - d) \\ - \lambda(1 - d_1)\log(\lambda(1 - d_1)) - (1 - \lambda)(1 - d_2)\log((1 - \lambda)(1 - d_2)). \quad (5.14)$$

Proof. The proof of Lemma 12 simply holds if we consider $\mathcal{C}_1$ in Fig. 5.1 as a deletion/substitution channel with parameters (d_1,f_1) and $\mathcal{C}_2$ as another one with parameters (d_2,f_2), then $\mathcal{C}$ becomes also a deletion/substitution channel with parameters $(\lambda d_1 + (1-\lambda)d_2, \lambda f_1 + (1-\lambda)f_2)$. Furthermore, replacing the deletion channel $\mathcal{C}_i$ with deletion probability d_i with a deletion/substitution channel with parameters (d_i,f_i) does not change the distribution of $\boldsymbol{N}_i$ and $\boldsymbol{M}_i$. Therefore, the proof of Lemma 13 holds for the deletion/substitution channel as well. □

Note that a deletion/substitution channel with parameters (d, f) can be considered as a series concatenation of two independent channels where the first one is a deletion only channel with deletion probability of d and the second one is a binary symmetric channel (BSC) with cross error probability $s = \frac{f}{1-d}$ ($1 - d - f \leq 1$ and if $d = 1$ then $s = 0$). If we define $C_s(d, s) = C(d, (1 - d)s)$, then for $d_2 = 1$ and $f_2 = 0$, we obtain

$$C_s(\lambda d_1 + 1 - \lambda, s) \leq \lambda C_s(d_1, s). \quad (5.15)$$

5.3.2 Parallel Concatenation of More Than Two Channels

So far, we considered the parallel concatenation of two independent deletion channels which is useful in improving upon the existing upper bounds. However, we can also consider the parallel concatenation of more than two deletion channels. If we define the deletion channel $\mathcal{C}$ as a parallel concatenation of P independent deletion

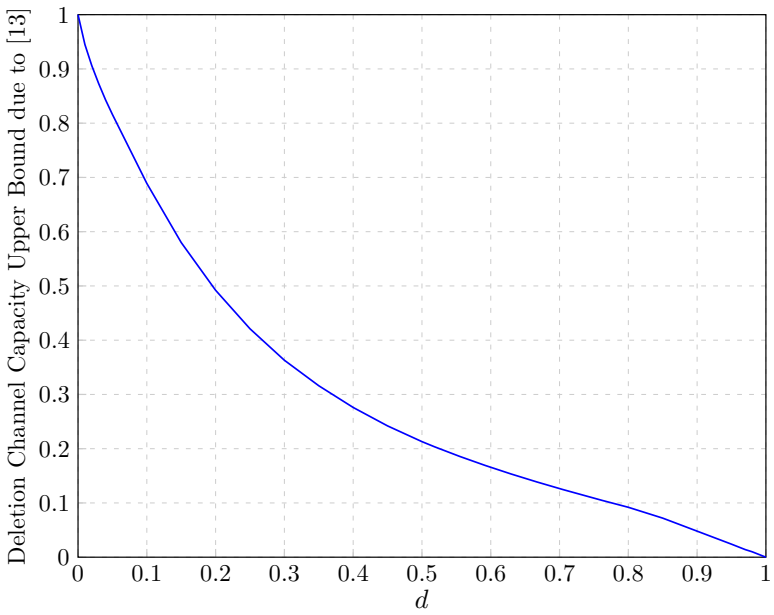

Figure 5.2: Previously best known upper bound on the i.i.d. deletion channel capacity.

channels $\mathcal{C}_p$ with deletion probability d_p ($p = \{1, \cdots, P\}$) where each input bit is transmitted with probability λ_p over $\mathcal{C}_p$, and modify the definition of $\boldsymbol{F}_y$ such that $f_y[i] \in \{1, \cdots, P\}$ denotes the index of the channel the i-th bit is coming from, then for $d = \sum_{p=1}^{P} \lambda_p d_p$, we have

$$C(d) \leq \sum_{p=1}^{P} \lambda_p C(d_p) + (1-d)\log(1-d) - \sum_{p=1}^{P} \lambda_p (1-d_p)\log(\lambda_p(1-d_p)), \qquad (5.16)$$

where $\sum_{p=1}^{P} \lambda_p = 1$. Note, however, that this result does not give any tighter upper bounds on the deletion channel capacity than the one obtained by considering the parallel concatenation of only two independent deletion channels.

5.4 Improved Upper Bounds on the Deletion Channel Capacity

An interesting application of the result (5.1) on the capacity of the deletion and deletion/substitution channels is in obtaining improved capacity upper bounds. For instance, the best known upper bound on the deletion channel capacity is not convex for $d \geq 0.65$ as shown in Fig. 5.2 (with values taken from the boldfaced values in Table IV of [13]). As clarified in the table, the best known values for small d are due

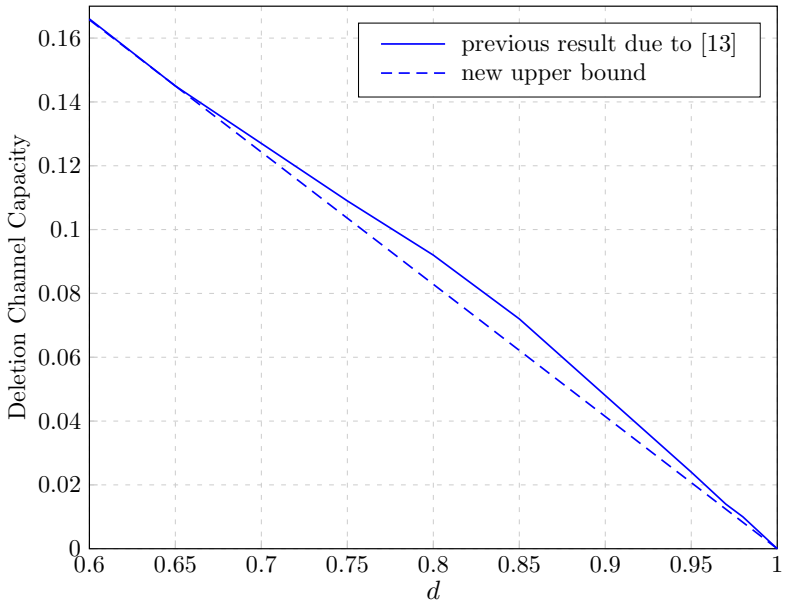

Figure 5.3: Improved upper bound on the deletion channel capacity employing $C(\lambda d + 1 - \lambda) \leq \lambda C(d)$.

to [31], for a wide range (up to $d \sim .8$) are due to the "fourth version" of the upper bound (named C_4 in [13]), and for large values of d are due to the "second version" named C_2^* in the same paper. Therefore, the deletion channel capacity upper bound can be improved for $d \in (0.65, 1)$ as $C(1 - 0.35\lambda) \leq \lambda C(0.65) \leq \lambda C_4(0.65)$ with $0 \leq \lambda \leq 1$. That is, we have $C(d) \leq 0.4143(1-d)$ for $d \in (0.65, 1)$. This is illustrated in Fig. 5.3.

We note that our result is a generalization of the one in [59] where it was shown that $C(d) \leq 0.4143(1-d)$ as $d \to 1$. We also note an earlier asymptotic result on a lower bound derived in [11] which states that $C(d) \geq 0.1185(1-d)$ as $d \to 1$.

As another application of the inequality derived in this chapter, we can consider the capacity of the deletion/substitution channel. The best known capacity upper bound for this case is given in [14], e.g., Fig. 1 of [14] presents several upper bounds for fixed $s = 0.03$ (see Fig. 5.4). It is clear that this bound is not a convex function of the deletion probability for $d \geq 0.6$, hence it can be improved. That is, applying

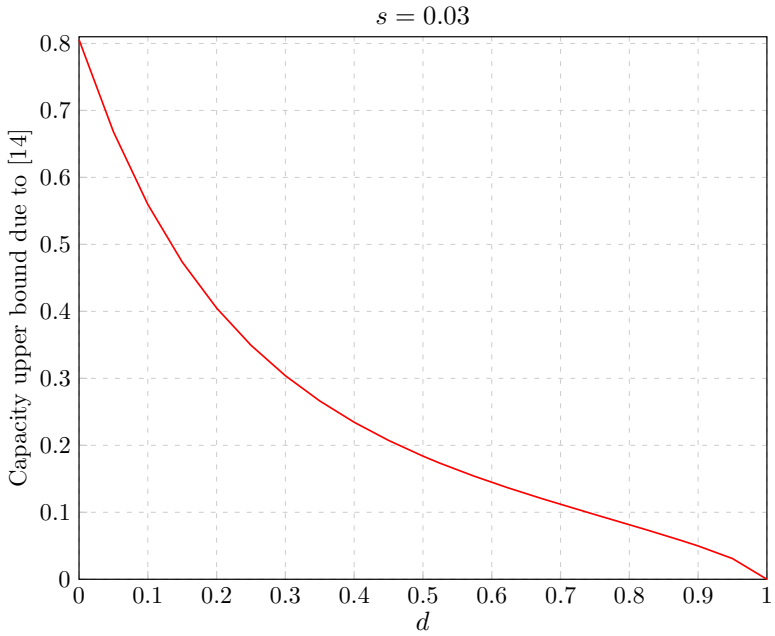

Figure 5.4: Previously best known upper bound on the deletion/substitution channel capacity for $s = 0.03$.

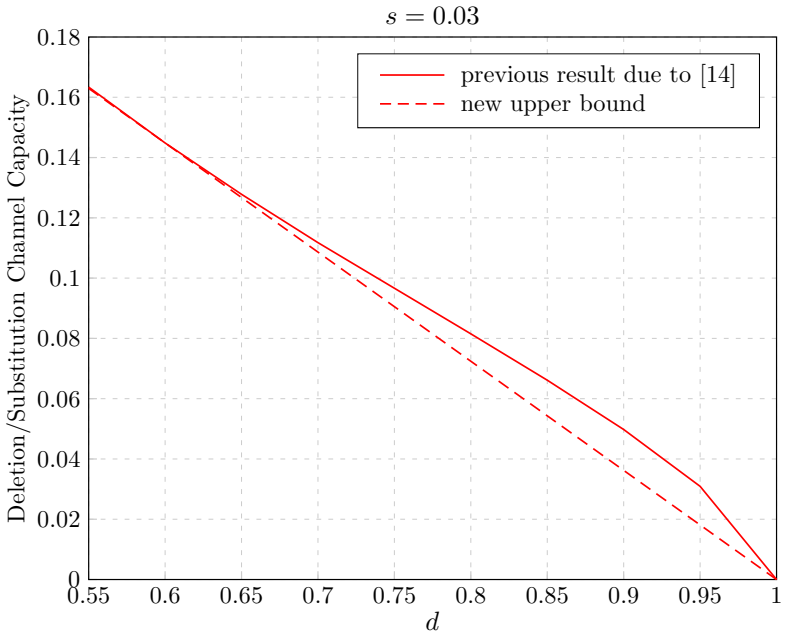

Figure 5.5: Improved upper bound on the deletion/substitution channel capacity for $s = 0.03$.

the result in (5.14), we obtain, for instance for $s = 0.03$, $C_s(d, 0.03) \leq 0.3621(1-d)$ for $d \geq 0.6$ which is a tighter bound as illustrated in Fig. 5.5.

5.5 Chapter Summary

In this chapter, an inequality relating the capacity of a deletion channel to two other deletion channels is found. The main idea is to consider parallel concatenation of two different independent deletion channels and relate the capacity of the resulting deletion channel with the capacity of the first two. An immediate application of this result is in obtaining improved upper bounds on the capacity of the deletion channel as the best available upper bounds are not convex in the deletion probability, and the derived inequality results in a tighter capacity characterization. For an i.i.d. deletion channel, we proved that $C(d) \geq 0.4143(1-d)$ for all $d \geq 0.65$. This is a stronger result than the earlier characterization in [59] which is valid only asymptotically as $d \to 1$. We also noted a generalization of the result to the case of a deletion/substitution channel and provided a tighter capacity upper bound for this case as well.

Chapter 6

An Upper Bound on the Capacity of the Non-Binary Deletion Channels

A simple upper bound on the capacity of non-binary deletion channels is derived in this chapter. Although binary deletion channels have received significant attention over the years, and many upper and lower bounds on their capacity have been derived, such studies for the non-binary case are largely missing. The state of the art is the following: as a trivial upper bound, capacity of an erasure channel with the same input alphabet as the deletion channel can be used, and as a lower bound the results of [22] are available. In this chapter, the fist non-trivial upper bound on the $2K$-ary deletion channel capacity is derived. We first show that any $2K$-ary input deletion channel can be considered as a parallel concatenation of K independent binary deletion channels which enables us to prove that $C_{2K}(d) \leq C_2(d) + (1-d)\log(K)$. Then by employing the existing upper bounds on the capacity of the binary deletion channel, we obtain upper bounds on the capacity of the $2K$-ary deletion channel. We illustrate via examples the use of the new bounds and discuss their asymptotic behavior as $d \to 0$.

The chapter is organized as follows. In Section 6.1, we give a brief introduction. In Section 6.2, we consider the $2K$-ary input deletion channel introduced in Chapter 2 as a parallel concatenation of K independent deletion channels (where each input is binary). In Section 6.3, we discuss the possible generalization of the existing Blahut-Arimoto algorithm (BAA) based upper bounding approaches (useful for the binary deletion channels) to the case of $2K$-ary deletion channels. In Section 6.4, we prove the main result of the chapter providing an upper bound on $C_{2K}(d)$ in terms of $C_2(d)$. In Section 6.5, several implications of the result are given: we compare the resulting capacity upper bounds with the existing capacity upper and lower bounds, and we provide a discussion of the channel capacity behavior as the deletion probability approaches zero. Finally, we conclude the chapter in Section 6.6.

6.1 Introduction

Non-binary independent and identically distributed (i.i.d.) deletion channels can be used to model information transmission over a finite buffer channel [22], where a packet (non-binary symbol) loss occurs whenever a packet arrives at a full buffer. Dobrushin [10] proved the existence of the Shannon's theorem for discrete memoryless channels with synchronization errors. As a result, Shannon's theorem holds in non-binary deletion channels and information and transmission capacities are equal.

In this chapter, we focus on a $2K$-ary deletion channel $\mathcal{C}$ in which every transmitted symbol is either lost through the transmission with probability of d or received correctly with probability of $1-d$. There is no information about the position of the lost symbols at either the transmitter or receiver. We present a non-trivial upper bound on the capacity of this channel. Clearly the capacity of a $2K$-ary erasure channel with erasure probability d is an upper bound on the capacity of the $2K$-ary deletion channel since by revealing information about the position of the lost symbols to the receiver, the corresponding genie-aided deletion channel is nothing but an erasure channel. Therefore, for the capacity of the $2K$-ary input deletion channel $C_{2K}(d)$, the relation $C_{2K}(d) \leq (1-d)\log(2K)$ holds. Besides this trivial upper bound, to the best of our knowledge, there are no other (tighter) upper bounds on the capacity of non-binary deletion channels.

To the best of our knowledge, the only non-trivial lower bounds on the capacity of the non-binary deletion channels are provided in [22] where two different bounds are derived. More precisely, the achievable rates of the $2K$-ary input deletion channel are computed for i.i.d. and Markovian codebooks by considering a simple decoder which decides in favor of a sequence if the received sequence is the subsequence of only one transmitted sequence. Considering i.i.d. codebooks, the derived achievable

rates are given by

$$C_{2K} \geq \log\left(\frac{2K}{2K-1}\right) + (1-d)\log(2K-1) - H_b(d), \tag{6.1}$$

and considering Markovian codebooks are given by

$$C_{2K} \geq \sup_{\gamma>0,\ 0<p<1} \left[-(1-d)\log\left((1-q)A + qB\right) - \gamma\log(e)\right] \tag{6.2}$$

with $q = \frac{1}{2K}\left(1 + \frac{(1-d)(2K-1)(2Kp-1)}{2K-1-d(2Kp-1)}\right)$, $A = \frac{e^{-\gamma}(1-p)}{(2K-1)(1-e^{-\gamma}(1-\frac{1-p}{2K-1}))}$ and $B = e^{-\gamma}\left((1-p)A + p\right)$. Non-binary input alphabet channels with synchronization errors are also considered in [36] where the capacity of memoryless synchronization error channels in the presence of noise and the capacity of channels with week synchronization errors (i.e., the transmitter and receiver are partly synchronized) have been studied. The main focus of the work in [36] is on asymptotic behavior of the channel capacity for large values of K.

Our main result is to relate the capacity of a $2K$-ary deletion channel with deletion probability d to the capacity of the binary deletion channel with deletion probability d by the inequality $C_{2K} \leq C_2(d) + (1-d)\log(K)$. As a result, any upper bound on the binary deletion channel capacity can be used to derive an upper bound on the $2K$-ary deletion channel capacity. For example, by using the result from Chapter 5, we obtain $C_{2K}(d) \leq (\log(K) + 0.4143)(1-d)$ for $d \geq 0.65$.

6.2 A Different Look at the $2K$-ary Deletion Channel

Any $2K$-ary input deletion channel with deletion probability d can be considered as a parallel concatenation of K independent binary deletion channels $\mathcal{C}_k$ ($k \in \{1, \ldots, K\}$) all with the same deletion probability d, as shown in Figure 6.1, in which the input symbols $2k-1$ and $2k$ travel through $\mathcal{C}_k$ and the surviving output symbols of the subchannels are combined based on the order in which they go through the subchannels. $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$ denote the input and output sequences of the k-th channel, respectively, and N_k and M_k denote the length of $\boldsymbol{X}_k$ and $\boldsymbol{Y}_k$, respectively.

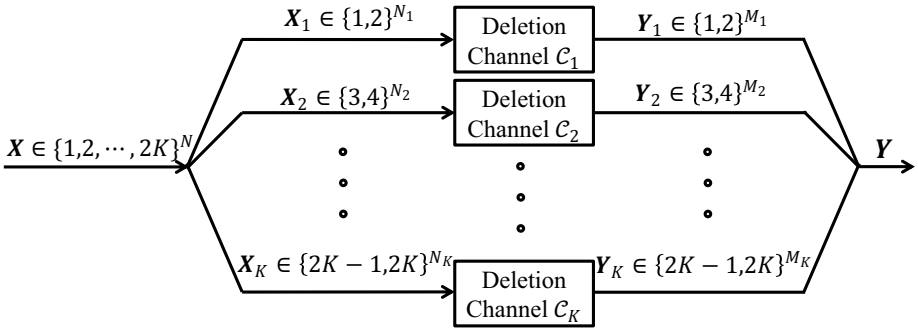

Figure 6.1: $2K$-ary deletion channel as a parallel concatenation of K independent binary input deletion channels.

To be able to relate the mutual information between the input and output sequences of the $2K$-ary deletion channel, $I(\boldsymbol{X};\boldsymbol{Y})$, with the mutual information between the input and output sequences of the considered binary deletion channels, $I(\boldsymbol{X}_k;\boldsymbol{Y}_k)$, we define two new random vectors $\boldsymbol{F}_x$ and $\boldsymbol{F}_y$. More precisely, $\boldsymbol{F}_x = (f_x[1],\ldots,f_x[N])$ and $\boldsymbol{F}_y = (f_y[1],\ldots,f_y[M])$ such that $f_x[n] \in \{1,\ldots,K\}$ and $f_y[m] \in \{1,\ldots,K\}$ denote the label of the subchannel the n-th input symbol and m-th output symbol belong to, respectively. Clearly, by knowing $\boldsymbol{X}$, one can determine $(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x)$ and by knowing $(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x)$ can determine $\boldsymbol{X}$. The same situation holds for $\boldsymbol{Y}$ and $(\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K,\boldsymbol{F}_y)$. Therefore, we have

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &= I(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x;\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K,\boldsymbol{F}_y) \\ &= \sum_{k=1}^{K} I_k + I_F, \end{aligned} \tag{6.3}$$

where $I_k = I(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x;\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1})$ and

$$I_F = I(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x;\boldsymbol{F}_y|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_K). \tag{6.4}$$

In Section 6.4, we will derive upper bounds on I_k and I_F which will enable us to relate the non-binary and binary deletion channels capacities, and lead to the main result of the chapter.

6.3 Discussion on the BAA Based Upper Bounds

The best known upper bounds on the capacity of the binary deletion channel for $d \leq 0.65$ are computed numerically in [13, 31]. In [31] a genie-aided channel is considered in which the receiver is provided by side information about the completely deleted runs, e.g., in transmitting $'110001'$ over the original channel by deleting the entire run of zeros, the sequence $'111'$ is received while in the considered genie-aided channel $'11{-}1'$ represents the possible reception. Then an upper bound on the capacity per unit cost of the genie-aided channel is computed by running the BAA algorithm. Fertonani and Duman [13], by considering several different genie-aided channels, are able to derive tighter upper bounds on the binary deletion channel capacity compared to the results in [31] for $d > 0.05$.

One approach to derive upper bounds on the capacity of the $2K$-ary deletion channel is to modify the numerical approaches in [13, 31] in which the decoder (and possibly the encoder) of the binary deletion channel is provided with some side information about the deletion process and the capacity (or an upper bound on the capacity) of the resulted genie-aided channel is numerically evaluated by running the BAA algorithm. Although this approach is useful for the binary channels (even when other type of impairments such as insertions and substitutions are considered [14]), for the non-binary case, running the BAA for large values of K is not computationally feasible. For example, one of the upper bounds in [13] is obtained by computing the capacity of the binary deletion channel with finite length of transmission $L = 17$. Obviously, by increasing the alphabet size, $2K$, the maximum possible value of L in running the BAA algorithm decreases. Therefore, to achieve reasonable upper bounds, L needs to be increased which makes the numerical computations infeasible.

The main contribution of this chapter is that we are able to relate the capacity of the $2K$-ary deletion channel to the binary deletion channel capacity through an inequality relation which enables us to upper bound the $2K$-ary deletion channel capacity avoiding computationally formidable BAA directly for the $2K$-ary deletion channel.

6.4 A Novel Upper Bound on $C_{2K}(d)$

As introduced in Section 6.2, a $2K$-ary deletion channel can be considered as a parallel concatenation of K independent binary deletion channels. This new look at a $2K$-ary deletion channel enables us to relate the $2K$-ary deletion channel capacity to the binary deletion channel capacity with the same deletion error probability which is given in the following theorem.

Theorem 9. *Let $C_{2K}(d)$ denote the capacity of a $2K$-ary deletion channel with deletion probability d, then*

$$C_{2K}(d) \leq C_2(d) + (1-d)\log(K). \tag{6.5}$$

As given in (6.3), the mutual information $I(\boldsymbol{X};\boldsymbol{Y})$ can be expanded in terms of several other mutual information terms, I_k for $k \in \{1,\ldots,K\}$ and I_F. To prove Theorem 9, we first derive upper bounds on I_k and I_F in the following two lemmas which enable us to complete the proof of Theorem 9.

Lemma 14. *For all the possible input distributions $P(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x)$, the mutual information I_k given in* (6.3) *can be upper bounded by*

$$I_k \leq E\{N_k\}C_2(d) + 2\log(N+1),$$

where $E\{.\}$ denotes the expected value.

Proof. For I_k, since $P(\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1},\boldsymbol{X}_k) = P(\boldsymbol{Y}_k|\boldsymbol{X}_k)$ and

$$P(\boldsymbol{Y}_k|\boldsymbol{X}_1,\ldots,\boldsymbol{X}_K,\boldsymbol{F}_x,\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1}) = P(\boldsymbol{Y}_k|\boldsymbol{X}_k),$$

we can write

$$\begin{aligned} I_k &= I(\boldsymbol{X}_k;\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1}) \\ &\quad +I(\boldsymbol{X}_1,...,\boldsymbol{X}_{k-1},\boldsymbol{X}_{k+1},...,\boldsymbol{X}_K,\boldsymbol{F}_x;\boldsymbol{Y}_k|\boldsymbol{Y}_1,...,\boldsymbol{Y}_{k-1},\boldsymbol{X}_k) \\ &= I(\boldsymbol{X}_k;\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1}) \\ &= H(\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1}) - H(\boldsymbol{Y}_k|\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1},\boldsymbol{X}_k) \\ &= H(\boldsymbol{Y}_k) - I(\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{k-1};\boldsymbol{Y}_k) - H(\boldsymbol{Y}_k|\boldsymbol{X}_k) \\ &\leq I(\boldsymbol{X}_k;\boldsymbol{Y}_k). \end{aligned} \tag{6.6}$$

Furthermore, $I(\boldsymbol{X}_k;\boldsymbol{Y}_k)$ can be written as

$$\begin{aligned} I(\boldsymbol{X}_k;\boldsymbol{Y}_k) &= I(\boldsymbol{X}_k;\boldsymbol{Y}_k,N_k) - I(\boldsymbol{X}_k;N_k|\boldsymbol{Y}_k) \\ &= I(\boldsymbol{X}_k;\boldsymbol{Y}_k|N_k) + I(\boldsymbol{X}_k;N_k) - I(\boldsymbol{X}_k;N_k|\boldsymbol{Y}_k). \end{aligned}$$

Since $H(N_k|\boldsymbol{X}_k)=0$ and $I(\boldsymbol{X}_k;N_k|\boldsymbol{Y}_k)\geq 0$, we arrive at

$$\begin{aligned} I(\boldsymbol{X}_k;\boldsymbol{Y}_k) &\leq I(\boldsymbol{X}_k;\boldsymbol{Y}_k|N_k) + H(N_k) \\ &\leq I(\boldsymbol{X}_k;\boldsymbol{Y}_k|N_k) + \log(N+1) \\ &= \sum_{n_k=0}^{N} P(N_k=n_k) I(\boldsymbol{X}_k;\boldsymbol{Y}_k|n_k) + \log(N+1), \end{aligned} \tag{6.7}$$

where the second inequality results since there are $N+1$ possibilities for N_k and as a result $H(N_k)\leq \log(N+1)$. Furthermore, as it is shown in [13], for a finite length transmission over the deletion channel, the mutual information rate between the transmitted and received sequences can be upper bounded in terms of the capacity of the channel after adding some appropriate term, which can be spelled out as [13, Eqn. (39)]

$$I(\boldsymbol{X}_k;\boldsymbol{Y}_k|N_k=n_k) \leq n_k C_2(d) + H(\boldsymbol{D}_k|N_k=n_k), \tag{6.8}$$

where $\boldsymbol{D}_k$ denotes the number of deletions through the transmission of N_k bits over the k-th channel and

$$\begin{aligned} H(\boldsymbol{D}_k|N_k = n_k) &= -\sum_{n=0}^{n_k} P(n_k, n, d) \log\left(P(n_k, n, d)\right) \\ &\leq \log(n_k + 1) \leq \log(N + 1), \end{aligned} \tag{6.9}$$

with $P(n_k, n, d) = \binom{n_k}{n} d^n (1-d)^{n_k - n}$. Substituting (6.9) and (6.8) into (6.7), we obtain

$$\begin{aligned} I(\boldsymbol{X}_k; \boldsymbol{Y}_k) &\leq \sum_{n_k=0}^{N} P(N_k = n_k)\left(n_k C_2(d)\right) + 2\log(N+1) \\ &= E\{N_k\} C_2(d) + 2\log(N+1). \end{aligned}$$

Finally, by substituting the above inequality in (6.6), the proof follows. □

Lemma 15. *For all the possible input distributions, the mutual information I_F given in* (6.4) *can be upper bounded by*

$$I_F \leq N(1-d)\log(K).$$

Proof. Using the definition of the mutual information, we can write

$$\begin{aligned} I_F &= H(\boldsymbol{F}_y|\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_K) - H(\boldsymbol{F}_y|\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_K, \boldsymbol{X}_1, \ldots, \boldsymbol{X}_K, \boldsymbol{F}_x) \\ &\leq H(\boldsymbol{F}_y|\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_K) \\ &\leq H(\boldsymbol{F}_y|M_1, \ldots, M_K), \end{aligned} \tag{6.10}$$

where the last inequality follows since $(M_1, \ldots, M_K)$ is a function of $(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_K)$, i.e., $H(M_1, \ldots, M_K|\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_K) = 0$. For fixed m_k such that $\sum_{k=1}^{K} m_k = m$, there are $\binom{m}{m_1, \ldots, m_K}$ possibilities for $\boldsymbol{F}_y$ leading to $H(\boldsymbol{F}_y|m_1, \ldots, m_K) \leq \log\binom{m}{m_1, \ldots, m_K}$. It follows from the inequality

$$\log\binom{m}{m_1} \leq m\log(m) - m_1\log(m_1) - (m - m_1)\log(m - m_1)$$

given in [60, p. 353] that

$$\begin{aligned}
\log \binom{m}{m_1, \ldots, m_K} &= \sum_{j=1}^{K-1} \log \binom{m - \sum_{k=1}^{j-1} m_k}{m_j} \\
&\leq \sum_{j=1}^{K-1} \left(m - \sum_{k=1}^{j-1} m_k \right) \log \left(m - \sum_{k=1}^{j-1} m_k \right) - m_j \log m_j \\
&\quad - \sum_{j=1}^{K-1} \left(m - \sum_{k=1}^{j} m_k \right) \log \left(m - \sum_{k=1}^{j} m_k \right) \\
&= m \log(m) - \sum_{k=1}^{K} m_k \log(m_k).
\end{aligned}$$

Therefore, we obtain $H(\boldsymbol{F}_y | m_1, \ldots, m_K) \leq m \log(m) - \sum_{k=1}^{K} m_k \log(m_k)$. Since

$$g([m_1, \ldots, m_k]) = \left(\sum_{k=1}^{K} m_k \right) \log \left(\sum_{k=1}^{K} m_k \right) - \sum_{k=1}^{K} m_k \log(m_k)$$

is a concave function of $[m_1, \ldots, m_K]$ (see Appendix I), employing the Jensen's inequality yields

$$I_F \leq \left(\sum_{k=1}^{K} E\{M_k\} \right) \log \left(\sum_{k=1}^{K} E\{M_k\} \right) - \sum_{k=1}^{K} E\{M_k\} \log(E\{M_k\}).$$

On the other hand, due to the fact that $\mathcal{C}_k$ are i.i.d. binary input deletion channels, we have $E\{M_k\} = N(1-d)\alpha_k$ where α_k depend on the input distribution $P(\boldsymbol{X})$ and $\sum_{k=1}^{K} \alpha_k = 1$. Hence, we obtain

$$\begin{aligned}
I_F &\leq N(1-d) \left(\log \left(N(1-d) \right) - \sum_{k=1}^{K} \alpha_k \log \left(N(1-d)\alpha_k \right) \right) \\
&= -N(1-d) \sum_{k=1}^{K} \alpha_k \log \alpha_k \\
&= N(1-d) H(\alpha_1, \ldots, \alpha_K) \\
&\leq N(1-d) \log(K), \qquad (6.11)
\end{aligned}$$

which concludes the proof. □

6.4.1 Proof of Theorem 9

Substituting the results of Lemmas 14 and 15 in (6.3), we obtain

$$\begin{aligned} I(\boldsymbol{X};\boldsymbol{Y}) &\leq E_{N_1,\ldots,N_K}\left\{\sum_{k=1}^{K} N_k\right\} C_2(d) + 2K\log(N+1) + N(1-d)\log(K) \\ &= NC_2(d) + 2K\log(N+1) + N(1-d)\log(K), \end{aligned}$$

where we have used the fact that $\sum_{k=1}^{K} N_k = N$ independent of the input distribution $P(\boldsymbol{X})$. Since the above inequality holds for any input distribution $P(\boldsymbol{X})$ and any value of N, we can write

$$\begin{aligned} C_{2K}(d) &= \lim_{N\to\infty} \max_{P(\boldsymbol{X})} \frac{1}{N} I(\boldsymbol{X};\boldsymbol{Y}) \\ &\leq C_2(d) + (1-d)\log(K), \end{aligned}$$

which concludes the proof of Theorem 9. □

6.5 Some Implications

The trivial upper bound on the capacity of the $2K$-ary deletion channel is given by $(1-d)\log(2K)$ which is the capacity of the $2K$-ary erasure channel. In fact, if we reveal the side information about the position of the dropped symbols to the receiver of a $2K$-ary deletion channel, the resulting genie-aided channel is nothing but a $2K$-ary erasure channel.

We have shown in the previous section that by substituting any upper bound on the capacity of the binary deletion channel into (6.5), an upper bound on the $2K$-ary deletion channel capacity results. Obviously, by employing $C_2(d) \leq 1-d$ which is the trivial upper bound on the binary deletion channel capacity, the erasure channel upper bound on the $2K$-ary deletion channel capacity is obtained. Therefore, any upper bound tighter than $1-d$ on the binary deletion channel capacity gives an upper bound tighter than $\log(2K)(1-d)$ on the $2K$-ary deletion channel capacity.

The amount of improvement is $1-d-C_2^{UB}(d)$, where C_2^{UB} denotes the upper bound on the binary deletion channel capacity.

As it is shown in [36] that $(1-d)\log(2K)-1 \leq C_{2K}(d) \leq (1-d)\log(2K)$, the existing trivial upper and lower bounds are tight enough for asymptotically large values of K, and i.i.d. distributed input sequences are sufficient to achieve the capacity. However, the importance of the result in Theorem 9 is for moderate values of K, where the amount of improvement in closing the gap between the existing upper and lower bounds is significant.

To demonstrate the improvement over the trivial erasure channel upper bound, we compare the upper bound $C_{2K}(d) \leq C_2^{UB}(d) + (1-d)\log(K)$ with the erasure channel upper bound $\log(2K)(1-d)$ and the tightest existing lower bound (6.2) (provided in [22]) in Figures 6.2 and 6.3 for 4-ary and 8-ary deletion channels, respectively. Here we utilize the binary deletion channel capacity upper bounds $C_2^{UB}(d)$ in [13] and Chapter 5, where for $d \leq 0.65$ we use the results in [13, Table III] and for $d \geq 0.65$ we use the upper bound $C_2(d) \leq 0.4143(1-d)$ given in Chapter 5.

Another implication of the result in Theorem 9 is in studying the asymptotic behavior of the $2K$-ary deletion channel capacity for $d \to 0$. It is shown in [61] that

$$C_2(d) = 1 + d\log(d) - A_1 d + A_2 d^2 + O(d^{3-\epsilon}), \tag{6.12}$$

for small d and any $\epsilon > 0$ with $A_1 \approx 1.15416377$, $A_2 \approx 1.78628364$ and $O(.)$ denoting the standard Landau (big-O) notation. Employing this result into (6.5), leads to an upper bound expansion for small values of d as

$$C_{2K}(d) \leq 1 + d\log(d) - (A_1 + \log(K))\, d + A_2 d^2 + \log(K) + O(d^{3-\epsilon}). \tag{6.13}$$

In Figure 6.4, we compare the above upper bound (by ignoring the $O(d^{3-\epsilon})$ term) with the lower bound (6.2) for $d \leq 0.1$ and for 4-ary and 8-ary input deletion channels. We observe that by employing the capacity expansion (6.12) in (6.5), a better

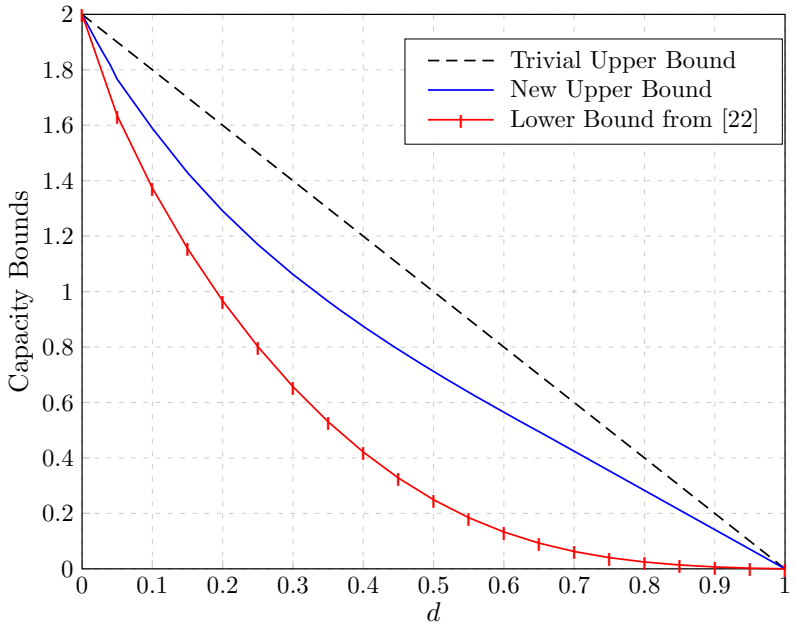

Figure 6.2: Comparison among the new upper bound (6.5), the lower bound (6.2) and the trivial erasure channel upper bound for the 4-ary deletion channel.

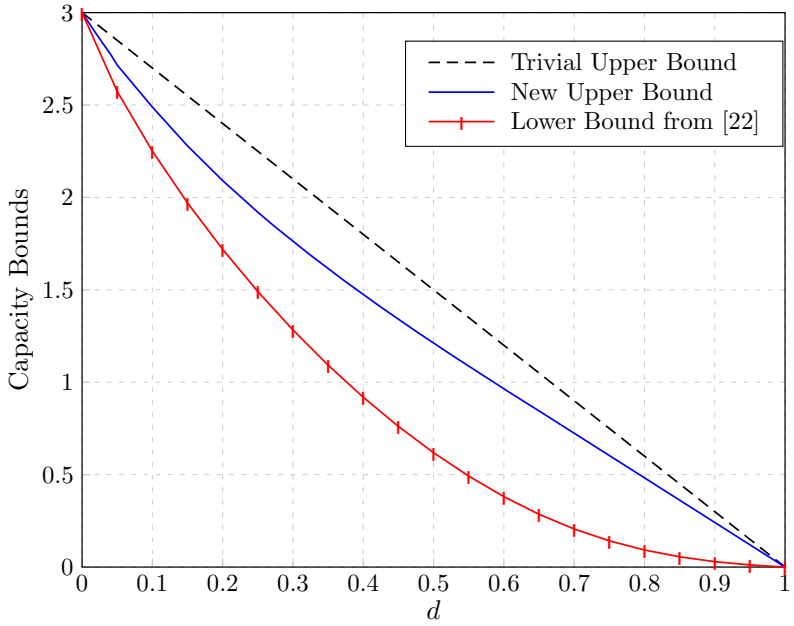

Figure 6.3: Comparison among the new upper bound (6.5), the lower bound (6.2) and the trivial erasure channel upper bound for the 8-ary deletion channel.

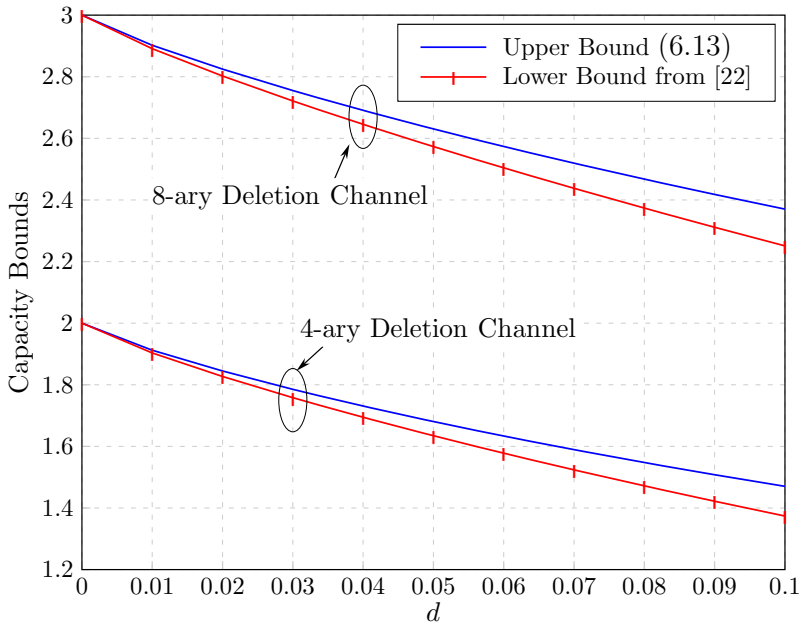

Figure 6.4: Comparison between the upper bound (6.13) (ignoring the $O(d^{3\epsilon})$ term) and the lower bound (6.2) for 4-ary and 8-ary deletion channels.

characterization for the asymptotic behavior of the $2K$-ary deletion channel capacity is obtained as $d \to 0$.

6.6 Chapter Summary

We have derived the first non-trivial upper bound on the $2K$-ary deletion channel capacity. We first considered the $2K$-ary deletion channel as a parallel concatenation of K independent binary deletion channels, all with the same deletion probability. We then related the capacity of the original channel to that of the binary deletion channel. By doing so we obtained an upper bound on the capacity of the $2K$-ary deletion channel in terms of the capacity of the binary deletion channel and as a result any upper bound on the capacity of the binary deletion channel. The provided upper bound results into tighter upper bounds than the trivial erasure channel upper bound for the entire range of the deletion probability d and all $K > 0$.

Chapter 7

Spectrally Efficient Alamouti Code Structure in Asynchronous Cooperative Systems

A cooperative communication system with two amplify and forward (AF) relays under flat fading channel conditions is considered in which the signals received from the relay nodes are not necessarily time aligned and both relays share the same time and frequency bands to communicate with the destination. We propose a new time-reversal (TR) based scheme providing Alamouti code structure with the objective of reducing the overhead needed to overcome the aynchronism and to obtain spatial diversity. The scheme is particularly useful when the delay between the two relay signals is large, e.g., in typical underwater acoustic (UWA) channels.

The chapter is organized as follows. In Section 7.1, we give an introduction on the problem addressed in the chapter. In Section 7.2, we present the considered asynchronous cooperative communication system model. In Section 7.3, we introduce the proposed relaying scheme. In Section 7.4, we first obtain the optimal detector structure, then present a simpler sub-optimal detector which results into the Alamouti fast ML detector for all the symbols. In Section 7.5, we numerically evaluate the performance of both optimal and sub-optimal detectors. Finally, the chapter is concluded in Section 7.6.

7.1 Introduction

As it is well known, the Alamouti coding [41] is a two transmit diversity scheme for multi-input multi-output (MIMO) communication systems. Although originally it was proposed for centralized MIMO systems, the idea was later extended to cooperative communications [62] where geographically separated nodes form virtual transmit antennas to provide spatial diversity. In most existing works, the actively cooperating nodes are assumed perfectly time aligned. However, due to the distributed nature of

a cooperative system, achieving perfect time alignment among the signals received from geographically separated nodes may not always be possible. Therefore, conventional MIMO transmission schemes designed for synchronized transmitting antennas may not be directly applicable.

There are several results in the literature to achieve diversity in asynchronous cooperative systems with decode and forward (DF) relaying, e.g., [2, 48]. In [2], time reversal space-time block coding (TR-STBC) is proposed where each transmitted block is preceded by a time guard greater than the maximum possible relative delay D_{max} among the relays. For the system with two relay nodes, the scheme in [2] gives an Alamouti coding structure at the receiver. There are also a few works on AF relaying, e.g., [3, 63]. In [3], an Alamouti space time transmission scheme employing orthogonal frequency division multiplexing (OFDM) is proposed in which the transmitted OFDM blocks are separated by adding a cyclic prefix (CP) longer than D_{max} and at the AF relays, time reversal and complex conjugation operations are implemented. However, D_{max} becomes large (as in a typical UWA channel), the effective transmission rate is diminished. That is, in transmitting binary phase shift keying (BPSK) modulated data blocks of length M, the maximum achievable data rate is $\frac{M}{M+D_{max}}$.

In this chapter, we consider an asynchronous cooperative system with two AF relays where only the receiver is aware of the relative time delay between the relays. We propose a new single carrier (SC)-TR-STBC scheme with the objective of adding a smaller overhead to combat the asynchronism issues and increase the effective transmission rate. We focus on the case of a single tap (frequency flat) channel, and utilize superposition of two suffixes generated for the first and second blocks in between the two blocks transmitted. This results in a more efficient way of obtaining the Alamouti structure at the receiver than the existing schemes (e.g. [2]), however

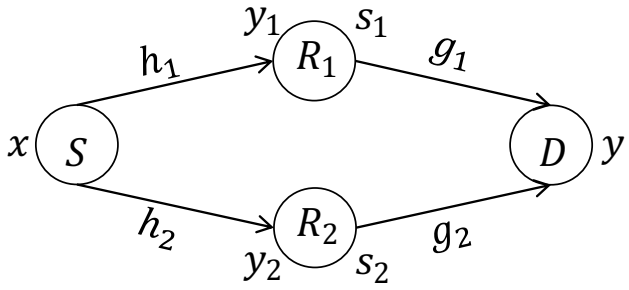

Figure 7.1: Relay channel with two relays.

some symbols get corrupted by intersymbol interference (ISI). The good news is that the interfering symbols are the ones with the Alamouti coding structure without ISI, therefore we can first detect these symbols then use the detected symbols to mitigate ISI and detect the rest of the symbols. To accomplish this goal, at the transmitter, both blocks are transmitted prefix-free and suffix-free, and they are separated by a time guard not less than D_{max}. The relay nodes produce the cyclic suffix for the first and CP for the second block and forward the superposition of the noisy CP and cyclic suffix to the destination while at the same time they perform appropriate time reversal and complex conjugation. The overhead is only D_{max} and as a result in transmitting BPSK modulated data blocks of length M the maximum receivable rate is $\frac{2M}{2M+D_{max}}$.

7.2 System Model

A cooperative relay system with two AF relays, shown in Fig. 7.1 is considered in which the channels from the source to the relays, i.e., h_i, and the relays to the destination, i.e., g_i, are all independent quasi-static Rayleigh fading (constant over two transmitted data blocks), and there is no direct link between the source and the destination. We assume that the relative time delay at the destination between the signals received from R_1 and R_2 is a multiple of the sampling time, and $\boldsymbol{s}_2$ is received D sampling times later than $\boldsymbol{s}_1$. We further assume that the maximum possible length

of the delay is D_{max} sampling times, i.e., $D \in \{-D_{max}, \ldots, 0, \ldots, D_{max}\}$, and D_{max} is known at the transmitter side.

7.3 Proposed Signaling Scheme

In this section, we first introduce the signaling approach at the source node, then the relaying strategy employed by the relay nodes.

7.3.1 Source Node Signaling Approach

Let us consider transmission of $2B$ blocks of length M over a channel with $2D_{max} \leq M$. At the transmitter, a single carrier (SC) transmission is performed for which two consecutive blocks, $\boldsymbol{x}_{2b+1}$ and $\boldsymbol{x}_{2b+2}$ ($b = \{0, \cdots, B-1\}$), are separated by a time guard of length M_p ($M_p \geq D_{max}$). Without loss of generality we focus on transmission of two blocks $\boldsymbol{x}_1 = [x_{1,0}, \cdots, x_{1,M-1}]^T$ and $\boldsymbol{x}_2 = [x_{2,0}, \cdots, x_{2,M-1}]^T$, therefore the objective is transmission of $\boldsymbol{x} = [\boldsymbol{x}_1^T, \boldsymbol{x}_2^T]^T$. However, the block $\bar{\boldsymbol{x}} = [\boldsymbol{x}_1^T, \boldsymbol{0}_{1\times M_p}, \boldsymbol{x}_2^T]^T$ is transmitted over the channel, where $\boldsymbol{0}_{n\times m}$ denotes the n by m all zero matrix (Instead of zero padding between the blocks, one can transmit the CP of the first block or cyclic suffix of the second block as a time guard, however we focus on zero padding approach in this chapter). The transmitted block $\bar{\boldsymbol{x}}$ can also be written as $\bar{\boldsymbol{x}} = \boldsymbol{Q}\boldsymbol{x}$ with

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{I}_M & \boldsymbol{0}_{M\times M} \\ \boldsymbol{0}_{M_p\times M} & \boldsymbol{0}_{M_p\times M} \\ \boldsymbol{0}_{M\times M} & \boldsymbol{I}_M \end{bmatrix}, \tag{7.1}$$

and $\boldsymbol{I}_M$ denoting the M by M identity matrix. For notational simplicity, we drop the size index of the matrices when there is no confusion.

7.3.2 Relaying Strategy

At the i-th relay, $\boldsymbol{y}_i = [\boldsymbol{y}_{i,1}^T, \boldsymbol{y}_{i,g}^T, \boldsymbol{y}_{i,2}^T]^T = h_i\bar{\boldsymbol{x}} + \boldsymbol{n}_i$ is received in which $\boldsymbol{n}_i$ are independent complex white Gaussian random vectors with zero mean and autocorrelation matrices $\sigma_i^2\boldsymbol{I}_{(2M+M_p)}$, $\boldsymbol{y}_{i,j} = h_i\boldsymbol{x}_j + \boldsymbol{n}_{i,j}$, $\boldsymbol{y}_{i,g} = \boldsymbol{n}_{i,g}$ and h_i are independent

Table 7.1: Relaying strategy of the relay nodes.

	First block $\boldsymbol{s}_{i,1}$	Infix block $\boldsymbol{s}_{i,g}$	Second block $\boldsymbol{s}_{i,2}$
R_1	$\boldsymbol{y}_{1,1}$	$\boldsymbol{y}_{1,1}[0:M_p-1]+\boldsymbol{y}_{1,2}[M-M_p:M-1]$	$\boldsymbol{y}_{1,2}$
R_2	$\zeta\left(\boldsymbol{y}_{2,2}^*\right)$	$-\zeta\left(\boldsymbol{y}_{2,1}^*[0:M_p-1]\right)+\zeta\left(\boldsymbol{y}_{2,2}^*[M-M_p:M]\right)$	$-\zeta\left(\boldsymbol{y}_{2,1}^*\right)$

circularly symmetric complex Gaussian random variables with zero mean and variance 0.5 per dimension. Then, based on the relaying strategy given in Table 7.1, $\boldsymbol{s}_i=[\boldsymbol{s}_{i,1}^T,\boldsymbol{s}_{i,g}^T,s_{i,2}^T]^T$ of length $2M+M_p$ is forwarded to the destination in which $\zeta(.)$ denotes the time reversal operation, i.e., $\zeta([z_0,\ldots,z_{N-1}])=[z_{N-1},\ldots,z_0]$, and $(.)^*$ denotes the complex conjugation operation. $\boldsymbol{s}_{i,1}$ and $\boldsymbol{s}_{i,2}$ are of length M and $\boldsymbol{s}_{1,g}$ is of length M_p. Note that, the infix block forwarded by the relay nodes corresponds to the noisy CP of the second block and noisy cyclic suffix of the first block.

We can also represent the forwarded signals from the relay nodes by $\boldsymbol{s}_1=\boldsymbol{R}_1\boldsymbol{y}_1$ and $\boldsymbol{s}_2=\boldsymbol{R}_2\boldsymbol{y}_2^*$, where

$$\boldsymbol{R}_1=\left[\begin{array}{c:c:c} \boldsymbol{I}_M & \boldsymbol{0}_{M\times M_p} & \boldsymbol{0} \\ \hdashline \boldsymbol{I}_{M_p}\ \boldsymbol{0} & \boldsymbol{0}_{M_p\times M_p} & \boldsymbol{0}\ \boldsymbol{I}_{M_p} \\ \hdashline \boldsymbol{0} & \boldsymbol{0}_{M\times M_p} & \boldsymbol{I}_M \end{array}\right], \tag{7.2}$$

and

$$\boldsymbol{R}_2=\left[\begin{array}{c:c:c} \boldsymbol{0} & \boldsymbol{0}_{M\times M_p} & \boldsymbol{J}_M \\ \hdashline -\boldsymbol{J}_{M_p}\ \boldsymbol{0} & \boldsymbol{0}_{M_p\times M_p} & \boldsymbol{0}\ \boldsymbol{J}_{M_p} \\ \hdashline -\boldsymbol{J}_M & \boldsymbol{0}_{M\times M_p} & \boldsymbol{0} \end{array}\right], \tag{7.3}$$

with $\boldsymbol{J}_M$ denoting an M by M anti-diagonal identity matrix, i.e., $J_{i,j}=\delta(i+j-M-1)$ $(i,j\in\{1,\ldots,M\}$ and $\delta(.)$ denoting the Kronecker delta function).

7.4 Signal Detection Techniques

In this section, we first provide the signaling structure at the destination of the considered cooperative system then present the optimal ML and sub-optimal detectors for the proposed signaling schemes.

7.4.1 Received Signal at the Destination

Without loss of generality, we focus on $D \geq 0$. At the destination node, for the received signal, we have

$$\begin{aligned}\boldsymbol{y} &= g_1\boldsymbol{s}_1 + g_2\boldsymbol{s}_2^D + \boldsymbol{n}_3 \\ &= g_1\boldsymbol{R}_1(h_1\bar{\boldsymbol{x}} + \boldsymbol{n}_1) + g_2\left(\boldsymbol{R}_2(h_2^*\bar{\boldsymbol{x}}^* + \boldsymbol{n}_2^*)\right)^D + \boldsymbol{n}_3 \\ &= g_1h_1\boldsymbol{R}_1\boldsymbol{Q}\boldsymbol{x} + g_2h_2^*(\boldsymbol{R}_2\boldsymbol{Q})^D\boldsymbol{x}^* + \boldsymbol{n}, \end{aligned} \tag{7.4}$$

in which $\boldsymbol{s}_2^D = [\boldsymbol{0}_{1\times D}, s_{2,0}, \cdots, s_{2,2M+M_p-D-1}]^T$, $\boldsymbol{n}_3$ is white Gaussian random vector with zero mean and autocorrelation matrix $\sigma_3^2\boldsymbol{I}_{(2M+M_p)}$, $\boldsymbol{n} = g_1\boldsymbol{R}_1\boldsymbol{n}_1 + g_2\boldsymbol{R}_2^D\boldsymbol{n}_2^* + \boldsymbol{n}_3$,

$$\boldsymbol{R}_1\boldsymbol{Q} = \left[\begin{array}{cc:cc} \multicolumn{2}{c:}{\boldsymbol{I}_M} & \multicolumn{2}{c}{\boldsymbol{0}_M} \\ \hdashline \boldsymbol{I}_{M_p} & \boldsymbol{0}_{M_p\times(M-M_p)} & \boldsymbol{0}_{M_p\times(M-M_p)} & \boldsymbol{I}_{M_p} \\ \hdashline \multicolumn{2}{c:}{\boldsymbol{0}_M} & \multicolumn{2}{c}{\boldsymbol{I}_M} \end{array}\right],$$

and

$$(\boldsymbol{R}_2\boldsymbol{Q})^D = \left[\begin{array}{cc:cc} \multicolumn{4}{c}{\boldsymbol{0}_{D\times 2M}} \\ \hdashline \multicolumn{2}{c:}{\boldsymbol{0}_M} & \multicolumn{2}{c}{\boldsymbol{J}_M} \\ \hdashline -\boldsymbol{J}_{M_p} & \boldsymbol{0}_{M_p\times(M-M_p)} & \boldsymbol{0}_{M_p\times(M-M_p)} & \boldsymbol{J}_{M_p} \\ \hdashline \boldsymbol{0}_{(M-D)\times D} & -\boldsymbol{J}_{M-D} & \multicolumn{2}{c}{\boldsymbol{0}_{(M-D)\times M}} \end{array}\right].$$

The first D samples in $\boldsymbol{y}$ are corrupted by inter block interference (IBI) and y_j for $j = \{M+D, \ldots, M+M_p-1\}$ are the overhead symbols, therefore we truncate the first D samples and y_{M+D} to y_{M+M_p-1}, and define a new vector as $\boldsymbol{y}' = [\boldsymbol{y}'^T_1, \boldsymbol{y}'^T_2]^T$ with $\boldsymbol{y}'_1 = [y_D, \ldots, y_{M-1}, y_M^*, \ldots, y_{M+D-1}^*]^T$ and

$$\boldsymbol{y}'_2 = [y_{M+M_p}, \ldots, y_{M+M_p+D-1}, y_{M+M_p+D}^*, \ldots, y_{2M+M_p-1}^*]^T.$$

Note that after truncation some information may be lost which can be extracted out by employing a more complex joint detector, however joint detection over more than

two successive blocks is out of the focus of this work. To simplify the derivations, we assume $M_p = D$ from now on by understanding that for $M_p > D$, after truncating y_j for $j = \{M + D, \ldots, M + M_p - 1\}$, the performance analysis would be the same. Furthermore, by defining $C_1 = g_1 h_1$, $C_2 = g_2 h_2^*$ and

$$\boldsymbol{H}' = \left[\begin{array}{c|c|c|cc} \mathbf{0} & C_1\boldsymbol{I}_{(M-D)} & \mathbf{0} & C_2\boldsymbol{J}_{M-D} \\ \hline C_1^*\boldsymbol{I}_D & \mathbf{0} & C_2^*\boldsymbol{J}_D & \mathbf{0} \quad C_1^*\boldsymbol{I}_D \\ \hline -C_2\boldsymbol{J}_D & \mathbf{0} & C_1\boldsymbol{I}_D & \mathbf{0} \quad C_2\boldsymbol{J}_D \\ \hline \mathbf{0} & -C_2^*\boldsymbol{J}_{M-D} & \mathbf{0} & C_1^*\boldsymbol{I}_{M-D} \end{array}\right], \tag{7.5}$$

we can write $\boldsymbol{y}' = \boldsymbol{H}' \begin{bmatrix} \boldsymbol{x}_1' \\ (\boldsymbol{x}_2')^* \end{bmatrix} + \boldsymbol{n}'$, where $\boldsymbol{x}_i' = [x_{i,0}^*, \ldots, x_{i,D-1}^*, x_{i,D}, \ldots, x_{i,M-1}]^T$, and $\boldsymbol{n}'$ is the colored Gaussian noise with the autocorrelation matrix $\boldsymbol{K}_n$ given as

$$\boldsymbol{K}_n = \left[\begin{array}{c|c|c|c|c|c} \sigma^2\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} & \sigma_2^2|g_2|^2\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \sigma^2\boldsymbol{I}_{M-2D} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & (\sigma^2+\sigma_1^2|g_1|^2)\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} & \sigma_1^2|g_1|^2\boldsymbol{I}_D \\ \hline \sigma_2^2|g_2|^2\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} & (\sigma^2+\sigma_2^2|g_2|^2)\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \sigma^2\boldsymbol{I}_{M-2D} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \sigma_1^2|g_1|^2\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} & \sigma^2\boldsymbol{I}_D \end{array}\right]$$

in which $\sigma^2 = |g_1|^2\sigma_1^2 + |g_2|^2\sigma_2^2 + \sigma_3^2$.

7.4.2 Optimal ML Detector

Since $\boldsymbol{n}'$ is colored Gaussian noise, in order to obtain the optimal ML detector, we first need to whiten the noise. To do so, we use the whitening filter $\boldsymbol{K}_n^{-\frac{1}{2}}$, where

$$\boldsymbol{K}_n^{-1} = \frac{1}{\sigma^2}\left[\begin{array}{c|c|c|c|c|c} (\alpha_2^2+\beta_2^2)\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} & -\alpha_2^2\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \boldsymbol{I}_{M-2D} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \beta_1^2\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} & -\alpha_1^2\boldsymbol{I}_D \\ \hline -\alpha_2^2\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} & \beta_2^2\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{I}_{M-2D} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & -\alpha_1^2\boldsymbol{I}_D & \mathbf{0} & \mathbf{0} & (\alpha_1^2+\beta_1^2)\boldsymbol{I}_D \end{array}\right], \tag{7.6}$$

in which $\alpha_i^2 = \frac{\sigma^2|g_i|^2\sigma_i^2}{\sigma^4+\sigma^2\sigma_i^2|g_i|^2-\sigma_i^4|g_i|^4}$ and $\beta_i^2 = \frac{\sigma^4}{\sigma^4+\sigma^2\sigma_i^2|g_i|^2-\sigma_i^4|g_i|^4}$. Employing the whitening filter $\boldsymbol{K}_n^{-\frac{1}{2}}$ yields $\boldsymbol{r} = \boldsymbol{K}_n^{-\frac{1}{2}}\boldsymbol{y}' = \boldsymbol{K}_n^{-\frac{1}{2}}\boldsymbol{H}'\boldsymbol{x}' + \boldsymbol{w}$, with $\boldsymbol{w}$ denoting a white Gaussian noise vector with zero mean and autocorrelation matrix $\boldsymbol{I}_{2M}$. Therefore, the ML detector is given by

$$\begin{aligned}\hat{\boldsymbol{x}'} &= \underset{\boldsymbol{x}'}{\arg\max}\, \mathrm{Re}\left\{\boldsymbol{r}^H\boldsymbol{K}_n^{-\frac{1}{2}}\boldsymbol{H}'\boldsymbol{x}'\right\} - \frac{1}{2}\eta_{x'} \\ &\equiv \underset{\boldsymbol{x}'}{\arg\max}\, \mathrm{Re}\left\{(\boldsymbol{H}\boldsymbol{y}')^H\boldsymbol{x}'\right\} - \frac{\sigma^2}{2}\eta_{x'}, \end{aligned} \tag{7.7}$$

with $\eta_{x'} = \boldsymbol{x}'^H\boldsymbol{H}'^H\boldsymbol{K}_n^{-1}\boldsymbol{H}'\boldsymbol{x}'$ and

$$\boldsymbol{H} = \sigma^2\boldsymbol{H}'^H\boldsymbol{K}_n^{-1}$$

$$= \left[\begin{array}{c|c|c|c|c|c} C_2^*\alpha_2^2\boldsymbol{J}_D & \boldsymbol{0} & C_1\beta_1^2\boldsymbol{I}_D & -C_2^*\beta_2^2\boldsymbol{J}_D & \boldsymbol{0} & -C_1\alpha_1^2\boldsymbol{I}_D \\ \hline C_1^*(\alpha_2^2+\beta_2^2)\boldsymbol{I}_D & \boldsymbol{0} & C_2\alpha_1^2\boldsymbol{J}_D & -C_1^*\alpha_2^2\boldsymbol{I}_D & \boldsymbol{0} & -C_2(\alpha_1^2+\beta_1^2)\boldsymbol{J}_D \\ \hline \boldsymbol{0} & C_1^*\boldsymbol{I}_{M-2D} & \boldsymbol{0} & \boldsymbol{0} & -C_2\boldsymbol{J}_{M-2D} & \boldsymbol{0} \\ \hline -C_1^*\alpha_2^2\boldsymbol{I}_D & \boldsymbol{0} & C_2\beta_1^2\boldsymbol{J}_D & C_1^*\beta_2^2\boldsymbol{I}_D & \boldsymbol{0} & -C_2\alpha_1^2\boldsymbol{J}_D \\ \hline \boldsymbol{0} & C_2^*\boldsymbol{J}_{M-2D} & \boldsymbol{0} & \boldsymbol{0} & C_1\boldsymbol{I}_{M-2D} & \boldsymbol{0} \\ \hline C_2^*\beta_2^2\boldsymbol{J}_D & \boldsymbol{0} & C_1(\beta_1^2-\alpha_1^2)\boldsymbol{I}_D & C_2^*(\beta_2^2-\alpha_2^2)\boldsymbol{J}_D & \boldsymbol{0} & C_1\beta_1^2\boldsymbol{I}_D \end{array}\right].$$

Due to the structure of $\boldsymbol{H}$, defining $\boldsymbol{y}_k = [y_{2D-k-1}, y_{M+k}^*, y_{M+2D-k-1}, y_{2M+k}^*]^T$ and $\boldsymbol{x}_k = [x_{1,k}^*, x_{1,2D-k-1}, x_{2,D-k-1}, x_{2,M-D+k}^*]^T$ leads to

$$\begin{aligned}\hat{\boldsymbol{x}'} = \underset{\boldsymbol{x}'}{\arg\max}\, \mathrm{Re}\Bigg\{ &\sum_{j=0}^{M-2D-1}\left(\boldsymbol{H}_1\begin{bmatrix} y_{2D+j} \\ y_{2M-j-1}\end{bmatrix}\right)^H\begin{bmatrix} x_{1,j+2D} \\ x_{2,M-j-D-1}^*\end{bmatrix} \\ &+ \sum_{k=0}^{D-1}\left(\boldsymbol{H}_2\boldsymbol{y}_k\right)^H\boldsymbol{x}_k\Bigg\} - \frac{\sigma^2}{2}\eta_{x'}, \end{aligned} \tag{7.8}$$

where $\boldsymbol{H}_1 = \begin{bmatrix} C_1^* & -C_2 \\ C_2^* & C_1\end{bmatrix}$ and

$$\boldsymbol{H}_2 = \begin{bmatrix} C_2^*\alpha_2^2 & C_1\beta_1^2 & -C_2^*\beta_2^2 & -C_1\alpha_1^2 \\ C_1^*(\alpha_2^2+\beta_2^2) & C_2\alpha_1^2 & -C_1^*\alpha_2^2 & -C_2(\alpha_1^2+\beta_1^2) \\ -C_1^*\alpha_2^2 & C_2\beta_1^2 & C_1^*\beta_2^2 & -C_2\alpha_1^2 \\ C_2^*\beta_2^2 & C_1(\beta_1^2-\alpha_1^2) & C_2^*(\beta_2^2-\alpha_2^2) & C_1\beta_1^2 \end{bmatrix}.$$

The above ML detection criterion can be split in several independent detection criteria. More precisely $[x_{1,j+2D}, x_{2,M-j-D-1}]^T$ for $j \in \{0,\ldots,M-2D-1\}$ and $\boldsymbol{x}_k$ for $k \in \{0,\cdots,D-1\}$, can be detected separately. For $j \in \{0,\ldots,M-2D-1\}$, the optimal ML detector yields to

$$\begin{aligned}\begin{bmatrix} \hat{x}_{1,j+2D} \\ \hat{x}_{2,M-j-D-1} \end{bmatrix} = &\underset{[x_1,x_2]^T}{\operatorname{argmax}} \operatorname{Re}\left\{ \left(\boldsymbol{H}_1 \begin{bmatrix} y_{2D+j} \\ y_{2M-j-1} \end{bmatrix} \right)^H \begin{bmatrix} x_1 \\ x_2^* \end{bmatrix} \right\} \\ &- \frac{\sigma^2}{2}(|C_1|^2+|C_2|^2)(|x_1|^2+|x_2|^2), \end{aligned} \tag{7.9}$$

which represents the Alamouti structure and as a result the system achieves full transmit diversity in detecting $[x_{1,j+2D}, x_{2,M-j-D-1}]^T$ ($j \in \{0,\ldots,M-2D-1\}$) independent of the delay value D. Furthermore, for $k \in \{0,\ldots,D-1\}$, the optimal detector is obtained as

$$\hat{\boldsymbol{x}}_k = \arg\max_{\boldsymbol{x}_k} \operatorname{Re}\left\{ (\boldsymbol{H}_2 \boldsymbol{y}_k)^H \boldsymbol{x}_k \right\} - \frac{\sigma^2}{2} \boldsymbol{x}_k^H \boldsymbol{H}_3 \boldsymbol{x}_k, \tag{7.10}$$

where $\boldsymbol{H}_3 = [\boldsymbol{H}_{3,1}, \boldsymbol{H}_{3,2}]$ with

$$\boldsymbol{H}_{3,1} = \left[\begin{array}{c:c} \beta_1^2|C_1|^2+\beta_2^2|C_2|^2 & C_1C_2^*(\alpha_1^2+\alpha_2^2) \\ (\alpha_1^2+\alpha_2^2)C_1^*C_2 & |C_1|^2(\alpha_2^2+\beta_2^2)+|C_2|^2(\alpha_1^2+\beta_1^2) \\ C_1^*C_2(\beta_1^2-\beta_2^2) & -\alpha_2^2|C_1|^2+\alpha_1^2|C_2|^2 \\ (\beta_1^2-\alpha_1^2)|C_1|^2-(\beta_2^2-\alpha_2^2)|C_2|^2 & C_1C_2^*(\beta_2^2-\beta_1^2) \end{array}\right],$$

and

$$\boldsymbol{H}_{3,2} = \left[\begin{array}{c:c} C_1C_2^*(\beta_1^2-\beta_2^2) & (\beta_1^2-\alpha_1^2)|C_1|^2-(\beta_2^2-\alpha_2^2)|C_2|^2 \\ \alpha_1^2|C_2|^2-\alpha_2^2|C_1|^2 & C_1^*C_2(\beta_2^2-\beta_1^2) \\ |C_1|^2\beta_2^2+|C_2|^2\beta_1^2 & C_1^*C_2(\beta_1^2-\alpha_1^2+\beta_2^2-\alpha_2^2) \\ C_1C_2^*(\beta_1^2-\alpha_1^2+\beta_2^2-\alpha_2^2) & |C_1|^2(2\beta_1^2-\alpha_1^2)+|C_2|^2(2\beta_2^2-\alpha_2^2) \end{array}\right].$$

Through numerical examples, in Section 7.5, we show that the optimal detector achieves full diversity in detection of all the symbols.

7.4.3 Sub-Optimal Detector

As explained in Section 7.4.2, to obtain the optimal detector we first need to whiten the noise which results in (7.9) and (7.10). In this section, we propose a sub-optimal detector with a smaller computational complexity in comparison with the optimal detector. We verify the performance of the proposed sub-optimal detector in achieving full diversity by numerical examples in Section 7.5. By considering only the first and last $M-D$ samples of $\boldsymbol{y}'$, we obtain ($j \in \{0,\ldots,M-D-1\}$)

$$\begin{bmatrix} y_{D+j} \\ y^*_{2M+D-j-1} \end{bmatrix} = \boldsymbol{H}_1^H \begin{bmatrix} x_{1,j+D} \\ x^*_{2,M-j-1} \end{bmatrix} + \begin{bmatrix} n_{D+j} \\ n_{2M+D-j-1} \end{bmatrix},$$

where n_{D+j} and $n_{2M+D-j-1}$ are i.i.d. complex Gaussian noise with zero mean and variance of σ^2. The above expression shows the Alamouti structure for $[x_{1,j+D}, x_{2,M-j-1}]^T$ ($j \in \{0,\ldots,M-D-1\}$), therefore we can employ fast ML detection to obtain the estimate $[\hat{x}_{1,j+D}, \hat{x}_{2,M-j-1}]^T$ for $j \in \{0,\ldots,M-D-1\}$. Note that by employing the optimal detector we arrive at the fast ML detection for $[x_{1,j+D}, x_{2,M-j-1}]^T$ ($j \in \{D,\ldots,M-2D-1\}$). Furthermore, for the remaining $2D$ symbols, we have ($k \in \{0,\ldots,D-1\}$)

$$\begin{bmatrix} y^*_{M+k} \\ y_{M+2D-k-1} \end{bmatrix} = \boldsymbol{H}_1^T \begin{bmatrix} x^*_{1,k} \\ x_{2,D-k-1} \end{bmatrix} + \begin{bmatrix} C_1^* \\ C_2 \end{bmatrix} x^*_{2,M-D+k} + \begin{bmatrix} n^*_{M+k} \\ n_{M+2D-k-1} \end{bmatrix}$$

where n_{M+k} and $n_{M+2D-k-1}$ are independent complex Gaussian noise with zero mean and variances of $\sigma^2+\sigma_1^2|g_1|^2$ and $\sigma^2+\sigma_2^2|g_2|^2$, respectively. If we employ the detected symbol $\hat{x}_{2,M-D+k}$ to mitigate the ISI effect of $x_{2,M-D+k}$, we obtain

$$\begin{bmatrix} y^*_{M+k} \\ y_{M+2D-k-1} \end{bmatrix} - \begin{bmatrix} C_1^* \\ C_2^* \end{bmatrix} \hat{x}^*_{2,M-D+k}$$

$$= \boldsymbol{H}_1^T \begin{bmatrix} x^*_{1,k} \\ x_{2,D-k-1} \end{bmatrix} + \begin{bmatrix} C_1^* \\ C_2 \end{bmatrix} (x_{2,M-D+k} - \hat{x}_{2,M-D+k})^* + \begin{bmatrix} n^*_{M+k} \\ n_{M+2D-k-1} \end{bmatrix},$$

which shows the Alamouti coding structure. Obviously, if $x_{2,M-D+k}$ is detected correctly, ISI is perfectly removed, otherwise decision signal is more corrupted. Therefore we expect that by increasing SNR the sub-optimal detector performance becomes closer to the optimal detector performance.

7.4.4 Comments on Implementation

Let us assume that the source node transmits waveforms of length T seconds. At the first relay, to generate the affix block, the first M_pT seconds and last M_pT seconds of the received signal need to be added together which can be implemented in the analog domain. To do so, R_1 suffers a latency of MT seconds. At the second relay, adding the sign changed version of the first M_pT seconds of the received signal to the last M_pT seconds of the received signal generates the affix block in analog domain, however we still need to do time reversal and complex conjugation. Time reversal introduces a time latency of $2MT$ seconds for which we first need to store the received signal corresponding to the two consecutive blocks along with the generated affix block, then forward the time reversed version of the signal. Complex conjugation may also be implemented in analog domain by first separating the in-phase and quadrature components of the signal then combining them by changing the sign of the quadrature component.

7.5 Simulation Results

In this section, we provide numerical examples to study the proposed scheme performance in combating asynchronism issues in cooperative communication systems. We consider BPSK modulated transmission of the data blocks of length $M = 128$ with an affix of length $M_p = 64$ where D is uniformly distributed over $\{0, \ldots, 64\}$ and h_i and g_i are i.i.d. complex Gaussian random variables with zero mean and unit variance. We also assume that $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \frac{N_0}{2}$ and define the average signal to noise ratio at the receiver as $E\{\frac{2(|g_1|^2|h_1|^2+|g_2|^2|h_2|^2)}{N_0(|g_1|^2+|g_2|^2+1)}\} \approx \frac{1.192694}{N_0}$.

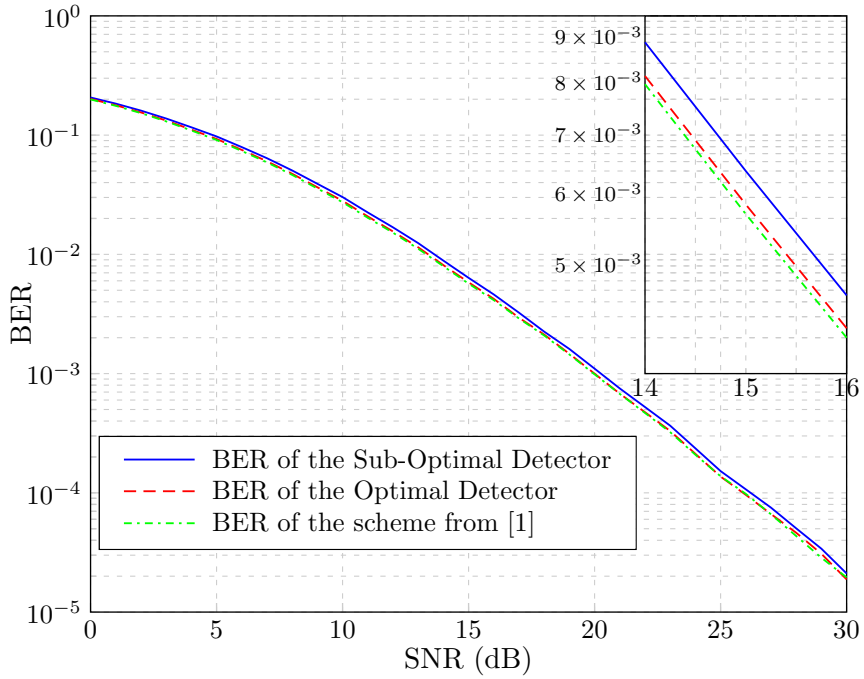

Figure 7.2: BER performance of the proposed scheme for both optimal and sub-optimal detectors and the scheme from [1].

In Fig. 7.2, we compare the BER performance of the proposed scheme for both optimal and sub-optimal detectors with the BER performance of the scheme proposed in [1] in which all the received symbols achieve perfect Alamouti structure similar to (7.9). We observe that the difference between the performance of the optimal detector and the scheme from [1] is less than 0.05 dB. Furthermore, the sub-optimal detector performs very close to the optimal detector and the scheme from [1] with a performance loss around 0.2 dB with respect to the scheme in [1] while the proposed scheme improves the transmission rate by 20%.

7.6 Chapter Summary

We proposed a new TR-STBC scheme useful in achieving asynchronous cooperative diversity under flat fading channel conditions which achieves a higher spectral efficiency than the existing STBC schemes in the literature. More precisely, to combat the asynchronism issues arising from relative delays among the signals received from

different relays, less overhead needs to be added in transmitting every two data blocks in comparison with the existing schemes. We obtained the optimal detector structure and proposed a sub-optimal detector with smaller computational complexity. Numerical examples show that the proposed scheme achieves full diversity for both optimal and sub-optimal detectors. In comparison with the scheme achieving perfect Alamouti structure for all the transmitted symbols, we experience a very small performance loss while providing higher transmission rate.

Chapter 8

Delay Diversity Relaying for Asynchronous Cooperative Communications with Large Relative Delays

In cooperative UWA systems, due to the low speed of sound, a node can experience significant time delays among the signals received from geographically separated nodes. One way to combat the asynchronism issues is to employ OFDM-based transmissions at the source node by preceding every OFDM block with an extremely long CP which reduces the transmission rates dramatically. One may increase the OFDM block length accordingly to compensate for the rate loss which also degrades the performance due to the significantly time-varying nature of the UWA channels. In this paper, we develop a new OFDM-based scheme to combat the asynchronism problem in cooperative UWA systems without adding a long CP (in the order of the long relative delays) at the transmitter in which by adding a much more manageable (short) CP at the source, we obtain a delay diversity structure at the destination for effective processing and exploitation of the spatial diversity by utilizing Viterbi decoder at the destination. We provide pairwise error probability (PEP) analysis of the system for both time-invariant and block fading channels showing that the system achieves full spatial diversity. We find through extensive simulations that the proposed scheme offers a significantly improved error rate performance for time-varying channels (typical in UWA communications) compared to the existing approaches.

The chapter is organized as follows. In Section 8.1, we start with an introduction on asynchronous cooperative UWA systems, motivation behind the work and our proposed OFDM-based signaling solution. In Section 8.2, the system model and the structure of the OFDM signals at the source, relays and destination are presented. The proposed signaling scheme which includes appropriate CP addition at the source and CP removal at the destination is explained in Section 8.3. It is shown that the

proposed scheme gives a delay diversity structure at the destination for large relative delays among the relays. In Section 8.4, the PEP analysis of the system under both quasi-static and block fading channel models is provided. In Section 8.5, the performance of the proposed scheme is evaluated through some numerical examples. Finally, we summarize the chapter in Section 8.6.

8.1 Introduction

Cooperative UWA communications which refers to a group of nodes, known as relays, helping the source to deliver its data to the destination is a promising physical layer solution to improve the performance of UWA systems [63–65]. In a UWA cooperative communication system, the time differences among signals received from geographically separated nodes can be excessive due to the low speed of sound in water. For instance, if the relative distance between two nodes with respect to another one is 500 m, then their transmissions experience a relative delay of 333 ms. Considering, for instance, that in an OFDM-UWA cooperative communication scheme with 512 sub-carriers over a total bandwidth of 8 kHz, the OFDM block duration is only 64 ms, the excessive delay of 333 ms becomes problematic. Furthermore, UWA channels are highly time varying due to the large Doppler spreads and Doppler shift effects (or Doppler scaling) [16]. Therefore, a practical non-centralized UWA cooperative communication system is asynchronous with large relative delays among the nodes under highly time-varying frequency selective channel conditions.

Our focus in this chapter is on asynchronous cooperative UWA communications where only the destination node is aware of the relative delays among the nodes. Existing signaling solutions for asynchronous radio terrestrial cooperative communications rely on quasi-static fading channels with limited delays among signals received from different relays at the destination, e.g., see [17] and references therein, in which every transmitted block (either OFDM transmission or single carrier transmission) is

preceded by a time guard not less than the maximum possible delay among the relays. Therefore, we cannot directly apply them for cooperative UWA communications. Our main objective is to develop new OFDM based signaling solutions to combat asynchronism issues arising from excessive large relative delays without preceding each OFDM block by a large CP in the order of the maximum possible relative delay.

In systems employing OFDM, e.g., [3, 18], the existing solutions are effective when the maximum length of the relative delays among signals received from various nodes are less than the length of an OFDM block which is not a practical assumption for the case of UWA communications. In [18], a space-frequency coding approach is proposed which is proved to achieve both full spatial and full multipath diversities. In [3], OFDM transmission is implemented at the source node and relays only perform time reversal and complex conjugation. A trivial generalization of existing OFDM-based results to compensate for large relative delays may be to increase the OFDM block lengths. The main drawback is that inter carrier interference (ICI) is increased due to the time variations of the UWA channels. Another trivial solution is to increase the length of the CP. This is not an efficient solution either, since it dramatically decreases the spectral efficiency of the system.

There are several single carrier transmission based solutions reported in the literature as well, e.g., [2, 51, 52, 63]. In [63] a time reversal distributed space time block code (DSTBC) is proposed for UWA cooperative communication systems under quasi-static multipath fading channel conditions. In [2], a DSTBC transmission scheme by decode and forward (DF) relaying is proposed which achieves both full spatial and multipath diversities. A distributed space time trellis code with DF relaying is proposed in [51, 52] which under sufficient conditions can achieve full spatial and multipath diversities.

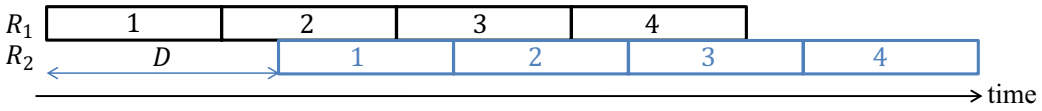

Figure 8.1: The structure of the received OFDM blocks from two different relays of the proposed delay diversity scheme for a relative delay of D seconds.

In this chapter, we focus on OFDM based cooperative UWA communication systems with full-duplex AF relays where all the nodes employ the same frequency band to communicate with the destination. We assume an asynchronous operation and potentially very large delays among different nodes (known only at the destination). We present a new scheme which can compensate for the effect of the long delays among the signals received from different nodes without adding an excessively long CP. We demonstrate that we can extract delay diversity out of the asynchronism among the cooperating nodes. The main idea is to add an appropriate (short) CP (much less than the long relative delays among the relays) to each OFDM block at the transmitter side to combat multipath effects of the channels and obtain a delay diversity structure at the destination. As an illustration, Fig. 8.1 shows the received OFDM block structure of the proposed delay diversity scheme from two different relays with a relative delay of D seconds. In Fig. 8.1, D is in a range that each block relayed through the relay R_1 is overlapped with its preceding block relayed through the relay R_2. E.g., under quasi-static fading scenario, each subcarrier of a received block is a summation of the corresponding subcarriers from two successively transmitted blocks which results into a delay diversity structure [66].

8.2 System and Signal Models

We consider a full-duplex AF relay system with two relays, shown in Fig. 8.2, in which there is no direct link between source (S) and destination (D), and the relays help the source deliver its data to the destination by using the AF method. No power allocation strategy is employed at the relay nodes and they use fixed power amplification factors.

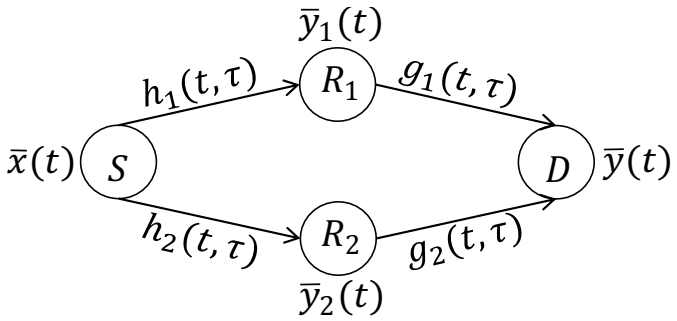

Figure 8.2: Relay channel with two relays.

Note that the model can be generalized to a system with arbitrary number of relays and a direct link between source and the destination, and optimal power allocation can be used in a straightforward manner. We assume that the channels from the source to the relays and the relays to the destination are time-varying multipath channels where $h_i(t,\tau)$ and $g_i(t,\tau)$ represent the source to the i-th relay and the i-th relay to the destination channel responses at time t to an impulse applied at time $t-\tau$, respectively.

At the transmitter, we employ a conventional OFDM transmission technique with N subcarriers over a total bandwidth of B Hz. We consider successive transmission of M data blocks of length N symbols. In discrete baseband signaling form, the m-th ($m \in \{1,\ldots,M\}$) data vector (in time) is denoted by $\boldsymbol{X}^m = [X_0^m,\ldots,X_{N-1}^m]^T$ and the samples of the m-th transmitted OFDM block are represented by $\boldsymbol{x}^m = \text{IFFT}(\boldsymbol{X}^m) = [x_0^m,\ldots,x_{N-1}^m]^T$, where $(.)^T$ denotes the transpose operation. Therefore, we have

$$x_n^m = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k^m e^{j\frac{2\pi k}{N}n}. \tag{8.1}$$

After adding a CP of length N_{CP} to $\boldsymbol{x}^m$, the CP-assisted transmission block $\bar{\boldsymbol{x}}^m$ results. By digital to analog (D/A) conversion of $\bar{\boldsymbol{x}}^m$ with sampling period of $T_s = \dfrac{1}{B}$ seconds, we obtain the continuous time signal $\bar{x}^m(t)$ with time duration of $T =$

$(N + N_{CP})T_s$ seconds which can be written as

$$\bar{x}^m(t) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X_k^m e^{j\frac{2\pi k}{NT_s}t} R(t), \tag{8.2}$$

where $x_n^m = \bar{x}^m(nT_s)$, $R(t) = u(t + N_{CP}T_s) - u(t - NT_s)$ and $u(t)$ denotes the unit step function. Furthermore, for the continuous time transmitted signal $\bar{x}(t)$, we can write $\bar{x}(t) = \sum_{m=1}^{M} \bar{x}^m(t - (m-1)T)$.

At the i-th relay ($i \in \{1, 2\}$), the signal $\bar{y}_i(t)$ is received, hence the part of $\bar{y}_i(t)$ corresponding to the m-th transmitted block, i.e., $\bar{y}_i^m(t) = \bar{y}_i(t + (m-1)T)R(t)$, can be written as

$$\bar{y}_i^m(t) = \int_{-\infty}^{\infty} \bar{x}^m(t-\tau) h_i^m(t,\tau) d\tau + \underbrace{\sum_{m' \neq m} \int_{-\infty}^{\infty} \bar{x}^{m'}(t - (m'-m)T - \tau) h_i^m(t,\tau) d\tau}_{\text{ISI}} + z_{1,i}^m(t), \tag{8.3}$$

where $z_{1,i}^m(t) = z_{1,i}(t + (m-1)T)R(t)$, $h_i^m(t,\tau) = h_i(t + (m-1)T, \tau)R(t)$ and $z_{1,i}^m(t)$ are independent complex Gaussian random processes with zero mean and power spectral density (PSD) of $\sigma_{1,i}^2$. By taking only the resolvable paths into account, we can write $h_i(t,\tau) = \sum_{l=1}^{L_{h_i}} h_{i,l}(t)\delta(\tau - \tau_{h_i,l})$, where L_{h_i} denotes the number of resolvable paths from the source to the i-th relay, $h_{i,l}(t)$ are independent zero-mean (for different i and l) complex Gaussian wide-sense stationary (WSS) processes with a total envelope power of $\sigma_{h_i,l}^2$ (i.e., independent time-varying Rayleigh fading channel tap gains) assuming that $\sum_{l=1}^{L_{h_i}} \sigma_{h_i,l}^2 = 1$, and $\tau_{h_i,l} \geq 0$ denotes the delay of the l-th resolvable path from the source to the i-th relay. Assuming $\tau_{h_i,L_{h_i}} \leq N_{CP}T_s$, i.e., the length of the CP overhead is greater than the delay spread of the channel (the main job of the CP to guarantee robustness against multipath), and defining $I_{1,i}^m(t) = \sum_{l=1}^{L_{h_i}} h_{i,l}^m(t)\bar{x}^{m-1}(t + T - \tau_{h_i,l})$, we can rewrite (8.3) as

$$\bar{y}_i^m(t) = \sum_{l=1}^{L_{h_i}} h_{i,l}^m(t)\bar{x}^m(t - \tau_{h_i,l}) + I_{1,i}^m(t) + z_{1,i}^m(t). \tag{8.4}$$

We assume that the signal passing through the second relay is received D seconds later than the signal passing through the first relay and we also assume $\tau_{h_i,1} = \tau_{g_i,1} = 0$ for $i \in \{1, 2\}$, i.e., the delay spread of the channel h_i (g_i) is $\tau_{h_i,L_{h_i}}$ ($\tau_{g_i,L_{g_i}}$). Therefore, by denoting the amplification factor of the i-th relay by $\sqrt{P_i}$, for the received signal at the destination $\bar{y}(t)$, we have

$$\bar{y}(t) = \int_{-\infty}^{\infty} \sqrt{P_1}\bar{y}_1(t-\tau)g_1(t,\tau)d\tau + \int_{-\infty}^{\infty} \sqrt{P_2}\bar{y}_2(t-D-\tau)g_2(t,\tau)d\tau + z_2(t), \tag{8.5}$$

where $z_2(t)$ is a Gaussian random processes with zero mean and PSD of σ_2^2. By employing $g_i(t,\tau) = \sum_{l=1}^{L_{g_i}} g_{i,l}(t)\delta(\tau - \tau_{g_i,l})$ in (8.5), we obtain

$$\bar{y}(t) = \sum_{l=1}^{L_{g_1}} \sqrt{P_1}g_{1,l}(t)\bar{y}_1(t-\tau_{g_1,l}) + \sum_{l=1}^{L_{g_2}} \sqrt{P_2}g_{2,l}(t)\bar{y}_2(t-\tau_{g_2,l}-D) + z_2(t).$$

Defining $z(t) = z_2(t) + \sum_{l=1}^{L_{g_1}} \sqrt{P_1}g_{1,l}(t)z_{1,1}(t-\tau_{g_1,l}) + \sum_{l=1}^{L_{g_2}} \sqrt{P_2}g_{2,l}(t)z_{1,2}(t-\tau_{g_2,l})$ which represents a Guassian random process conditioned on known $g_{i,l}(t)$ for all i and l, we can write

$$\begin{aligned}\bar{y}(t) =& \sum_{l=1}^{L_{g_1}} \sqrt{P_1}g_{1,l}(t) \sum_{q=1}^{L_{h_1}} h_{1,q}(t-\tau_{g_1,l})\bar{x}(t-\tau_{g_1,l}-\tau_{h_1,q}) \\ &+ \sum_{l=1}^{L_{g_2}} \sqrt{P_2}g_{2,l}(t) \sum_{q=1}^{L_{h_2}} h_{2,q}(t-D-\tau_{g_2,l})\bar{x}(t-D-\tau_{g_2,l}-\tau_{h_2,q}) + z(t).\end{aligned}$$

Note that without conditioning on $g_{i,l}(t)$, $z(t)$ represents a complex random process with zero mean and PSD of $\sigma^2 = P_1\sigma_{1,1}^2 + P_2\sigma_{1,2}^2 + \sigma_2^2$. Therefore, we define the received signal to noise ratio (SNR) as $\frac{P_1+P_2}{\sigma^2}$. We also define $L = \left\lceil \frac{\max_i(\tau_{h_i,L_{h_i}} + \tau_{g_i,L_{g_i}})}{T_s} \right\rceil$, where $\tau_{h_i,L_{h_i}} + \tau_{g_i,L_{g_i}}$ is the delay spread of the overall channel experienced at the destination through R_i.

8.3 Delay Diversity Structure

To achieve a delay diversity structure and overcome ISI at the destination, we need to add an appropriate CP at the source and perform CP removal at the destination.

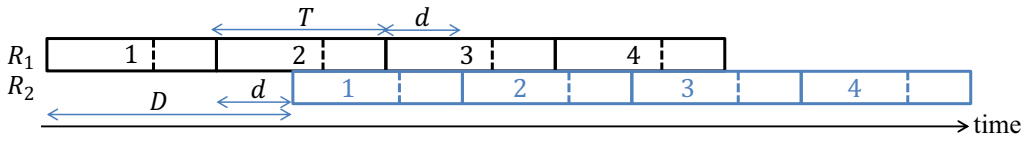

Figure 8.3: The structure of the received signal.

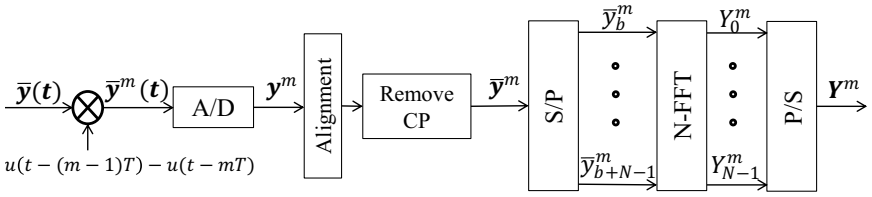

Figure 8.4: The structure of the receiver.

8.3.1 Appropriate CP Length

In a conventional OFDM system, if we have a window of length $(N+L)T_s$ seconds corresponding to one OFDM block, then by removing the first L samples of the considered window and feeding the remaining N samples to the FFT block, the ISI is completely removed. Therefore, in our scheme, to guarantee robustness of the system against ISI, we need to have an overlap of length $(N+L)T_s$ seconds between two blocks received from two different relays at the destination. Fig. 8.3 shows the structure of the received signal at the destination for the case that the blocks relayed by R_2 are received D seconds later than the blocks relayed by R_1, where $T < D < 2T$ and $d = \text{mod}(D, T)$ with $d = \text{mod}(D, T)$ denoting the remainder of division of D by T. To obtain the appropriate overlap structure, we need to have $T - d \geq (N+L)T_s$ or $d \geq (N+L)T_s$ or both which results into $T \geq 2(N+L)T_s$, i.e., $N_{CP} \geq N + 2L$.

8.3.2 Received Signal at the Destination

The baseband signalling structure of the receiver is shown in Fig. 8.4, where $\bar{\boldsymbol{y}}^m = [\bar{y}_0^m, \ldots, \bar{y}_{N+N_{CP}-1}^m]$ denotes the sampled vector of the received signal in the m-th signaling interval and b is the starting point of the m-th FFT window which is decided

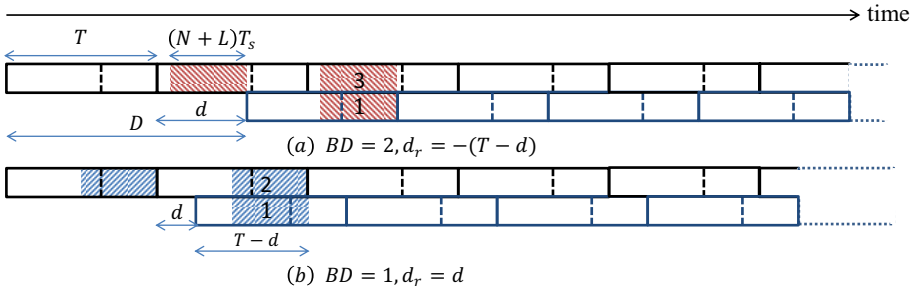

Figure 8.5: Example of different situations for BD and d_r.

by the destination based on the delay value D. Since $N_{CP} \geq N + 2L$, by defining $\bar{y}^m(t) = \bar{y}(t + (m-1)T)R(t)$, we can write

$$\begin{aligned}\bar{y}^m(t) = &\sum_{l=1}^{L_{g_1}} \sqrt{P_1} g_{1,l}^m(t) \sum_{q=1}^{L_{h_1}} h_{1,q}^m(t - \tau_{g_1,l}) \bar{x}^m(t - \tau_{g_1,l} - \tau_{h_1,q}) + I_1^m(t) + I_2^m(t) \\ &+ \sum_{l=1}^{L_{g_2}} \sqrt{P_2} g_{2,l}^m(t) \sum_{q=1}^{L_{h_2}} h_{2,q}^{m-BD}(t - d_r - \tau_{g_2,l}) \bar{x}^{m-BD}(t - d_r - \tau_{g_2,l} - \tau_{h_2,q}) + z^m(t),\end{aligned}$$

where $I_1^m(t) = \sum_{l=1}^{L_{g_1}} \sqrt{P_1} g_{1,l}^m(t) \sum_{q=1}^{L_{h_1}} h_{1,q}^m(t - \tau_{g_1,l}) \bar{x}^{m-1}(t + T - \tau_{g_1,l} - \tau_{h_1,q})$ and

$$\begin{aligned}I_2^m(t) \quad = \quad &\sum_{l=1}^{L_{g_2}} \sqrt{P_2} g_{2,l}^m(t) \sum_{q=1}^{L_{h_2}} h_{2,q}^{m-BD}(t - d_r - \tau_{g_2,l}) \bar{x}^{m-BD-1}(t + T - d_r - \tau_{g_2,l} - \tau_{h_2,q}) \\ &+ \sum_{l=1}^{L_{g_2}} \sqrt{P_2} g_{2,l}^m(t) \sum_{q=1}^{L_{h_2}} h_{2,q}^{m-BD}(t - d_r - \tau_{g_2,l}) \bar{x}^{m-BD+1}(t - T - d_r - \tau_{g_2,l} - \tau_{h_2,q})\end{aligned}$$

represent the ISI, and BD and d_r, as shown in Fig. 8.5, denote the effective OFDM block delay and effective residual delay observed at the destination, respectively. For BD and d_r, we have

$$BD = \begin{cases} \lfloor \frac{D}{T} \rfloor & , \quad d \leq (N+L)T_s \\ \lceil \frac{D}{T} \rceil & , \quad d > (N+L)T_s \end{cases}, \tag{8.6}$$

and

$$d_r = \begin{cases} d & , \quad d \leq (N+L)T_S \\ d - T & , \quad d > (N+L)T_s \end{cases}, \tag{8.7}$$

respectively (note that when $m - BD < 0$, $\bar{x}^{m-BD}(t) = 0$ for all values of t). More precisely, BD represents the number of block delays between two received OFDM blocks which have at least an overlap of length $(N + L)T_s$ seconds (necessary to combat the ISI). As discussed in Section 8.3.1, by choosing $N_{CP} \geq N + 2L$, achieving the appropriate overlap between the received OFDM blocks is guaranteed. By appropriate CP removal (whose details are explained in Section 8.3.3), $\boldsymbol{y}^m$ is obtained as $\boldsymbol{y}^m = [\bar{y}^m(bT_s), \ldots, \bar{y}^m((b+N-1)T_s)]$. By taking FFT of $\boldsymbol{y}^m$, we have $\boldsymbol{Y}^m = [Y_0^m, \ldots, Y_{N-1}^m] = \text{FFT}(\boldsymbol{y}^m)$, i.e.,

$$
\begin{aligned}
Y_k^m &= \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} y_n^m e^{-j\frac{2\pi n}{N}k} \\
&= \frac{1}{\sqrt{N}} \sum_{n=b}^{b+N-1} \bar{y}^m(nT_s) e^{-j\frac{2\pi n}{N}k} \\
&= \frac{1}{\sqrt{N}} \sum_{n=b}^{b+N-1} \Bigg[\sum_{l=1}^{L_{g_1}} \sqrt{P_1} g_{1,l}^m(nT_s) \sum_{q=1}^{L_{h_1}} h_{1,q}^m(nT_s - \tau_{g_1,l}) \bar{x}^m(nT_s - \tau_{h_1,q} - \tau_{g_1,l}) + z^m(nT_s) \\
&+ \sum_{l=1}^{L_{g_2}} \sqrt{P_2} g_{2,l}^m(nT_s) \sum_{q=1}^{L_{h_2}} h_{2,q}^{m-BD}(nT_s - d_r - \tau_{g_2,l}) \bar{x}^{m-BD}(nT_s - d_r - \tau_{h_2,q} - \tau_{g_2,l}) \Bigg] e^{-j\frac{2\pi n}{N}k}.
\end{aligned}
$$

Furthermore, by substituting the result of (8.1) in the above equation, we obtain

$$
Y_k^m = \boldsymbol{GH}_1^m(k) \boldsymbol{X}^m + \boldsymbol{GH}_2^{m-BD}(k) \boldsymbol{X}^{m-BD} + Z_k^m, \tag{8.8}
$$

where $\boldsymbol{GH}_i^m(k) = [GH_i^m[k,0], \cdots, GH_i^m[k,N-1]]$ with

$$
GH_1^m[k,k'] = \frac{\sqrt{P_1}}{N} \sum_{n=b}^{b+N-1} \sum_{l=1}^{L_{g_1}} g_{1,l}^m(nT_s) \sum_{q=1}^{L_{h_1}} h_{1,q}^m(nT_s - \tau_{g_1,l}) e^{j\frac{2\pi n}{N}(k'-k)} e^{-j\frac{2\pi k'}{NT_s}(\tau_{g_1,l} + \tau_{h_1,q})},
$$

$$
\begin{aligned}
GH_2^{m-BD}[k,k'] = &\frac{\sqrt{P_2}}{N} \sum_{n=b}^{b+N-1} \sum_{l=1}^{L_{g_2}} g_{2,l}^m(nT_s) \times \\
&\times \sum_{q=1}^{L_{h_2}} h_{2,q}^{m-BD}(nT_s - d_r - \tau_{g_2,l}) e^{j\frac{2\pi}{NT_s}\left[n(k'-k)T_s - k'(d_r + \tau_{g_2,l} + \tau_{h_2,q})\right]},
\end{aligned}
$$

and $Z_k^m = \frac{1}{\sqrt{N}} \sum_{n=b}^{b+N-1} z^m(nT_s) e^{-\frac{j2\pi n}{N}k}$ conditioned on channel state information are complex Gaussian random variables with zero mean. Hence, by defining $\boldsymbol{X}^m = \boldsymbol{0}_N$

for $m < 1$ and $m > M$ and $\boldsymbol{X}^m = [X_0^m, X_1^m, \cdots, X_{N-1}^m]^T$ for $1 \leq m \leq M$, we can write

$$\boldsymbol{Y}^m = \boldsymbol{GH}_1^m \boldsymbol{X}^m + \boldsymbol{GH}_2^{m-BD} \boldsymbol{X}^{m-BD} + \boldsymbol{Z}^m, \tag{8.9}$$

where $\boldsymbol{GH}_i^m = \left[\boldsymbol{GH}_i^m(0)^T, \ldots, \boldsymbol{GH}_i^m(N-1)^T\right]^T$. In fact, $\boldsymbol{GH}_i^m$ represents the effective $S-R_i-D$ channel seen by the destination which depends on both $S-R_i-D$ channel and the position of the FFT window.

8.3.3 Appropriate CP Removal at the Destination

To take FFT at the destination, we need to choose the FFT window by appropriate CP removal. Since the received OFDM blocks are not synchronized, we align the receiver FFT window with one of the relays. By precise alignment, an overlap of length $(N+L)T_s$ seconds between the OFDM blocks received through R_1 and R_2 can be achieved which is determined with the value of d. Note that an overlap of at least $N+L$ samples is necessary to guarantee robustness of the transmission against ISI. As shown in Fig. 8.6, for $d \geq (N+L)T_s$, the receiver FFT window is aligned with R_2 and for $d < (N+L)T_s$ it is aligned with R_1. The only effect of unaligned FFT windowing in time at the destination, as long as appropriate CP removal is done, is phase shift at the frequency domain covered in the definition of $\boldsymbol{GH}_i^m$.

8.3.4 Detection by Viterbi Algorithm

For the time-invariant channel scenario the noise samples Z_k^m are independent complex Gaussian random variables for all m and k and i.i.d. for any specific k. Therefore, for time-invariant channel conditions, N parallel Viterbi detectors with M^{BD} states (assuming M-PSK modulation) can be employed for ML detection of the transmitted symbols, where the k-th Viterbi detector gets $\boldsymbol{Y}_k$ as input to detect the transmitted symbols over the k-th subcarrier. On the other hand, for the time-varying channel scenarios, the received noise samples at each OFDM block are dependent complex Gaussian random variables conditioned on known channel state information.

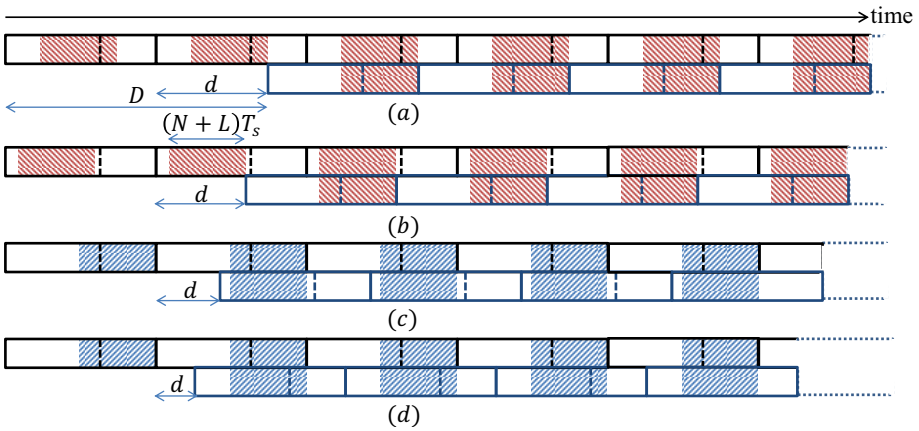

Figure 8.6: Different possible FFT windowings for different ranges of d (a) $d \geq (N+2L)T_s$, (b) $(N+L)T_s \leq d < (N+2L)T_s$, (c) $NT_s < d < (N+L)T_s$, and (d) $d \leq NT_s$.

The complexity of the Viterbi algorithm for the time varying case is prohibitive due to the ICI effects. On the other hand, the noise samples corresponding to different FFT windows at the destination are independent but not necessarily identically distributed. However, we implement a suboptimal detector in which we ignore the ICI effects and assume that Z_k^m are i.i.d. for any given subcarrier k, and employ the same structure as in the time-invariant case. Hence, $\boldsymbol{Y}'_k = [Y_k'^1, \ldots, Y_k'^M]$ is given to the k-th Viterbi decoder where $[Y_0'^m, \ldots, Y_{N-1}'^m] = diag(\boldsymbol{Y}_m)$ and $diag(\boldsymbol{A})$, with $\boldsymbol{A}$ being a square matrix, denotes a vector of diagonal elements of $\boldsymbol{A}$.

8.4 PEP Analysis

Design of the space-time codes is out of focus of this work; however, we present the PEP performance analysis of the system under quasi-static frequency selective and block fading frequency selective channels in this section which can be useful in diversity order analysis of the proposed scheme and future space-time code designs. In the following, we present the PEP analysis for the quasi-static and block fading frequency selective channels, respectively.

8.4.1 Quasi-Static Frequency-Selective Channels

In this section, we consider the PEP performance analysis for ML detection presented in Section 8.3.4. We provide the result under the condition that the channels from the source to the relays have significantly higher SNRs than the channels from the relays to the destination, i.e., $\frac{1}{\sigma_{1,i}^2} \gg \frac{P_i}{\sigma_2^2}$. We assume that the channels are quasi-static Rayleigh fading, i.e., the channel gains in time domain are random variables but fixed for the transmission of M consecutive OFDM blocks. We denote $h_{i,l}^m(nT_s) = h_{i,l}$ and $g_{i,l}^m(nT_s) = g_{i,l}$ for $n = \{0, \ldots, N-1\}$ and $m = \{1, \ldots, M\}$, where $h_{i,l}$ and $g_{i,l}$ are zero mean circularly symmetric complex Gaussian random variables with variances of $\sigma_{h_i,l}^2$ and $\sigma_{g_i,l}^2$, respectively, with $\sum_{l=1}^{L_{h_i}} \sigma_{h_i,l}^2 = 1$ and $\sum_{l=1}^{L_{g_i}} \sigma_{g_i,l}^2 = 1$. Therefore, we can write

$$GH_1[k,k'] = \sqrt{P_1}\left[\sum_{l=1}^{L_{g_1}} g_{1,l} e^{-2\pi j \frac{k'\tau_{g_1,l}}{NT_s}}\right]\left[\sum_{q=1}^{L_{h_1}} h_{1,q} e^{-2\pi j \frac{k'\tau_{h_1,q}}{NT_s}}\right]\delta(k-k'), \tag{8.10}$$

and

$$GH_2[k,k'] = \sqrt{P_2}\left[\sum_{l=1}^{L_{g_2}} g_{2,l} e^{-2\pi j \frac{k'\tau_{g_2,l}}{NT_s}}\right]\left[\sum_{q=1}^{L_{h_2}} h_{2,q} e^{-2\pi j \frac{k'\tau_{h_2,q}}{NT_s}}\right] e^{-2\pi j \frac{k'}{NT_s} d_r}\delta(k-k'), \tag{8.11}$$

where $\delta(.)$ denotes the Dirac delta function, and for a fixed k, $G_{i,k} = \sum_{l=1}^{L_{g_i}} g_{i,l} e^{-2\pi j \frac{k\tau_{g_i,l}}{NT_s}}$ and $H_{i,k} = \sum_{q=1}^{L_{h_i}} h_{i,q} e^{-2\pi j \frac{k\tau_{h_i,q}}{NT_s}}$ are independent complex Gaussian random variables with zero mean and unit variance. Hence, for the received signal on the k-th subcarrier, we have

$$Y_k^m = GH_1[k,k]X_k^m + GH_2[k,k]X_k^{m-BD} + Z_k^m, \tag{8.12}$$

where conditioned on $g_{i,l}$ for all $l \in \{1, \ldots, L_{g_i}\}$ and $i \in \{1,2\}$, Z_k^m are i.i.d. complex Gaussian random variables with zero mean and variance of $\sigma_2^2 + P_1|G_{1,k}|^2\sigma_{1,1}^2 + P_2|G_{2,k}|^2\sigma_{1,2}^2$. The above relation for $BD > 0$ is a delay diversity structure which

can be used to extract spatial diversity out of the relay system shown by the PEP analysis. If we define $\boldsymbol{Y}_k = [Y_k^1, \ldots, Y_k^{M+BD}]$, $\boldsymbol{Z}_k = [Z_k^1, \ldots, Z_k^{M+BD}]$, $\boldsymbol{GH}(k) = [GH_1[k,k], GH_2[k,k]]$ and

$$\boldsymbol{X}_k = \begin{bmatrix} X_k^1 & \cdots & X_k^{1+BD} & \cdots & X_k^M & \cdots & 0 \\ 0 & \cdots & X_k^1 & \cdots & X_k^{M-BD} & \cdots & X_k^M \end{bmatrix}, \tag{8.13}$$

we can write $\boldsymbol{Y}_k = \boldsymbol{GH}(k)\boldsymbol{X}_k + \boldsymbol{Z}_k$. Note that our focus is on extracting spatial diversity out of the asynchronous cooperative system which is attained in the form of the delay diversity. In fact, we assume no explicit channel coding is employed across different subcarriers and as a result no multipath diversity is attained; however, it does not mean that the system does not achieve multipath diversity. Now, let us focus on a given subcarrier, e.g., k-th subcarrier, where conditioned on given $g_{i,l}$ for $l \in \{1, \ldots, L_{g_i}\}$ and $i \in \{1, 2\}$, $\boldsymbol{Z}_k$ is a complex white Gaussian vector with zero mean and autocorrelation matrix of $\left(\sigma_2^2 + P_1|G_{1,k}|^2\sigma_{1,1}^2 + P_2|G_{2,k}|^2\sigma_{1,2}^2\right)\boldsymbol{I}_{M+BD}$ with $\boldsymbol{I}_M$ denoting the M by M identity matrix. Therefore, the Viterbi algorithm proposed in Section 8.3.4 can be used as an ML detection scheme on symbols transmitted over the k-th subcarrier. Furthermore, by employing ML detection at the destination for the conditional PEP over the k-th subcarrier, $P\left(\boldsymbol{X}_k \to \boldsymbol{X}_k' | \boldsymbol{GH}(k)\right)$ which shows the probability of deciding in favor of $\boldsymbol{X}_k'$ at the receiver while $\boldsymbol{X}$ is the actual transmitted symbol conditioned on the channel realizations, we have

$$P\left(\boldsymbol{X}_k \to \boldsymbol{X}_k' | \boldsymbol{GH}(k)\right) \leq \frac{1}{2} e^{-\frac{d_k^2}{4\left(\sigma_2^2 + P_1|G_{1,k}|^2\sigma_{1,1}^2 + P_2|G_{2,k}|^2\sigma_{1,2}^2\right)}}, \tag{8.14}$$

where in deriving the last inequality the Chernoff bound is employed [67, p. 58], $d_k = ||\boldsymbol{GH}(k)(\boldsymbol{X}_k - \boldsymbol{X}_k')||$ and $||\boldsymbol{e}||$ denotes the Euclidean length of vector $\boldsymbol{e}$. Therefore, we have

$$P\left(\boldsymbol{X}_k \to \boldsymbol{X}_k'\right) \leq \frac{1}{2} E_{d_k}\{e^{-\frac{d_k^2}{4\sigma^2}}\}, \tag{8.15}$$

where $E_{d_k}\{.\}$ denotes the expected value with respect to the random variable d_k. Under the assumption that $\frac{1}{\sigma_{1,i}^2} \gg \frac{P_i}{\sigma_2^2}$, we obtain $\frac{d_k^2}{\sigma^2} \simeq \frac{d_k^2}{\sigma_2^2}$. Furthermore, due to

the definition of the Euclidean distance, by defining $\alpha_k^2 = \frac{P_1}{4\sigma_2^2}\sum_{m=1}^{M}|X_k^m - X_k'^m|^2$, $\beta_k^2 = \frac{P_2}{4\sigma_2^2}\sum_{m=1}^{M}|X_k^m - X_k'^m|^2$, and

$$\gamma_k^2 e^{-j\phi_{\gamma,k}} = \frac{\sqrt{P_1P_2}}{4\sigma_2^2}\sum_{m=BD+1}^{M}(X_k^m - X_k'^m)(X_k^{m-BD} - X_k'^{m-BD})^*,$$

we can write

$$\begin{aligned}\frac{d_k^2}{4\sigma_2^2} =& \frac{\boldsymbol{GH}(k)(\boldsymbol{X}_k - \boldsymbol{X}_k')(\boldsymbol{X}_k - \boldsymbol{X}_k')^H\boldsymbol{GH}^H(k)}{4\sigma_2^2} \\ =& \alpha_k^2\frac{|GH_1[k,k]|^2}{P_1} + \beta_k^2\frac{|GH_2[k,k]|^2}{P_2} \\ &+ 2Re\left\{\frac{|GH_1[k,k]|}{\sqrt{P_1}}e^{-j2\pi\phi_{1,k}}\frac{|GH_2[k,k]|}{\sqrt{P_2}}e^{j2\pi\phi_{2,k}}\gamma_k^2 e^{-j2\pi\phi_{\gamma,k}}\right\}. \end{aligned} \tag{8.16}$$

We also have $|GH_i[k,k]| = P_i|H_{i,k}||G_{i,k}|$, where $|H_{i,k}| \sim Rayleigh(\frac{\sqrt{2}}{2})$ and $|G_{i,k}| \sim Rayleigh(\frac{\sqrt{2}}{2})$, i.e., $|H_{i,k}|$ and $|G_{i,k}|$ are Rayleigh distributed random variables, and $\phi_{i,k}$ are uniformly distributed random variables over $[0, 2\pi]$. If we define $a_i = |H_{i,k}|$ and $b_i = |G_{i,k}|$, then $\frac{d_k^2}{4\sigma_2^2} = \alpha_k^2a_1^2b_1^2 + \beta_k^2a_2^2b_2^2 + 2a_1a_2b_1b_2\gamma_k^2\cos(\phi)$, where $\phi = \angle e^{-j(\phi_{1,k}-\phi_{2,k}+\phi_{\gamma,k})}$ is uniformly distributed over $[0, 2\pi]$. Therefore, we have

$$\begin{aligned} E_{d_k}\{e^{-\frac{d_k^2}{4\sigma_2^2}}\} &= E_{a_1,a_2,b_1,b_2,\phi}\left\{e^{-(\alpha_k^2a_1^2b_1^2+\beta_k^2a_2^2b_2^2+2a_1a_2b_1b_2\gamma_k^2\cos(\phi))}\right\} \\ &= E_{a_1,a_2,b_1,b_2}\left\{\frac{1}{2\pi}\int_0^{2\pi}e^{-(\alpha_k^2a_1^2b_1^2+\beta_k^2a_2^2b_2^2+2a_1a_2b_1b_2\gamma_k^2\cos(\phi))}d\phi\right\} \\ &=^a E_{a_1,a_2,b_1,b_2}\left\{e^{-(\alpha_k^2a_1^2b_1^2+\beta_k^2a_2^2b_2^2)}I_0(2a_1a_2b_1b_2\gamma_k^2)\right\} \\ &= E_{a_2,b_1,b_2}\left\{e^{-\beta_k^2a_2^2b_2^2}\int_0^{\infty}e^{-\alpha_k^2a_1^2b_1^2}I_0(2a_1a_2b_1b_2\gamma_k^2)2a_1e^{-a_1^2}da_1\right\} \\ &=^b E_{a_2,b_1,b_2}\left\{\frac{e^{-\beta_k^2a_2^2b_2^2}}{\alpha_k^2b_1^2+1}e^{\frac{a_2^2b_1^2b_2^2\gamma_k^4}{\alpha_k^2b_1^2+1}}\right\} \\ &= E_{b_1,b_2}\left\{\int_0^{\infty}\frac{e^{-\beta_k^2a_2^2b_2^2}}{\alpha_k^2b_1^2+1}e^{\frac{a_2^2b_1^2b_2^2\gamma_k^4}{\alpha_k^2b_1^2+1}}2a_2e^{-a_2^2}da_2\right\} \\ &= E_{b_1,b_2}\left\{\frac{1}{1+\alpha_k^2b_1^2+b_1^2b_2^2\alpha_k^2\beta_k^2+\beta_k^2b_2^2-b_1^2b_2^2\gamma_k^4}\right\}, \end{aligned} \tag{8.17}$$

where the equality a holds due to the definition of the modified Bessel function of the first kind $I_0(.)$ and the equality b results from [68, p. 294. Eq. 2. 15. 1. 2.]. Furthermore it follows from the inequality $\alpha_k^2\beta_k^2 \geq \gamma_k^4$ that $\frac{1}{1+\alpha_k^2 b_1^2+b_1^2 b_2^2\alpha_k^2\beta_k^2+\beta_k^2 b_2^2-b_1^2 b_2^2\gamma_k^4} \leq \frac{1}{\left(\sqrt{\frac{\alpha_k^2\beta_k^2-\gamma_k^4}{\alpha_k^2\beta_k^2}}\alpha_k^2 b_1^2+1\right)\left(\sqrt{\frac{\alpha_k^2\beta_k^2-\gamma_k^4}{\alpha_k^2\beta_k^2}}\beta_k^2 b_2^2+1\right)}$ which leads to (since $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are independent)

$$\begin{aligned}
E_{d_k}\{e^{-\frac{d_k^2}{4\sigma_2^2}}\} \leq & E_{b_1}\left\{\frac{1}{\sqrt{\frac{\alpha_k^2\beta_k^2-\gamma_k^4}{\alpha_k^2\beta_k^2}}\alpha_k^2 b_1^2+1}\right\} E_{b_2}\left\{\frac{1}{\sqrt{\frac{\alpha_k^2\beta_k^2-\gamma_k^4}{\alpha_k^2\beta_k^2}}\beta_k^2 b_2^2+1}\right\} \\
=&\frac{1}{\alpha_k^2\beta_k^2-\gamma_k^4} e^{\frac{\beta_k}{\alpha_k\sqrt{\alpha_k^2\beta_k^2-\gamma_k^4}}} E_1\left(\frac{\beta_k}{\alpha_k\sqrt{\alpha_k^2\beta_k^2-\gamma_k^4}}\right) e^{\frac{\alpha_k}{\beta_k\sqrt{\alpha_k^2\beta_k^2-\gamma_k^4}}} E_1\left(\frac{\alpha_k}{\beta_k\sqrt{\alpha_k^2\beta_k^2-\gamma_k^4}}\right) \\
\leq&\frac{1}{\alpha_k^2\beta_k^2-\gamma_k^4}\log\left(1+\alpha_k^2\sqrt{\frac{\alpha_k^2\beta_k^2-\gamma_k^4}{\alpha_k^2\beta_k^2}}\right)\log\left(1+\beta_k^2\sqrt{\frac{\alpha_k^2\beta_k^2-\gamma_k^4}{\alpha_k^2\beta_k^2}}\right), \qquad (8.18)
\end{aligned}$$

where $E_1(a)$ denotes the exponential integral function which is given as $E_1(a) = \int_a^\infty \frac{e^{-x}}{x}dx$ and the last inequality follows sine $e^a E_1(a) \leq \log(1+\frac{1}{a})$. Invoking the result of (8.18) in (8.15), and defining $s_k^2 = \sum_{m=1}^{M} |X_k^m - X_k'^m|^2$ and

$$f_k^2 = \left|\sum_{m=BD+1}^{M}(X_k^m - X_k'^m)(X_k^{m-BD} - X_k'^{m-BD})^*\right|$$

yields

$$P\left(\boldsymbol{X}_k \to \boldsymbol{X}_k'\right) \leq \frac{8\sigma_2^4}{P_1P_2\left(s_k^4-f_k^4\right)}\log\left(1+\frac{P_1}{4\sigma_2^2}\sqrt{s_k^4-f_k^4}\right)\log\left(1+\frac{P_2}{4\sigma_2^2}\sqrt{s_k^4-f_k^4}\right), \qquad (8.19)$$

under the assumption that the channels from the source to the relays have higher SNR ratios than the channels from the relays to the destination. We observe from (8.19) that the system achieves the diversity order of 2. For instance, for $P_1 = P_2$, we have $SNR = \frac{2P_1}{\sigma_2^2}$ and $P\left(\boldsymbol{X}_k \to \boldsymbol{X}_k'\right) \leq \frac{32SNR^{-2}}{(s_k^4-f_k^4)}\log\left(1+\frac{SNR}{8}\sqrt{s_k^4-f_k^4}\right)^2$.

8.4.2 Block Fading Frequency-Selective Channels

In this section, we analyze the PEP performance of the proposed scheme under block fading frequency selective channels. Similar to the analysis for the quasi-static fading

scenario, we assume that the channels from the source to the relays have significantly higher SNR ratios than the channels from the relays to the destination. We first give the considered block fading channel model. We then provide a discussion on the discrete noise samples at the destination under the block fading channels and at the end provide the PEP analysis for which similar to the quasi-static channel conditions, we assume no coding is employed over the subcarriers and focus on the spatial diversity analysis of the system.

8.4.2.1 Block Fading Frequency-Selective Channel Model

Here, we follow the same channel model and procedure as taken in [69] in which the PEP performance analysis of spacetime coded OFDM-MIMO system over correlated block fading channels has been considered. The main difference between the system model in [69] and the one in this chapter is in the effective channel model and the noise experienced at the destination. In fact, we need to use some simplifying approximations to be able to derive a closed form upper bound on the PEP of the system. By approximating the received noise samples Z_k^m as complex Gaussian random variables (see Section 8.4.2.2), we provide a PEP analysis for the system under block fading channel scenario in which channel coefficients are fixed during each block transmission and change block by block based on the following Fourier expansion relation [70] (for ease of presentation we assume that $L_{g_i} = L_{h_i} = L$, $\tau_{h_i,l} = \tau_{g_i,l} = \tau_l$ and all the channels experiencing the same Doppler frequency shift f_d)

$$h_{i,l}^m(t) = h_{i,l}^m \approx \sum_{n=-\frac{L_t-1}{2}}^{\frac{L_t-1}{2}} \alpha_{i,l}[n] e^{j\frac{2\pi n(mT)}{MT}}, \tag{8.20}$$

and

$$g_{i,l}^m(t) = g_{i,l}^m \approx \sum_{n=-\frac{L_t-1}{2}}^{\frac{L_t-1}{2}} \beta_{i,l}[n] e^{j\frac{2\pi n(mT)}{MT}}, \tag{8.21}$$

in which $\alpha_{i,l}[n]$ and $\beta_{i,l}[n]$ are independent circularly symmetric complex Gaussian random variables with zero mean and variance of $\frac{\sigma^2_{h_{i,l}}}{L_t}$ and $\frac{\sigma^2_{h_{i,l}}}{L_t}$, respectively, with $L_t = \lceil 2f_dMT + 1 \rceil$. $h^m_{i,l}$ and $g^m_{i,l}$ can also be represented as $h^m_{i,l} = \boldsymbol{\alpha}_i(l)^T \boldsymbol{w}_t(m)$ and $g^m_{i,l} = \boldsymbol{\beta}_i(l)^T \boldsymbol{w}_t(m)$, where $\boldsymbol{w}_t(m) = [e^{-j2\pi M f_d T}, \ldots, 1, \ldots, e^{j2\pi M f_d T}]^T$, $\boldsymbol{\alpha}_i(l) = \left[\alpha_{i,l}[-\frac{L_t-1}{2}], \ldots, \alpha_{i,l}[\frac{L_t-1}{2}]\right]^T$ and $\boldsymbol{\beta}_i(l) = \left[\beta_{i,l}[-\frac{L_t-1}{2}], \ldots, \beta_{i,l}[\frac{L_t-1}{2}]\right]^T$.

Since all the channels are block fading, for $GH^m_1[k,k']$ and $GH^m_2[k,k']$, we have

$$GH^m_1[k,k'] = \sqrt{P_1} \sum_{l=1}^{L_{g_1}} g^m_{1,l} e^{-2\pi j \frac{k'\tau_{g_1,l}}{NT_s}} \sum_{q=1}^{L_{h_1}} h^m_{1,q} e^{-2\pi j \frac{k'\tau_{h_1,q}}{NT_s}} \delta(k-k'),$$

$$GH_2[k,k'] = \sqrt{P_2} e^{-2\pi j \frac{k'}{NT_s} d_r} \sum_{l=1}^{L_{g_2}} g^m_{2,l} e^{-2\pi j \frac{k'\tau_{g_2,l}}{NT_s}} \sum_{q=1}^{L_{h_2}} h^m_{2,q} e^{-2\pi j \frac{k'\tau_{h_2,q}}{NT_s}} \delta(k-k').$$

We can also write $GH^m_1[k,k] = G^m_{1,k} H^m_{1,k}$ and $GH^m_2[k,k] = G^m_{2,k} H^m_{2,k} e^{-2\pi j \frac{k'}{NT_s} d_r}$ with $H^m_{i,k} = \sum_{l=1}^{L} h_{i,l} e^{-j\frac{2\pi k \tau_l}{NT_s}} = \boldsymbol{h}^T_i(m) \boldsymbol{w}_f(k)$, $G^m_{i,k} = \sum_{l=1}^{L} g_{i,l} e^{-j\frac{2\pi k \tau_l}{NT_s}} = \boldsymbol{g}^T_i(m) \boldsymbol{w}_f(k)$ where $\boldsymbol{h}_i(m) = [h^m_{i,1}, \cdots, h^m_{i,L}]^T$, $\boldsymbol{g}_i(m) = [g^m_{i,1}, \cdots, g^m_{i,L}]^T$ and $\boldsymbol{w}_f(k) = [e^{-j\frac{2\pi k \tau_1}{NT_s}}, \ldots, e^{-j\frac{2\pi k \tau_L}{NT_s}}]^T$. On the other hand, by defining $\boldsymbol{v}_i = [\boldsymbol{\alpha}^T_i(1), \ldots, \boldsymbol{\alpha}^T_i(L)]^T$, $\boldsymbol{q}_1 = [\boldsymbol{\beta}^T_1(1), \ldots, \boldsymbol{\beta}^T_1(L)]^T$, $\boldsymbol{q}_2 = [\boldsymbol{\beta}^T_2(1), \ldots, \boldsymbol{\beta}^T_2(L)]^T e^{-2\pi j \frac{k'}{NT_s} d_r}$ and $\boldsymbol{W}_t(m) = diag\{\boldsymbol{w}_t(m), \ldots, \boldsymbol{w}_t(m)\}_{LL_t \times L}$, we obtain $H^m_{i,k} = \boldsymbol{v}^T_i \boldsymbol{W}_t(m) \boldsymbol{w}_f(k)$ and $G^m_{i,k} = \boldsymbol{q}^T_i \boldsymbol{W}_t(m) \boldsymbol{w}_f(k)$.

8.4.2.2 Discussion on the Statistic of the Noise Samples

For the block fading scenario, we have $h^m_{i,l}(t) = h^m_{i,l}$ and $g^m_{i,l}(t) = g^m_{i,l}$, hence we can write

$$Z^m_k = \frac{1}{\sqrt{N}} \sum_{n=b}^{b+N-1} \left[z^m_2(nT_s) + \sum_{l=1}^{L_{g_1}} \sqrt{P_1} g^m_{1,l} z^m_{1,1}(nT_s - \tau_{g_1,l}) + \sum_{l=1}^{L_{g_2}} \sqrt{P_2} g^m_{2,l} z_{1,2}(nT_s - \tau_{g_2,l}) \right] e^{-\frac{j2\pi n}{N} k}.$$

Since the noise samples from the OFDM block durations m and m' ($m \neq m'$) are independent then obviously $E\{Z^m_k Z^{m'}_k\} = 0$. Furthermore, for $E\{|Z^m_k|^2\}$, we can

write

$$E\{|Z_k^m|^2\} =$$
$$\frac{1}{N}E\Bigg\{\sum_{n=b}^{b+N-1}\sum_{v=b}^{b+N-1}\left[z_2^m(nT_s)+\sum_{l=1}^{L_{g_1}}\sqrt{P_1}g_{1,l}^m z_{1,1}^m(nT_s-\tau_{g_1,l})+\sum_{l=1}^{L_{g_2}}\sqrt{P_2}g_{2,l}^m z_{1,2}^m(nT_s-\tau_{g_2,l})\right]$$
$$\times\left[z_2^{m*}(vT_s)+\sum_{i=0}^{L_{g_1}}\sqrt{P_1}g^{m*}{}_{1,i}z_{1,1}^{m\ *}(vT_s-\tau_{g_1,i})+\sum_{i=0}^{L_{g_2}}\sqrt{P_2}g^{m*}{}_{2,i}z_{1,2}^{m\ *}(vT_s-\tau_{g_2,i})\right]e^{-\frac{j2\pi(n-v)}{N}k}\Bigg\}.$$

By using the facts that $z_{j,i}(t)$ are zero mean independent Gaussian random processes (as a result $E\{z_{1,1}(t)z_{1,2}(t')\}=0$ for all t and t') and $E\{z_{j,i}(t)z_{j,i}(t')\}=\sigma_{j,i}^2\delta(t-t')$, we obtain

$$E\{|Z_k^m|^2\} =\frac{1}{N}\sum_{n=b}^{b+N-1}\sum_{v=b}^{b+N-1}\Bigg[\sigma_2^2\delta(n-v)+P_1\sum_{l=1}^{L_{g_1}}\sum_{i=1}^{L_{g_1}}g_{1,l}^m g^{m*}{}_{1,i}\sigma_{1,1}^2\delta(nT_s-\tau_{g_1,l}-vT_s+\tau_{g_1,i})$$
$$+P_2\sum_{l=1}^{L_{g_2}}\sum_{i=1}^{L_{g_2}}g_{2,l}^m g^{m*}{}_{2,i}\sigma_{1,2}^2\delta(nT_s-\tau_{g_2,l}-vT_s+\tau_{g_2,i})\Bigg]e^{-\frac{j2\pi(n-v)}{N}k}$$

which leads to

$$E\{|Z_k^m|^2\} =\sigma_2^2+P_1\sigma_{1,1}^2\sum_{l=1}^{L_{g_1}}|g_{1,l}^m|^2+P_2\sigma_{1,2}^2\sum_{l=1}^{L_{g_2}}|g_{2,l}^m|^2+2\sigma_{1,1}^2Re\left[\sum_{f_1,\tau_{g_1,l}-\tau_{g_1,i}=f_1T_s}g_{1,l}^m g_{1,i}^{m\ *}e^{-\frac{j2\pi k}{N}f_1}\right]$$
$$+2\sigma_{1,2}^2Re\left[\sum_{f_2,\tau_{g_2,l}-\tau_{g_2,i}=f_2T_s}g_{2,l}^m g_{2,i}^{m\ *}e^{-\frac{j2\pi k}{N}f_2}\right], \tag{8.22}$$

where f_1 and f_2 only take positive integer values. Since Z_k^m are independent for a specific k but not identically distributed, the optimal ML detection can be obtained by employing Viterbi detector over the normalized received signals according to $E\{|Z_k^m|^2\}$. However, in the following, we provide the PEP analysis by approximating the received noise samples Z_k^m as i.i.d. complex Gaussian random variables with zero mean and variance of σ^2 to match with the sub-optimal detector we considered for the general time-varying channel case in Section 8.3.4.

8.4.2.3 PEP Analysis

Conditioned on known channel state information at the receiver, for the considered Viterbi detector[1], we have

$$\boldsymbol{X}' = \arg\min_{\boldsymbol{X}} \sum_{m=1}^{M+BD} \sum_{k=0}^{N-1} \left| Y_k^m - G_{1,k}^m H_{1,k}^m X_k^m - G_{2,k}^m H_{2,k}^m e^{-2\pi j \frac{k'}{NT_s} d_r} X_k^{m-Bd} \right|^2, \quad (8.23)$$

where $X_k^m = 0$ for $m < 1$ or $m > M$. Therefore, similar to (8.14) under the assumption that $\frac{1}{\sigma_{1,i}^2}$ are sufficiently larger than $\frac{P_i}{\sigma_2^2}$, we can write

$$P(\boldsymbol{X}_k \to \boldsymbol{X}'_k | \boldsymbol{H}_i, \boldsymbol{G}_i \, i \in \{1,2\}) \leq \frac{1}{2} e^{-\frac{d^2\left(\boldsymbol{X}_k, \boldsymbol{X}'_k\right)}{4\sigma_2^2}}, \quad (8.24)$$

where $d^2(\boldsymbol{X}_k, \boldsymbol{X}'_k) = \sum_{m=1}^{M+BD} \left| G_{1,k}^m H_{1,k}^m d_k^m + G_{2,k}^m H_{2,k}^m d_k^{m-BD} \right|^2$. By defining $\boldsymbol{W}_{t,f}(m,k) = diag\{\boldsymbol{W}_t(m)\boldsymbol{w}_f(k), \boldsymbol{W}_t(m)\boldsymbol{w}_f(k)\}$, $d_k^m = X_k^m - X_k^{'m}$ and $\boldsymbol{d}_k(m) = [d_k^m, d_k^{m-BD}]^T$, we can write

$$\begin{aligned} d^2(\boldsymbol{X}_k, \boldsymbol{X}'_k) = \sum_{m=1}^{M+BD} & \boldsymbol{w}_f^T(k) \boldsymbol{W}_t^T(m) [\boldsymbol{v}_1 \boldsymbol{q}_1^T, \boldsymbol{v}_2 \boldsymbol{q}_2^T] \boldsymbol{W}_{t,f}(m,k) \boldsymbol{d}_k(m) \times \\ & \times \boldsymbol{d}_k^H(m) W_{t,f}^H(m,k) [\boldsymbol{v}_1 \boldsymbol{q}_1^T, \boldsymbol{v}_2 \boldsymbol{q}_2^T]^H \boldsymbol{W}_t^*(m) \boldsymbol{w}_f^*(k). \end{aligned}$$

On the other hand, we have

$$\begin{aligned} & \boldsymbol{w}_f^T(k) \boldsymbol{W}_t^T(m) [\boldsymbol{v}_1 \boldsymbol{q}_1^T, \boldsymbol{v}_2 \boldsymbol{q}_2^T] \\ & \quad = \boldsymbol{w}_f^T(k) \boldsymbol{W}_t^T(m) \left[[\boldsymbol{\alpha}_1^T(1), \ldots, \boldsymbol{\alpha}_1^T(L)]^T \boldsymbol{q}_1^T, [\boldsymbol{\alpha}_2^T(1), \ldots, \boldsymbol{\alpha}_2^T(L)]^T \boldsymbol{q}_2^T \right] \\ & \quad = \boldsymbol{w}_f^T(k) \begin{bmatrix} \boldsymbol{w}_t^T(m) \boldsymbol{\alpha}_1(1) \boldsymbol{q}_1^T & , & \boldsymbol{w}_t^T(m) \boldsymbol{\alpha}_2(1) \boldsymbol{q}_2^T \\ \vdots & \vdots & \vdots \\ \boldsymbol{w}_t^T(m) \boldsymbol{\alpha}_1(L) \boldsymbol{q}_1^T & , & \boldsymbol{w}_t^T(m) \boldsymbol{\alpha}_2(L) \boldsymbol{q}_2^T \end{bmatrix} \\ & \quad = \left[\boldsymbol{q}_1^T \sum_{l=1}^{L} e^{-j\frac{2\pi k \tau_l}{NT_s}} \boldsymbol{\alpha}_1^T(l) \boldsymbol{w}_t(m), \boldsymbol{q}_2^T \sum_{l=1}^{L} e^{-j\frac{2\pi k \tau_l}{NT_s}} \boldsymbol{\alpha}_2^T(l) \boldsymbol{w}_t(m) \right]. \end{aligned}$$

[1]For the considered block fading channels, the optimal ML detection is obtained by normalizing the received signals over each subcarrier with variance of its corresponding noise, however, we present the result for the case that there is no normalization to match with the Viterbi detection of the general time-varying case.

By defining

$$\boldsymbol{A}_i(k) = diag\left\{\sum_{l=1}^{L} e^{-j\frac{2\pi k\tau_l}{NT_s}}\boldsymbol{\alpha}_i^T(l), \ldots, \sum_{l=1}^{L} e^{-j\frac{2\pi k\tau_l}{NT_s}}\boldsymbol{\alpha}_i^T(l)\right\}_{LL_t\times LL_t^2},$$

and $\boldsymbol{W}_{\alpha,t}(m) = diag\{\boldsymbol{w}_t(m), \ldots, \boldsymbol{w}_t(m)\}_{LL_t^2\times LL_t}$, we obtain

$$\boldsymbol{w}_f^T(k)\boldsymbol{W}_t^T(m)[\boldsymbol{v}_1\boldsymbol{q}_1^T, \boldsymbol{v}_2\boldsymbol{q}_2^T] = \left[\boldsymbol{q}_1^T\boldsymbol{A}_1(k)\boldsymbol{W}_{\alpha,t}(m), \boldsymbol{q}_2^T\boldsymbol{A}_2(k)\boldsymbol{W}_{\alpha,t}(m)\right], \tag{8.25}$$

Note that $\boldsymbol{A}_i(k)\boldsymbol{W}_{\alpha,t}(m) = \sum_{l=1}^{L} e^{-j\frac{2\pi k\tau_l}{NT_s}}\boldsymbol{\alpha}_{1,1}^T(l)\boldsymbol{w}_t(m)\boldsymbol{I}_{L_tL}$. Furthermore, by defining $\boldsymbol{q}(k) = \left[\boldsymbol{q}_1^T\boldsymbol{A}_1(k), \boldsymbol{q}_2^T\boldsymbol{A}_2(k)\right]^H$, and $\boldsymbol{W}_{A,t}(m) = diag\{\boldsymbol{W}_{\alpha,t}(m), \boldsymbol{W}_{\alpha,t}(m)\}$, we can write

$$\boldsymbol{w}_f^T(k)\boldsymbol{W}_t^T(m)[\boldsymbol{v}_1\boldsymbol{q}_1^T, \boldsymbol{v}_2\boldsymbol{q}_2^T] = \boldsymbol{q}(k)^H\boldsymbol{W}_{A,t}(m). \tag{8.26}$$

Therefore, defining

$$\boldsymbol{D}_A(\boldsymbol{X}_k, \boldsymbol{X}'_k) = \sum_{m=1}^{M+BD} \boldsymbol{W}_{A,t}(m)\boldsymbol{W}_{t,f}(m,k)\boldsymbol{d}_k(m)\boldsymbol{d}_k^H(m)\boldsymbol{W}_{t,f}^H(m,k)\boldsymbol{W}_{A,t}^H(m), \tag{8.27}$$

yields to

$$d^2(\boldsymbol{X}_k, \boldsymbol{X}'_k) = \boldsymbol{q}(k)^H\boldsymbol{D}_A(\boldsymbol{X}_k, \boldsymbol{X}'_k)\boldsymbol{q}(k). \tag{8.28}$$

Since $\boldsymbol{D}_A(\boldsymbol{X}_k, \boldsymbol{X}'_k)$ is a positive semi-definite matrix, we can write

$$\boldsymbol{D}_A(\boldsymbol{X}_k, \boldsymbol{X}'_k) = \boldsymbol{U}_k\boldsymbol{\Lambda}_k\boldsymbol{U}_k^H, \tag{8.29}$$

where $\boldsymbol{U}_k$ is a unitary matrix and $\boldsymbol{\Lambda} = diag\{\lambda_{k,1}, \ldots, \lambda_{k,r}, 0, \ldots, 0\}$ with $\lambda_{k,i}$ being the positive eigenvalues of $\boldsymbol{D}_A(\boldsymbol{X}_k, \boldsymbol{X}'_k)$. We define $D(\boldsymbol{X}_k, \boldsymbol{X}'_k)$ as the number of values of m where either $X_k^m \neq X_k'^m$ or $X_k^{m-BD} \neq X_k'^{m-BD}$, i.e.,

$$D(\boldsymbol{X}_k, \boldsymbol{X}'_k) = M + BD - \sum_{m=1}^{M+BD} \delta(X_k^m - X_k'^m)\delta(X_k^{m-BD} - X_k'^{m-BD}). \tag{8.30}$$

Since $\boldsymbol{d}_k(m)\boldsymbol{d}_k^H(m)$ is a rank one matrix whenever $X_k^m \neq X_k'^m$ and/or $X_k^{m-BD} \neq X_k'^{m-BD}$ and a zero matrix otherwise, we can write

$$r_k = rank(D_A(\boldsymbol{X}_k, \boldsymbol{X}'_k)) \leq \min(D(\boldsymbol{X}_k, \boldsymbol{X}'_k), 2LL_t^2),$$

and as a result $r = \min_{\boldsymbol{X}_k, \boldsymbol{X}_k'} r_k \leq \min(D_{eff}(k), 2LL_t^2)$ with $D_{eff}(k) = \min_{\boldsymbol{X}_k, \boldsymbol{X}_k'} D(\boldsymbol{X}_k, \boldsymbol{X}_k')$. Conditioned on known channel coefficients, it follows from (8.29) that

$$P(\boldsymbol{X}_k \to \boldsymbol{X}_k' | \boldsymbol{H}_i[k], \boldsymbol{G}_i[k]) \leq \frac{1}{2} e^{-\frac{1}{4\sigma_2^2} \sum_{c=1}^{r_k} \lambda_{k,c} \left| \boldsymbol{U}_{k,c}^H \boldsymbol{q}(k) \right|^2}, \tag{8.31}$$

where $\boldsymbol{U}_{k,c}$ denotes the c-th column of $\boldsymbol{U}_k$. Furthermore, if we define $\boldsymbol{\mu}_i(k) = [\mu_{k,i,1}, \ldots, \mu_{k,i,L_t}] = \sum_{l=1}^{L} e^{-j\frac{2\pi k \tau_l}{NT_s}} \boldsymbol{\alpha}_i(l)$, since $\alpha_{i,l}[n]$ are i.i.d. complex Gaussian random variables with zero mean, $\mu_{k,i,p}$ are also i.i.d. complex Gaussian random variables with zero mean and variance of $\sigma_{\mu,i}^2 = \sum_{l=1}^{L} \frac{\sigma_{h_i,l}^2}{L_t}$. Furthermore, by denoting $\boldsymbol{q}_1$ as $\boldsymbol{q}_1 = [q_1, \ldots, q_{LL_t}]^H$ and $\boldsymbol{q}_2$ as $\boldsymbol{q}_2 = [q_{LL_t+1}, \ldots, q_{2LL_t}]^H$, we can write

$$\begin{aligned} \boldsymbol{q}(k) = [&q_1\mu_{k,1,1}, \ldots, q_1\mu_{k,1,L_t}, \ldots, q_{LL_t}\mu_{k,1,1}, \ldots, q_{LL_t}\mu_{k,1,L_t}, q_{LL_t+1}\mu_{k,2,1}, \ldots \\ &\ldots, q_{LL_t+1}\mu_{k,2,L_t}, \ldots, q_{2LL_t}\mu_{k,2,1}, \ldots, q_{2LL_t}\mu_{k,2,L_t}]. \end{aligned} \tag{8.32}$$

Therefore, $\boldsymbol{U}_{k,c}^H \boldsymbol{q}(k)$ can be written as

$$\begin{aligned} \boldsymbol{U}_{k,c}^H \boldsymbol{q}(k) &= \sum_{p=1}^{LL_t} \sum_{t=1}^{L_t} U_{k,c,(p-1)L_t+t}^* q_p \mu_{k,1,t} + \sum_{p=1}^{LL_t} \sum_{t=1}^{L_t} U_{k,c,LL_t^2+(p-1)L_t+t}^* q_{LL_t+p} \mu_{k,2,t} \\ &= \sum_{i=1}^{2} \sum_{p=1}^{LL_t} q_{(i-1)LL_t+p} \chi_{k,(i-1)LL_t+p}, \end{aligned} \tag{8.33}$$

where $\chi_{k,(i-1)LL_t+p} = \sum_{t=1}^{L_t} U_{k,c,(i-1)LL_t^2+(p-1)L_t+t}^* \mu_{k,i,t}$ are zero mean complex Gaussian random variables with variance of $\sigma_{\chi,k,(i-1)LL_t+p}^2 = \sum_{t=1}^{L_t} |U_{k,c,(i-1)LL_t^2+(p-1)L_t+t}|^2 \sigma_{\mu,i}^2$. To calculate the PEP of the system, we make two simplifying assumptions:

- $\boldsymbol{U}_{k,i}^H \boldsymbol{q}(k)$ and $\boldsymbol{U}_{k,j}^H \boldsymbol{q}(k)$ are independent for different values of i and j ($i \neq j$). (Note that for $\boldsymbol{q}_k$ being a complex Gaussian random vector with zero mean, $\boldsymbol{U}_{k,c}^H \boldsymbol{q}_k$ and $\boldsymbol{U}_{k,j}^H \boldsymbol{q}_k$ are independent complex Gaussian random variables.)

- $\chi_{k,p}$ are i.i.d. complex Gaussian random variables with zero mean and variance of $\sigma_{\chi,k,p}^2$.

For a fixed c, by defining $R_p = \left| \sum_{l=p+1}^{2LL_t} q_l \chi_{k,l} \right|$ and $\theta_p = \arccos\left(\frac{\sum_{l=p+1}^{2LL_t} |q_l \chi_{k,l}| \cos(\phi_l)}{R_p} \right)$ in which ϕ_l is the angle between the complex valued random vectors q_l and $\chi_{k,l}$ and uniformly distributed over $[0, 2\pi]$, we can write

$$\begin{aligned} |\boldsymbol{U}_{k,c}^H \boldsymbol{q}(k)|^2 &= \left| \sum_{p=1}^{2LL_t} q_p \chi_{k,p} \right|^2 = \sum_{p=1}^{2LL_t} |q_p \chi_{k,p}|^2 + 2Re\left\{ \sum_{p=1}^{2LL_t} \sum_{l=p+1}^{2LL_t} q_p \chi_{k,p} q_l^* \chi_l^* \right\} \\ &= \sum_{p=1}^{2LL_t} |q_p \chi_{k,p}|^2 + 2 \sum_{p=1}^{2LL_t} |q_p \chi_{k,p}| R_p \cos(\phi_p - \theta_p). \end{aligned} \tag{8.34}$$

Due to the fact that ϕ_p are uniformly distributed over $[0, 2\pi]$, $|\chi_p|$ are Rayleigh distributed and by following the same procedure as in (8.17), we can write

$$\begin{aligned} &E_{\phi_1, |\chi_{k,1}|} \left\{ e^{-\frac{\lambda_{k,c} R_0^2}{4\sigma_2^2}} \right\} \\ &= E_{|\chi_{k,1}|} \left\{ e^{-\frac{\lambda_{k,c} \sum_{p=1}^{2LL_t} |q_p \chi_{k,p}|^2}{4\sigma_2^2}} e^{-\frac{\lambda_{k,c} \sum_{p=2}^{2LL_t} |q_p \chi_{k,p}| R_p \cos(\phi_p - \theta_p)}{2\sigma_2^2}} I_0\left(\frac{\lambda_{k,c} |q_1 \chi_{k,1}| R_1}{2\sigma_2^2} \right) \right\} \\ &= \frac{2\sigma_2^2 e^{-\frac{\lambda_{k,c} R_1^2}{4\sigma_2^2}}}{\lambda_{k,c} \sigma_{\chi,k,1}^2 |q_1|^2 + 2\sigma_2^2} e^{\frac{\lambda_{k,c}^2 \sigma_{\chi,k,1}^2 |q_1|^2 R_1^2}{4\sigma_2^2 (\lambda_{k,c} \sigma_{\chi,k,1}^2 |q_1|^2 + 2\sigma_2^2)}} \\ &= \frac{2\sigma_2^2}{\lambda_{k,c} \sigma_{\chi,k,1}^2 |q_1|^2 + 2\sigma_2^2} e^{-\frac{\lambda_{k,c} R_1^2}{2\lambda_{k,c} \sigma_{\chi,k,1}^2 |q_1|^2 + 4\sigma_2^2}}. \end{aligned} \tag{8.35}$$

Therefore, by taking the expected value with respect to $\phi_1, |\chi_{k,1}|, \ldots, \phi_{2LL_t}, |\chi_{k,2LL_t}|$, we arrive at

$$\begin{aligned} &E_{\phi_1, |\chi_{k,1}|, \ldots, \phi_{2LL_t}, |\chi_{k,2LL_t}|} \left\{ e^{-\frac{\lambda_{k,c} R_0^2}{4\sigma_2^2}} \right\} \\ &= \frac{2\sigma_2^2}{2\sigma_2^2 + \lambda_{k,c} \sigma_{\chi,k,1}^2 |q_1|^2} \frac{2\sigma_2^2 + \lambda_{k,c} \sigma_{\chi,k,1}^2 |q_1|^2}{2\sigma_2^2 + \lambda_{k,c} (\sigma_{\chi,k,1}^2 |q_1|^2 + \sigma_{\chi,k,2}^2 |q_2|^2)} \cdots \frac{2\sigma_2^2 + \lambda_{k,c} \sum_{p=1}^{2LL_t - 1} \sigma_{\chi,k,p}^2 |q_p|^2}{2\sigma_2^2 + \lambda_{k,c} \sum_{p=1}^{2LL_t} \sigma_{\chi,k,p}^2 |q_p|^2} \\ &= \frac{2\sigma_2^2}{2\sigma_2^2 + \lambda_{k,c} \sum_{p=1}^{2LL_t} \sigma_{\chi,k,p}^2 |q_p|^2} \end{aligned} \tag{8.36}$$

To obtain the above expected value, we first define $V = \sum_{p=1}^{2LL_t} \sigma_{\chi,k,p}^2 |q_p|^2$ in which $|q_p| \sim Rayliegh(\frac{\sigma_{q,p}}{\sqrt{2}})$ and $|q_p|$ are independent for all p, then obtain the expected value over the new defined random variable. Let us define

$$S_0 = \left\{ p \middle| \sigma_{\chi,k,p}^2 \sigma_{q,p}^2 \neq \sigma_{\chi,k,l}^2 \sigma_{q,l}^2 \forall l \neq p \right\},$$

i.e., there are $|S_0|$ different values of $\sigma^2_{\chi,k,p}\sigma^2_{q,p}$ such that $\sigma^2_{\chi,k,p}\sigma^2_{q,p} \neq \sigma^2_{\chi,k,l}\sigma^2_{q,l}$ and $\forall l \neq p$. Furthermore, assume that there are J distinct values a_j^2 for $j \in \{1,\ldots,J\}$ for which there are p and l ($p \neq l$ and $p,l \in \{1,\ldots,2LL_t\}$) such that $\sigma^2_{\chi,k,p}\sigma^2_{q,p} = \sigma^2_{\chi,k,l}\sigma^2_{q,l} = a_j^2$. We also define $S_j = \left\{ p \middle| \sigma^2_{\chi,k,p}\sigma^2_{q,p} = a_j^2 \right\}$ for $j \in \{1,\ldots,J\}$ and $N_j = |S_j|$. It follows from [67, p. 876] and the definition of the Gamma distribution that

$$p_V(v) = \sum_{p\in S_0} \frac{\pi_p}{\sigma^2_{\chi,k,p}\sigma^2_{q,p}} e^{-\frac{v}{\sigma^2_{\chi,k,p}\sigma^2_{q,p}}} + \sum_{j=1}^{J}\sum_{p\in S_j} \frac{v^{N_j-1}}{(N_j-1)!(\sigma^2_{\chi,k,p}\sigma^2_{q,p})^{N_j}} e^{-\frac{v}{\sigma^2_{\chi,k,p}\sigma^2_{q,p}}}, \quad (8.37)$$

where $\pi_p = \prod_{l\in S_0, l\neq p} \frac{\sigma^2_{\chi,k,p}\sigma^2_{q,p}}{\sigma^2_{\chi,k,p}\sigma^2_{q,p} - \sigma^2_{\chi,k,l}\sigma^2_{q,l}}$. Therefore, we can write

$$\begin{aligned} E_{|q_1|,\ldots,|q_{2LL_t}|}\left\{ \frac{2\sigma_2^2}{2\sigma_2^2 + \lambda_{k,c}\sum_{p=1}^{2LL_t}\sigma^2_{\chi,k,p}|q_p|^2} \right\} &= E_V\left\{ \frac{2\sigma_2^2}{2\sigma_2^2 + \lambda_{k,c}V} \right\} \\ &= \sum_{p\in S_0} \frac{\pi_p}{2\sigma^2_{\chi,k,p}\sigma^2_{q,p}} \int_0^\infty \frac{2\sigma_2^2}{2\sigma_2^2+\lambda_{k,c}v} e^{-\frac{v}{2\sigma^2_{\chi,k,p}\sigma^2_{q,p}}} \\ &\quad + \sum_{j=1}^{J}\sum_{p\in S_j} \frac{1}{(N_j-1)!(\sigma^2_{\chi,k,p}\sigma^2_{q,p})^{N_j}} \int_0^\infty \frac{2\sigma_2^2 v^{N_j-1}}{2\sigma_2^2+\lambda_{k,c}v} e^{-\frac{v}{\sigma^2_{\chi,k,p}\sigma^2_{q,p}}}. \end{aligned}$$

By using the integral calculated in [71, p. 325. Eq. 2. 3. 6. 14.] and due to the definition of the exponential integral function $E_1(a) = \int_a^\infty \frac{e^{-z}}{z} dz$, we obtain

$$\begin{aligned} PEP_{k,c} &= E_V\left\{ \frac{2\sigma_2^2}{2\sigma_2^2+\lambda_{k,c}V} \right\} \\ &= \sum_{p\in S_0} \frac{\pi_p}{\sigma^2_{\chi,k,p}\sigma^2_{q,p}} \frac{2\sigma_2^2}{\lambda_{k,c}} e^{\frac{2\sigma_2^2}{\lambda_{k,c}\sigma^2_{\chi,k,p}\sigma^2_{q,p}}} E_1\left(\frac{2\sigma_2^2}{\lambda_{k,c}\sigma^2_{\chi,k,p}\sigma^2_{q,p}} \right) \\ &\quad + \sum_{j=1}^{J}\sum_{p\in S_j} \frac{(2\sigma_2^2)^{N_j}(-1)^{N_j-1}}{(N_j-1)!(\lambda_{k,c}\sigma^2_{\chi,k,p}\sigma^2_{q,p})^{N_j}} \left[e^{\frac{2\sigma_2^2}{\lambda_{k,c}\sigma^2_{\chi,k,p}\sigma^2_{q,p}}} E_1\left(\frac{2\sigma_2^2}{\lambda_{k,c}\sigma^2_{\chi,k,p}\sigma^2_{q,p}} \right) \right. \\ &\qquad \left. + \sum_{k=1}^{N_j-1} (k-1)! \left(\frac{-\lambda_{k,c}\sigma^2_{\chi,k,p}\sigma^2_{q,p}}{2\sigma_2^2} \right)^k \right] \quad (8.38) \end{aligned}$$

Finally, for the PEP of the system under block fading channel conditions, we obtain

$$P(\boldsymbol{X}_k \to \boldsymbol{X}'_k) \leq \frac{1}{2} E\left\{ e^{-\frac{1}{4\sigma_2^2}\sum_{c=1}^{r_k}\lambda_{k,c}\left|\boldsymbol{U}_{k,c}^H \boldsymbol{q}(k)\right|^2} \right\} \leq \frac{1}{2}\prod_{c=1}^{r_k} PEP_{k,c}, \quad (8.39)$$

where $PEP_{k,c}$ are given in (8.38).

Table 8.1: Parameters of two different scenarios used to compare the proposed scheme with the scheme in [3].

Scenario	Scheme	N	M	D	$\sigma_{h_i} = \sigma_{g_i}$ $(j, i \in \{1, 2\})$	N_{CP}	T (ms)	Data Rate (kb/s)
S_1	Proposed Scheme	512	100	$1039T_s$	$[1,\ 0.8,\ 0.6]/\sqrt{2}$	522	129.25	7.9226
	Scheme from [3]	1024	2	$1039T_s$	$[1,\ 0.8,\ 0.6]/\sqrt{2}$	1044	258.5	7.9226
S_2	Proposed Scheme	256	100	$527T_s$	$[0.8,\ 0,\ 0.6]$	266	65.25	7.8467
	Scheme from [3]	512	2	$527T_s$	$[0.8,\ 0,\ 0.6]$	532	130.5	7.8467

8.5 Simulation Results

For numerical evaluations we assume that the total occupied bandwidth is 8 kHz (over the frequency band from 12 kHz to 20 kHz). We define f_d as the Doppler frequency shift observed at the destination in Hz, and $\sigma_{h_i} = [\sigma_{h_i,0}, \ldots, \sigma_{h_i,L_{h_i}}]$. We assume $P_i = 1$, $\tau_{h_i,l} = \tau_{g_i,l} = lT_s = 125l$ μs, and $\sigma_{i,1}^2 = 2\sigma_2^2$ $(i \in \{1, 2\})$.

In Figs. 8.7 and 8.8, we compare the performance of the proposed scheme with the performance of the scheme proposed in [3] for different values of f_dT_s under two different scenarios where quadrature phase-shift keying (QPSK) modulated symbols are transmitted over N subcarriers. The parameters of the two scenarios are reported in Table 8.1 in which to make a fair comparison, both schemes are set to the same data transmission rate. We generate time varying Rayleigh fading channel tap gains following the Jakes' model [72]. We chose [3] for comparison since it also considers an OFDM based cooperative transmission with full-duplex AF relays. However, in [3], the relays perform time reversal and symbol complex conjugation as well. In Figs. 8.7 and 8.8, we observe that for the time-invariant channel case ($f_dT_s = 0$), the performance of both schemes are identical. However, for time varying scenarios, the proposed scheme outperforms the scheme proposed in [3]. The reason is that, for the range of the relative delay considered, to attain the same data rate for both schemes, the scheme proposed in [3] transmits longer OFDM blocks (larger N) and as a result more ICI is experienced over the received subcarriers. Obviously, by increas-

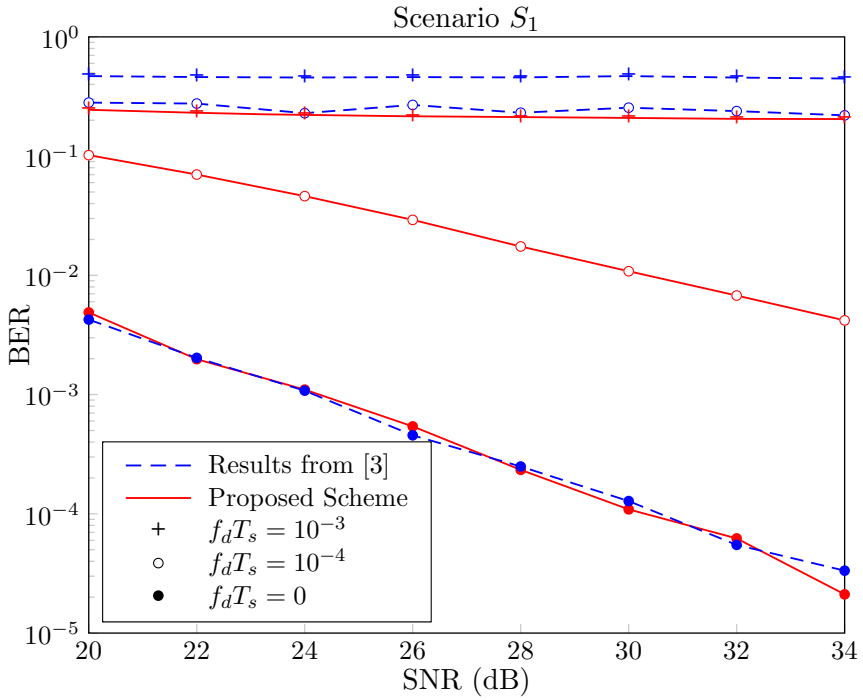

Figure 8.7: Comparison between the performance of the proposed scheme with the scheme proposed in [3] under the scenario S_1.

ing f_d, i.e., faster fading conditions, more ICI are experienced over the subcarriers and the performance becomes worse. We observe that in the SNR range considered, the bit error rate (BER) of the fast fading scenario ($f_dT_s = 10^{-3}$) reaches an error floor. We expect that for higher SNR values, the slow fading scenario ($f_dT_s = 10^{-4}$) converges to an error floor as well. In both cases, we employ a sub-optimal Viterbi decoder to detect the transmitted signal, which assumes that the noise samples over the same subcarrier of different blocks are i.i.d. and ignores the ICI effects to reduce complexity.

In Fig. 8.9, we compare the PEP performance of the proposed scheme with the derived upper bounds for the quasi-static frequency selective channel. We consider transmission of $M = 10$ OFDM blocks with $N = 64$ binary phase-shift keying (BPSK) modulated subcarriers over multipath channels with $\sigma_{h_i} = \sigma_{g_i} = \frac{[1,0.8,0.6]}{\sqrt{2}}$ where there is a relative delay of 145 T_s seconds between the signals received from the two relay

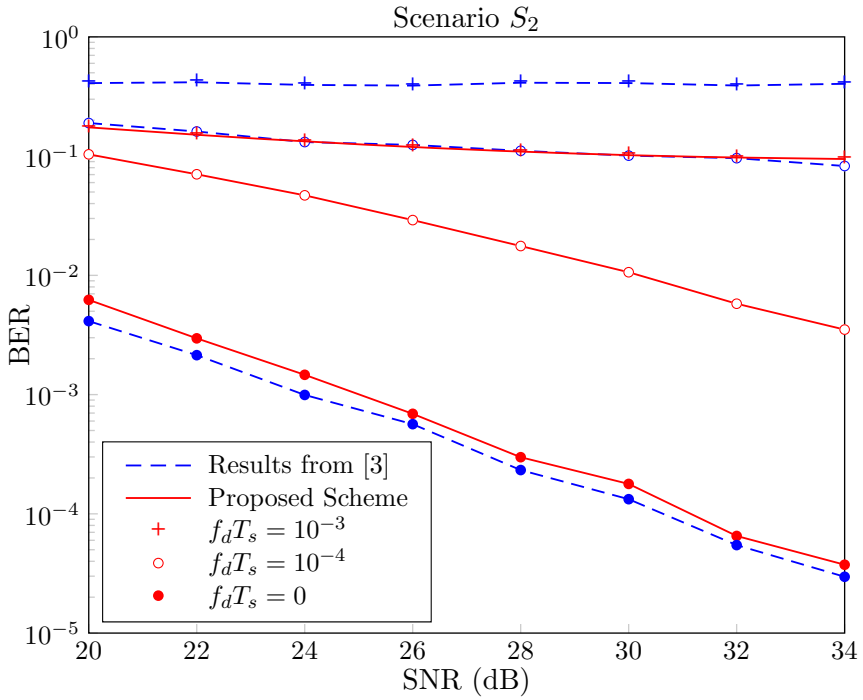

Figure 8.8: Comparison between the performance of the proposed scheme with the scheme proposed in [3] under the scenario S_2.

nodes, i.e., $BD = 1$. Furthermore, we assume that $P_1 = P_2 = 1$, $\sigma^2_{1,1} = \sigma^2_{1,2} = \frac{\sigma_2^2}{10}$. Therefore, the SNR of the system at the receiver is $\frac{5}{3\sigma_2^2}$. We consider $\boldsymbol{X}_k = \mathbf{1}_M$, where $\mathbf{1}_M$ represents an all one vector, and $\boldsymbol{X}'_k = [\mathbf{1}^T_{\frac{M}{2}-1}, -1, \mathbf{1}^T_{\frac{M}{2}}]^T$. Note that the considered case is the worst case scenario, i.e., gives the maximum PEP among all the possible pairwise error events ($s_k^4 - f_k^4$ is minimized). We observe in Fig. 8.9 that by increasing SNR we achieve a tighter upper bound on PEP. Furthermore, as we expect for high SNR values, the diversity order of the system is 2 which is due to the delay diversity structure of the system.

In Fig. 8.10, we compare the derived upper bound on $P(\boldsymbol{X}_k \rightarrow \boldsymbol{X}'_k)$, for $\boldsymbol{X}_k = \mathbf{1}_{10}$ and several different $\boldsymbol{X}'_k$ as given in Table 8.2 under the block fading channel conditions considered in Section 8.4.2.1 with $f_dT_s = 0.01$. We consider the same transmission specs as considered in the study given in Fig. 8.9. As we expected, larger $D(\boldsymbol{X}_k, \boldsymbol{X}'_k)$ (as defined in (8.30)) results in a higher diversity order and a better

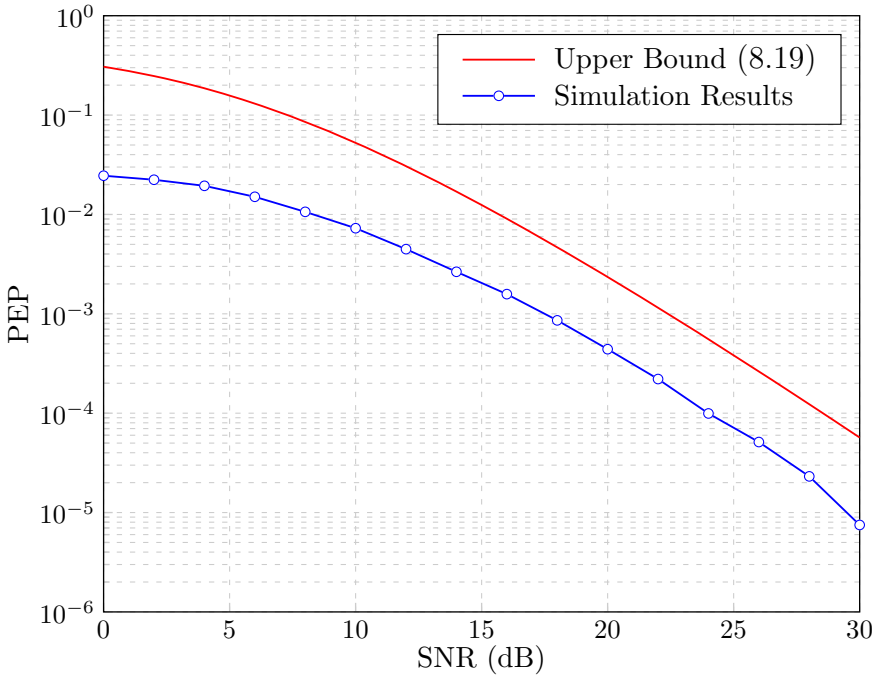

Figure 8.9: Comparison between the upper bound (8.19) and actual PEP for $\boldsymbol{X}_k = \mathbf{1}_M$ and $\boldsymbol{X}'_k = [-1, \mathbf{1}^T_{M-2}, -1]^T$ under quasi-static frequency selective channels.

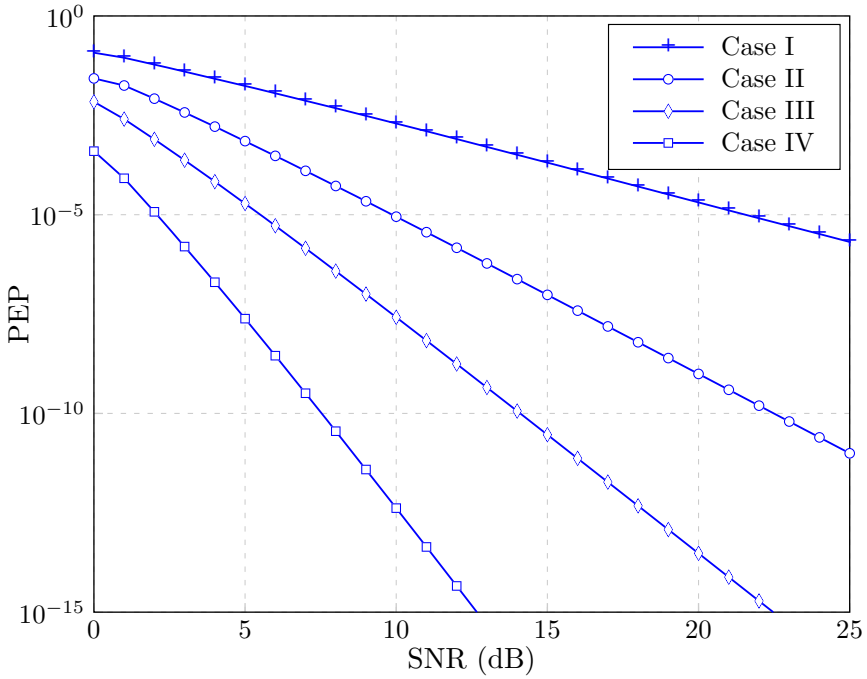

Figure 8.10: Comparison between the upper bound (8.39) for $\boldsymbol{X}_k = \mathbf{1}_{10}$ and $\boldsymbol{X}'_k$ as given in Table 8.2.

Table 8.2: Different cases considered in Fig. 8.10 with $\boldsymbol{X}_k = \mathbf{1}_{10}$.

	Case I	Case II	Case III	Case IV
$\boldsymbol{X}'_k$	$[-1, \mathbf{1}_9^T]^T$	$[-1, \mathbf{1}_8^T, -1]^T$	$[-\mathbf{1}_5^T, \mathbf{1}_5^T]$	$[-1, 1, -1, 1, -1, 1, -1, 1, -1, 1]^T$
$D(\boldsymbol{X}_k, \boldsymbol{X}'_k)$	2	4	6	10

performance which shows the importance of designing appropriate codes to extract the maximum possible diversity out of the system.

8.6 Chapter Summary

We developed a new OFDM transmission scheme for UWA cooperative communication systems suffering from asynchronism among the relays by considering possibly large relative delays among the relays (typical in UWA systems) and time-varying frequency selective channels among the cooperating nodes. The main advantage of the proposed scheme is in managing asynchronism issues arising from excessively large delays among the relays without adding time guards (CP in OFDM-based transmissions) in the order of the maximum possible delay, which increases the spectral efficiency of the system and improves the performance in time-varying channel conditions compared with the existing solutions in the literature. In fact, we showed that independent of the maximum possible delay between the relays, by adding an appropriate CP at the transmitter and appropriate CP removal at the receiver, a delay diversity structure can be obtained at the receiver, where full-duplex AF scheme is applied at the relays. Through numerical examples, we evaluated the performance of the proposed scheme for time-varying multipath channels with Rayleigh fading channel taps, modeling UWA channels. We compared our results with those of the existing schemes and found that while for time invariant channels, the performance is similar, for time varying cases (typical in UWA communications) the proposed scheme is significantly superior.

Chapter 9

Summary and Conclusions

In this work, we studied two different classes of communication systems suffering from synchronization issues namely point-to-point links and cooperative communication scenarios. We first focused on an information theoretic study of P2P channels with synchronization errors. We then turned our attention to design physical layer solutions for asynchronous cooperative communication systems. Our motivation to consider asynchronous cooperative communication systems is derived from cooperative UWA communications applications in which the relative delays among the relays can be excessive.

We developed several analytical lower bounds on the capacity of the insertion/deletion channels by lower bounding the mutual information rate for i.u.d. input sequences. We derived the first analytical lower bound on the capacity of the deletion-AWGN channel and improved the existing analytical lower bound. The presented lower bound on the capacity of the deletion-AWGN channel is very close to the existing simulation based lower bounds for small values of deletion probability. We also improved the existing lower bounds on the random insertion and sticky channel capacities for small values of insertion probability. The lower bound on the capacity of the i.i.d. deletion channel, for small values of deletion probability, is very close to the tightest presented lower bounds, and is in agreement with the first order expansion of the channel capacity for $d \to 0$, while our result is a strict lower bound for all range of d. The derived analytical lower bound results were presented in [73, 74] and a full version is accepted for publication in [75].

We provided several lower bounds on the capacity of binary input symmetric output channels with synchronization errors suffering from other types of impairments

such as substitutions, erasures errors or AWGN. We showed that the capacity of any channel with synchronization errors which can be considered as a cascade of two independent channels (where only the first one suffers from synchronization errors and the second one is a memoryless channel) can be lower bounded in terms of the capacity of the first channel and the parameters of both channels. We considered two classes of channels: binary input symmetric q-ary output channels (e.g., for $q = 3$ a binary input channel with substitutions and erasures) with synchronization errors and BI-AWGN channels with synchronization errors. We gave the first lower bound on the capacity of substitution/erasure channel with synchronization errors and the first analytical result on the capacity of BI-AWGN channel with synchronization errors. We also demonstrated that the lower bounds developed on the capacity of the deletion-AWGN channel for small σ^2 values and the deletion-substitution channel for small values of s improve the existing results. Part of the results was presented in [76] and a full version is submitted for possible publication in [77].

As another contribution in characterizing the binary deletion channel capacity, we provided tighter upper bounds on the binary deletion channel capacity for $d \geq 0.65$. To do so, we first considered a deletion channel with deletion probability $d = \lambda d_1 + (1 - \lambda)d_2$ $(0 \leq \lambda \leq 1)$ as a parallel concatenation of two different deletion channels with deletion probabilities d_1 and d_2, then derived an upper bound on $C_2(d)$ in terms of $C_2(d_1)$, $C_2(d_2)$ and the parameters of the three considered channels. The presented upper bound for $d_2 = 1$ simplifies to $C_2(\lambda d + (1 - \lambda)) \leq \lambda C_2(d)$ where by considering $d = 0.65$ and employing the best existing upper bound in the literature for $d = 0.65$ results into a tighter upper bound for $d \geq 0.65$ compared with the best existing results. The results are submitted for possible publication in [78].

To better characterize the non-binary deletion channel capacity, we presented the first non-trivial upper bound on the $2K$-ary deletion channel capacity. We showed

that for the input symbol set $\{1, \ldots, 2K\}$, a parallel concatenation of K binary deletion channels ($\mathcal{C}_1$, $\ldots$, $\mathcal{C}_K$) all with the deletion probability d, in a way that input symbols $2k-1$ and $2k$ travel only through $\mathcal{C}_k$, is nothing but a $2K$-ary deletion channel with deletion probability d. This new consideration enables us to prove $C_{2K}(d) \leq C_2(d) + (1-d)\log(K)$ in which substituting any upper bound on the binary deletion channel capacity gives an upper bound on the $2K$-ary deletion channel capacity as well. We showed that the new derived upper bounds improve upon the existing trivial upper bound for the entire range of d. The new derived upper bound result on the $2K$-ary input deletion channel capacity is accepted for presentation in [79].

In this thesis, we also considered asynchronous UWA cooperative systems as another class of communication systems suffering from synchronization issues in which the relative delay among the signals received from geographically separated nodes may be excessive. To increase the spectral efficiency of the asynchronous cooperative communication systems with two relays, we proposed a new TR-STBC scheme which needs a smaller overhead to be added in transmitting every two data blocks in comparison with the existing schemes. We obtained the optimal detector structure and proposed a sub-optimal detector with a reduced computational complexity. Through numerical examples, we showed that the proposed scheme achieves full spatial diversity for both optimal and sub-optimal detectors. We observed that our scheme increases the transmission rate significantly in comparison with the scheme achieving perfect Alamouti code structure for all the transmitted symbols in exchange of a small performance loss.

As another signaling solution to combat the timing errors in asynchronous UWA cooperative systems, we developed a new OFDM transmission scheme by considering possibly large relative delays among the relays. The main advantage of

the new proposed scheme is in managing asynchronism issues arising from excessively large delays among the relays without adding time guards (CP in OFDM-based transmissions) in the order of the maximum possible delay. In fact, we showed that independent of the maximum possible delay between the relays, by adding an appropriate CP at the transmitter and performing an appropriate CP removal at the receiver, a delay diversity structure can be obtained at the receiver, where full-duplex AF scheme is applied at the relays. We provided the PEP analysis of the system under both quasi-static frequency selective and time varying (block fading) frequency selective channels showing that the system achieves full spatial diversity under both channel conditions. Furthermore, the PEP analysis for the time varying block fading case shows the ability of the system in providing time diversity by employing appropriate space-time coding structure. We evaluated the performance of the scheme by numerical evaluations for both time varying and static frequency selective fading channels. By comparing our scheme with the existing ones, we observed that for time varying channels our scheme is significantly superior and its performance is similar to that of the existing approaches for the time invariant cases. The manuscript on the results for the proposed delay diversity scheme is submitted for possible publication in [80].

There are several directions of research which can be pursued following the investigation provided in this thesis.

In providing analytical lower bounds on the capacity of the insertion and deletion channels, we focused on i.u.d. input distributions. It appears that, the approach taken to derive the capacity lower bounds can be modified to obtain tighter lower bounds by considering other types of input distributions. For example, we expect that by considering input sequences with geometrically distributed run lengths one can improve upon the existing lower bounds on the capacity of the insertion/deletion channels.

Another possible line of research is the information theoretic study of the MIMO communication systems suffering from synchronization errors. As a starting point, one needs to prove that the Shannon's theorem applies in such a channel. To do so, the first step is to prove the information stability of the channel for which two results needs to be established: existence of the information capacity of MIMO channels with synchronization errors and existence of an information stable sequence which achieves the information capacity of the channel. For the case of the MIMO systems with deletion errors, i.e., the channels from each transmit antenna to each receive antenna experience independent deletion processes, we are able to prove the former statement (not included in the thesis); however, the proof of the latter is not easy and requires development of some new tools. After proving that Shannon's theorem applies in the MIMO channels with synchronization errors, one can extend some existing upper and lower bounds on the capacity of the single input single output channels with synchronization errors to the MIMO case.

In the case of the proposed spectrally efficient Alamouti scheme, one possible extension is to modify the scheme for frequency selective channels in which to combat the multipath effects of the channel along with the asynchronism issues, one needs to add longer overhead to cover both the maximum possible relative delay and delay spread of the channels. Then, the proposed optimal and sub-optimal detectors for flat fading channels need to be modified to the case of the multipath channels along with considering different equalization algorithms.

Another possible extension of our work in this thesis is in the design of space-time codes for the proposed OFDM-based delay diversity scheme for the case of asynchronous cooperative systems. The provided pairwise error probability analysis under time-varying block fading frequency selective channel conditions can be used as the main code design criteria. In Chapter 8, we provided several examples of the

upper bound on the pairwise error probability of the system for different pairs of codewords which show the importance of the appropriate code design for the considered time-varying channel conditions. Based on these initial thoughts, we expect that by designing specific space-time codes, the maximum possible diversity can be extracted out of the asynchronous systems for the considered OFDM-based delay diversity scheme.

A second line of research following the OFDM-based delay diversity scheme is to extend the proposed scheme to the case of asynchronous two way relaying systems in which asynchronism among the signals occurs both at the relay nodes and source nodes. That is, the signals transmitted by the source nodes are received unaligned at the relay nodes and the signals broadcast by relay nodes are received unaligned at the source nodes. One possible extension in the case with two full-duplex amplify and forward relays and two active users is to consider the same transmission strategy as introduced in Chapter 8 for both users. Then, each each by knowing the relative delays among the signals can remove the effect of its own signal which leads to a delay diversity structure.

REFERENCES

[1] Z. Li and X.-G. Xia, "A simple Alamouti space-time transmission scheme for asynchronous cooperative systems," *IEEE Signal Processing Letters*, vol. 14, no. 11, pp. 804–807, Nov. 2007.

[2] X. Li, "Space-time coded multi-transmission among distributed transmitters without perfect synchronization," *IEEE Signal Processing Letters*, vol. 11, no. 12, pp. 948–951, Dec. 2004.

[3] Z. Li, X.-G. Xia, and M. H. Lee, "A simple orthogonal space-time coding scheme for asynchronous cooperative systems for frequency selective fading channels," *IEEE Trans. Comms.*, vol. 58, no. 8, pp. 2219–2224, Aug. 2010.

[4] R. Gallager, "Sequential decoding for binary channels with noise and synchronization errors," *Tech. Rep., MIT Lincoln Lab. Group Report*, Oct. 1961.

[5] E. Drinea and M. Mitzenmacher, "Improved lower bounds for the capacity of i.i.d. deletion and duplication channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 8, pp. 2693–2714, Aug. 2007.

[6] J. Hu, T. M. Duman, M. F. Erden, and A. Kavcic, " Achievable information rates for channels with insertions, deletions and intersymbol interference with i.i.d. inputs," *IEEE Trans. Comms.*, vol. 58, no. 4, pp. 1102–1111, Apr. 2010.

[7] E. Ratzer and D. MacKay, "Codes for channels with insertions, deletions and substitutions," in *2nd International Symposium on Turbo Codes and Related Topics*, 2000, pp. 149–156.

[8] X. Zhang, M. Tao, and C. S. Ng, " Utility-based wireless resource allocation for variable rate transmission," *IEEE Trans. Wireless Comms.*, vol. 7, no. 9, pp. 3292–3296, Sep. 2008.

[9] G. Hughes, "Patterned Media," *The Physics of Ultra-High-Density Magnetic Recording*, M. L. Plumer, J. V. Ek, and D. Weller, Eds. Berlin, Germany: Springer-Verlag, 2001, ch.7.

[10] R. L. Dobrushin, "Shannon's theorems for channels with synchronization errors," *Probs. Inf. Transm.*, vol. 3, no. 4, pp. 11–26, 1967.

[11] M. Mitzenmacher and E. Drinea, "A simple lower bound for the capacity of the deletion channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4657–4660, Oct. 2006.

[12] A. Kirsch and E. Drinea, "Directly lower bounding the information capacity for channels with i.i.d. deletions and duplications," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 86 –102, Jan. 2010.

[13] D. Fertonani and T. M. Duman, "Novel bounds on the capacity of the binary deletion channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2753–2765, June 2010.

[14] D. Fertonani, T. M. Duman, and M. F. Erden, "Bounds on the capacity of channels with insertions, deletions and substitutions," *IEEE Trans. Comms.*, vol. 59, no. 1, pp. 2–6, Jan. 2011.

[15] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Trans. Inf. Theory*, vol. 50, no. 12, pp. 3062– 3080, Dec. 2004.

[16] M. Stojanovic, "Underwater acoustic communications," *Wiley Encyclopedia of Electrical and Electronics Engineering*, vol. 22, pp. 688–698, 1999.

[17] H. M. Wang and X.-G. Xia, "Asynchronous cooperative communication systems: A survey on signal designs," *SCIENCE CHINA Inf. Sciences*, vol. 54, no. 8, pp. 1547–1531, Aug. 2011.

[18] Y. Li, W. Zhang, and X.-G. Xia, "Distributive high-rate space-frequency codes achieving full cooperative and multipath diversities for asynchronous cooperative communications," *IEEE Trans. on Vehicular Tech.*, vol. 58, no. 1, pp. 207–217, Jan. 2009.

[19] M. Mitzenmacher, "Capacity bounds for sticky channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, pp. 72–77, Jan. 2008.

[20] Z. Liu and M. Mitzenmacher, "Codes for deletion and insertion channels with segmented errors," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 224–232, Jan. 2010.

[21] W. Zeng, J. Tokas, R. Motwani, and A. Kavcic, "Bounds on mutual information rates of noisy channels with timing errors," in *Proc. of IEEE Int. Symp. on Inf. Theory (ISIT)*, 2005, pp. 709–713.

[22] S. Diggavi and M. Grossglauser, "On information transmission over a finite buffer channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1226–1237, March 2006.

[23] R. G. Gallager, *Information Theory and Reliable Communication*. Wiley, 1968.

[24] S. Z. Stambler, "Memoryless channels with synchronization errors: The general case," *Probs. Inf. Transm.*, vol. 6, no. 3, pp. 223–229, 1970.

[25] N. D. Vvedenskaya and R. L. Dobrushin, "The computation on a computer of the channel capacity of a line with symbol drop-out," *Problems of Information Transmission*, vol. 4, no. 3, pp. 92–95, 1968.

[26] S. N. Diggavi and M. Grossglauser, "On transmission over deletion channels," in *Proc. of the Annual Allerton Conf. on Communication, Control and Computing*, vol. 39, no. 1, 2001, pp. 573–582.

[27] E. Drinea and M. Mitzenmacher, "On lower bounds for the capacity of deletion channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 10, pp. 4648–4657, Oct. 2006.

[28] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, July 1972.

[29] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[30] A. Kavcic and R. Motwani, "Insertion/deletion channels: Reduced-state lower bounds on channel capacities," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, 2004, p. 229.

[31] S. Diggavi, M. Mitzenmacher, and H. Pfister, "Capacity upper bounds for deletion channels," in *Proceedings of the International Symposium on Information Theory (ISIT)*, 2007, pp. 1716–1720.

[32] Y. Kanoria and A. Montanari, "On the deletion channel with small deletion probability," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, June 2010, pp. 1002–1006.

[33] A. Kalai, M. Mitzenmacher, and M. Sudan, "Tight asymptotic bounds for the deletion channel with small deletion probabilities," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, June 2010, pp. 997 –1001.

[34] M. Ramezani and M. Ardakani, "On the capacity of duplication channels," *accepted for publication in IEEE Trans. on Comms.*, 2013.

[35] F. Wang, T. Duman, and D. Aktas, "Capacity bounds and concatenated codes over segmented deletion channels," *accepted for publication in IEEE Transactions on Communications*, 2013.

[36] H. Mercier, V. Tarokh, and F. Labeau, "Bounds on the capacity of discrete memoryless channels corrupted by synchronization and substitution errors," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4306–4330, July 2012.

[37] T. M. Duman and A. Ghrayeb, *Coding for MIMO Communication Systems*. Wiley, 2007.

[38] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath, "Capacity limits of MIMO channels," *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 5, pp. 684–702, June 2003.

[39] A. J. Paulraj, D. A. Gore, R. U. Nabar, and H. Bolcskei, "An overview of MIMO communications– a key to Gigabit wireless," *Proceedings of the IEEE*, vol. 92, no. 2, pp. 198–218, Feb. 2004.

[40] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, March 1998.

[41] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE Journal on Selected Areas in Comms.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[42] K. Azarian, H. El Gamal, and P. Schniter, "On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4152–4172, Dec. 2005.

[43] E. C. van der Mulen, "Three-terminal communication channels," *Advances in applied Probability*, vol. 3, pp. 120–154, 1971.

[44] T. M. Cover and A. El Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, pp. 572–584, Sep. 1979.

[45] G. Susinder Rajan and B. Sundar Rajan, "OFDM based distributed space time coding for asynchronous relay networks," in *IEEE International Conference on Communications (ICC)*, 2008, pp. 1118–1122.

[46] O.-S. Shin, A. M. Chan, H. T. Kung, and V. Tarokh, " Design of an OFDM cooperative space-time diversity system," *IEEE Trans. on Vehicular Tech.*, vol. 56, no. 4, pp. 2203–2215, 2007.

[47] Y. Mei, Y. Hua, A. Swami, and B. Daneshrad, "Combating synchronization errors in cooperative relays," in *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, 2005, pp. 369–372.

[48] Y. Li and X.-G. Xia, "Full diversity distributed space-time trellis codes for asynchronous cooperative communications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2005, pp. 911–915.

[49] ——, "A family of distributed space-time trellis codes with asynchronous cooperative diversity," *IEEE Transactions on Communications*, vol. 55, no. 4, pp. 790–800, April 2007.

[50] M. O. Damen and A. R. Hammons, " Delay-tolerant distributed-TAST codes for cooperative diversity," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3755–3773, Oct. 2007.

[51] Z. Zhong, S. Zhu, and A. Nallanathan, "Distributed space-time trellis code for asynchronous cooperative communications under frequency-selective channels," *IEEE Trans. on Wireless Comms.*, vol. 8, no. 2, pp. 796–805, Feb. 2009.

[52] ——, "Delay tolerant distributed linear convolutional space-time code with minimum memory length under frequency-selective channels," *IEEE Trans. on Wireless Comms.*, vol. 8, no. 8, pp. 3944–3949, Aug. 2009.

[53] Y. Shang and X.-G. Xia, "Space-time trellis codes with asynchronous full diversity up to fractional symbol delays," *IEEE Trans. on Wireless Comms.*, vol. 7, no. 7, pp. 2473–2479, July 2008.

[54] A. R. Hammons Jr and H. El-Gamal, "On the theory of space-time codes for PSK modulation," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 524–542, March 2000.

[55] H.-F. Lu and P. V. Kumar, " Rate-diversity tradeoff of space-time codes with fixed alphabet and optimal constructions for PSK modulation," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2747–2751, Oct. 2003.

[56] H. El-Gamal and M. O. Damen, "Universal space-time coding," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, pp. 1097–1119, May 2003.

[57] A. R. Bahai, B. R. Saltzberg, and M. Ergen, *Multi-carrier digital communications: Theory and applications of OFDM.* Springer, 2004.

[58] R. L. Dobrushin, "General formulation of Shannon's main theorem on information theory," *American Math. Soc. Trans.*, vol. 33, pp. 323–438, 1963.

[59] M. Dalai, "A new bound on the capacity of the binary deletion channel with high deletion probabilities," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2011, pp. 499–502.

[60] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* Wiley, 2006.

[61] Y. Kanoria and A. Montanari, "Optimal coding for the deletion channel with small deletion probability," *ArXiv e-prints:1104.5546[cs.IT]*, Apr. 2011.

[62] J. Laneman and G. Wornell, "Distributed space-time-coded protocols for exploiting cooperative diversity in wireless networks," *IEEE Trans. on Inf. Theory*, vol. 49, no. 10, pp. 2415–2425, Oct. 2003.

[63] M. Vajapeyam, S. Vedantam, U. Mitra, J. C. Preisig, and M. Stojanovic, "Distributed space-time cooperative schemes for underwater acoustic communications," *IEEE Journal of Oceanic Engineering*, vol. 33, no. 4, pp. 489–501, Oct. 2008.

[64] Z. Han, Y. L. Sun, and H. Shi, "Cooperative transmission for underwater acoustic communications," in *IEEE Int. Conf. on Comms. (ICC)*, May 2008, pp. 2028–2032.

[65] K. Tu, T. M. Duman, J. G. Proakis, and M. Stojanovic, " Cooperative MIMO-OFDM communications: Receiver design for Doppler-distorted underwater acoustic channels," in *Proc. of Asilomar Conf. on Signals, Systems and Computers*, Nov. 2010, pp. 1335–1339.

[66] D. Gore, S. Sandhu, and A. Paulraj, "Delay diversity codes for frequency selective channels," in *IEEE Int. Conf. on Comms. (ICC)*, vol. 3, 2002, pp. 1949–1953.

[67] J. G. Proakis and M. Salehi, *Digital Communications*. McGraw-Hill, 2007.

[68] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series*. Gordon and Breach Science Publishers, 1986, vol. 2.

[69] B. Lu, X. Wang, and K. Narayanan, "LDPC-based space-time coded OFDM systems over correlated fading channels: Performance analysis and receiver design," *IEEE Transactions on Communications*, vol. 50, no. 1, pp. 74–88, Jan. 2002.

[70] S. G. Wilson, *Digital Modulation and Coding*. New York: Prentice Hall, 1996.

[71] A. P. Prudnikov, Y. A. Brychkov, and O. I. Marichev, *Integrals and Series*. Gordon and Breach Science Publishers, 1986, vol. 1.

[72] W. C. Jakes, *Microwave Mobile Communication*. New York, NY: Wiley, 1974.

[73] M. Rahmati and T. M. Duman, "Analytical lower bounds on the capacity of deletion channels," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2011.

[74] ——, "Achievable rates over insertion channels," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2011.

[75] ——, "Bounds on the capacity of random insertion and deletion-additive noise channels," *accepted for publication in IEEE Trans. Inf. Theory, ArXiv e-prints:1101.1310v3[cs.IT].*

[76] ——, "On the capacity of binary input symmetric q-ary output channels with synchronization errors," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2012, pp. 691–695.

[77] ——, "Achievable rates for noisy channels with synchronization errors," *submitted to IEEE Trans. Inf. Theory, ArXiv e-prints:1203.6396[cs.IT]*, Mar. 2012.

[78] ——, "A note on the deletion channel capacity," *submitted to IEEE Trans. Inf. Theory, ArXiv e-prints:1211.2497v1[cs.IT]*, Nov. 2012.

[79] ——, "An upper bound on the capacity of non-binary deletion channels," *accepted for presentation in IEEE Int. Symp. Inf. Theory (ISIT) 2013, ArXiv e-prints:1301.6599v1[cs.IT].*

[80] ——, "Achieving delay diversity in asynchronous underwater acoustic (UWA) cooperative communication systems," *submitted to IEEE Trans. on Wireless Communications*, March 2013.

APPENDIX A

PART OF PROOF OF PROPOSITION 2

For an i.i.d. deletion channel, for a given input sequence $\boldsymbol{x}(b; n_1, ..., n_K)$, we have

$$
\begin{aligned}
&H\left(\boldsymbol{Y}\middle|\boldsymbol{x}(b;n;K^x)\right) \\
&\quad\le -\sum_{j=0}^{n}\sum_{j_1+...+j_{K^x}=j}\binom{n_1^x}{j_1}...\binom{n_{K^x}^x}{j_{K^x}}d^j(1-d)^{n-j}\log\left(\binom{n_1^x}{j_1}...\binom{n_{K^x}^x}{j_{K^x}}d^j(1-d)^{n-j}\right) \\
&\quad= -\sum_{j=0}^{n}d^j(1-d)^{n-j}\log\left(d^j(1-d)^{n-j}\right)\sum_{j_1+...+j_{K^x}=j}\binom{n_1^x}{j_1}...\binom{n_{K^x}^x}{j_{K^x}} \\
&\qquad-\sum_{j=0}^{n}d^j(1-d)^{n-j}\sum_{j_1+...+j_{K^x}=j}\binom{n_1^x}{j_1}...\binom{n_{K^x}^x}{j_{K^x}}\log\left(\binom{n_1^x}{j_1}...\binom{n_{K^x}^x}{j_{K^x}}\right) \\
&\quad=nH_b(d)-\sum_{j=0}^{n}d^j(1-d)^{n-j}\sum_{k=1}^{K^x}\sum_{j_k=0}^{j}\binom{n_k^x}{j_k}\binom{n-n_k^x}{j-j_k}\log\binom{n_k^x}{j_k}, \qquad \text{(A.1)}
\end{aligned}
$$

where, we have used the generalized Vandermonde's identity, that is,

$$\sum_{j_1+...+j_{K^x}=j}\binom{n_1^x}{j_1}...\binom{n_{K^x}^x}{j_{K^x}}=\binom{n}{j},$$

and the results

$$
\begin{aligned}
&\sum_{j_1+...+j_{K^x}=j}\binom{n_1^x}{j_1}...\binom{n_{K^x}^x}{j_{K^x}}\log\left(\binom{n_1^x}{j_1}...\binom{n_{K^x}^x}{j_{K^x}}\right) \\
&\qquad=\sum_{j_1+...+j_{K^x}=j}\binom{n_1^x}{j_1}...\binom{n_{K^x}^x}{j_{K^x}}\sum_{k=1}^{K^x}\log\binom{n_k^x}{j_k} \\
&\qquad=\sum_{k=1}^{K^x}\sum_{j_k=0}^{j}\binom{n_k^x}{j_k}\binom{n-n_k^x}{j-j_k}\log\binom{n_k^x}{j_k},
\end{aligned}
$$

and $-\sum_{j=0}^{n}\binom{n}{j}d^j(1-d)^{n-j}\log\left(d^j(1-d)^{n-j}\right)=nH_b(d)$.

APPENDIX B

PROOF OF PROPOSITION 4

In the proof of Proposition 2, we considered outputs of the deletion channel resulting from different deletion patterns $D(n;K;j)$ for a given $\boldsymbol{x}(b;n;K)$ as if they are distinct output sequences which results in an upper bound on the conditional output entropy of the deletion channel. Here, the output bits of the deletion channel are input to the BSC, thus we use a similar approach, that is, for a given $\boldsymbol{x}(b;n;K)$, we consider outputs of the deletion channel from different deletion patterns $(D(n;K;j) * \boldsymbol{x}(b;n;K))$ as distinct input sequences into the BSC and also consider the sequences resulting from such sequences at the output of the BSC as distinct output sequences. Employing the result given in Eqn. (3.17), an upper bound on the conditional output entropy is found. For a given $\boldsymbol{x} = (b; n_1, n_2, \ldots, n_k)$, we have

$$P\Bigg(D(n;K;j) = (j_1,\ldots,j_K)\Bigg|\boldsymbol{x}(b;n_1,\ldots,n_K)\Bigg) = \binom{n_1}{j_1}\ldots\binom{n_K}{j_K} d^j(1-d)^{n-j}, \quad \text{(B.1)}$$

and for every $D(n;K;j)$, we can write

$$P(\boldsymbol{y}'|D(n;K;j)*\boldsymbol{x}(n;K)) = \begin{cases} s^e(1-s)^{n-j-e} & if \quad |\boldsymbol{y}'| = n-j \\ 0 & otherwise \end{cases}, \quad \text{(B.2)}$$

where $e = d_H\left(\boldsymbol{y}'; D(n;K;j)*\boldsymbol{x}(n;K)\right)$, and $d_H = (\boldsymbol{a};\boldsymbol{b})$ is the Hamming distance between two sequences $\boldsymbol{a}$ and $\boldsymbol{b}$. On the other hand, for every output sequence of length $n-j$, conditioned on a given input $\boldsymbol{x}(n;K)$, we have

$$P\left(\boldsymbol{y}'(n-j)\middle|\boldsymbol{x}(n;K)\right) = \sum_{D\in\mathcal{D}_K^n(j)} P\left(\boldsymbol{y}'(n-j)\middle|D,\boldsymbol{x}(n;K)\right) P\left(D\middle|\boldsymbol{x}(n;K)\right) \quad \text{(B.3)}$$

and we can write

$$\begin{aligned} &-P(\boldsymbol{y}'|\boldsymbol{x})\log\left(P(\boldsymbol{y}'|\boldsymbol{x})\right) \\ &\qquad = -\sum_{D\in\mathcal{D}_K^n(j)} P(\boldsymbol{y}'|D*\boldsymbol{x})P(D|\boldsymbol{x})\log\left(\sum_{D'\in\mathcal{D}_K^n(j)} P(\boldsymbol{y}'|D'*\boldsymbol{x})P(D'|\boldsymbol{x})\right) \\ &\qquad \leq -\sum_{D\in\mathcal{D}_K^n(j)} P(\boldsymbol{y}'|D*\boldsymbol{x})P(D|\boldsymbol{x})\log\left(P(\boldsymbol{y}'|D*\boldsymbol{x})P(D|\boldsymbol{x})\right), \quad \text{(B.4)} \end{aligned}$$

where the inequality follows from the result in Eqn. (3.17). Hence, for $\boldsymbol{x}(b;n;K^x) = (b;n_1^x,\ldots,n_{K^x}^x)$, we can write

$$H\left(\boldsymbol{Y}'\middle|\boldsymbol{x}(b;n;K^x)\right) = -\sum_{j=0}^{n}\sum_{\boldsymbol{y}'\in\mathcal{Y}_{-j}^d} P(\boldsymbol{y}'(n-j)|\boldsymbol{x})\log\left(P(\boldsymbol{y}'(n-j)|\boldsymbol{x})\right)$$

$$\leq -\sum_{j=0}^{n}\sum_{\boldsymbol{y}'\in\mathcal{Y}_{-j}^d}\sum_{D\in\mathcal{D}_K^n(j)} P(\boldsymbol{y}'|D*\boldsymbol{x})P(D|\boldsymbol{x})\log\left(P(\boldsymbol{y}'|D*\boldsymbol{x})P(D|\boldsymbol{x})\right), \quad \text{(B.5)}$$

where the inequality is obtained from the expression in (B.4). From Eqns. (B.1) and (B.2), we have

$$H\left(\boldsymbol{Y}'\middle|\boldsymbol{x}(b;n;K^x)\right) \leq -\sum_{j=0}^{n}\sum_{e=0}^{n-j}\binom{n-j}{e}\sum_{j_1+\ldots+j_K=j} s^e(1-s)^{n-j-e}\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}$$
$$\times d^j(1-d)^{n-j}\log\left(\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}d^j(1-d)^{n-j}s^e(1-s)^{n-j-e}\right), \quad \text{(B.6)}$$

where we used the fact that there are $\binom{n-j}{e}$, distinct output sequences of length $n-j$ resulting from e substitution errors into a given input $\boldsymbol{x}$, i.e.,

$$e = d_H\left(\boldsymbol{y}'(n-j); D(n;K;j)*\boldsymbol{x}(n;K)\right).$$

We then obtain

$$H\left(\boldsymbol{Y}'\middle|\boldsymbol{x}(b;n;K^x)\right)$$
$$\leq -\sum_{j=0}^{n}\sum_{j_1+\cdots+j_K=j}\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}d^j(1-d)^{n-j}\Bigg[-(n-j)H_b(s)$$
$$+\log\left(\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}d^j(1-d)^{n-j}\right)\Bigg]$$
$$= nH_b(d) - \sum_{j=0}^{n}\sum_{j_1+\cdots+j_K=j}\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}d^j(1-d)^{n-j}\Bigg[-n(1-d)H_b(s)$$
$$+\log\left(\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}\right)\Bigg]$$
$$= nH_b(d) + n(1-d)H_b(s)$$
$$-\sum_{j=0}^{n}d^j(1-d)^{n-j}\sum_{k=1}^{K}\sum_{j_k=0}^{j}\binom{n_k}{j_k}\binom{n-n_k}{j-j_K}\log\binom{n_k}{j_k}, \quad \text{(B.7)}$$

where the details are similar to the steps leading to Eqn. (A.1). By considering i.u.d. input sequences and due to the fact that the last term in Eqn. (B.6) is same as the last term of Eqn. (A.1), we obtain

$$H(\boldsymbol{Y}'|\boldsymbol{X}) \leq nH_b(d) - \sum_{j=1}^{n} W_j(n) \binom{n}{j} d^j (1-d)^{n-j} + n(1-d)H_b(s), \tag{B.8}$$

which concludes the proof.

APPENDIX C

PROOF OF PROPOSITION 6

For an i.i.d. deletion-AWGN channel, for a given $\boldsymbol{x}(b;n;K)$ and a fixed j, we have

$$\begin{aligned} f_{\widetilde{\boldsymbol{y}}}(\eta|\boldsymbol{x}(b;n;K),j) &= \sum_{D\in\mathcal{D}_K^n(j)} f_{\widetilde{\boldsymbol{y}}}(\eta|\boldsymbol{x}(b;n;K),D)P(D|\boldsymbol{x}(b;n;K)) \\ &= \sum_{D\in\mathcal{D}_K^n(j)} f_{\widetilde{\boldsymbol{y}}}(\eta|\alpha(D,\boldsymbol{x}))P(D|\boldsymbol{x}(b;n;K)) \\ &= \sum_{D\in\mathcal{D}_K^n(j)} f_{\widetilde{y}_1\ldots\widetilde{y}_j}(\eta_1\ldots\eta_j|\alpha_1\ldots\alpha_j)P(D|\boldsymbol{x}(b;n;K)) \\ &= \sum_{D\in\mathcal{D}_K^n(j)} f_{\widetilde{y}_1}(\eta_1|\alpha_1)\ldots f_{\widetilde{y}_j}(\eta_j|\alpha_j)P(D|\boldsymbol{x}(b;n;K)), \quad \text{(C.1)} \end{aligned}$$

where $\alpha(D,\boldsymbol{x}) = 1-2(D*\boldsymbol{x})$, i.e., $\alpha_i(D,\boldsymbol{x}) \in \{1,-1\}$, and the last equality follows the fact that the noise samples $\boldsymbol{z}_i$'s are independent and $\alpha_i(D,\boldsymbol{x})$'s are also independent. By employing

$$f_{\widetilde{y}_i}(\eta_i|\alpha_i(D,\boldsymbol{x})) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}\right),$$

and

$$P\left(D(n;K;j)\middle|\boldsymbol{x}(b;n;K),j\right) = \frac{\binom{n_1}{j_1}\ldots\binom{n_K}{j_K}}{\binom{n}{j}},$$

we can write

$$\begin{aligned} f_{\widetilde{\boldsymbol{y}}}(\eta|\boldsymbol{x}(b;n;K),j) &= \frac{1}{(\sqrt{2\pi}\sigma)^j}\sum_{D\in\mathcal{D}_K^n(j)}\prod_{i=1}^{j}\exp\left(\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}\right)P\left(D|\boldsymbol{x}(b;n;K),j\right) \\ &= \frac{1}{(\sqrt{2\pi}\sigma)^j}\sum_{j_1+\ldots+j_K=j}\frac{\binom{n_1}{j_1}\ldots\binom{n_K}{j_K}}{\binom{n}{j}}\prod_{i=1}^{j}\exp\left(\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}\right), \end{aligned} \quad \text{(C.2)}$$

Therefore, we obtain

$$
\begin{aligned}
&h(\widetilde{\boldsymbol{Y}}|\boldsymbol{x},j)\\
&=-\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{(\sqrt{2\pi}\sigma)^j}\sum_{j_1+\ldots+j_K=j}\frac{\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}}{\binom{n}{j}}\prod_{i=1}^{j}\exp\left(\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}\right)\times\\
&\quad\times\log\left(\frac{1}{(\sqrt{2\pi}\sigma)^j}\sum_{j_1'+\ldots+j_K'=j}\frac{\binom{n_1}{j_1'}\cdots\binom{n_K}{j_K'}}{\binom{n}{j}}\prod_{i=1}^{j}\exp\left(\frac{-(\eta_i-\alpha_i(j',\boldsymbol{x}))^2}{2\sigma^2}\right)\right)d\eta_1\ldots d\eta_j\\
&=j\log(\sqrt{2\pi}\sigma)+\log\binom{n}{j}\\
&\quad-\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{(\sqrt{2\pi}\sigma)^j}\sum_{j_1+\ldots+j_K=j}\frac{\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}}{\binom{n}{j}}\prod_{i=1}^{j}\exp\left(\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}\right)\times\\
&\quad\times\left[\log\left(\sum_{j_1'+\ldots+j_K'=j}\binom{n_1}{j_1'}\cdots\binom{n_K}{j_K'}\prod_{i=1}^{j}\exp\left(\frac{-(\eta_i-\alpha_i(j',\boldsymbol{x}))^2}{2\sigma^2}\right)\right)\right]d\eta_1\ldots d\eta_j,
\end{aligned}
\tag{C.3}
$$

where we used the result of the generalized Vandermonde's identity and also the fact that $\int_{-\infty}^{\infty}f_{\widetilde{y}_i}(\eta_i|\bar{y}_i)d\eta_i=1$. By using the inequality

$$
\begin{aligned}
&\sum_{j_1'+\ldots+j_K'=j}\binom{n_1}{j_1'}\cdots\binom{n_K}{j_K'}\prod_{i=1}^{j}\exp\left(\frac{-(\eta_i-\alpha_i(j',\boldsymbol{x}))^2}{2\sigma^2}\right)\\
&\qquad\qquad\geq\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}\prod_{i=1}^{j}\exp\left(\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}\right),
\end{aligned}
$$

which holds for every $j_1+\ldots+j_K=j$, we can write

$$
\begin{aligned}
&h(\widetilde{\boldsymbol{Y}}|\boldsymbol{x},j)\leq j\log(\sqrt{2\pi}\sigma)+\log\binom{n}{j}\\
&\quad-\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\frac{1}{(\sqrt{2\pi}\sigma)^j}\sum_{j_1+\ldots+j_K=j}\frac{\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}}{\binom{n}{j}}\prod_{i=1}^{j}\exp\left(\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}\right)\times\\
&\quad\times\left[\log\left(\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}\prod_{i=1}^{j}\exp\left(\frac{-(\eta_i-\alpha_i(D,\boldsymbol{x}))^2}{2\sigma^2}\right)\right)\right]d\eta_1\ldots d\eta_j\\
&=(j)\log(\sqrt{2\pi e}\sigma)+\log\binom{n}{j}-\sum_{j_1+\ldots+j_K=j}\frac{\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}}{\binom{n}{j}}\log\left(\binom{n_1}{j_1}\cdots\binom{n_K}{j_K}\right).
\end{aligned}
\tag{C.4}
$$

By considering i.u.d. input sequences, we have

$$h(\widetilde{\boldsymbol{Y}}|\boldsymbol{X},T) \leq \sum_{j=0}^{n} \binom{n}{j} d^j (1-d)^j \sum_{\boldsymbol{x}\in\mathcal{X}} \frac{1}{2^n} h(\widetilde{\boldsymbol{Y}}|\boldsymbol{x},d)$$

$$= n(1-d)\log(\sqrt{2\pi e}\sigma) + \sum_{j=0}^{n} \binom{n}{j} d^j (1-d)^j \left[\log \binom{n}{j} - W_j(n)\right], \quad \text{(C.5)}$$

where $W_j(n)$ is given in Eqn. (3.5), and the result is obtained by following the same steps as in the computation leading to (3.19). Therefore, by substituting Eqn. (C.5) into Eqn. (3.37), Eqn. (3.36) is obtained which concludes the proof.

APPENDIX D

OUTPUT SEQUENCE DISTRIBUTION FOR THE STICKY CHANNEL

Here, we obtain the exact probability of output sequences resulting from at most two duplications in transmitting i.u.d. input sequences. Obviously, any output sequence of length n results only from the same transmitted sequence with no duplications, and with probability of $(1-i)^n$. We can write

$$Q\left(\boldsymbol{y}(n)|\boldsymbol{x}\right) = \begin{cases} (1-i)^n & \boldsymbol{y}(n) = \boldsymbol{x} \\ 0 & \boldsymbol{y}(n) \neq \boldsymbol{x} \end{cases}, \tag{D.1}$$

hence, we have

$$P(\boldsymbol{y}(n)) = \sum_{\boldsymbol{x}\in\mathcal{X}} Q(\boldsymbol{y}(n)|\boldsymbol{x})P(\boldsymbol{x}) = \frac{(1-i)^n}{2^n}. \tag{D.2}$$

Now, we consider the output sequences of length $n+1$. For output sequences of length $n+1$ and with K runs ($\boldsymbol{y}(b; n+1, K) = (b; m_1, m_2, \ldots, m_K)$ and $\sum_{k=1}^{K} m_k = n+1$), we have

$$Q\left(\boldsymbol{y}(b; n+1; K)|\boldsymbol{x}\right) = \begin{cases} (m_k - 1)i(1-i)^{n-1} & \boldsymbol{x} = (b; m_1, \ldots, m_k - 1, \ldots, m_K) \\ 0 & \textit{otherwise} \end{cases}. \tag{D.3}$$

We then have

$$\begin{aligned} P\left(\boldsymbol{y}(b; n+1; K)\right) &= \sum_{\boldsymbol{x}\in\mathcal{X}} Q(\boldsymbol{y}(b; n+1; K)|\boldsymbol{x})P(\boldsymbol{x}) \\ &= \sum_{k=1}^{K} \frac{(m_k - 1)i(1-i)^{n-1}}{2^n} \\ &= \frac{(n+1-K)i(1-i)^{n-1}}{2^n}, \end{aligned} \tag{D.4}$$

hence, all the output sequences of length $n+1$ and with K runs are equiprobable which simplifies output entropy calculation.

By a similar argument, for output sequences with length $n+2$, we can write

$$Q\left(\boldsymbol{y}(b; n+2; K)|\boldsymbol{x}\right) = \begin{cases} (m_k - 1)(m_r - 1)i^2(1-i)^{n-2} & \boldsymbol{x} = (b; \cdots, m_k - 1, \cdots, m_r - 1, \cdots) \\ \binom{m_k - 2}{2} i^2(1-i)^{n-2} & \boldsymbol{x} = (b; \cdots, m_k - 2, \cdots), m_k \geq 4 \\ 0 & \textit{otherwise} \end{cases} \tag{D.5}$$

hence, we obtain

$$
\begin{aligned}
&P\left(\boldsymbol{y}(b;n+2;K;l)\right)=\sum_{\boldsymbol{x}\in\mathcal{X}}Q(\boldsymbol{y}(b;n+2;K)|\boldsymbol{x})P(\boldsymbol{x})\\
=&\frac{i^2(1-i)^{n-2}}{2^n}\Bigg(\sum_{k=1,m_k\geq 4}^{K}\binom{m_k-2}{2}+\sum_{k=1}^{K-1}\sum_{r=k+1}^{K}(m_k-1)(m_r-1)\Bigg)\\
=&\frac{i^2(1-i)^{n-2}}{2^n}\Bigg(\sum_{k=1}^{K}\left(\frac{(m_k-2)(m_k-3)}{2}\right)-l+\sum_{k=1}^{K-1}\sum_{r=k+1}^{K}(m_k-1)(m_r-1)\Bigg)\\
=&\frac{i^2(1-i)^{n-2}}{2^n}\Bigg(\sum_{k=1}^{K}\left(\frac{(m_k-1)^2-3m_k+5}{2}\right)-l+\sum_{k=1}^{K-1}\sum_{r=k+1}^{K}(m_k-1)(m_r-1)\Bigg)\\
=&\frac{i^2(1-i)^{n-2}}{2^{n+1}}\Big((n-K)^2+n+K-2-2l\Big).
\end{aligned}
\tag{D.6}
$$

where l is the number of runs in $\boldsymbol{y}$ with length of one ($l=\sum_{k=1}^{K}\delta(m_k-1)$), and the last equality is obtained by using the fact that $\sum_{k=1}K(m_k-1)^2+2\sum_{k=1}^{K-1}\sum_{r=k+1}^{K}(m_k-1)(m_r-1)=\left(\sum_{k=1}^{K}(m_k-1)\right)^2=(n-K+2)^2$. Based on the above calculation, all the output sequences with length $n+2$, K runs and l runs of length 1 are equiprobable.

APPENDIX E

PROOF OF THEOREM 4

We first give a lower bound on the output entropy of the binary input q-ary output channel with synchronization errors related to the output entropy of the binary synchronization error channel, then give an upper bound on the conditional output entropy of the binary input q-ary output channel with synchronization errors related to the conditional output entropy of the binary synchronization error channel.

Lemma 16. *For a binary input q-ary output channel with synchronization errors, for any input distribution and any odd q, we have*

$$H(\boldsymbol{Y}^{(q)}) \geq H(\boldsymbol{Y}) - E\{\boldsymbol{M}\} \log \left(2p_0^2 + \sum_{k=1}^{\frac{q-1}{2}} (p_k + p_{-k})^2 \right), \tag{E.1}$$

where $\boldsymbol{Y}$ denotes the output sequence of the synchronization error channel and input sequence of the binary input symmetric q-ary output channel, and $\boldsymbol{Y}^{(q)}$ denotes the output sequence of the binary input symmetric q-ary output channel.

Proof. For $p(\boldsymbol{y}^{(q)}|\boldsymbol{y}, \boldsymbol{M} = m)$, we have $p(\boldsymbol{y}^{(q)}|\boldsymbol{y}, \boldsymbol{M} = m) = \prod_{k=-\frac{q-1}{2}}^{\frac{q-1}{2}} p_k^{j_k}$, where j_k denotes the number of transitions $b \to \frac{k}{b}$. E.g., in a binary input 5-ary output channel we have $p(-1102|1111) = p_{-1}p_1p_0p_2$. Therefore, for a fixed output sequence $\boldsymbol{y}^{(q)}$ of length m with j_k symbols of k, since there are $2^{j_0} \prod_{k=1}^{\frac{q-1}{2}} \binom{j_k}{i_k} \binom{j_{-k}}{i_{-k}}$ possibilities for $\boldsymbol{y}$ such that $d(\boldsymbol{y}, \boldsymbol{y}^{(q)})_{b\to 0} = j_0$ and $d(\boldsymbol{y}, \boldsymbol{y}^{(q)})_{b\to \frac{k}{b}} = i_k$, we can write

$$\begin{aligned}
\sum_{\boldsymbol{y}, p(\boldsymbol{y}\neq 0)} p(\boldsymbol{y}^{(q)}|\boldsymbol{y}, \boldsymbol{M} = m) &\leq 2^{j_0} p_0^{j_0} \prod_{q=1}^{\frac{q-1}{2}} \sum_{i_k=0}^{j_k} \binom{j_k}{i_k} p_k^{i_k} p_{-k}^{j_k - i_k} \sum_{i_{-k}=0}^{j_{-k}} \binom{j_{-k}}{i_{-k}} p_{-k}^{i_{-k}} p_k^{j_{-k}-i_{-k}} \\
&= 2^{j_0} p_0^{j_0} \prod_{k=1}^{\frac{q-1}{2}} (p_k + p_{-k})^{j_k + j_{-k}} \\
&= 2^{m_0} p_0^{m_0} \prod_{k=1}^{\frac{q-1}{2}} (p_k + p_{-k})^{m_k(\boldsymbol{y}^{(q)})},
\end{aligned} \tag{E.2}$$

where $m_k(\boldsymbol{y}^{(q)}) = \#\{t \leq m | y_t^{(q)} \in \{k, -k\}\}$, i.e., the number of the times $Y_t^{(q)} = k$ or

$Y_t^{(q)} = -k$. Hence,

$$\begin{aligned}
&\sum_{\boldsymbol{y}^{(q)}} p(\boldsymbol{y}^{(q)}|\boldsymbol{M}=m) \sum_{\boldsymbol{y},p(\boldsymbol{)}\neq 0} p(\boldsymbol{y}^{(q)}|\boldsymbol{y},\boldsymbol{M}=m) \\
&\quad \leq \sum_{\boldsymbol{y}^{(q)}} p(\boldsymbol{y}^{(q)}|\boldsymbol{M}=m)(2p_0)^{m_0} \prod_{k=1}^{\frac{q-1}{2}} (p_k+p_{-k})^{m_k(\boldsymbol{y}^{(q)})} \\
&\quad = \sum_{m_0+\cdots+m_{\frac{q-1}{2}}=m} \binom{m}{m_0,\cdots,m_{\frac{q-1}{2}}} p_0^{m_0} \prod_{l=1}^{\frac{q-1}{2}} (p_l+p_{-l})^{m_l} \left((2p_0)^{m_0} \prod_{k=1}^{\frac{q-1}{2}} (p_k+p_{-k})^{m_k} \right) \\
&\quad = \left(2p_0^2 + \sum_{k=1}^{\frac{q-1}{2}} (p_k+p_{-k})^2 \right)^m . \qquad \text{(E.3)}
\end{aligned}$$

By substituting the result of (E.3) in the result of Lemma 6, we obtain

$$\begin{aligned}
H(\boldsymbol{Y}^{(q)}) &\geq H(\boldsymbol{Y}) - E\{\boldsymbol{M}\} \log \left(2p_0^2 + \sum_{k=1}^{\frac{q-1}{2}} (p_k+p_{-k})^2 \right) \\
&= -E\{\boldsymbol{M}\} \log \left(2p_0^2 + \sum_{k=1}^{\frac{q-1}{2}} (p_k+p_{-k})^2 \right), \qquad \text{(E.4)}
\end{aligned}$$

which concludes the proof. □

Lemma 17. *For a binary input q-ary output channel with synchronization errors, for any odd q and any input distribution, we have*

$$H(\boldsymbol{Y}^{(q)}|\boldsymbol{X}) \leq H(\boldsymbol{Y}|\boldsymbol{X}) + E\{\boldsymbol{M}\} H(p_{-\frac{q-1}{2}}, \cdots, p_{\frac{q-1}{2}}). \qquad \text{(E.5)}$$

Proof. By using the result of Lemma 7, we can write

$$\begin{aligned}
H(\boldsymbol{Y}^{(q)}|\boldsymbol{X}) &\leq E\{\boldsymbol{M}\} H(Y_j^{(q)}|Y_j) + H(\boldsymbol{Y}|\boldsymbol{X}) \\
&= E\{\boldsymbol{M}\} H(p_{-\frac{q-1}{2}}, \cdots, p_{\frac{q-1}{2}}) + H(\boldsymbol{Y}|\boldsymbol{X}). \qquad \text{(E.6)}
\end{aligned}$$

□

Obviously, by employing the results of Lemmas 16 and 17 and using the same approach as in the proof of Theorem 2, the proof of Theorem 4 is complete.

APPENDIX F

PROOF OF THEOREM 5

We need the following two lemmas to proof Theorem 5. In the first one, a lower bound on the output entropy of the binary input q-ary output channel with synchronization errors is derived relating with the output entropy of the binary synchronization error channel. In the second one, we give an upper bound on the conditional output entropy of the binary input q-ary output channel with synchronization errors related to the conditional output entropy of the binary synchronization error channel. By employing the result of two following lemmas and using the same approach as in the proof of Theorem 3, Theorem 5 is proved.

Lemma 18. *For a binary input q-ary output channel with synchronization errors, for any input distribution and any even q, we have*

$$H(\boldsymbol{Y}^{(q)}) \geq H(\boldsymbol{Y}) - E\{\boldsymbol{M}\} \log \left(\sum_{k=1}^{\frac{q}{2}} (p_k + p_{-k})^2 \right), \tag{F.1}$$

where $\boldsymbol{Y}$ denotes the output sequence of the synchronization error channel and input sequence of the binary input symmetric q-ary output channel, and $\boldsymbol{Y}^{(q)}$ denotes the output sequence of the binary input q-ary output channel.

Proof. Due to the result of Lemma 6, we have

$$H(\boldsymbol{Y}^{(q)}) - H(\boldsymbol{Y}) \geq -E_{\boldsymbol{M}} \left\{ \log \left(\sum_{\boldsymbol{y}^{(q)}} \sum_{\boldsymbol{y}, p(\boldsymbol{y} \neq 0)} p(\boldsymbol{y}^{(q)} | \boldsymbol{y}, \boldsymbol{M} = m) p(\boldsymbol{y}^{(q)} | \boldsymbol{M} = m) \right) \right\}. \tag{F.2}$$

On the other hand for $p(\boldsymbol{y}^{(q)} | \boldsymbol{y}, \boldsymbol{M} = m)$, we have $p(\boldsymbol{y}^{(q)} | \boldsymbol{y}, \boldsymbol{M} = m) = \prod_1^{\frac{q}{2}} p_k^{j_k} p_{-k}^{j_{-k}}$, where j_k denotes the number of transitions $b \to \frac{k}{b}$. For instance, in a binary input 6-ary output channel we have $p(-11-32|1111) = p_{-1} p_1 p_{-3} p_2$. On the other hand, for a fixed output sequence $\boldsymbol{y}^{(q)}$ of length m with j_k symbols of k, there are $\prod_{k=1}^{\frac{q}{2}} \binom{j_k}{i_k} \binom{j_{-k}}{i_{-k}}$ possibilities for $\boldsymbol{y}$ such that $d(\boldsymbol{y}, \boldsymbol{y}^{(q)})_{b \to \frac{k}{b}} = i_k$. By defining $m_k(\boldsymbol{y}^{(q)}) = \#\{t \leq m | y_t^{(q)} \in \{k, -k\}\}$, i.e., the number of the times $Y_t^{(q)} = -k$ or

$Y_t^{(q)} = k$, we can write

$$\begin{aligned}\sum_{\boldsymbol{y},p(\boldsymbol{y}\neq 0)} p(\boldsymbol{y}^{(q)}|\boldsymbol{y},\boldsymbol{M}=m) &\leq \prod_{k=1}^{\frac{q}{2}} \sum_{i_k=0}^{j_k} \binom{j_k}{i_k} p_k^{i_k} p_{-k}^{j_k-i_k} \sum_{i_{-k}=0}^{j_{-k}} \binom{j_{-k}}{i_{-k}} p_{-k}^{i_{-k}} p_{-k}^{j_{-k}-i_{-k}} \\ &= \prod_{k=1}^{\frac{q}{2}} (p_k+p_{-k})^{j_k+j_{-k}} = \prod_{k=1}^{\frac{q}{2}} (p_k+p_{-k})^{m_k(\boldsymbol{y}^{(q)})}, \quad \text{(F.3)}\end{aligned}$$

Furthermore, by taking the summation over all the possibilities of $\boldsymbol{y}^{(q)}$ in (F.3), we obtain

$$\begin{aligned}&\sum_{\boldsymbol{y}^{(q)}} p(\boldsymbol{y}^{(q)}|\boldsymbol{M}=m) \sum_{\boldsymbol{y},p(\boldsymbol{y})\neq 0} p(\boldsymbol{y}^{(q)}|\boldsymbol{y},\boldsymbol{M}=m) \\ &\qquad \leq \sum_{\boldsymbol{y}^{(q)}} p(\boldsymbol{y}^{(q)}|\boldsymbol{M}=m) \prod_{k=1}^{\frac{q}{2}} (p_k+p_{-k})^{m_k} \\ &\qquad = \sum_{m_1+\cdots+m_{\frac{q}{2}}=m} \binom{m}{m_1,\ldots,m_{\frac{q}{2}}} \prod_{l=1}^{\frac{q}{2}} (p_l+p_{-l})^{m_l} \prod_{k=1}^{\frac{q}{2}} (p_k+p_{-k})^{m_k} \\ &\qquad = \left(\sum_{k=1}^{\frac{q}{2}} (p_k+p_{-k})^2\right)^m. \quad \text{(F.4)}\end{aligned}$$

By substituting the result of (F.4) in (F.2), we obtain

$$\begin{aligned}H(\boldsymbol{Y}^{(q)}) - H(\boldsymbol{Y}) &\geq -\log\left(\sum_{k=1}^{\frac{q}{2}} (p_k+p_{-k})^2\right) \sum_m m p(m) \\ &= -E\{\boldsymbol{M}\} \log\left(\sum_{k=1}^{\frac{q}{2}} (p_k+p_{-k})^2\right), \quad \text{(F.5)}\end{aligned}$$

which concludes the proof. □

Lemma 19. *In any binary input q-ary output channel with synchronization errors, for any input distribution and any even q, we have*

$$H(\boldsymbol{Y}^{(q)}|\boldsymbol{X}) \leq H(\boldsymbol{Y}|\boldsymbol{X}) + E\{\boldsymbol{M}\} H(p_{-\frac{q}{1}}, \cdots, p_{-1}, p_1, \cdots, p_{\frac{q}{2}}). \quad \text{(F.6)}$$

Proof. The proof is similar to the proof of Lemma 17. □

It follows from the results in Lemmas 18 and 19 that

$$I(\boldsymbol{X};\boldsymbol{Y}^{(q)}) \geq I(\boldsymbol{X};\boldsymbol{Y}) - E\{\boldsymbol{M}\}\left[\log\left(\sum_{k=1}^{\frac{q}{2}}\left(p_k + p_{-k}\right)^2\right) + H(p_{-\frac{q}{1}},\cdots,p_{-1},p_1,\cdots,p_{\frac{q}{2}})\right].$$

Furthermore, following the same approach as in the proof of Theorem 3 concludes the proof of Theorem 5.

APPENDIX G

PROOF OF THEOREM 6

We first compute $H_\Delta = H(p_{-M}, \ldots, p_M) + \log(\Delta)$ and $\sum_{m=1}^{M} \frac{1}{\Delta} (p_m + p_{-m})^2$ for $M \to \infty$ and $\Delta \to 0$. Then by employing the result of Theorem 5, we prove the theorem.

For large M, we have $p_m \cong f(1 - m\Delta)\Delta$ with the understanding that the approximation becomes exact as $\Delta \to 0$ where $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$. Therefore, for $H_\Delta = H(p_{-M}, \ldots, p_{-1}, p_1, \ldots, p_M) + \log(\Delta)$, we can write

$$
\begin{aligned}
&\lim_{M\to\infty,\Delta\to 0} H_\Delta \\
&= \lim_{M\to\infty,\Delta\to 0} -\sum_{m=1}^{M} \Big[f(1-m\Delta)\log(f(1-m\Delta)) + f(1+m\Delta)\log(f(1+m\Delta)) \Big] \Delta \\
&= \int_0^\infty \Bigg[f(1-x) \left(\log(\sqrt{2\pi}\sigma) + \frac{(1-x)^2}{2\sigma^2} \log(e) \right) \\
&\qquad + f(1+x) \left(\log(\sqrt{2\pi}\sigma) + \frac{(1+x)^2}{2\sigma^2} \log(e) \right) \Bigg] dx \\
&= \int_{-\infty}^\infty f(1-x) \left(\log(\sqrt{2\pi}\sigma) + \frac{(1-x)^2}{2\sigma^2} \log(e) \right) dx \\
&= \log(\sqrt{2\pi}\sigma) + \frac{\log(e)}{2}. \qquad \text{(G.1)}
\end{aligned}
$$

On the other hand, for $\sum_{m=1}^{M} \frac{1}{\Delta} (p_m + p_{-m})^2$, by letting $M \to \infty$ and $\Delta \to 0$, we obtain

$$
\begin{aligned}
&\lim_{M\to\infty,\Delta\to 0} \sum_{m=1}^{M} \frac{1}{\Delta} (p_m + p_{-m})^2 \\
&\qquad = \lim_{M\to\infty,\Delta\to 0} \sum_{m=1}^{M} \left(f(1-m\Delta) + f(1+m\Delta) \right)^2 \Delta \\
&\qquad = \int_0^\infty \left(f(1-x) + f(1+x) \right)^2 dx \\
&\qquad = \frac{1}{\sqrt{2\pi}\sigma} \int_0^\infty \left(f(\sqrt{2}(1-x)) + f(\sqrt{2}(1+x)) + e^{-\frac{1}{\sigma^2}} f(\sqrt{2}x) \right) dx \\
&\qquad = \frac{1}{2\sqrt{\pi}\sigma} (1 + e^{-\frac{1}{\sigma^2}}). \qquad \text{(G.2)}
\end{aligned}
$$

Using the results of (G.1) and (G.2), we can write

$$\begin{aligned}\lim_{M\to\infty,\Delta\to 0}\left(H(p_{-M},\ldots,p_{-1},p_1,\ldots,p_M)+\log\left(\sum_{m=1}^{M}(p_m+p_{-m})^2\right)\right)\\ =\lim_{M\to\infty,\Delta\to 0}\left(H_\Delta+\log\left(\sum_{m=1}^{M}\frac{1}{\Delta}(p_m+p_{-m})^2\right)\right)\\ =\log\left(\sqrt{\frac{e}{2}}(1+e^{-\frac{1}{\sigma^2}})\right).\end{aligned}\tag{G.3}$$

Finally, by substituting this result into (4.38), the proof follows.

APPENDIX H

STOCHASTIC PROPERTIES OF $\boldsymbol{M}_1$ AND $\boldsymbol{M}_2$

For $P(\boldsymbol{M}_1, \boldsymbol{M}_2)$, we can write

$$
\begin{aligned}
&P(\boldsymbol{M}_1 = M_1, \boldsymbol{M}_2 = M_2) \\
&= \sum_{N_1=M_1}^{N-M_2} P(\boldsymbol{M}_1 = M_1, \boldsymbol{M}_2 = M_2 | \boldsymbol{N}_1 = N_1) P(\boldsymbol{N}_1 = N_1) \\
&= \sum_{N_1=M_1}^{N-M_2} P(\boldsymbol{M}_1 = M_1 | \boldsymbol{N}_1 = N_1) P(\boldsymbol{M}_2 = M_2 | \boldsymbol{N}_1 = N_1) P(\boldsymbol{N}_1 = N_1) \\
&= \sum_{N_1=M_1}^{N-M_2} \binom{N_1}{M_1} d_1^{N_1-M_1} (1-d_1)^{M_1} \binom{N-N_1}{M_2} d_2^{N-N_1-M_2} (1-d_2)^{M_2} \binom{N}{N_1} \lambda^{N_1} (1-\lambda)^{N-N_1} \\
&= \binom{N-M_2}{M_1} \binom{N}{M_2} (\lambda(1-d_1))^{M_1} ((1-\lambda)(1-d_2))^{M_2} \times \\
&\quad \times \sum_{N_1=M_1}^{N-M_2} \binom{N-(M_1+M_2)}{N_1-M_1} (\lambda d_1)^{N_1-M_1} ((1-\lambda)d_2)^{N-N_1-M_2} \\
&= \binom{N-M_2}{M_1} \binom{N}{M_2} (\lambda(1-d_1))^{M_1} ((1-\lambda)(1-d_2))^{M_2} (\lambda d_1 + (1-\lambda)d_2)^{N-M_1-M_2}.
\end{aligned}
$$

Furthermore, due to the structure of the channel $\mathcal{C}'$, $\boldsymbol{M}_2$ is binomially distributed, i.e., $P(\boldsymbol{M}_2 = M_2) = \binom{N}{M_2} ((1-\lambda)(1-d_2))^{M_2} (\lambda + (1-\lambda)d_2)^{N-M_2}$, and as a result $E\{\boldsymbol{M}_2\} = N(1-\lambda)(1-d_2)$. On the other hand, to obtain $E_{\boldsymbol{M}_1}\{\boldsymbol{M}_1 | \boldsymbol{M}_2\}$, we first need to obtain $P(\boldsymbol{M}_1 | \boldsymbol{M}_2)$, for which we can write

$$
\begin{aligned}
&P(\boldsymbol{M}_1 = M_1 | \boldsymbol{M}_2 = M_2) \\
&\quad = \frac{P(\boldsymbol{M}_1, \boldsymbol{M}_2)}{P(\boldsymbol{M}_2)} \\
&\quad = \binom{N-M_2}{M_1} (\lambda(1-d_1))^{M_1} (\lambda d_1 + (1-\lambda)d_2)^{N-M_1-M_2} (\lambda + (1-\lambda)d_2)^{M_2-N}.
\end{aligned}
$$

Therefore, we obtain

$$
\begin{aligned}
&E_{\boldsymbol{M}_1}\{\boldsymbol{M}_1 | \boldsymbol{M}_2\} \\
&= \sum_{M_1=0}^{N-M_2} M_1 \binom{N-M_2}{M_1} (\lambda(1-d_1))^{M_1} (\lambda d_1 + (1-\lambda)d_2)^{N-M_1-M_2} (\lambda + (1-\lambda)d_2)^{M_2-N} \\
&= (N-M_2) \frac{\lambda(1-d_1)}{\lambda + (1-\lambda)d_2}.
\end{aligned}
$$

APPENDIX I

CONCAVITY OF $g([M_1, \ldots, M_k])$

For the Hessian of the the function $g([M_1, \ldots, M_k])$, we have

$$\nabla^2 g([M_1, \ldots, M_k]) = \frac{1}{\sum_{k=1}^{K} M_k} \mathbf{1}\mathbf{1}^T - diag([\frac{1}{M_1}, \ldots, \frac{1}{M_K}]),$$

where $\mathbf{1}$ is an all one vector of length K, i.e., $\mathbf{1} = [1, \ldots, 1]^T$, and $diag([\frac{1}{M_1}, \ldots, \frac{1}{M_K}])$ denotes a diagonal matrix whose k-th diagonal element is $\frac{1}{M_k}$. Furthermore, by defining $\boldsymbol{a} = [a_1, \ldots, a_K]^T$, we can write

$$\begin{aligned}
\boldsymbol{a}\nabla^2 g\boldsymbol{a}^T =& \frac{(\sum_{k=1}^{K} a_k)^2}{\sum_{k=1}^{K} M_k} - \sum_{k=1}^{K} \frac{a_k^2}{M_k} \\
=& \frac{1}{\sum_{k=1}^{K} M_k} \left(\sum_{k=1}^{K} a_k^2 + 2 \sum_{k=1}^{K-1} \sum_{j=k+1}^{K} a_k a_j - \sum_{k=1}^{K} a_k^2 - \sum_{k=1}^{K} \frac{\sum_{j \neq k} M_j}{M_k} a_k^2 \right) \\
=& \frac{1}{\sum_{k=1}^{K} M_k} \sum_{k=1}^{K-1} \sum_{j=k+1}^{K} \left(2a_k a_j - \frac{M_j}{M_k} a_k^2 - \frac{M_k}{M_j} a_j^2 \right) \\
=& \frac{-1}{\sum_{k=1}^{K} M_k} \sum_{k=1}^{K-1} \sum_{j=k+1}^{K} \frac{M_j}{M_k} (a_k - \frac{M_k}{M_j} a_j)^2,
\end{aligned}$$

which is negative for all $M_k,\ M_j > 0$. Therefore, $\nabla^2 g([M_1, \ldots, M_k])$ is a negative semi-definite matrix and as a result $g([M_1, \ldots, M_k])$ is a concave function of $[M_1, \ldots, M_k]$.