Information Theoretical Studies of Non-Gaussian Noise Channels and Markovian Constrained Relay Channels

Linyun Huang

Department of Electrical and Computer System Engineering
Monash University

This dissertation is submitted for the degree of
Doctor of Philosophy

May 2019

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

Linyun Huang
May 2019

Acknowledgements

I would like to express my sincere appreciation to all who provided me the opportunity, their time and inspiration to complete my research project. Without them, the completion of this dissertation would have been impossible.

First and foremost, I am deeply grateful to my PhD supervisors, Dr. Yi Hong, Professor Emanuele Viterbo and Dr. Shuiyin Liu for offering me their endless guidance, continuous support and encouragement during the hard time in my life and PhD studies. To them, my gratitude is beyond words.

I would also like to thank my research group members and former members Dr. Lakshmi Prasad Natarajan, Dr. Amin Sakzad and Dr. Son Hoang Dau for inspiring discussions and constructive criticism. I was lucky enough to have Daniel Ryan as my desk mate in my office at Monash for chit-chatting with me during tea breaks. I would also like to thank Dr. Ricky Wong for sharing with me his ideas on machine learning and Dr. Zhengyu Zhou for mentoring me throughout my undergrad and PhD studies.

And finally, my deepest gratitude extends to my family for their endless love and care and to Biwen Tao for her spiritual support throughout this journey and my life.

List of Publications

Conference Papers:

[HHV14] L.Y. Huang, Y. Hong and E. Viterbo, "On Parameter Estimation of the Envelope Gaussian Mixture Model", in *Proc. Australia Commun. Theory Workshop*, pp. 33–38, Sydney, Australia, Feb. 2014.

[HHV15] L.Y. Huang, Y. Hong and E. Viterbo, "On the Capacity of the Gaussian Mixture Channel Under Average and Peak Power Constraints", *Int. Conf. On Telecommun. (ICT) 2015*, pp. 146–150, Sydney, Australia, Apr. 2015.

Journal Papers:

[HH18] L.Y. Huang and Y. Hong, "On the Capacity of a Two-Hop Half-Duplex Relay Channel with a Markovian Constrained Relay", *submitted to IEEE Access*.

Abstract

Communications in non-Gaussian noise channel and communication network with memory are two important but difficult frontiers of information theory. In this thesis, I studied these two areas. In the first part of this thesis, the Gaussian mixture distribution is adopted to model the non-Gaussian noise behaviour, typically found in powerline communications, man-made electromagnetic interferences, and underwater communications. Here, the capacity of a Gaussian mixture noise channel and its capacity-achieving input distribution are investigated. In the second part, I studied the capacity of a Markovian constrained relay channel and the maxentropic state transition probabilities for relay transmitter are derived. The derived results have been verified via a number of simulations.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

In information theory, communication in non-Gaussian noise channels and communication networks with memory are among the most essential and intriguing research directions that attract generations of researchers. Comparing against the idea of single user point-to-point discrete-time memoryless channels with additive white Gaussian noise, assumptions based on non-Gaussian noise and relay communication network with memory indeed provide more accurate models for practical communication systems, but these two frontiers are difficult fields of study. With continuing effort of researchers, numerous advances have been made in these fields of study. However, many questions still remain unsolved.

This two-parted thesis aims at establishing their mathematical models and information theoretical limits, i.e., the capacities. In the first part of this thesis, the point-to-point channel with non-Gaussian noise is studied. In particular, the Gaussian mixture model is used to model non-Gaussian noise. In the second part of this thesis, a relay channel in which the relay is Markovian constrained is considered. For both channels, we study their theoretical limits and thereby reveal potential rate gains in today's products.

1.1 Motivation

Non-Gaussian Noise Channel

In recent years, a non-Gaussian noise channel that is widely used to model real-world scenarios becomes an emerging topic and attracts much research attention. This is due to two main factors. First, preliminary researches on many newly developed communication systems, such as power line communication [39] and underwater communication [93], reveal that their noise distributions are, in general, non-Gaussian and therefore conventional models are insufficient to model such complicated communication environments. Second, knowledge of channel behaviour is an essential prerequisite before introducing sophisticated communication schemes.

A typical example of a non-Gaussian noise channel is the powerline communication channel which first been studied around World War II but never become main stream of communication research until 1990s [39]. The revival interest in powerline communication has mainly focused on low voltage power distribution network which has geographically widest spread. The driving force of utilising powerline communication systems includes load management, meter reading, home automation and intelligent building, collecting power user statistics and communication during disaster recovery (due to power down of wireless communication system). Powerline systems are also considered as an already implemented networks and thus there will be no additional infrastructure cost for communication. In addition, Low Voltage power line networks cover large area and Medium/High Voltage networks cover long distances, both of which make power line systems independent communication networks.

In general, powerline communication is a communication technology that enables sending data over existing power cables [39]. It can be viewed as powerline communication for AC lines and DC lines, respectively, with wide applications as mentioned above. Different from traditional communications engineering, the powerline communication systems deal with harsh channels, since the power line network differs considerably from conventional media such as twisted pair and coaxial, particularly in terms of topology, structure, and physical properties. Further, the noise in the powerline communication is typically modelled by a non-Gaussian noise model, rather than the additive

white Gaussian noise (AWGN) model in many other communication systems. The good understanding of channel characteristics and physical parameters lay the foundation for deriving the the optimal transmission rate (channel capacity). Deriving the optimal transmission rate encourages the research community to search for optimal codes of it. These motivates my research in the first part of this thesis.

Markovian Constrained Relay Channel

In the second part of the thesis, I focus on two-hop communication via a relay node, which is a promising solution for hot-spot capacity enhancement, network coverage extension, and gap filling in the next-generation cellular systems [74]. Driven by the increasing demand of high spectral efficiency, broad coverage, and high quality of service (QoS) of wireless communications, a relay node can be used to assist transmission by forwarding message from source to destination, where several cooperative protocols are usually introduced, such as amplify-and-forward (AF) [71], compress-and-forward [67] and decode-and-forward (DF) schemes [69]. Most of the transmission schemes are based on the assumption of half-duplex mode [71], and more recently, full-duplex mode has been also considered [89].

Capacity of a two-hop relay channel (source-relay-destination) with a half-duplex relay has been extensively studied in the literature (see e.g., [25]). To achieve the capacity, various strategies that relay forwards information to destination have been proposed [69, 115]. However, the following practical situations in the relay, which may effect its capacity, have never been considered in the previous study:

1. the effect of switching noise, i.e., intersymbol interference (ISI), due to logic transition in relay hardware [92];

2. the joint energy and information requirements in low-power wireless relay networks [40].

Switching noise is an induced ISI when switching a digital signal, caused by the relay's switching between reception mode and transition mode. It can be avoided by better hardware design or introducing a short guard time after

switching occurs [90]. In order to reduce the effect of switching noise and fulfil the joint energy and information requirements, constrained codes such as runlength limited codes [40], may be adopted. On the other hand, the performance in terms of join energy and information transfer is measured by the probabilities of overflow and underflow of the battery at the receiver and information rates. Classical codes, which are designed with the purpose of maximizing the information rate, are unstructured (i.e., random-like). Hence they do not allow the controlling of the timing of the energy transfer, and hence to optimize the probability of energy overflow and underflow. The constraint sequences enable us to characterise relay's reception or transmission mode in the switching-noise case, as well as data transmission with or without energy in the second case. By introducing constrained codes to the relay networks, memory is also introduced to the relay system. The capacity of relay network with memory still remains an open problem. To limit the scope of the study, we consider the two-hop relay channel with a half-duplex relay, where the source to the destination has no direct link, and the relay adopts constraint codes.

In addition to being an exciting theoretical challenge, understanding a two-hop relay channel is the prerequisite of understanding larger relay networks since the two-hop relay channel is the fundamental building block of a larger communication network, these results may be instrumental in deriving capacity when constrained sequences are introduced to a more complex relay network. These motivates my research in the second part of this thesis.

1.2 Informaton Prelimimaries

Figure 1.1 A point-to-point channel.

Discrete-Time Memoryless Channel

A scalar point-to-point discrete-time memoryless channel, as depicted in Figure 1.2, is characterised by

$$Y = X + N \tag{1.1}$$

where X denotes the channel input random variable, N the additive noise random variable, and Y the channel output random variable. The noise random variable N is assumed to be independent of the capacity achieving input X. For such channel, Shannon first develops the idea of reliable communication in his landmarking paper [99], in which the channel capacity is defined by

$$C = \max_{X} I(X;Y) \tag{1.2}$$

where $I(X;Y)$ is the mutual information between the input random variable X and the output random variable Y which is defined as

$$I(X;Y) = \iint p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dx\, dy\ . \tag{1.3}$$

The maximisation on mutual information is taken over all possible input distributions and was first tackled in [81]. A two-step iterative algorithm, known as Blahut-Arimoto Algorithm for computing the capacity of arbitrary discrete-time memoryless channels was proposed independently by Arimoto [3] and Blahut [14].

The most common discrete-time memoryless channel is the Gaussian channel, in which the noise random variable N is assumed to be Gaussian distributed. In Shannon's landmarking paper [99], Shannon obtained the famous formula for the capacity of average power constrained Gaussian noise channel

$$C = \frac{1}{2} \log_2 \Big(1 + \text{SNR}\Big) \ \text{ bits/channel use} \tag{1.4}$$

where SNR is the average power signal-to-noise ratio. In [99], Shannon also proved that for Gaussian channel, under average power constraint, the capacity-

achieving input is also Gaussian. While assuming that the input is under amplitude constraint other than average power constraint, Shannon provides upper and lower bounds on the capacity and he further noticed that for low SNR, the capacity is exactly given by (1.4). Smith [42, 101] studied the channel capacity of additive white Gaussian noise channel under peak and average power constraints in which he showed that the capacity can be achieved by discrete inputs with finite number of mass points.

In the case that the noise is non-Gaussian, Shannon [99] also provides an upper bound on the capacity. It is shown that the Gaussian noise is least favourable since the Gaussian noise has the maximum entropy. Also, Das showed in [28] that, for a scalar additive channel under average power constraint, if the noise distribution is "heavy-tailed" (i.e., has a tail which decays at a rate slower than the Gaussian), the capacity-achieving input will have a finite support. In addition, Tchamkerten extended this research to a class of non-Gaussian noise distributions and showed that for noise in this class, the input is discrete with finite number of mass points under average and peak power constraints [102]. More recently, Fahs *et. al* showed that under solely average power constraint, the capacity-achieving input for a class of non-Gaussian noise channels is also discrete [38]. Based on the above results, various works analyzed specific channels with Guassian noises, such as the Rayleigh fading channel [1], non-coherent additive white Gaussian noise channel [62], Rician fading channel [45] and quadrature Gaussian channel [97]. In [112], Zhang studied additive white Gaussian noise channel channel under duty-cycle constraints. Many of these results are built upon Smith's original work and the derived capacity-achieving distributions are of discrete nature. Readers may refer to [22] for a list of channels whose capacity-achieving inputs are discrete under either or both average power and peak power constraints.

Discrete-time Channel with Memory

The information rate between the input process $\boldsymbol{X} = (X_1, X_2, ..., X_N)$ and the output process $\boldsymbol{Y} = (Y_1, Y_2, ..., Y_N)$ of a discrete channel with memory m and

some well-defined starting state S_0 is given in the limit as

$$\mathcal{I}(\boldsymbol{X};\boldsymbol{Y}|S_0 = s_0) \triangleq \lim_{N\to\infty} \frac{1}{N} I(X_1, ..., X_N; Y_1, ..., Y_N|S_0 = s_0) \tag{1.5}$$

General speaking, this limitation is only well-defined in certain cases and/or may depending on the starting state s_0 [4]. In [41], Gallager defined finite-state channel and the information rate for finite-state channel is well-defined. Verdú and Han [105] showed that (1.5) may be adopted as the general form of mutual information if input process and output process are jointly ergodic and hence we have the general form of the capacity

$$C = \lim_{N\to\infty} \sup_{X_1^N} \frac{1}{N} I(X_1, ..., X_N; Y_1, ..., Y_N) \ . \tag{1.6}$$

For such case, the choice of initial state does not affect mutual information and is therefore ignored [98].

In this thesis, we assume that the input process $\boldsymbol{X}$ is a Markov process and is processed through a noiseless or noisy channel. The output process $\boldsymbol{Y}$ is therefore a Markov or a hidden Markov process. We further assume that the input alphabet is finite, i.e., the set of possible values of X_n is finite.

Entropy of Markov process were first considered by Shannon in his landmarking paper [99], in which Shannon computed the maximal entropy rate of a discrete-time Markov source, or equivalently the noise-free capacity of constrained sequences. In the presence of the noise, the computation of capacity of Markovian inputs sending over a noisy channel has long been an open problem[111, 94, 106]. Zehavi and Wolf [111], and independently, Shamai and Kofman [98] considered the noisy channel with run-length limit sequences and derived a set of analytical lower and upper bounds on the capacity. Arnold *et al.* [5], Sharma and Singh [100], Pfister *et al.* [86] proposed a Monte Carlo method for computing the exact mutual information of a finite-state machine channel with Markovian inputs. In order to maximize mutual information of Markovian inputs transmitted over a finite-state machine channel, a Blahut-Arimoto Algorithm for Markovian inputs was proposed in [63]. Vontobel *et al.* [106] extend this to a generalised Blahut-Arimoto Algorithm considering local convergence properties. The global convergence of the generalised Blahut-

Arimoto Algorithm requires the concavity of the mutual information between input and output; and the concavity of certain conditional entropy. Han [46] proposed a randomised algorithm to compute the capacity of a finite-state channel which only requires the concavity of mutual information. Han also showed in [47], that the concavity conjuncture of mutual information in finite-state machine channel does not hold in some cases. Hence, despite the good performance of these algorithms in many practical examples, the convergence of these algorithms, in general, is not guaranteed.

1.3 Objective of the Thesis

The objective of this thesis is to find the channel capacity of a Gaussian mixture noise channel and the capacity for a two-hop relay channel with a Markovian constrained relay. Either the closed form or the tight bounds of channel capacity will be derived. Of particular interest is which input distribution will maximise the channel capacity of a certain statistic model. The key elements of research is addressed below:

1. **New statistic models that better describe the non-Gaussian noise behaviour and relay constraints:**
    - As for non-Gaussian noise channel, these models should be simple to implement such that ideas can be quickly tested and verified.
    - As for a relay system, a simple finite-state machine should be introduced that fully describes the behaviour of a constrained relay. Such finite-state finite machine is fully specified by the state-transition probabilities and parameterized conditional output distribution.
2. **Parameters estimation of the proposed models**: A common question after a new model being proposed is how to estimate the parameters given some observations. The proposed estimator should be a sufficient statistic and a simple function of the observations.
3. **Capacity results for Gaussian mixture noise channels and Markovian constrained relay channel:** Deriving the channel capacity encourages the research community to search for optimal codes of non-Gaussian noise

channels and two-hop half-duplex Markovian constrained relay channels. For the Markovian constrained relay channel, the following property of finite-state machine makes this possible:

- *Stationary and ergodicity:* The output process of a stationary and ergodic Markovian process is also stationary and ergodic. Thus, by Shannon-McMillan-Breiman theorem, the mutual information between the input Markovian sequences and output sequences is determined by the probability of a typical sequence and the noise distribution.

4. **Computation methods:** Other than a theoretical result, what is more important is the practical computation method. The purpose of computing is insight and the numbers are often the best road to insight. The following key properties makes this possible:

   - *Discreteness and finiteness:* For a point-to-point non-Gaussian noise channel, under mild assumptions, the capacity achieving input is discrete and of finite number of mass points [102]. This allows to maximize the mutual information by finding the optimal allocation of weights and positions on probability mass points.
   - *Markov property:* For a Markovian constrained relay channel, the Markov structure of the finite-state machine allows to factor the probability measure of the states and the output process. This enables efficient computation methods.

1.4 Overview of Thesis and Original Contributions

This thesis reports the theoretical studies in two areas. Each part of the thesis has their individual acronym and symbols. Each chapter is self-contained with detailed introduction, system model, and conclusion. In the first part of this research, we consider a point-to-point memoryless channel with Gaussian mixture noise. The main contributions are reported in Chapters 2 and 3.

- **Chapter 2**: This chapter provides an comprehensive overview and background of types of non-Gaussian noise and non-Gaussian noise models.

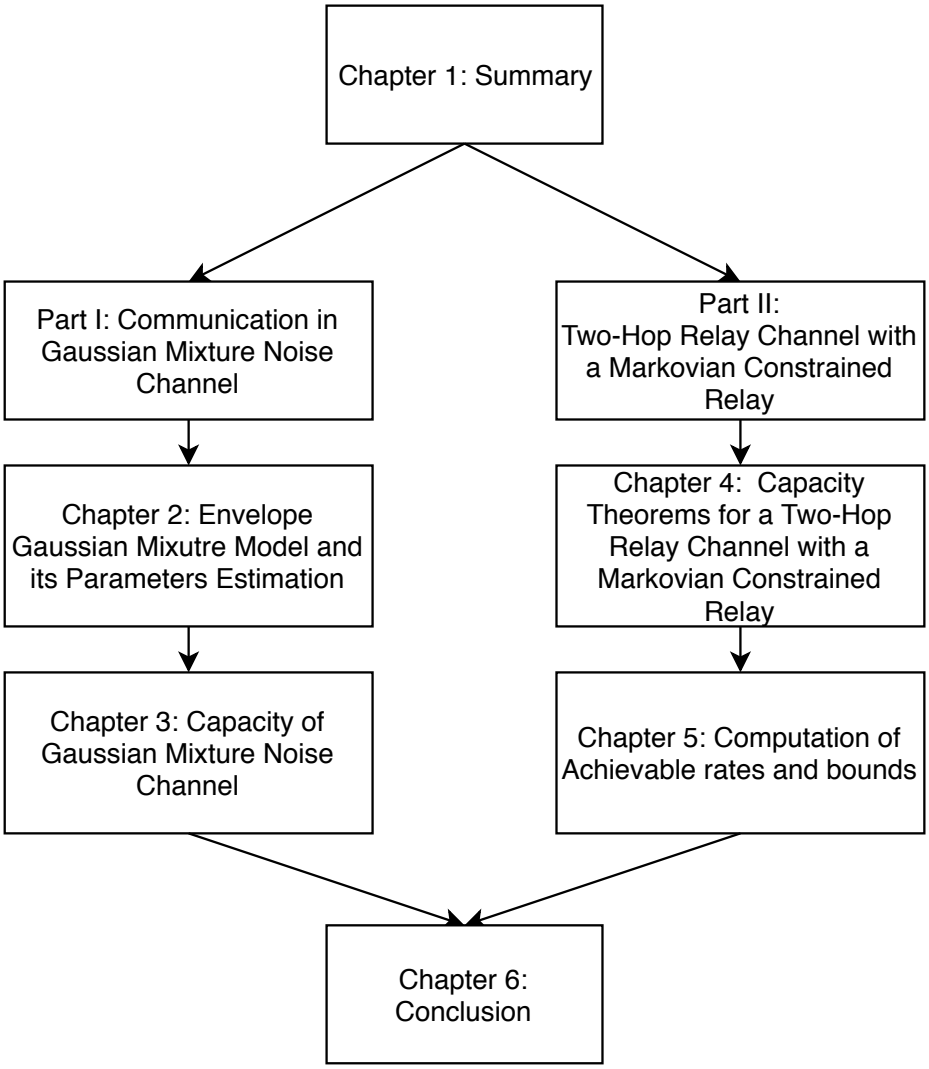

Figure 1.2 Thesis structure and relationship between chapters.

It then introduces a *simple* and *exact* closed form probability density function, called the envelope Gaussian mixture model which describes the non-Gaussian behaviour in powerline communications channels. A natural question that arises after the model has been proposed is the problem of estimation of the envelope Gaussian mixture parameters. In this chapter, the problem of parameters estimation is addressed and the proposed estimator of weights and variances is based upon the Expectation-Maximization (EM) algorithm. Finally, the simulation shows that the

estimated envelope Gaussian mixture model using EM exhibits excellent agreement with the histograms. **This result is reported in** [HHV14].

- **Chapter 3:** This chapter considers the capacity of the Gaussian mixture noise channel and its capacity-achieving input. In particular, we consider the symmetric (e.g. powerline) and asymmetric (e.g. NAND flash memory) Gaussian mixture noise cases. It is shown that, under average and peak power constraints, the capacity-achieving input is discrete with finitely many mass points. Furthermore, some properties of the capacity-achieving distribution are proved and demonstrated by simulations. **This result is reported in** [HHV15].

In the second part of this research, we consider two-hop half-duplex relay channel with a Markovian constrained relay, The main contributions are reported in Chapters 4 and 5.

- **Chapter 4:** This chapter provides introduction to two cornerstones on which the second part of the thesis builds: the relay channels and finite-state models. In this chapter, capacity theorems for a two-hop half-duplex relay channel with a Markovian constrained relay are studied. It first focuses on deriving the cut-set bound, i.e., an upper bound on the capacity. The timing strategy is introduced which satisfies the half-duplex constraint and it is shown to achieve this bound. This leads to the general capacity formula for a two-hop half-duplex relay channel with a Markovian constrained relay. **This result is reported in** [HH18].

- **Chapter 5:** This chapter first introduces a new type of constrained sequence, called the hold time constrained sequence, which captures the switching noise behaviour in relay hardware. It then simplifies the capacity formula and introduced the relay adjacency matrix such that the capacity can be easily computed in the case that relay-to-destination link is noiseless. The relay adjacency matrix can be considered as an extended version of adjacency matrix by Shannon for the two-hop relay channel, which is commonly used to compute the largest eigenvalue and corresponding largest eigenvector of non-negative primitive matrices. Moreover, for the case that the relay-to-destination link is noisy,

noisy-relay adjacency matrix is introduced. Algorithms are presented to compute the noisy-relay adjacency matrix and compute the optimised information rate which serves as a natural lower bound on the capacity. The bounds are shown to be very tight while comparing against upper bounds. **This result is reported in** [HH18].

Chapter 6: This chapter provides conclusions and discusses potential directions for future work.

Part I

Communication over Gaussian mixture noise channels

List of Acronyms and Symbols

List of Acronyms

EM . . . . . Expectation Maximization

ML . . . . . Maximum Likelihood

i.i.d. . . . . . independent and identically distributed

pdf . . . . . probability density function

cdf . . . . . cumulative density function

PLC . . . . . Powerline communication

AC . . . . . alternative current

CRTTV . . . . cathode ray tube television

PSD . . . . . power spectrum density

MSE . . . . . mean square error

KKT . . . . . Kuhn-Tucker Conditions

AWGN . . . . additive white gaussian noise

SNR . . . . . Signal-to-Noise Ratio

List of Symbols

σ_k . . . . . standard deviation of k-th Gaussian component

π_k . . . . . mixing coefficient of k-th Gaussian component

$\underline{\pi}$ . . . . . vector of mixing coefficients

a . . . . . mixing coefficient of two-term Gaussian mixture model

B	impulsive index
Γ	mean power ratio between impulsive noise power and Gaussian noise power
P_l	noise power for l-th noise component
T_{AC}	time period of AC mains
β_l	phase shift of noise relative to AC mains
n_l	indicator which indicates the occurrence of types of noise
$I_0(\cdot)$	modified Bessel function of the first kind of $zero$-th order
$I_1(\cdot)$	modified Bessel function of the first kind of $first$ order
γ	ratio of two variances
$\mathcal{L}(\cdot\|\cdot)$	log likelihood function
$\hat{\theta}$	estimation of parameters θ
ξ	soft assignment
p_N	distribution of noise
$\mathcal{N}(\cdot;\cdot,\cdot)$	Gaussian distribution
P	average input power
A	peak amplitude
$\mathbb{N}$	set of all natural numbers
$\mathbb{R}$	set of all real numbers
$\mathbb{C}$	set of all complex numbers
Ω	space of all possible probability distribution on $\mathbb{R}$
$I(X;Y)$	mutual information
$H(X)$	entropy
C	capacity
$D_{KL}(p(\cdot)\|\|q(\cdot))$	divergence between distribution $p(\cdot)$ and $q(\cdot)$
$\nabla \log p_N(w)$ . .	score of the noise distribution p_N

Chapter 2

Gaussian Mixture Noise and Its Envelope Distribution

In many communication systems, the Gaussian mixture model is widely used to characterize non-Gaussian man-made and natural interference. The envelope distribution of such noise model is often expressed as a weighted sum of Rayleigh if in-phase and quadrature components of the noise are dependent. Instead, in this chapter, a simple and exact closed form probability density function (pdf) of the envelope Gaussian mixture model (i.e. the envelope of independent in-phase and quadrature components of complex non-Gaussian noise) is obtained. Furthermore, the problem of estimating of the envelope Gaussian mixture parameters is addressed. The proposed estimator of weights and variances is based upon the Expectation-Maximization (EM) algorithm.

This chapter is organized as follows. In section 2.1, the models for noise in non-Gaussian channels are discussed. In section 2.2, the envelope Gaussian mixture density function is derived. In section 2.3, the maximum likelihood (ML) estimator via the EM algorithm is presented. The performance of the EM algorithm is compared to that of conventional ML estimator using the quasi-Newton method in section 2.4. The conclusion is drawn in section 2.5.

2.1 Introduction

In wireless communication systems, classical white Gaussian noise is often assumed to be a very accurate model. However, in other cases, such as the

underwater [93] and the powerline communications (PLC) [33], the noise may exhibit non-Gaussian behaviour and thus it is important to consider different versatile and robust noise/interference models. In 1977, Middleton [78] proposed the Middleton's Class A noise model, a Gaussian mixture density model with Poisson selection, to describe the electromagnetic interference from a variety of noise sources. Arzberger *et al.* [33] also suggested that the noise in powerline channels can also be modelled by the Gaussian mixture model, a parametric pdf expressed as summation of weighted Gaussian pdfs. The envelope distribution of this mixture density is often represented as weighted sum of Rayleigh under the assumption that the in-phase components and quadrature components of the noise are dependent [39]. In our work, we assume that both in-phase and quadrature components are independent and identically distributed (i.i.d.) random variables as motivated by [93, 33]. Hence, the envelope distribution of in-phase and quadrature noise components will not result in the Rayleigh mixture model, but give rise to the envelope Gaussian mixture model.

Parameters of the envelope Gaussian mixture model can be estimated by ML estimation (MLE). When closed form expression cannot be found for MLE, iterative methods, such as the Newton-based and the EM algorithm, are used. These algorithms iteratively maximize the log likelihood function. In an earlier work by Sari *et al.* [93], a conventional ML estimator using the quasi-Newton method is used. In this chapter, we consider the ML estimator via the EM algorithm, a widely used method popularized by Dempster, Laird and Rubin in 1977 [30]. The EM algorithm has been used in many parameters estimation problems, especially in dealing with the curved exponential densities [108], such as the Middleton's Class A model [110], the Rayleigh mixture [95] and the Gaussian mixture density [13]. This chapter is based on the research results reported in [56] by the author.

2.1.1 Introduction to Non-Gaussian Noise

Non-Gaussian noise channels have become emerging topics and have attracted much attention. This is due to two main factors. First, preliminary research on many newly developed communication systems, such as PLC [33], NAND flash

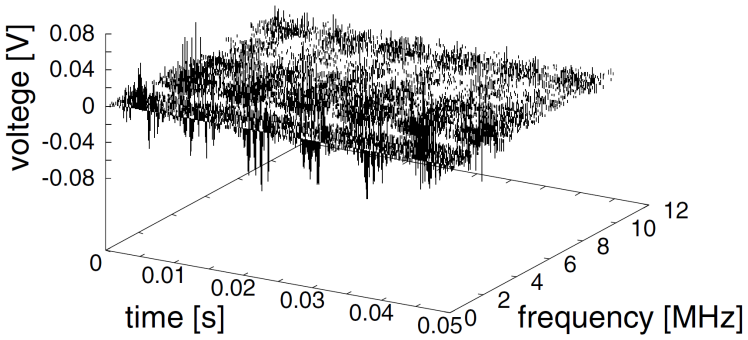

Figure 2.1 Noise amplitude distribution in time-frequency domain [52].

memories [66], channels with man-made electromagnetic interferences [78], and underwater communications [93], reveals that their noise distributions are, in general, non-Gaussian and therefore conventional models are insufficient to model such complicated communication environments. Second, a knowledge of channel behaviour is an essential prerequisite before introducing sophisticated communication schemes. The good understanding of channel characteristics and physical parameters lay the foundation for deriving the optimal transmission rate (channel capacity). Deriving the optimal transmission rate enables the research community to search for optimal codes of it.

The non-Gaussian noise was first introduced in PLC in which the statistic behaviour of the electronic appliances is indeed different from that of a conventional radio communication system. In wireless systems, noise is usually modelled as an additive white Gaussian. Due to the fact that the communication environment is complicated in PLC, it is vital to derive a statistic model which matches with some features of the measurements. Basic research which classifies PLC channel noise was conducted by O. Hooijen [53] and extended by M. Zimmerman and K. Dostert [113, 114]. Later, H. Meng *et al.* [77] classified current models into two main categories: frequency-domain and time-domain approaches. In time domain, the noise is non-Gaussian since PLC contains both periodic noise which synchronises with half cycle of AC power supply [113, 52]

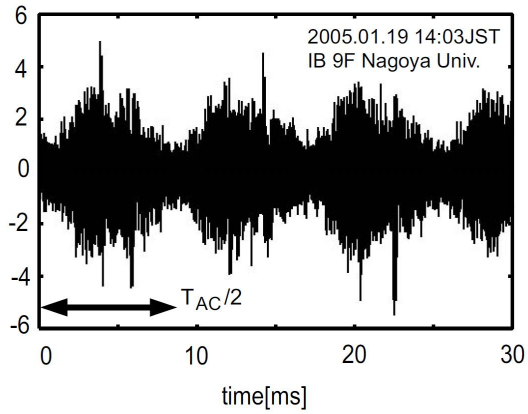

Figure 2.2 Snapshot of normalized overall noise waveform [61].

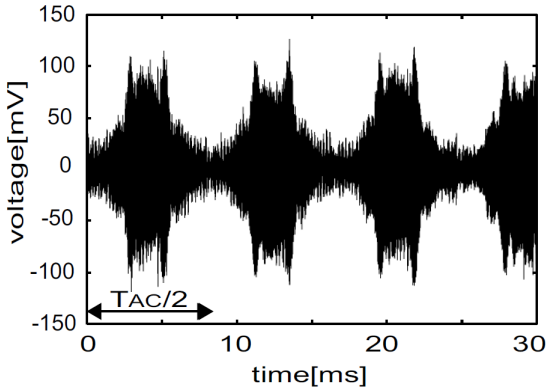

Figure 2.3 Noise waveform of fluorescent Lamp [61].

as well as non-periodic impulsive noise that are generated by electronic appliances and switches [113, 52, 20]. Measurements also show that, in frequency domain, background noise of PLC is coloured such that its amplitude varies with frequency [77, 17, 73]. Fig. 2.1 shows noise amplitude distribution of PLC noise in time-frequency domain.

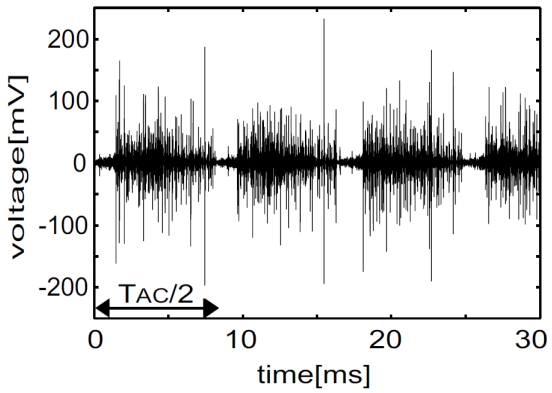

Figure 2.4 Noise waveform of Vacuum cleaner [61].

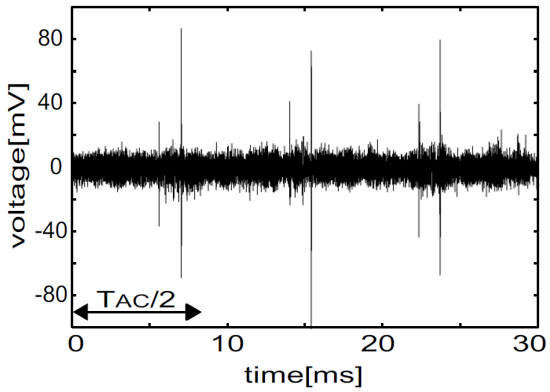

Figure 2.5 Noise waveform of CRTTV [61].

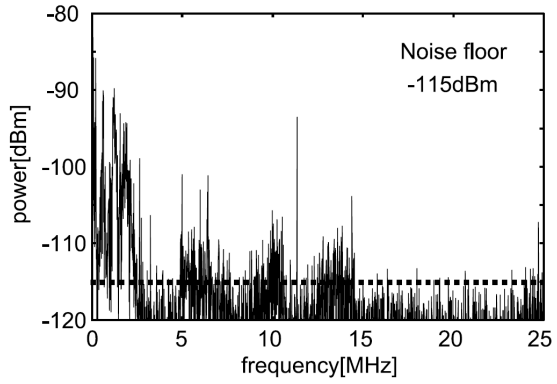

Figure 2.6 Measured noise in frequency domain [61].

Classes of noise	**Description**
Coloured Background Noise	Noise with low power spectral density (PSD), including the thermal noise by front-end amplifier of the receiver.
Narrowband Noise	Amplitude modulated signals generated by radio broadcast stations due to sharing bands with broadcasting and wireless signals. This noise is also known as *tone jammer*.
Time-Variant Continuous Noise	Envelope of this noise changes synchronously to the AC mains. Hence period of the noise is half of the AC mains duration, $T_{AC}/2$. This noise often appears when appliance which have oscillator whose power supply is rectified but not smoothed is connected to the line.
Periodic Impulsive Noise Synchronous to AC mains	Noise generated by the switching action of silicon controlled rectifier diodes in electronic appliances. Hence period of this noise is the period or half of the period of AC mains.
Periodic Impulsive Noise Asynchronous to AC mains	Noise generated by switched-mode power supply with a frequency higher than that of the AC mains.
Isolated (Aperiodic) Impulsive Noise	Noise generated at random time with long intervals (>1 second), mainly caused by switches in the network.

Table 2.1 Classification of non-Gaussian noise in PLC [39].

These results give the insight of the fundamental properties of power line noise in the frequency range upto 30 MHz [77]. Hence one can derive useful statistical models based on these measurements and parameters characterised from it. Classification of PLC noise is listed in Table 2.1 with explanations for each type of noise. Similar classification can be found in [39, 77, 113, 43]. Fig. 2.3, 2.4 and 2.5 shows the waveform of the noise generated by three electronic appliances including fluorescent lamp, vacuum cleaner and CRT TV. On average, they have higher noise power spectrum density (PSD) than other

appliances, hence the noise they generated may dominate the channel when they are attached to the line. However, in general, PLC overall noise waveform is the summation of noise generated by all noise sources as shown in Fig. 2.2. Similarly, noise in frequency domain is the summation of colour background noise PSD with other noise PSDs as shown in Fig. 2.6.

2.1.2 Non-Gaussian noise models

Non-Gaussian noise models are usually analyzed in two different categories: frequency-domain approaches and time-domain approaches [77]. In 2006, cyclo-stationary noise model [61] was proposed to describe the periodic behaviour of PLC noise. Another important model for PLC is the Middleton's Class A noise model [78–80]. However, Middleton's Class A noise model does not fall in either category since it only describes the randomness of the impulsive noise without considering the noise behaviour in time domain and attenuation in frequency domain. Such a model can also be approximated by a two-term Gaussian mixture model which is described later in this section.

Due to the fact that, in PLC, higher frequency noise attenuates faster than lower frequency noise and power of appliances are concentrated at lower frequencies, background PLC noise is non-white as shown in Fig. 2.6. These phenomenons exist in narrowband PLC channel which utilises kHz band as well as wideband PLC channel. However, noise spectrum in wideband PLC is more complicated than that of the narrowband due to the multipath effect (which results in frequency selective spectrum) in PLC channels and due to the presence of narrowband noise at certain frequency bands.

Frequency-domain approaches are therefore adopted to analyse coloured noise. In [53], Hooijen adopted the spectrum fitting technique to model the non-white noise PSD in PLC. Later, Benyoucef [10] studied the case in which narrowband noise is present. The major flaw of spectrum fitting technique is that it does not provide the information of randomness of noise at each frequency band (or just assume it to be Gaussian distributed). Arzberger *et al.* [33], therefore, used sum of two Rayleigh to model the PLC noise magnitude at individual frequency band. Similar concept was utilized by H. Meng *et al.* [77] who used the Nakagami-m pdf to model the PLC noise at different

frequencies. Some researchers also used the log-normal distribution [17]. After parameters estimation being performed for properly selected statistical models, one can consequently describe the randomness of PLC noise in frequency domain. However, a major drawback of these approaches is that they do not describe any of the periodic behaviour of PLC noise in time domain.

Meanwhile, time-domain approaches are performed to analyze the time behaviour, especially periodic behaviour, of the PLC noise. In time-domain approaches, two general questions are answered [113]:

1. When do impulsive events occur?
2. How strong are these impulses?

Two parameters are characterised to answer the first question, i.e., impulse width and inter-arrival time. The second question is answered by measuring the impulse amplitudes. Distribution of these parameters are derived in [21, 29]. In [113], Markov model is used to model the randomness of isolated impulsive noise. However, it does not describe the frequency behaviour of the non-Gaussian noise.

In 2005, Katayama *et al.* [61] proposed the cyclo-stationary noise model which models the PLC background noise and impulsive noise as a whole in time domain. This approach can be integrated with spectrum fitting models such that PLC noise can be described in both time and frequency domain. Its pdf is non-Gaussian and therefore can be modelled by the Gaussian mixture model.

In 1979, Middleton proposed the Middleton's Class A, Class B and Class C noise models [78–80] that describe man-made inference communication environment. Among these three models, Class A noise model is widely used due to its great simplicity of implementation. Class B model considers the case in which the spectrum of noise is boarder than the bandwidth of the receiver. However, this pdf turns out to be too complicated to implement. Class C model is the linear combination of Class A and Class B model which is also ineffective to use. Middleton's Class A model is expressed as infinite summation of weighted Gaussian to be selected by Poisson distribution. Since summation of infinite weighted Gaussian can be approximated by finite-term Gaussian mixture model, Middleton's Class A noise model can be consider to be a special case of Gaussian mixture model. In the Gaussian mixture model and the Middleton's Class A model, the occurrence of impulsive event is

considered to be a random variable whereas in cyclo-stationary noise model, it is deterministic. Summary of existing models are listed in Table 2.2. Later in this chapter, the Middleton's Class A model, sum of two Rayleigh, the cyclo-stationary noise model and the Gaussian mixture model are explained in detail as they are widely used in modelling PLC and other non-Gaussian noises [11]. Our research is also based upon the Gaussian mixture model due to two main reasons:

1. It is an appropriate approximation of many other models such as Middleton's Class A model.

2. It has great simplicity since it is a linear combination of finite number of weighted Gaussian distribution.

PLC non-Gaussian noise modelling approaches		
Frequency-domain	Time-domain	Other
• Spectrum Fitting [53, 10] • Nakagami-m pdf [77] • Gaussian pdf [10] • Log-normal pdf [17] • Sum of two Rayleigh [33]	• Markov Chain [113] • Pure Measurements and experimental derivation [21, 29]	• Middleton's Class A [78–80] • Gaussian Mixture [33, 11] • Cyclo-stationary Noise [61]

Table 2.2 Summary of existing PLC noise modelling aprroaches

2.1.3 Gaussian Mixture Model

The Gaussian mixture model is a pdf expressed as summation of weighted Gaussian pdfs. A simple two-term Gaussian mixture model is described as follow [79, 11, 33]

$$p_N(r; a, \sigma_1^2, \sigma_2^2) = \frac{a}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{r^2}{2\sigma_1^2}} + \frac{(1-a)}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{r^2}{2\sigma_2^2}} \tag{2.1}$$

where $0 \leq a \leq 1$ is the mixing coefficient and σ_1^2, σ_2^2 are the variances of the two Gaussian components. An example where the above model can be used is when background noise is always present and impulsive noise events occur with probability $(1-a)$. The first Gaussian component can be seen as the nominal

background noise with variance σ_1^2. The second component represents the combination of the background noise and the impulsive noise, when impulsive noise events occur. The two-term Gaussian mixture model was also considered as an approximation of Middleton's Class A noise model [78–80] and has been used extensively in both modelling the powerline noise [33, 11] and the underwater communications noise [93]. In many practical cases, a small number K of Gaussian components (e.g., 2 or 3) are sufficient to accurately model the noise without overfitting. where $\underline{\pi} = \{\pi_k\}_{k=1}^K$, are mixing coefficients of each Gaussian density and $\sum_{k=1}^K \pi_k = 1$. The variances of the Gaussian pdfs are $\underline{\sigma}^2 = \{\sigma_k^2\}_{k=1}^K$ Fig. 2.7 shows that the Gaussian mixture is indeed different from the conventional single Gaussian noise model given the same variance.

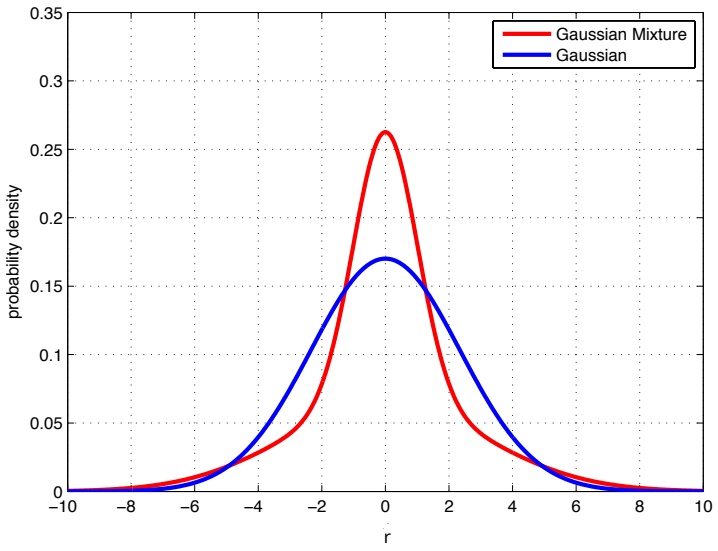

Figure 2.7 Comparison between two-term Gaussian mixture model and Gaussian model having the same variance.

Sum of two Rayleigh

By assuming that the in-phase and quadrature components follows equation (2.1) and are dependent to each other, we get the sum of two Rayleigh model

which is expressed as

$$p_{RM}(r; a, \sigma_1^2, \sigma_2^2) = a \cdot \frac{r}{\sigma_1^2} e^{-\frac{r^2}{2\sigma_1^2}} + (1-a) \cdot \frac{r}{\sigma_2^2} e^{-\frac{r^2}{2\sigma_2^2}} , \tag{2.2}$$

where $0 \leq a \leq 1$ is the mixing coefficient and σ_1, σ_2 are the scale parameters of the two Rayleigh components. Fig. 2.8 shows that sum of two Rayleigh is different from single Rayleigh model given the same variance. This model is derived from the Gaussian mixture model and it models the statistic behaviour of the noise magnitude.

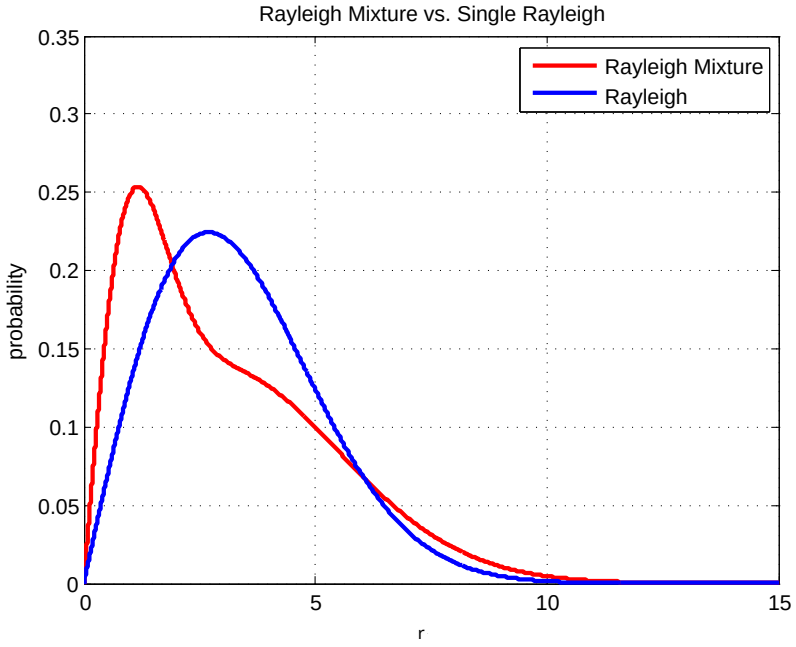

Figure 2.8 Comparison between two-term Rayleigh mixture model and Rayleigh model having the same variance.

Middleton's Class A Noise Model

One of the most important models of non-Gaussian noise is the Middleton's Class A noise model [78–80]. Among the three models Middleton proposed (i.e. Class A, Class B and Class C), Class A noise model is widely used due to its great simplicity of implementation. Middleton's Class A model is expressed as

infinite summation of weighted Gaussian to be selected by Poisson distribution

$$p_N(r) = \sum_{m=0}^{\infty} p_m \frac{1}{\sqrt{2\pi\sigma_m^2}} \exp -\frac{r^2}{2\sigma_m^2} \ , \tag{2.3}$$

where

$$p_m = \frac{e^{-B} B^m}{m!} \ , \tag{2.4}$$

$$\sigma_m^2 = \sigma^2 \cdot \frac{m/B + \Gamma}{1 + \Gamma} = \sigma_i^2 \frac{m}{B} + \sigma_g^2 \ , \tag{2.5}$$

where B is called the impulsive index. σ_i^2 and σ_g^2 are the impulsive noise power and Gaussian noise power respectively. For given m-th component, when B becomes large, Gaussian noise power becomes dominant as shown in (2.5). The noise becomes impulsive when B is small. Γ is the mean power ratio between impulsive noise power and Gaussian noise power. $\sigma^2 = \sigma_i^2 + \sigma_g^2$ is the overall noise power. The weighted coefficients of Middleton's Class A noise model is expressed as Poisson distribution p_m such that $\sum_{m \in \mathbb{N}} p_m = 1$. If we summarise the remainder terms that have low weighted coefficients and approximate them with single Gaussian, the Middleton's Class A model becomes the finite-term Gaussian mixture model.

Cyclo-stationary Noise Model

Cyclo-stationary noise model [61] was proposed to model noise waveform of three types of noise including background (time-invariant) noise, time-variant continuous noise and periodic impulsive noise. Power of cyclo-stationary noise can be expressed as:

$$\sigma^2(\tau) = \sum_{l=0}^{L-1} P_l |\sin(2\pi\tau/T_{AC} + \beta_l)|^{n_l} \tag{2.6}$$

where P_l is noise power for l-th noise component. β_l is the phase shift of certain type of noise relative to AC mains and n_l is the indicator which indicates the occurrence of different types of noise. When n_l is 0, the l-th noise component

always occurs and when n_l goes to very large, the l-th noise only appears when $\sin$ function equals 1. Typical values of parameters P_l, β_l and n_l are listed in Table 2.3. Each l-th component corresponds to each type of noise. One can modify this model by deleting or adding new components. Fig. 2.9 and 2.12 show simulated cyclo-stationary noise power and noise waveform in time and time-frequency domain.

l	P_l	β_l[deg]	n_l	Types of noise
0	0.230	–	0	Background noise
1	1.38	-6	1.91	Time-variant continuous noise
2	7.17	-35	1.57×10^5	Periodic impulsive noise

Table 2.3 Parameters of cyclo-stationary noise model as depicted in Figure 2.9 – 2.12.

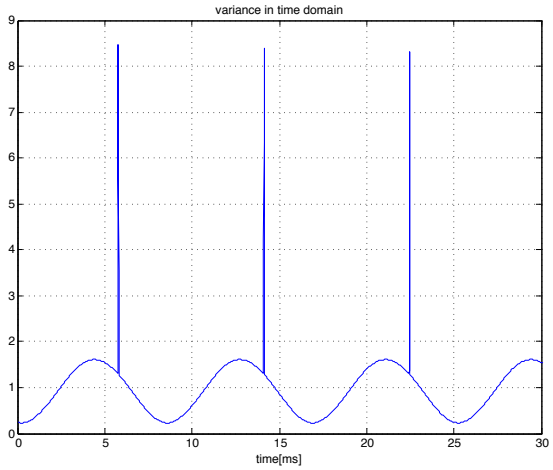

Figure 2.9 Simulated cyclo-stationary noise power waveform in time domain.

2.2 Envelope Gaussian Mixture Distribution

We first discuss the widely used two-term Gaussian mixture model for the in-phase and quadrature noise amplitudes R and Q. Both in-phase and quadrature noise amplitudes are considered as independent random variables with the following two-term Gaussian mixture density function in equation (2.1), in which $0 \leq a \leq 1$ is the mixing coefficient and σ_1^2, σ_2^2 are the variances of the

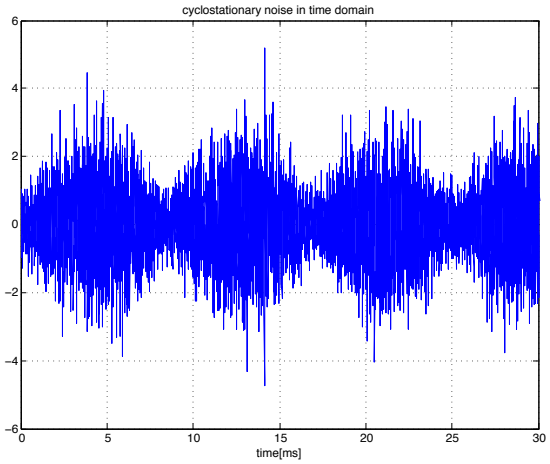

Figure 2.10 Simulated cyclo-stationary noise waveform in time domain.

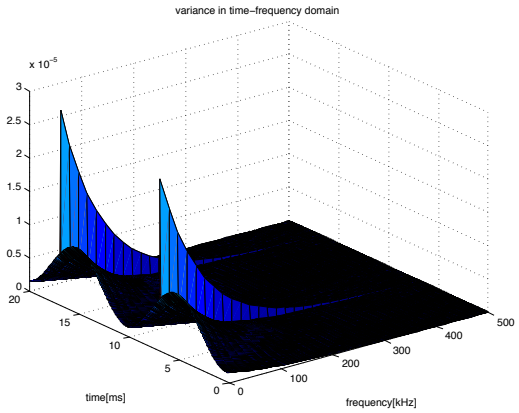

Figure 2.11 Simulated cyclo-stationary noise power in time frequency domain.

two Gaussian components. An example where the above model can be used is when background noise is always present and impulsive noise events occur with probability $(1-a)$. The first Gaussian component can be seen as the nominal background noise with variance σ_1^2. The second component represents the combination of the background noise and the impulsive noise, when impulsive noise events occur. Since both background noise and impulsive noise are assumed to be Gaussian random variables, the sum of the two will also be a Gaussian random variable.

The two-term Gaussian mixture model was also considered as an approximation of Middleton's Class A noise model [78, 79] and has been used extensively

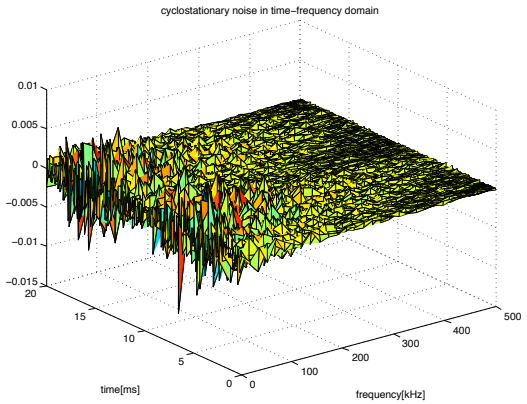

Figure 2.12 Simulated cyclo-stationary noise waveform in time frequency domain.

in both modelling the powerline noise [33] and the underwater communications noise [93]. In many practical cases, a small number K of Gaussian components (e.g., 2 or 3) are sufficient to accurately model the noise without overfitting. In general we have

$$p_N(r;\underline{\pi},\underline{\sigma}^2) \;=\; \sum_{k=1}^{K} \pi_k \cdot \frac{1}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{r^2}{2\sigma_k^2}} \tag{2.7}$$

where $\underline{\pi} = \{\pi_k\}_{k=1}^K$, are mixing coefficients of each Gaussian density and $\sum_{k=1}^K \pi_k = 1$. The variances of the Gaussian pdf's are $\underline{\sigma}^2 = \{\sigma_k^2\}_{k=1}^K$, and we assume that $\sigma_k^2 > \sigma_{k-1}^2 > \sigma_{k-2}^2 > ... > \sigma_1^2$.

Let us now consider the noise envelope random variable W as a function of the in-phase and quadrature noise components R and Q. We start with the case where $f_R(r)$ and $f_Q(q)$ are two-term Gaussian mixtures as in (2.1) and R and Q are i.i.d. random variables. Then W can be expressed as:

$$W = \sqrt{R^2 + Q^2}\ . \tag{2.8}$$

Given the joint pdf $f_{R,Q}(r,q)$, the cumulative density function (cdf), $F_W(w)$, is defined as:

$$F_W(w) \;=\; \int\int_{\sqrt{r^2+q^2}\le w} f_{R,Q}(r,q)drdq\ . \tag{2.9}$$

We can then find $f_W(w)$ by differentiating $F_W(w)$ directly using differentiation rule due to Leibnitz [85], we have:

$$f_W(w) = \int_{-w}^{w} \frac{w}{\sqrt{w^2-q^2}} \Big(f_{R,Q}(\sqrt{w^2-q^2}, q) + f_{R,Q}(-\sqrt{w^2-q^2}, q) \Big) dq \ . \tag{2.10}$$

Since R and Q are assumed to be independent, the two terms $f_{R,Q}(\sqrt{w^2-q^2}, q)$ and $f_{R,Q}(-\sqrt{w^2-q^2}, q)$ will result in:

$$\begin{aligned} f_{R,Q}(\sqrt{w^2-q^2}, q) &= f_{R,Q}(-\sqrt{w^2-q^2}, q) \\ &= \frac{a^2}{2\pi\sigma_1^2} e^{-\frac{w^2-q^2+q^2}{2\sigma_1^2}} + \frac{a(1-a)}{2\pi\sigma_1\sigma_2} e^{-\frac{w^2-q^2}{2\sigma_1^2}-\frac{q^2}{2\sigma_2^2}} \\ &+ \frac{a(1-a)}{2\pi\sigma_1\sigma_2} e^{-\frac{w^2-q^2}{2\sigma_2^2}-\frac{q^2}{2\sigma_1^2}} + \frac{(1-a)^2}{2\pi\sigma_2^2} e^{-\frac{w^2-q^2+q^2}{2\sigma_2^2}} \ . \end{aligned} \tag{2.11}$$

Substituting (2.11) into (2.10), yields:

$$\begin{aligned} f_W(w) &= \int_{-w}^{w} \frac{2w}{\sqrt{w^2-q^2}} \Bigg[\frac{a^2}{2\pi\sigma_1^2} e^{-\frac{w^2}{2\sigma_1^2}} + \frac{(1-a)^2}{2\pi\sigma_2^2} e^{-\frac{w^2}{2\sigma_2^2}} \\ &+ \frac{a(1-a)}{2\pi\sigma_1\sigma_2} \left(e^{-\frac{w^2-q^2}{2\sigma_1^2}-\frac{q^2}{2\sigma_2^2}} + e^{-\frac{w^2-q^2}{2\sigma_2^2}-\frac{q^2}{2\sigma_1^2}} \right) \Bigg] dr \ . \end{aligned} \tag{2.12}$$

The integral (2.12) can be computed in closed form by letting $q = w\sin(\theta)$ and $dq = w\cos(\theta)d\theta$ to yield:

$$\begin{aligned} f_W(w) &= \frac{a^2 w}{\sigma_1^2} e^{-\frac{w^2}{2\sigma_1^2}} + \frac{(1-a)^2 w}{\sigma_2^2} e^{-\frac{w^2}{2\sigma_2^2}} \\ &+ \frac{2a(1-a)w}{\sigma_1\sigma_2} e^{\left(-\frac{1}{4}\left(\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}\right)w^2\right)} I_0\left(\frac{1}{4}\left|\frac{1}{\sigma_1^2}-\frac{1}{\sigma_2^2}\right| w^2\right) \ . \end{aligned} \tag{2.13}$$

where

$$I_0(\eta) = \frac{1}{\pi}\int_0^{\pi} e^{\eta\cos(\theta)} d\theta \ .$$

is the modified Bessel function of the first kind of *zero*-th order. In (2.13), first two terms follow a Rayleigh distribution. The term with modified Bessel function distinguished the envelope distribution of two-term Gaussian mixture

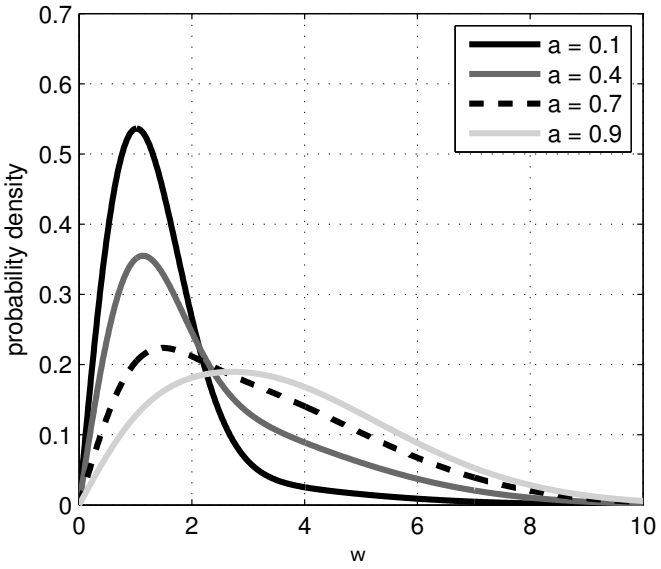

Figure 2.13 Envelope of two-term Gaussian miwture pdf with mixing coefficient $0.1 \leq a \leq 0.9$ and $c = \sigma_2^2/\sigma_1^2 = 10$.

from the two-term Rayleigh mixture. The envelope of two-term Gaussian mixture is shown in Fig. 2.13 for different values of a and and the ratio of two variances $c = \sigma_2^2/\sigma_1^2 = 10$. One can observe that when the value of mixing coefficient $a = 0.7$, the tail of the pdf is approximately linearly decaying from the peak value (dashed line in Fig. 2.13). When the ratio of the two variances $\sigma_2^2/\sigma_1^2 \to 1$, similar to the case of a two-term Rayleigh mixture model, the envelope of two-term Gaussian mixture (2.13) turns into a single Rayleigh distribution. The effect of varying the ratio of σ_2^2/σ_1^2 is illustrated in Fig. 2.14, in which the dashed grey line is the Rayleigh pdf with variance $\sigma^2 = 1$. However, when $\sigma_2^2 \gg \sigma_1^2$, the envelope Gaussian mixture is significantly different from the Rayleigh mixture as shown in Fig. 2.15. The two-term Rayleigh mixture model often exhibits a more pronounced multi-modal behaviour than the envelope of two-term Gaussian mixture.

Assuming that the in-phase and quadrature components are not identically distributed and follow different Gaussian mixture model with K and L

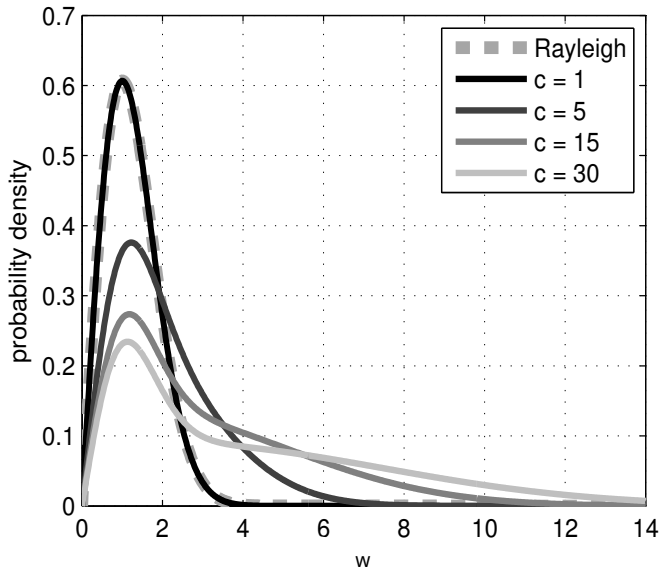

Figure 2.14 Envelope of two-term Gaussian mixture pdf with mixing coefficient $a = 0.5$ and $c = \sigma_2^2/\sigma_1^2 = 1, 5, 15, 30$. The envelope Gaussian mixture converges to the Rayleigh distribution as $c \to 1$ (dashed grey curve).

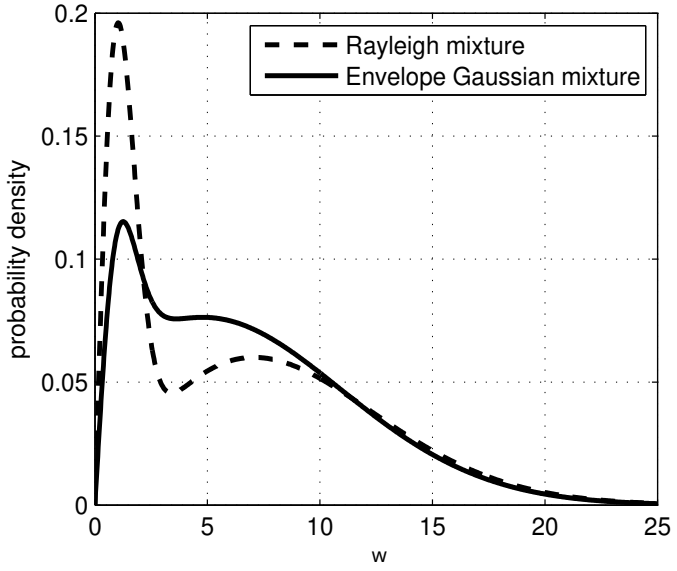

Figure 2.15 Comparison of envelope of two-term Gaussian mixture pdf with two-term Rayleigh mixture both with mixing coefficient $a = 0.7$ and $c = \sigma_2^2/\sigma_1^2 = 50$.

components, i.e,

$$f_R(r) = \sum_{i=1}^{K} [\pi_r]_i \cdot \frac{1}{\sqrt{2\pi[\sigma_r^2]_i}} e^{-\frac{r^2}{2[\sigma_r^2]_i}} \tag{2.14}$$

$$f_Q(q) = \sum_{k=1}^{L} [\pi_q]_k \cdot \frac{1}{\sqrt{2\pi[\sigma_q^2]_k}} e^{-\frac{q^2}{2[\sigma_q^2]_k}} . \tag{2.15}$$

Then a more general envelope Gaussian mixture distribution $f_W(w)$ can be written as:

$$\begin{aligned} f_W(w) = \sum_{k=1}^{L}\sum_{i=1}^{K} \Bigg\{ & \frac{[\pi_r]_i \cdot [\pi_q]_k \cdot w}{[\sigma_r]_i [\sigma_q]_k} \cdot e^{\left(-\frac{1}{4}\left(\frac{1}{[\sigma_q^2]_k}+\frac{1}{[\sigma_r^2]_i}\right)w^2\right)} \\ & \cdot I_0\left(\frac{1}{4}\left|\frac{1}{[\sigma_q^2]_k} - \frac{1}{[\sigma_r^2]_i}\right| w^2\right) \Bigg\} . \end{aligned} \tag{2.16}$$

However, in many practical cases, the in-phase and quadrature noise components can be described by the same distribution with the same parameters, and the envelope Gaussian mixture density simplifies to:

$$\begin{aligned} f_W(w) = \sum_{k=1}^{K}\sum_{i=1}^{K} \Bigg\{ & \frac{\pi_i \pi_k w}{\sigma_i \sigma_k} \cdot e^{\left(-\frac{1}{4}\left(\frac{1}{\sigma_k^2}+\frac{1}{\sigma_i^2}\right)w^2\right)} \\ & \cdot I_0\left(\frac{1}{4}\left|\frac{1}{\sigma_k^2} - \frac{1}{\sigma_i^2}\right| w^2\right) \Bigg\} . \end{aligned} \tag{2.17}$$

2.3 Parameters Estimation of the Envelope Gaussian Mixture Model

In this section, we adopt the well-known two-step iterative method called the EM Algorithm that finds the ML or maximum a posteriori estimates of parameters in statistical models in which observations are treated as “incomplete data" [13]. Parameters of mixture densities, such as the Rayleigh mixture model, the Middleton’s Class A model and the Gaussian mixture can be estimated by using the EM algorithm. Hence it is natural to predict that parameters of the envelope Gaussian mixture density can be estimated by using the same algorithm.

Given a data set $W = \{w_1, ..., w_N\}$, we assume that all data samples are i.i.d.. Let $p(w|\theta)$ be the pdf that is governed by the set of parameters, θ, to be estimated. We have:

$$p(W|\theta) = \prod_{n=1}^{N} p(w_n|\theta) \ ,$$

$$\mathcal{L}(\theta|W) = \log p(W|\theta) = \sum_{n=1}^{N} \log p(w_n|\theta) \tag{2.18}$$

where $\mathcal{L}(\theta|W)$ is the log likelihood function. Our purpose is to find the parameters θ which maximize $\mathcal{L}(\theta|W)$ such that:

$$\hat{\theta}_{ML} = \arg\max_{\theta} \mathcal{L}(\theta|W) \ . \tag{2.19}$$

Under the assumption that the envelope Gaussian mixture model is taken, $p(w_n|\theta)$ is replaced with (2.17) and the incomplete data log likelihood function is given by:

$$\begin{aligned} \mathcal{L}(\theta|W) \quad = \quad & \sum_{n=1}^{N} \log \sum_{k=1}^{K} \sum_{i=1}^{K} \left\{ \frac{\pi_i \pi_k w_n}{\sigma_i \sigma_k} \cdot e^{\left(-\frac{1}{4} \left(\frac{1}{\sigma_k^2} + \frac{1}{\sigma_i^2} \right) w_n^2 \right)} \right. \\ & \left. \cdot I_0 \left(\frac{1}{4} \left| \frac{1}{\sigma_k^2} - \frac{1}{\sigma_i^2} \right| w_n^2 \right) \right\} . \end{aligned} \tag{2.20}$$

In order to solve this equation, the EM algorithm is utilized by introducing the latent variables. Here we employ two sets of latent variables $U = \{u_i\}_{i=1}^{K}$ and $V = \{v_k\}_{k=1}^{K}$ as binary indicator variables (i.e. $u_i \in \{0, 1\}$, $v_k \in \{0, 1\}$, $\sum_i u_i = 1$ and $\sum_k v_k = 1$). The value u_i indicates which Gaussian component in (2.14) generates the i-th in-phase noise sample and similarly v_k for the quadrature noise samples. It is important to note that since in-phase noise samples and quadrature noise samples are assumed to be independently generated, u_i and v_k are also independent. The product of these two latent variables, $u_i \cdot v_k$, is binary and forms a 2-dimensional indicator function (i.e. $u_i \cdot v_k \in \{0, 1\}$ and $\sum_i \sum_k u_i \cdot v_k = 1$).

We use the EM algorithm to estimate the parameters $\theta = \{\underline{\pi}, \underline{\sigma}^2\}$. The *expectation* step and *maximization* step are defined as follows:

- *E*-step: Compute $\mathcal{Q}(\theta|\theta^{(p)}) \triangleq \mathbb{E}[\mathcal{L}(\theta|W)|U, V, \theta^{(p)}]$

- *M*-step: Determine $\theta = \theta^{(p+1)}$ maximizing $\mathcal{Q}(\theta|\theta^{(p)})$

where $\theta^{(p)}$ is the estimation of θ at p-th iteration of the EM algorithm. We call W the 'incomplete data' and we assume that the complete data $\mathcal{S} = (W, U, V)$ includes the binary latent variables U and V. Then the joint density function $p(w, u, v)$ is:

$$p(w, u, v) = p(u, v)p(w|u, v) = p(u)p(v)p(w|u, v) \tag{2.21}$$

since U and V are independent. The proportion of the noise samples that are generated by k-th or i-th Gaussian component is π_k or π_i and therefore the joint distributions over U and V are specified in terms of the mixing coefficient π_i and π_k, such that $p(u_i = 1) = \pi_i$ and $p(v_k = 1) = \pi_k$. Both u_i and v_k are indicator variables, therefore we can write joint distribution in the following form:

$$p(u, v) = \prod_{i=1}^{K} \prod_{k=1}^{K} [\pi_i \pi_k]^{u_i \cdot v_k} \ . \tag{2.22}$$

Similarly, the conditional distribution of W given particular values for U and V is the envelope Gaussian mixture component that is:

$$p(w|u, v) = \prod_{i=1}^{K} \prod_{k=1}^{K} [p(w|\sigma_i^2, \sigma_k^2)]^{u_i \cdot v_k} \ . \tag{2.23}$$

Then we have:

$$\begin{aligned} p(w) &= \sum_u \sum_v p(u, v)p(w|u, v) \\ &= \sum_u \sum_v \prod_{i=1}^{K} \prod_{k=1}^{K} [\pi_i \pi_k \cdot p(w|\sigma_i^2, \sigma_k^2)]^{u_i \cdot v_k} \ . \end{aligned} \tag{2.24}$$

This is an equivalent formulation of the mixture model involving two explicit latent variables. By dealing with the complete observation W, U and V, we can simplify the log likelihood function using (2.24). Replacing $p(w_n|\theta)$ in (2.20)

with the equivalent formulation found in (2.24), we have:

$$\begin{aligned}\mathcal{L}(\theta|W) &= \sum_{n=1}^{N}\sum_{i=1}^{K}\sum_{k=1}^{K} u_i \cdot v_k\{\log \pi_i + \log \pi_k \\ &\quad + \log p(w_n|\sigma_i^2, \sigma_k^2)\} \ . \end{aligned} \tag{2.25}$$

Using (2.25), the $\mathcal{Q}$ function, becomes:

$$\begin{aligned}\mathcal{Q}(\theta|\theta^{(p-1)}) &\triangleq \mathbb{E}[\mathcal{L}(\theta|W)|U, V, \theta^{(p)}] \\ &= \sum_{n=1}^{N}\sum_{i=1}^{K}\sum_{k=1}^{K} \mathbb{E}[u_i^{(n)} \cdot v_k^{(n)}]\{\log \pi_i + \log \pi_k \\ &\quad + \log p(w_n|[\sigma_i^{(p-1)}]^2, [\sigma_k^{(p-1)}]^2)\} \end{aligned} \tag{2.26}$$

where $\mathbb{E}[u_i^{(n)} \cdot v_k^{(n)}] = \xi_{n,i,k}$ is the conditional probability of W given U and V, we also call it "soft" assignment (or responsibility) which can be found by using Bayes' theorem:

$$\begin{aligned}\xi_{n,i,k}^{(p)} &= \mathbb{E}[u_i^{(n)} \cdot v_k^{(n)}] = p(u_i = 1, v_k = 1|w_n) \\ &= 0 \times p(u_i = 0, v_k = 0|w) + 0 \times p(u_i = 1, v_k = 0|w) \\ &\quad +0 \times p(u_i = 0, v_k = 1|w) + 1 \times p(u_i = 1, v_k = 1|w) \\ &= \frac{p(u_i = 1)p(v_k = 1)p(w|u_i = 1, v_k = 1)}{p(w_n)} \\ &= \frac{\pi_i^{(p-1)}\pi_k^{(p-1)}p(w_n|[\sigma_i^{(p-1)}]^2, [\sigma_k^{(p-1)}]^2)}{\sum_{s=1}^{K}\sum_{t=1}^{K}\pi_s\pi_t p(w_n|[\sigma_s^{(p-1)}]^2, [\sigma_t^{(p-1)}]^2)} \ . \end{aligned} \tag{2.27}$$

The *maximization* step of the EM algorithm finds the expression for $\pi_i^{(p)}$ and $\pi_k^{(p)}$. We introduce the Lagrange multiplier λ with the constraint $\sum_i \pi_i = \sum_k \pi_k = 1$ and solve the following set of equations:

$$\frac{\partial}{\partial \pi_k^{(p)}}\left\{\sum_{n=1}^{N}\sum_{i=1}^{K}\sum_{k=1}^{K}\xi_{n,i,k}\log \pi_k^{(p)} + \lambda\left(\sum_k \pi_k^{(p)} - 1\right)\right\} = 0 \tag{2.28}$$

for $k = 1, \ldots, K$. Summing left side over n and i (or k) gives $\lambda = -N$ which results in:

$$\pi_k^{(p)} = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K}\xi_{n,i,k}^{(p)} \ . \tag{2.29}$$

The variances $[\sigma_k^{(p)}]^2$ and $[\sigma_i^{(p)}]^2$ can be found by solving the set of K equations:

$$\frac{\partial}{\partial[\sigma_k^{(p)}]^2}\left\{\sum_{n=1}^{N}\sum_{i=1}^{K}\sum_{k=1}^{K}\xi_{n,i,k}^{(p)}\left[\log\frac{w_n}{\sigma_k^{(p)}\sigma_k^{(p)}}-\frac{1}{4}\left(\frac{1}{[\sigma_k^{(p)}]^2}+\frac{1}{[\sigma_i^{(p)}]^2}\right)w_n^2\right.\right.$$
$$\left.\left.+\log I_0\left(\frac{1}{4}\left|\frac{1}{[\sigma_k^{(p)}]^2}-\frac{1}{[\sigma_i^{(p)}]^2}\right|w_n^2\right)\right]\right\}=0\quad k=1,\ldots,K\ . \tag{2.30}$$

We get:

$$[\sigma_k^{(p)}]^2=\frac{\sum_{n=1}^{N}\sum_{i=1}^{K}\xi_{n,i,k}\frac{w_n}{2}(1-\phi(\frac{1}{4}|\frac{1}{[\sigma_k^{(p)}]^2}-\frac{1}{[\sigma_i^{(p)}]^2}|w_n^2))}{\sum_{n=1}^{N}\sum_{i=1}^{K}\xi_{n,i,k}^{(p)}}\ . \tag{2.31}$$

where $\phi(\cdot)=\frac{I_1(\cdot)}{I_0(\cdot)}$ and $I_1(\cdot)$ is the modified Bessel function of first kind of *first* order. Equations (2.31) are non-linear equations and therefore can only be solved numerically.

In the following, we provide the EM algorithm for envelope Gaussian mixture model.

Algorithm 1: EM ALGORITHM FOR THE ENVELOPE GAUSSIAN MIXTURE MODEL

Initialization: Initialize the mixing coefficients $\pi_i^{(0)}$ and $\pi_k^{(0)}$, together with variances $[\sigma_i^{(0)}]^2$ and $[\sigma_k^{(0)}]^2$ with random values. **Repeat** until convergence

Step 1: (**E step**) Given the data sample w_n, evaluate the responsibility using the current parameters $\pi^{(p-1)}$ and $[\sigma^{(p-1)}]^2$ for all i and k from 1 to K

$$\begin{aligned}\xi_{n,i,k}^{(p)} &= \mathbb{E}[u_i^{(n)}\cdot v_k^{(n)}]=p(u_i=1,v_k=1|w_n)\\ &= 0\times p(u_i=0,v_k=0|w)+0\times p(u_i=1,v_k=0|w)\\ &\quad +0\times p(u_i=0,v_k=1|w)+1\times p(u_i=1,v_k=1|w)\\ &= \frac{p(u_i=1)p(v_k=1)p(w|u_i=1,v_k=1)}{p(w_n)}\\ &= \frac{\pi_i^{(p-1)}\pi_k^{(p-1)}p(w_n|[\sigma_i^{(p-1)}]^2,[\sigma_k^{(p-1)}]^2)}{\sum_{s=1}^{K}\sum_{t=1}^{K}\pi_s\pi_t p(w_n|[\sigma_s^{(p-1)}]^2,[\sigma_t^{(p-1)}]^2)}\ .\end{aligned} \tag{2.32}$$

Step 2: (**M step**) Update the parameters using the current responsibility

$$\pi_k^{(p)} = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{K}\xi_{n,i,k} \ . \tag{2.33}$$

$$[\sigma_k^{(p)}]^2 = \frac{\sum_{n=1}^{N}\sum_{i=1}^{K}\xi_{n,i,k}^{(p)}\frac{w_n}{2}(1-\phi(\frac{1}{4}\left|\frac{1}{[\sigma_k^{(p)}]^2}-\frac{1}{[\sigma_i^{(p)}]^2}\right|w_n^2))}{\sum_{n=1}^{N}\sum_{i=1}^{K}\xi_{n,i,k}^{(p)}} \ . \tag{2.34}$$

for all k from 1 to K. $\phi(\cdot)=\frac{I_1(\cdot)}{I_0(\cdot)}$ and $I_1(\cdot)$ is the modified Bessel function of first kind of *first* order due to the fact that the derivative of $I_0(\cdot)' = -I_1(\cdot)$. . Equation (2.34) are non-linear equations and therefore can only be solved numerically.

Step 3: Evaluate the likelihood function

$$\mathcal{L}(\theta^{(p)}|W) = \sum_{n=1}^{N}\log\left\{\sum_{i=1}^{K}\sum_{k=1}^{K}\pi_k^{(p)}\pi_i^{(p)}\cdot p(w_n|[\sigma_k^{(p)}]^2,[\sigma_i^{(p)}]^2)\right\} \tag{2.35}$$

and check the convergence of either the parameters or the likelihood function. If the convergence criterion is not satisfied, return to Step 1 (**E step)**.

end

2.4 Simulation Results

A simulation of the proposed estimator via the EM algorithm is performed, in which we consider an envelope Gaussian mixture model with unknown variances and mixing coefficients, however we assume that the number of components is known. Random data samples are randomly generated from the distribution with known parameters. An alternative method in estimating parameters of the envelope Gaussian mixture apart from the EM algorithm is the quasi-Newton method (See [93] for detail). The performance of the EM algorithm will be compared with the performance of the quasi-Newton

method with the BGFS step update for the envelope Gaussian mixture model [107]. The EM algorithm and the quasi-Newton method will terminate when the change in log likelihood function is less than stopping criterion, ϵ. The EM algorithm is more often used in parameter estimation of the curved exponential families, since the quasi-Newton method is more complicated to implement. On the other hand, the EM algorithm enjoys greater simplicity and stability but has slower convergence speed. Here we focus on comparing their convergence speed and accuracy in terms of iteration numbers and Mean Square Error (MSE) between the estimated pdf and normalised data histograms.

Our first example takes 50000 data samples which are generated by the envelope of a two-term Gaussian mixture distribution with parameters $\pi_1 = 0.3$, $\pi_2 = 0.7$, $\sigma_1^2 = 1$ and $\sigma_2^2 = 10$. Fig. 4 illustrates the normalised data histogram with $N = 25$ bins of generated data samples, together with the estimated envelope Gaussian mixture pdf and true envelope Gaussian mixture pdf. The MSE between the envelope Gaussian mixture pdf and the normalised data histogram is defined as follows:

$$\mathbf{MSE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{D}_i - D_i)^2 \tag{2.36}$$

where D_i is the frequency density of the i-th bin and $\hat{D}_i$ is the estimated probability density taken at the midpoint of i-th bin. The MSE between the estimated pdf and histogram is 3.12779×10^{-4} and the parameters estimated are $\hat{\pi}_1 = 0.303502$, $\hat{\pi}_2 = 0.696498$, $\hat{\sigma}_1^2 = 1.04399$ and $\hat{\sigma}_2^2 = 9.946461$, which show that the ML estimator via the EM algorithm yields good performance. Example 2 shown in Fig. 5 also illustrates that parameters of the envelope Gaussian mixture model have been correctly estimated by the EM algorithm. In fact, for examples with distributions that are well separated (i.e. $\sigma_2^2 \gg \sigma_1^2$), both the EM algorithm and the quasi-Newton method perform well in terms of reproducing the true parameters. However, in some ill-conditioned cases (i.e. example 4, 5, 7 and 8) where too many components have been included in the model due to overfitting or to similar values of variances, both the EM algorithm and the quasi-Newton method perform poorly in terms of attaining the true values of parameters as shown in Table 5.2. In addition, example 6 does not converge to the true value due to small sample size. Fortunately, both methods may

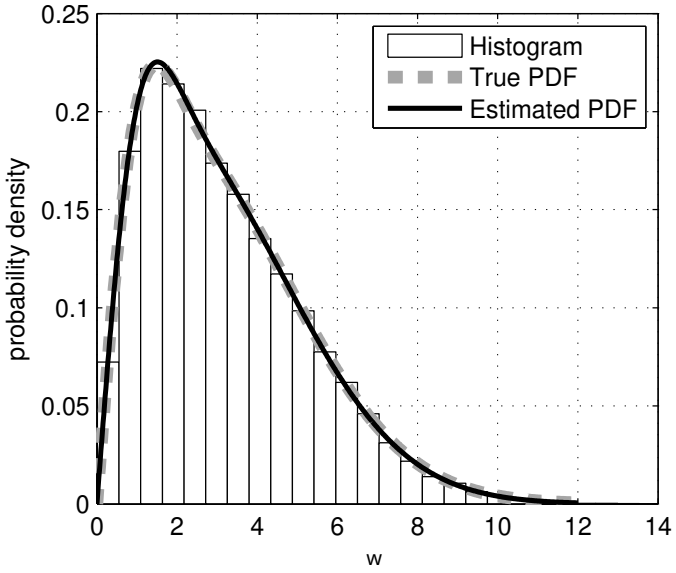

Figure 2.16 Comparison of normalised data histogram, true envelope Gaussian mixture pdf and estimated envelope Gaussian mixture pdf.

achieve some local optima of the log likelihood functions in all cases, therefore the estimated envelope Gaussian mixture pdf's using both methods still exhibit excellent agreement with the histograms.

Despite that the EM algorithm performs well in terms of maximizing the log likelihood function, it has relatively slower convergence rate depending on the models and data size [107]. If we only compare the numbers of iterations to convergence, quasi-Newton would take fewer iterations, since the convergence speed of the quasi-Newton algorithm is super linear, whereas the EM algorithm converges linearly [107]. Table 5.2 shows that the computation iterations for estimation of the parameters of the envelope of two-term Gaussian mixture model with stopping criterion, $\epsilon = 10^{-6}$. It is clear in the Table 5.2 that the quasi-Newton method converges much faster than the EM algorithm. The quasi-Newton method, on average, took less than $1/8$ in iteration numbers of the EM algorithm. For example 4, 6 and 7, number of iteration for the EM algorithm can go over 1000. Moreover, the computation time per iteration of the EM algorithm is much longer than that of the quasi-Newton method, since the variances, σ_k^2 and σ_i^2 in (2.31) have to be found numerically in each EM

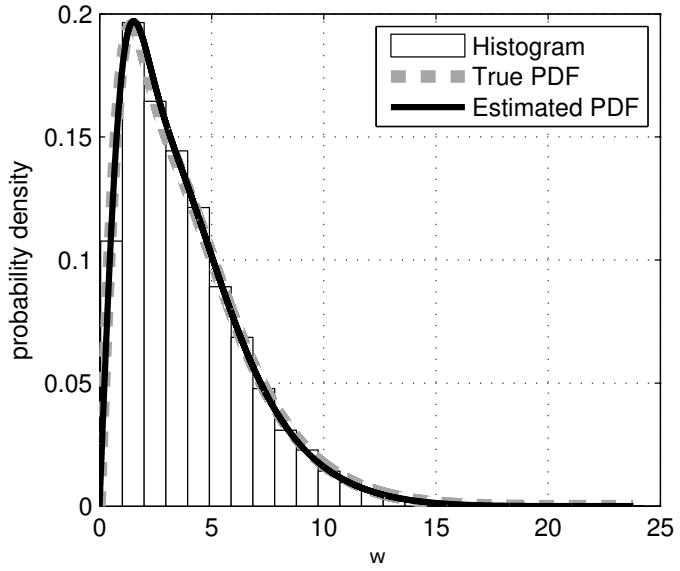

Figure 2.17 normalized data histogram generated with mixing coefficient $\pi_1 = 0.3$, $\pi_2 = 0.4$, $\pi_3 = 0.3$, $\sigma_1^2 = 1$, $\sigma_2^2 = 9$ and $\sigma_3^2 = 25$. Stopping criterion is set to be 10^{-6}. Number of iteration taken for the EM algorithm to converge is 1491. Parameters estimated are $\hat{\pi}_1 = 0.301052$, $\hat{\pi}_2 = 0.406479$, $\hat{\pi}_3 = 0.292469$, $\hat{\sigma}_1^2 = 1.109467$, $\hat{\sigma}_2^2 = 8.970086$ and $\hat{\sigma}_3^2 = 24.172147$.

iteration. For this reason, we conclude that, in terms of iteration numbers, the quasi-Newton method should be preferred to the EM algorithm in estimation parameters of the envelope Gaussian mixture density functions. However, reader should take note that the EM algorithm is still an attractive method in this estimation problem due to greater simplicity (i.e. automatic satisfaction of probability constraints and monotonic convergence without the need to set a step size). On the other hand, implementation of the quasi-Newton method is indeed complicated. As Jamshidian and Jennrich [59] have pointed out, to choose between the EM algorithm and the quasi-Newton method is more or less a personal choice.

2.5 Conclusion

In this chapter, we have derived the general expression for the envelope Gaussian mixture model. Such model is different from the Rayleigh mixture and may be applied to describe the envelope of powerline noise and the envelope of underwater communication noise. We proposed the EM algorithm to estimate the parameters of the envelope Gaussian mixture. Finally, we discussed the convergence speed and accuracy of the EM algorithm by simulations. This was compared with the quasi-Newton algorithm. However, in ill-conditioned cases, the parameters cannot be estimated correctly. Yet The EM algorithm is still able to attain some local optimums which maximize the log likelihood function. The performance of the EM algorithm is compared with conventional quasi-Newton method in terms of iteration number. The results show that, for the envelope Gaussian mixture model, quasi-Newton is generally preferred to the EM algorithm due to a reduced iteration number and computation time.

Table 2.4 Comparison of the performance of the EM algorithm with the performance of the Quasi-Newton Method for The Envelope Gaussian Mixture Model

	True Parameters						Estimated parameters via EM						Estimated parameters via QN						EM	QN
#	π_1	π_2	π_3	σ_1^2	σ_2^2	σ_3^2	π_1	π_2	π_3	σ_1^2	σ_2^2	σ_3^2	π_1	π_2	π_3	σ_1^2	σ_2^2	σ_3^2	Iter	Iter
1	0.3	0.7	-	1	10	-	0.29992	0.70008	-	1.01055	10.00381	-	0.29998	0.70002	-	1.01863	10.00449	-	168	20
2	0.95	0.05	-	1	10	-	0.94084	0.05916	-	0.97562	9.37184	-	0.94085	0.05915	-	0.97564	9.37315	-	145	22
3	0.5	0.5	-	1	10	-	0.50771	0.49229	-	1.00778	10.04186	-	0.50775	0.49225	-	1.00779	10.04249	-	121	19
4	0.5	0.5	-	1	1.5	-	0.40913	0.59087	-	1.01156	1.40103	-	0.365771	0.634229	-	1.10931	1.53166	-	$1000 <$	21
5	0.95	0.05	-	1	1.5	-	0.25011	0.74989	-	0.83423	1.09010	-	0.36550	0.63450	-	0.95190	1.18866	-	103	56
6	0.3	0.4	0.3	1	9	25	0.30305	0.42639	0.27056	1.03800	9.49385	26.098378	0.30265	0.42510	0.27225	1.03661	9.45800	26.04114	$1000 <$	63
7	0.4	0.4	0.2	1	2	25	0.40117	0.40481	0.19402	0.94892	2.11022	26.19052	0.27070	0.53347	0.19583	0.79183	1.87935	26.03970	$1000 <$	74
8	0.2	0.4	0.4	1	9	10	0.18559	0.33575	0.47866	0.80414	9.45651	9.45735	0.18555	0.65777	0.15668	0.80420	9.45740	9.45147	196	56

Chapter 3

Capacity over Gaussian mixture noise channels

In many communications systems, the Gaussian mixture model is extensively used to characterize non-Gaussian man-made and natural interference. We study the capacity of the Gaussian mixture noise channel and its capacity-achieving input. In particular, we consider the symmetric and asymmetric Gaussian mixture noise cases. It is shown that, under average and peak power constraints, the capacity-achieving input is discrete with finitely many mass points. Furthermore, some properties of the capacity-achieving distribution are proved and demonstrated by simulations.

This chapter is organised as follows. Introduction is provided in section 3.1 and we formulate our problems in section 3.2. We introduce the Kuhn-Tucker conditions in section 3.3 that are of fundamental importance to our optimisation problem. In Section 3.4, we present our main result together with some properties of the capacity-achieving input distribution for both symmetric and asymmetric Gaussian mixture noise channels. Section 3.5 is a general consideration for the slope of capacity curve of non-Gaussian noise channel and section 3.6 is the numerical evaluation. Finally, conclusions are drawn in section 3.7.

3.1 Introduction

In wireless communications systems, classic white Gaussian noise is a popular model that has been extensively studied since the origins of communication theory. However, for many practical cases such as powerline communications [33], NAND flash memories [66, 32, 18], man-made electromagnetic interferences [78], and underwater communications [93], the conventional Gaussian model is not sufficient to characterise these environments. One of the most common models for the non-Gaussian noise environments is the Gaussian mixture model. Hence, it is of interest for us to study the theoretic limit of the Gaussian mixture channels.

Many research papers in the literature have focused on the study of what is the optimal signalling for non-Gaussian channels under different constraints. The earliest work was conducted by Smith [42, 101], who studied the channel capacity of AWGN channel under peak and average power constraints. Also, Das showed in [28] that, for a scalar additive channel under average power constraint, if the noise distribution is "heavy-tailed" (i.e., has a tail which decays at a rate slower than the Gaussian), the capacity-achieving input will have a finite support. In addition, Tchamkerten extended this research to a class of non-Gaussian noise distributions and showed that for noise in this class, the input is discrete with finite number of mass points under average and peak power constraints [102]. More recently, Fahs *et. al* showed that under solely average power constraint, the capacity-achieving input for a class of non-Gaussian noise channels is also discrete [38].

Based on the above results, various works analysed specific channels with non-Guassian noises, such as the Rayleigh fading channel [1], non-coherent AWGN channel [62], Rician fading channel [45] and quadrature Gaussian channel [97]. In [112], Zhang studied AWGN channel under duty-cycle constraints. Many of these results are built upon Smith's original work and the derived capacity-achieving distributions are of discrete nature. Readers may refer to [22] for a list of channels which their capacity-achieving inputs are discrete under either or both average power and peak power constraints. This chapter is based on the research results reported in [57] by the author.

3.2 System Model

We consider the following discrete-time, real, and memoryless additive non-Gaussian channel model

$$Y = X + N \tag{3.1}$$

where Y, X, and N are real random variables representing channel output, input, and noise, respectively. The noise random variable, N, is independent of the input and its probability density function $p_N(t;\gamma,a)$ follows a two-term Gaussian mixture distribution

$$p_N(t;\gamma,a) = a\mathcal{N}(t;\mu_1,1) + (1-a)\mathcal{N}(t;\mu_2,\gamma) \tag{3.2}$$

where $\mathcal{N}(\cdot)$ denotes a Gaussian distribution, that is

$$\mathcal{N}(t;\mu,\gamma) = \frac{1}{\sqrt{2\pi\gamma}} e^{-\frac{(t-\mu)^2}{2\gamma}}.$$

The parameters μ_1 and μ_2 are the means of the first and the second Gaussian components, respectively. The variance of the first Gaussian component is normalised to be one and $\gamma > 1$ denotes the ratio of first and second noise variances. In (3.2), $0 < a < 1$ is the *mixing coefficient*. The mean μ_N and the variance σ_N^2 of the random variable N are given by

$$\mu_N = a\mu_1 + (1-a)\mu_2, \tag{3.3}$$

$$\sigma_N^2 = a(\mu_1^2+1) + (1-a)(\mu_2^2+\gamma) - \mu_N^2. \tag{3.4}$$

The Gaussian mixture model is widely utilised in modelling non-Gaussian noise in various areas, including powerline communications, underwater communications, and NAND flash memory channels. Among them, we specifically investigate two types of noise distributions:

- Type I: the noise distribution is symmetric and its Gaussian components have zero mean (e.g. the powerline communications channel [33]);
- Type II: the noise distribution is asymmetric (skewed) and has a non-zero mean (e.g. NAND flash memory channel [66]).

Type I: Powerline communications channel

The signature of the powerline communications channel is that its noise is impulsive. A widely utilised model to describe the powerline noise channel is the Gaussian mixture, in which background noise with unit variance is always present and the impulsive noise events occur with probability $(1-a)$. The first Gaussian component can be seen as the nominal background noise with normalised variance and the second Gaussian component represents the combination of background noise and impulsive noise. Hence, the means of both the Gaussian components in (3.2) are zero, $\mu_1 = \mu_2 = 0$, i.e.,

$$p_N(t, \gamma, a) = a\mathcal{N}(t; 0, 1) + (1-a)\mathcal{N}(t; 0, \gamma). \tag{3.5}$$

Type II: NAND flash memory channel

Unlike many conventional channels where noise is assumed to be zero mean Gaussian, NAND flash memory noise has non-zero mean and is skewed. Therefore, we use the following Gaussian mixture model to approximate the NAND flash memory noise,

$$p_N(t, \gamma, a) = a\mathcal{N}(t; 0, \gamma) + (1-a)\mathcal{N}(t; \mu_2, 1) \tag{3.6}$$

by adjusting μ_2 and γ.

Furthermore, we impose an average input power constraint $\mathbb{E}[X^2] \leq P$ and a peak power constraint $-A \leq X \leq A$ where P denotes the average input power and A denotes the peak amplitude of X. The cumulative density function (cdf) F of input random variable X is in the space of all possible probability distribution function on $\mathbb{R}$

$$\Omega \triangleq \left\{ F(x) : \int_{-A}^{+A} dF(x) = 1, \int_{-A}^{+A} x^2 dF(x) \leq P \right\}.$$

If we normalise the noise variance, then P is the normalised signal-to-noise ratio (SNR).

3.3 The Kuhn-Tucker Conditions

The Kuhn-Tucker (KKT) conditions proposed by [42, 101] are of fundamental importance for this chapter and are used widely to verify the optimality of the inputs for variety of channels. Denoting

$$i(x;F) = \int_{-\infty}^{\infty} p_N(y-x) \log \frac{p_N(y-x)}{p_Y(y;F)} dy \tag{3.7}$$

yields the mutual information

$$I_F(X;Y) = \int_{-A}^{A} i(x;F) dF(x). \tag{3.8}$$

where $p_Y(y;F) \triangleq \int_{-A}^{+A} p_N(y-x) dF(x)$ is the output pdf.. We note that $I_F(X;Y)$ is a function of F. The channel capacity of this memoryless channel, C, is the supremum of the mutual information over the set of all possible input cdfs Ω under the average input power and peak power constraints.

$$\begin{aligned} C &\triangleq \sup_{F \in \Omega} I_F(X;Y) \\ &= \sup_{F \in \Omega} \int_x \int_{\mathbb{R}} p_N(y-x) \log \frac{p_N(y-x)}{p_Y(y;F)} dy dF(x) \end{aligned} \tag{3.9}$$

The KKT conditions serve as necessary and sufficient conditions for an average power and peak power constrained input cdf $F^* \in \Omega$ to achieve the capacity.

Theorem 1 (KKT conditions [101]). *F^* is a capacity-achieving input distribution if and only if there exists a $\lambda > 0$ such that,*

$$\lambda(x^2 - a) + C - i(x;F) \geq 0 \quad \forall x \in [-A, A] \tag{3.10}$$

with equality if x is a point of increase of F^*.[1]

[1]We say that a point, x, is a point of increase of F^* if for every $\epsilon > 0$,

$$F^*(x+\epsilon) - F^*(x-\epsilon) > 0.$$

In order to validate the usage of the KKT conditions, the following Lemmas have to hold. Their proofs are given in the references.

Lemma 1 ([101]). *Ω is a convex and compact space in the Levy metric.*

Lemma 2 ([102]). *The mutual information $I_F(X;Y)$ is strictly concave, continuous, and weakly differentiable[2] on Ω.*

3.4 The Capacity-achieving Distribution

In this section, we demonstrate the properties of the capacity-achieving input distribution under average and peak power constraints. Let p_X be the pdf of a random variable X such that $p_X = dF(x)$. We have the following theorem.

Theorem 2. *The capacity of the Gaussian mixture noise channel under average power constraint P and peak power constraint $|X| \leq A$ is*

$$C(P) = \max_{F \in \Omega} I_F(X;Y).$$

Furthermore, the capacity-achieving distribution exhibits the following properties:

(i) *$C(P)$ is achieved by a unique input distribution $F^* \in \Omega$;*

(ii) *$p_X^* = dF^*$ is discrete and has finite number of probability mass points; and*

(iii) *p_X^* is symmetric if p_N is symmetric.*

Sketch of the proof of (i). By Lemma 1 and Lemma 2, mutual information $I_F(X;Y)$ is continuous and Ω is compact on Levy metric. Then $I_F(X;Y)$ is a compact subset of $\mathbb{R}$ which contains its maximum for some F. Since $I_F(X;Y)$ is strictly concave and Ω is convex, uniqueness of the capacity-achieving input is established. Detailed proof of Property (i) can be found in [101]. □

[2] Let f be a function from Ω into $\mathbb{R}$. Let $\theta \in [0,1]$ and fix $x_0 \in \Omega$. If for all $x \in \Omega$ the limit

$$\lim_{\theta \downarrow 0} \frac{f(x_0 + \theta(x - x_0)) - f(x_0)}{\theta}$$

exists, f is said to be weakly differentiable in Ω at x_0.

Proof of property (ii). Property (ii) states that the capacity-achieving distribution p_X^* is of discrete nature and has finitely many probability mass point. To prove the discrete nature of the input, we recall the condition stated in [102], i.e., for noise pdf $p_N(\cdot)$ that satisfies the following conditions, the capacity-achieving distribution is discrete with finite number of mass points,

1. $p_N(t) > 0, \forall t \in \mathbb{R}$;
2. there exists $\alpha > 0$ such that $\mathbb{E}[|N|^\alpha] < \infty$;
3. there exists, a $\delta > 0$ such that $p_N(\cdot)$ admits an analytic extension on $\mathcal{D}_\delta = \{z \in \mathbb{C} : |\Im(z)| < \delta\}$; [3]
4. There exist, $k \geq 0$ and two non-increasing functions $\rho_L : [k, \infty) \to \mathbb{R}_+$ and $\rho_U : [k, \infty) \to \mathbb{R}_+$ such that:
   (a) for all $z \in \mathcal{D}_\delta$ with $|z| \geq k$,
   $$0 < \rho_L(|\Re(z)|) \leq |p_N(z)| \leq \rho_U(|\Re(z)|) \leq 1; \tag{3.11}$$
   (b) $H(\rho_U)|_k < \infty$;
   (c) $\int_{y \geq k+x} \frac{(U(y-x))^3}{(L(y))^2} dy < \infty$ for all $x \in \mathbb{R}_+$,

where $\rho_L(|\Re(z)|)$ and $\rho_U(|\Re(z)|)$ are the lower and upper bounds of $|p_N(z)|$. $H(\rho_U)|_k$ is defined as

$$H(\rho_U)|_k \triangleq -\int_{y \leq k} \rho_U(y) \log \rho_U(y) dy.$$

Properties 1 and 2 are straightforward for the Gaussian mixture distribution. Also, the $p_N(z)$ is an entire function (i.e. analytic in the whole complex plane) [2], hence Property 3 holds. In order to prove Property 4(a), for some positive real constants $c_1 = \max p_N(t)$, c_2, μ_1, μ_2 and, without loss of generality, assuming $\mu_1 \leq 0 \leq \mu_2$, the tails of the Gaussian mixture have the following upper

[3] Let $\mathcal{E}$ be a complex open connected set including the real line. A mapping $f : \mathbb{R} \to \mathbb{R}$ is said to have an analytic extension on $\mathcal{E}$ if there exists a mapping $g : \mathbb{C} \to \mathbb{C}$ which is analytic on $\mathcal{E}$ and so that $g(x) = f(x)$ for any $x \in \mathbb{R}$.

bound for $\Re(z) < \mu_1$,

$$\begin{aligned} |p_N(z)| &\leq |c_1 e^{-\frac{(z-\mu_1)^2}{c_2}}| = c_1 e^{\frac{|\Im(z)|^2}{c_2}} e^{-\frac{|\Re(z)-\mu_1|^2}{c_2}} \\ &< c_1 e^{\frac{\delta^2}{c_2}} e^{-\frac{|\Re(z)-\mu_1|^2}{c_2}} = c_1 e^{\frac{\delta^2}{c_2}} e^{-\frac{(|\Re(z)|+\mu_1)^2}{c_2}}. \end{aligned} \tag{3.12}$$

Similarly, for $\Re(z) > \mu_2$,

$$|p_N(z)| < c_1 e^{\frac{\delta^2}{c_2}} e^{-\frac{(|\Re(z)|-\mu_2)^2}{c_2}}. \tag{3.13}$$

Hence for any $z \in \mathcal{D}_\delta$, $|p_N(z)|$ is upper bounded by

$$\rho_U(|\Re(z)|) = \begin{cases} c_1 e^{\frac{\delta^2}{c_2}} e^{-\frac{(|\Re(z)|+\mu_1)^2}{c_2}} & \text{if } \Re(z) < \mu_1 \\ c_1 e^{\frac{\delta^2}{c_2}} & \text{if } \mu_1 \leq \Re(z) \leq \mu_2 \\ c_1 e^{\frac{\delta^2}{c_2}} e^{-\frac{(|\Re(z)|-\mu_2)^2}{c_2}} & \text{if } \Re(z) > \mu_2 \end{cases}. \tag{3.14}$$

As for the lower bound, $p_N(z)$ can be lower bounded by one of the Gaussian component. Hence we have, for some positive real constants c_3, c_4 and $\mu_1 \leq 0 \leq \mu_2$,

$$\begin{aligned} |p_N(z)| &\geq |c_3 \cdot e^{-\frac{(z-\mu_1)^2}{c_4}}| \geq c_3 \cdot e^{-\frac{|z-\mu_1|^2}{c_4}} \\ &\geq c_3 \cdot e^{-\frac{(|\Re(z)|-\mu_1+\delta)^2}{c_4}} = \rho_L(|\Re(z)|). \end{aligned} \tag{3.15}$$

Properties 4(b) and 4(c) follow the proof in [102]. Hence the capacity-achieving pdf p_X^* is of discrete nature and has finite number of mass points if its support is bounded. □

Note that under the average power constraint only, the capacity-achieving input is also discrete by showing that p_N satisfies the conditions listed in [38]. Property (iii) states that the capacity-achieving input distribution is symmetric about zero if the noise is symmetric.

Sketch of the proof of (iii). This property is justified by the fact that, for symmetric noise, the mirror reflection of the capacity-achieving input is also capacity-achieving. However, by Property (i), the capacity-achieving input is unique which implies that the capacity-achieving input must be symmetric. Hence for

Type I noise model, the capacity-achieving input is also symmetric. Detailed proof of this property can be found in [101]. □

Remark: Property (iii) states that if the noise distribution is symmetric, then the capacity-achieving input distribution is also symmetric. On the other hand, if the noise is asymmetric (i.e. Type II), the symmetry of the input may not hold. However, what will hold is that the capacity-achieving input distribution will still have a zero mean. This is simply justified by the fact that the input with a mean shift does not increase the capacity but consumes more input power. This implies that if the capacity-achieving distribution is asymmetric, the mean of the capacity-achieving input is always zero. Furthermore, the second moment of the capacity-achieving input is P, which is complied with the KKT conditions.

3.5 Concavity, monotonicity and the slope of non-Gaussian capacity curve

By Property (ii), the optimal input has finitely many mass points. Hence, we assume the number of mass points M to be finite and solve the following optimisation problem,

$$\underset{p_X}{\text{maximize}} \quad I(X;Y) - \lambda \mathbb{E}[X^2]$$

where $p_X(x) = \sum_{i=1}^{M} \pi_i \delta(x - x_i)$, $\{x_i\}_{i=1}^{M}$ are the mass point locations with probability $\pi_i \in [0,1]$ and $\sum_{i=1}^{M} \pi_i = 1$. Note that $\lambda > 0$ is the Lagrange multiplier given the input power P, which can be interpreted as the slope of capacity curve $C(P)$, $\lambda(P) = \frac{d}{dP} C(P)$. By concavity and monotonicity of the mutual information, $\lambda < \lambda(0)$ for any normalised SNR, where $\lambda(0)$ is the slope of capacity curve $C(P)$ at $P = 0$. Since there exists a unique $C(P)$ for given $\lambda(P)$, by varying λ between 0 and $\lambda(0)$ (equivalently varying P between high SNR and low SNR), the capacity curve $C(P)$ can be plotted for different P. The closed form of $\lambda(0)$ is derived as follows.

Assuming X' is a normalized power input, i.e., $P = \mathbb{E}[X^2] = P\mathbb{E}[X'^2]$, where $\mathbb{E}[X'^2] = 1$. Without losing generality, the channel model becomes

$Y = \sqrt{P}X' + N$. For this channel model, the derivative of mutual information has the following asymptotic behaviour for $P \downarrow 0$ [44]

$$\lim_{P\downarrow 0} \frac{d}{dP} I(X;Y) = \frac{\sigma_{X'}^2}{2} \mathbb{E}\{[\nabla \log p_N(w)]^2\} \tag{3.16}$$

where $\sigma_{X'}^2$ is the variance of scaled input random variable X'. $\nabla \log p_N(w)$ is the score of the noise distribution p_N which always has zero mean [26] and $\mathbb{E}\{[\nabla \log p_N(w)]^2\}$ is known as the Fisher Information of the noise distribution $p_N(w)$.

As stated in Section 3.4, the capacity-achieving input X^* has a zero mean, therefore X' also has zero mean and $\sigma_{X'}^2 = \mathbb{E}[X'^2] = 1$. The slope of the channel capacity $C(P)$ when $P \downarrow 0$ is then given by

$$\lim_{P\downarrow 0} \frac{d}{dP} I(X;Y) = \frac{1}{2} \mathbb{E}\{[\nabla \log p_N(w)]^2\}. \tag{3.17}$$

In fact, the above formula gives the rate of increase with respect to the input power P for the non-Gaussian channel capacity around $P = 0$. This formula also indicates that the slope of mutual information at zero input power only depends on the noise distribution $p_N(w)$. For example, for type I noise model, we have

$$\nabla \log p_N(w) = -a \cdot w - (1-a)\frac{w}{\gamma},$$
$$\mathbb{E}\{[\nabla \log p_N(w)]^2\} = \left(a + \frac{(1-a)}{\gamma}\right)^2 (a + (1-a)\gamma).$$

Hence, for Type I noise channel, the slope of the capacity curve at $P = 0$ is exactly given by

$$\lambda(0) = (1+\epsilon) \cdot \left(\frac{a}{2} + \frac{(1-a)}{2\gamma}\right) \tag{3.18}$$

where $\epsilon = (\gamma + \frac{1}{\gamma} - 2) \cdot a(1-a) > 0$.

Furthermore, for the capacity $C(P) = \max_{p_X}[H(Y) - H(N)]$, since the noise entropy $H(N)$ is a constant, maximizaximizing mutual information is the same as maximizing output entropy $H(Y)$. For a fixed output power P_{out},

the Gaussian output maximizes the entropy and we have the following upper bound

$$C(P) < \frac{1}{2}\log\left(2\pi e P_{out}\right) - H(N) \quad \text{(nats)}. \tag{3.19}$$

It should be noted that, by Cramer's Decomposition Theorem [27], a Gaussian output is achieved only if both the input and noise are Gaussian. Hence for a non-Guassian Channel, Gaussian entropy is always an upper bound of the output entropy, which cannot be achieved.

We further show that the mutual information imposed by the Gaussian input is a good approximation to the channel capacity by simulation.

3.6 Numerical Evaluation

In this section, we numerically compute the optimal input, i.e., the optimal mass point locations and their probabilities, by using the interior point method and we start from random initial mass point locations and probabilities. Note that, although the original problem itself is convex, optimisation with respect to the number of points, the mass point locations and the probabilities are in general non-convex. Hence, the numerical solutions maybe sub-optimal.

3.6.1 Type I: Symmetric Gaussian Mixture Noise

For Type I noise model (3.5), we choose $a = 0.7$ and $\gamma = 10$ which are typical values for noise in the powerline channel [33]. The program is terminated when the improvement in mutual information is less than 10^{-6} for a fixed SNR. Figure 3.1(a) illustrates the Capacity with average and peak power constraints. Upper bound is asymptotically close to capacity for large SNRs. The performance of the Guassian input is also close to the performance of the optimal input as shown in Figure 3.1(a).

Fig 3.2 illustrates the optimal mass point locations with their corresponding probabilities under average power constraint P and amplitude constraint $A \leq 2.5$. At low SNRs (< -6dB), equiprobable binary input is optimal. An additional point is required as SNR increases to -6dB as shown in Figure 3.2. At high

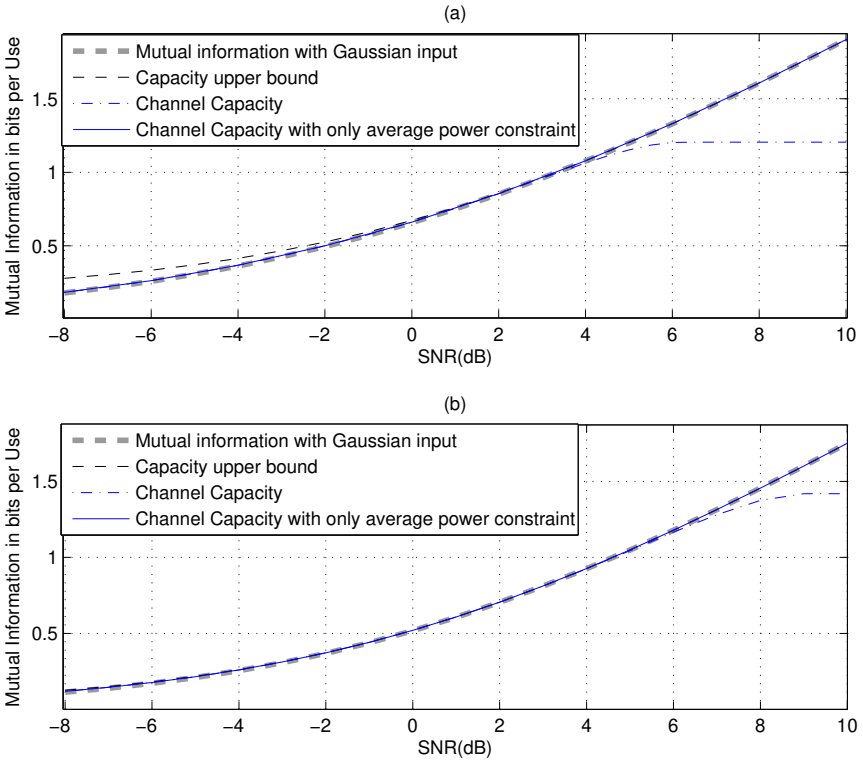

Figure 3.1 (a): Capacity C(SNR) (bits per channel use) of symmetric Gaussian mixture noise channel vs. normalised SNR (dB). (b): Capacity C(SNR) (bits per channel use) of asymmetric Gaussian mixture noise channel vs. normalised SNR (dB).

SNRs, where peak power constraint is solely active, more weights are allocated to the points near the boundary.

3.6.2 Type II: Asymmetric Gaussian Mixture Noise

For type II Gaussian mixture model (3.6), we choose $a = 0.5$, $\gamma = 2$ and $\mu_1 = 1$ such that the unimodality still holds for the asymmetric Gaussian mixture model. We set the peak amplitude constraint to be $A = 3.5$ so that the peak and average power are both active at low SNRs. Figure 3.1(b) shows the capacity under average and peak power constraints. The upper and Gaussian approximation are asymptotically close to capacity for large SNRs.

In Figure 3.4, binary input is optimal at low SNRs, before -5dB, two mass points are not enough to achieve optimal rate and more points are required. In

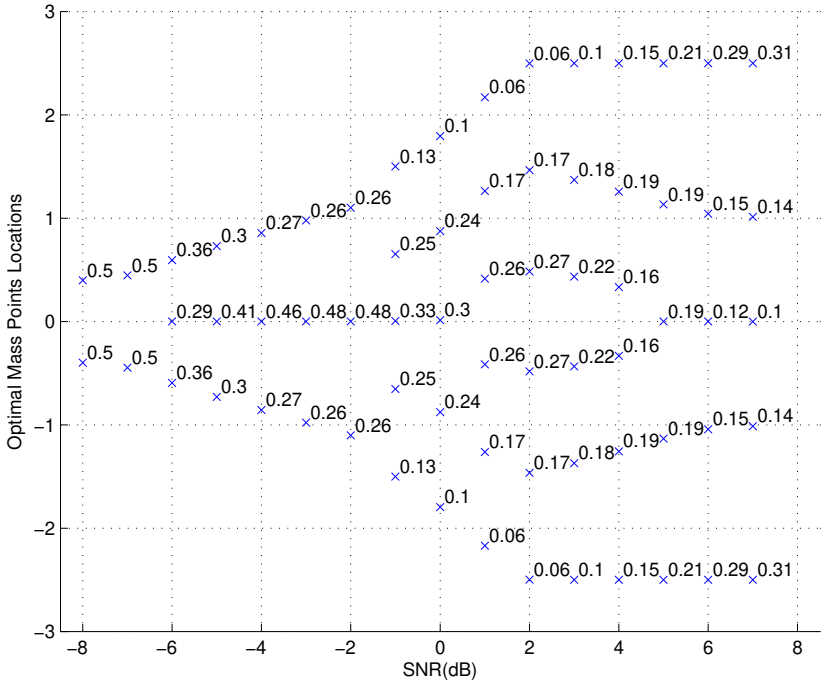

Figure 3.2 Optimal mass points of symmetric Gaussian mixture channel vs. normalised SNR(dB).

addition, the symmetry of the input in general does not hold for asymmetric noise channel especially at low SNRs. As SNR increases, the input tends symmetric as shown in Figure 3.4.

Figure 3.5 illustrates an interesting point that if the noise is negative-skewed, the input will be positive-skewed. A reason for this is that since Gaussian entropy is always an upper bound of the output entropy, maximizing the output entropy is equivalent to minimising the Kullback-Leibler divergence between some output $p_Y(y;F)$ and the Gaussian output (i.e. $\min D_{KL}(p_Y||\mathcal{N})$) [28]. Therefore, the input tries to retain a symmetric Gaussian output (though not achieved) by skewing itself towards the opposite direction as shown in Figure 3.5.

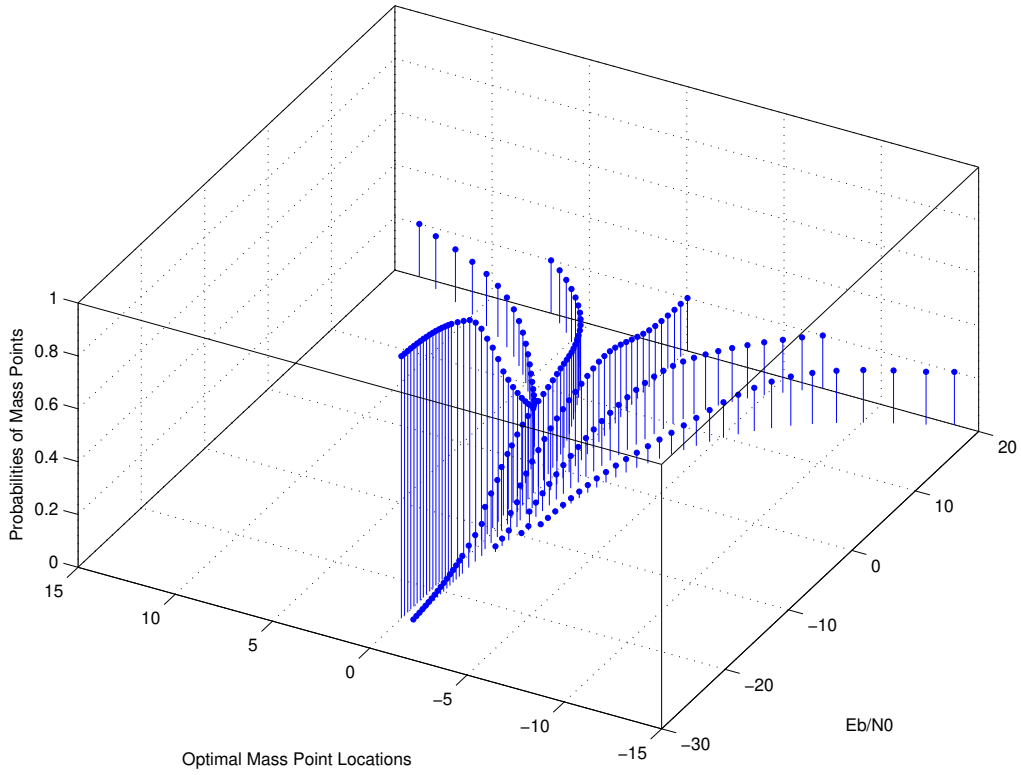

Figure 3.3 Optimal mass points of asymmetric Gaussian mixture channel in 3 Dimensions for $A = 20$.

3.7 Conclusion

In this chapter, we considered the symmetric and asymmetric additive Gaussian mixture noise channels under both average and peak power constraints. The unique capacity-achieving input for the Gaussian mixture noise model is of discrete nature with finite number of mass points. Furthermore, we showed that the capacity-achieving input is symmetric for the symmetric noise distribution.

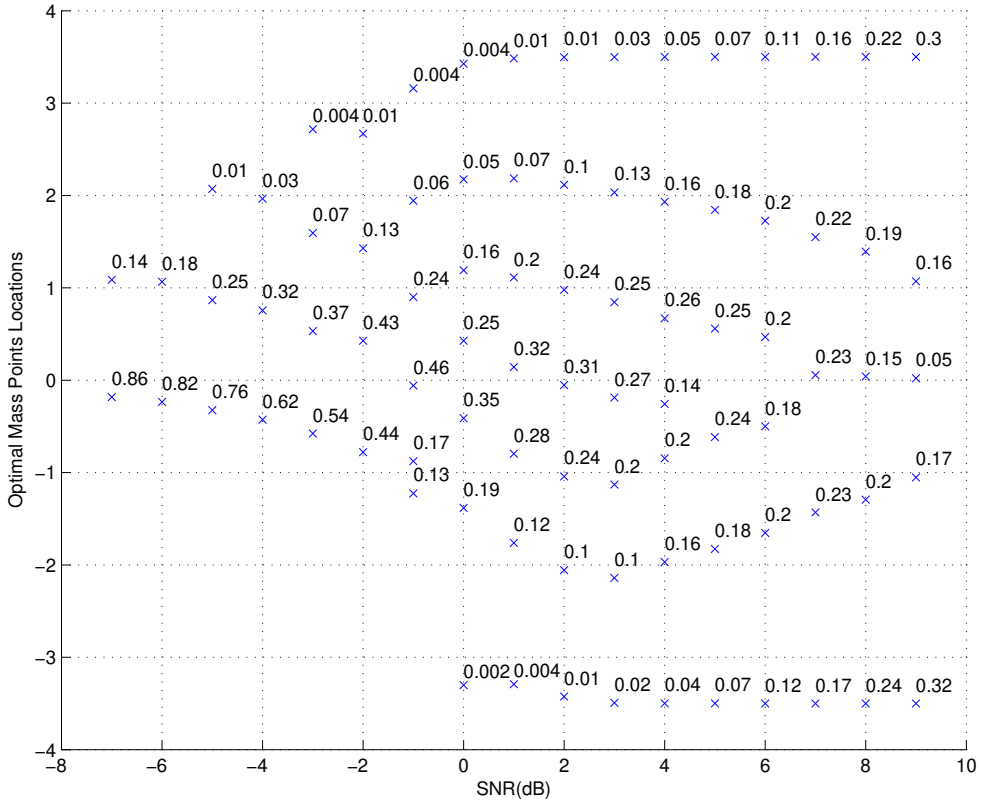

Figure 3.4 Optimal mass points of asymmetric Gaussian mixture channel vs. normalised SNR(dB).

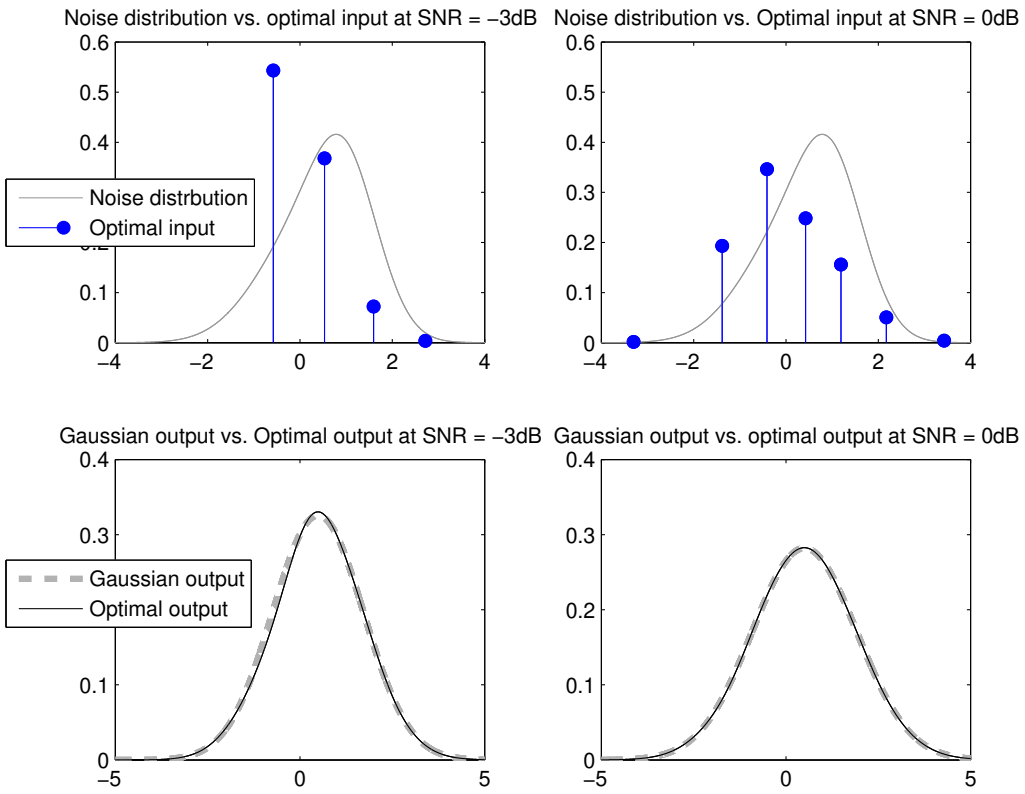

Figure 3.5 On top: Capacity-achieving input vs. noise distribution at -3dB and 0dB, respectively. On bottom: Optimal output vs. Gaussian with the same mean and variance at -3dB and 0dB, respectively.

Part II

Communication over Markovian Constrained Relay Channels

List of Acronyms and Symbols

List of Acronyms

i.i.d. . . . . . independent and identically distributed

pdf . . . . . probability density function

cdf . . . . . cumulative density function

RLL . . . . . runlength limited

GBAA . . . . . Generalised Blahut-Arimoto Algorithm

DF . . . . . decode-and-forward

CF . . . . . compress-and-forward

AF . . . . . amplify-and-forward

AEP . . . . . Asymptotic Equipartition Property

BSC . . . . . binary symmetric channel

AWGN . . . . . additive white Gaussian noise

List of Symbols

X_n . . . . . symbol transmitted by source node at time n

x_n . . . . . realisation of X_n

V_n . . . . . symbol received by relay node at time n

U_n . . . . . symbol transmitted by relay node at time n

Y_n . . . . . symbol received by destination node at time n

$\boldsymbol{X}$ . . . . . sequence transmitted by source node of length N

$\boldsymbol{V}$ . . . . . sequence received by relay node of length N

$\boldsymbol{U}$ . . . . . sequence transmitted by relay node of length N

$\boldsymbol{Y}$ . . . . . sequence received by destination node of length N

S_1^N . . . . . state sequence of length N

$\mathcal{S}$ . . . . . state space of S_n

W_b . . . . . codeword sent in transmission block b

$\mathbf{P}$ . . . . . probability transition matrix

$\boldsymbol{\mu}$ . . . . . stationary distribution on states

P_{ij} . . . . . i,j-th entry of $\mathbf{P}$

$\hat{w}$ . . . . . estimation of w

R . . . . . rate

$I(X;Y)$ . . . . mutual information

$H(X)$ . . . . . entropy

C . . . . . capacity

$C_{\mathbf{nf}}$ . . . . . capacity with noise-free relay-to-destination link

$\mathcal{I}(U_1^N;Y_1^N)$ . . mutual information rate between sequence U_1^N and Y_1^N

$\mathcal{L}_\theta$ . . . . . set of admissible runlengths of symbol θ

$\mathbf{I}$ . . . . . relay mutual information matrix

$\mathbf{G}$ . . . . . noisy adjacency matrix

$\mathbf{A}(\boldsymbol{\zeta})$ . . . . . relay adjacency matrix

$\tilde{\mathbf{A}}(\zeta)$ . . . . . noisy-relay adjacency matrix

$\boldsymbol{l}(\zeta)$ . . . . . left eigenvector of $\mathbf{A}(\boldsymbol{\zeta})$

$\boldsymbol{r}(\zeta)$ . . . . . left eigenvector of $\mathbf{A}(\boldsymbol{\zeta})$

$\lambda_{\max}(\zeta)$ . . . . largest eigenvalue of $\mathbf{A}(\boldsymbol{\zeta})$

$c(\zeta)$ . . . . . normalisation constant for $\boldsymbol{l}(\zeta)$ and $\boldsymbol{r}(\zeta)$

$\mathbf{T}$ . . . . . a-posteriori state transition weight matrix

$\mathcal{T}$ . . . . . set of all valid transitions from state i to j

$\mathcal{Q}$ . . . . . set of all valid transitions from state i to j that are labeled with 0

$\mathcal{A}$ . . . . . set of jointly typical sequences

$\mathcal{E}$ . . . . . error event

Chapter 4

Capacity over a Two-Hop Half-Duplex Relay Channel with a Markovian Constrained Relay

In this chapter, we consider a two-hop half-duplex relay channel (source-relay-destination) with a Markovian constrained relay. Suppose the source and the relay wish to transmit information to the destination under the condition that the relay is half-duplex constrained. The capacity of such channel is shown to be equal to the well-known cut-set upper bound.

This chapter is organised as follows. In section 4.1, we provide literature review on the relay channels and finite-state models. In section 4.2, the system model is introduced in which we give half-duplex constraint and Markovian constraint in detail. Cut-set bound on the capacity is given in section 4.3 and we show that timing scheme achieves the cut-set bound in 4.4.

4.1 Introduction

THE three cornerstones of this part of the thesis are relay channels, finite-state model and information theory. Relay channels serve as the channel model of the problem, in which the relay transmitters are modelled by finite-state models. This introduces memories into the relay system. We wish to compute the mutual information between the transmitter at the source and receiver at the destination. The first difficulty of this problem comes from the

fact that the capacities of relay channels are in general not known, the problem becomes harder when memory is introduced into the system. The second difficulty comes from the fact that the output of such relay channel is the noisy version of a finite-state model (i.e., a hidden Markov chain), and computation of the entropy of hidden Markov chain remains an open problem. To overcome these obstacles, we relate them to information theory via stationarity and ergodicity, i.e., the Shannon-McMillan-Breiman theorem. This allows us to accurately estimate and maximise the information rates for such channel. This chapter is based on the research results reported in [HH18] by the author.

4.1.1 Relay channel

The relay channel models a communication scenario, where a wireless link is aided by relays to increase the spectral efficiency of a wireless communication system. A relay channel consists of one sender, one receiver with a number of intermediate relays which helps communicate from the sender to the receiver in order to meet some of the demands in next generations of wireless system [55, 74, 84, 109]. Particularly, in modern wireless networks, relay nodes may be used to extend and improve the coverage, and hence the reliability of the base stations. This can be achieved without suffering the high cost of adding extra base station. Hence, relaying is one of the key features that has been considered in several wireless standards [74, 109]. The simplest relay channel has one sender, one receiver and one relay node as depicted in Figure 4.1

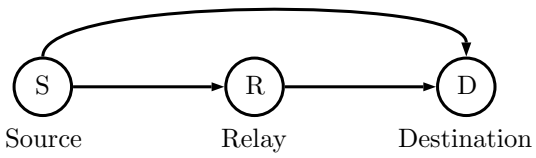

Figure 4.1 The relay channel.

The general model for a relay channel that considers a source communicating with a destination via a relay was first studied by Van der Meulen [103] and later by Sato [24]. Further advances in theoretical limits and coding of the relay channel were made by Cover and El Gamal in [25]. Kramer *et al.* gave a comprehensive survey on single/multiple-relay channels in [69].

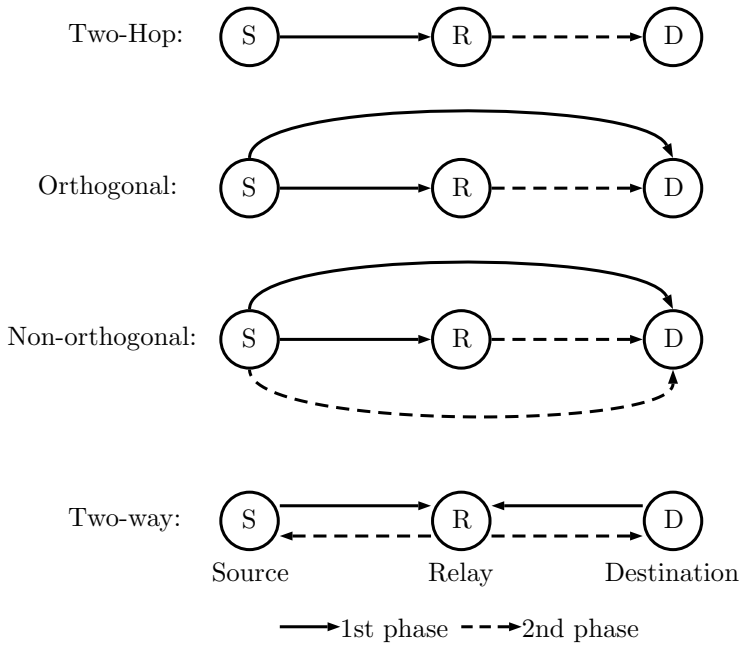

Figure 4.2 Half-duplex relay protocols.

Half-duplex Relaying

In many modern wireless communication systems, the relay is not able to operate in full-duplex mode, i.e. to transmit and receive signals simultaneously, hence it is natural to consider half-duplex relay models [16, 49, 51, 50]. In half-duplex protocols, information is sent from source node to destination node via the relay in two phases as shown in Figure 4.2. In the first phase, the source node communicate to the relay node and the relay node remains silent, whereas in the second phase, the relay node talks to the destination node and the source node remains silent.

The advantage of the half-duplex relaying protocol is that it can be easily implemented in practice to extend coverage of the base station and improve reliability. Half duplex assumption also makes the capacity problem much simpler than the general relay counterpart as it decouples the problem into source-relay cooperation and relay-destination cooperation [70]. However, half-duplex protocols are not spectrally-efficient as the source node is only allows to transmit every other phase.

When the direct source-to-destination link is available, cooperative relaying protocols are introduced to achieve higher information rates comparing to its

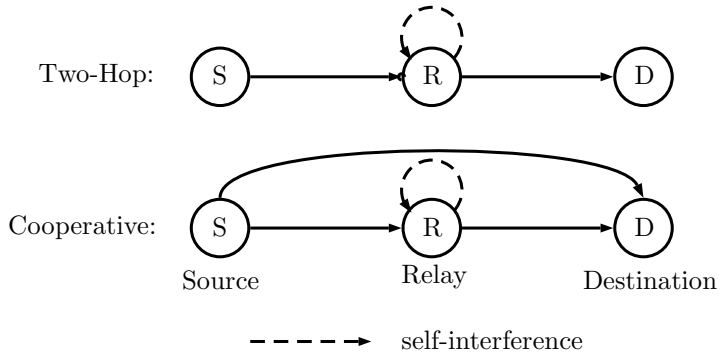

Figure 4.3 Full-duplex relay protocols.

dual-hop counterpart [7, 70, 72, 82, 83, 96]. Pioneer works [36] on cooperative relaying focuses on orthogonal schemes in which the destination node also listens to the source node in the first phase as shown in Figure 4.2. To further increase the information rate, non-orthogonal schemes are proposed in [7, 82]. For non-orthogonal schemes as depicted in Figure 4.2, the source is able to communicate with destination in both phases, however computation of the capacity and analysis of such scheme are much more challenging than dual-hop and orthogonal schemes [7, 82, 31].

Analysis and optimisation of information rate in half-duplex dual-hop and orthogonal schemes was conducted in past two decades. El Gamal [36] considered the general capacity expression for relay channels with orthogonal components. Kramer [68] developed the half-duplex model and he showed that higher information rates can be achieved when the relay random switching between transmission mode and reception mode since the silent symbols may also convey information to the destination. Capacity achieving input distribution for Gaussian channel with half-duplex/duty cycle constraint was proved to be discrete by Lei *et al.* [112]. Zlatanov *et al.* [115] extended these results and computed the capacity for two-hop half-duplex relay channel.

Full-duplex relaying

In order to counter the drawbacks of half-duplex relaying, full-duplex relaying is proposed which provides higher spectral efficiency comparing against the half-duplex counterpart. Full-duplex was believed to be impractical since the relay transmitting power is usually orders of magnitude larger than the received

signal power which results in heavy self-interference at the relay node. In the past decades, self-interference mitigation techniques [6, 12, 15, 23, 35, 34, 37, 48, 58, 65, 87, 88, 91] are proposed which make full-duplex implementation possible. These techniques are divided into two main categories: analog and digital techniques. For example, basic analog cancellation techniques include antenna separation [35, 34, 48, 88], orientation [88, 91] and directionality [37, 87]. These analog mitigation schemes can be combined with digital techniques such as time domain subtraction for further mitigation [12, 35, 34, 37, 58]. However, the self-interference cannot be completely removed which results in the residual self-interference. The residual self-interference shall always be taken into account while modelling the full-duplex relaying as illustrated in Figure 4.3. Zlatanov *et. al.* [116] consider the capacity of a two-hop full-duplex relay channel using similar approaches used in [115, 112].

Protocols and Capacity-achieving Transmission Strategy

To this day, the capacity of relay channel with its respective optimal relay function are only known in few cases. Common relay protocols include *decode-and-forward* (DF), *compress-and-forward* (CF) and *amplify-and-forward* (AF) [67, 69]. In DF relaying scheme, the relay decode the information sent by the source, encodes it then forward it to the destination in the next time block. In CF relaying scheme, the relay quantises the received symbols and then forward them to the destination. Last but not least, AF scheme amplifies the the signal and forward it the destination.

Motivated by DF protocols, the following papers studied how relay shall forward information to destination in order to approach the capacity. In [68], Kramer showed that if the relay switches among the three sleep-reception-or-transmission modes, each occupying a certain fraction of the time, then higher information rates can be achieved, when compared to that of a fixed mode structure, i.e., all terminals know at all times which mode (receive or transmit) every terminal is using. Timing strategy is also proved to be capacity achieving for deterministic half-duplex relay networks [75]. Furthermore, capacity achieving input distribution for Gaussian channel with half-duplex/idealised duty cycle constraint was proved to be discrete by Lei *et al.* [112]. Zlatanov *et*

al. [115] extended these results to a two-hop half-duplex relay channel and computed the capacity.

All the above results were obtained for memoryless relay channels. In [76], Marina *et al.* carried out pioneering work on capacity theorems for relay channels with finite length intersymbol interference (ISI), or equivalently with memory. For such channels, the authors showed the capacity-achieving transmission strategy of the relay and proved a special structure for the capacity achieving distributions of the source and relay signals.

4.1.2 Finite-State Models

Markov sources were first considered by Shannon in his landmarking paper [99], in which Shannon computed the maximal entropy rate of a discrete-time Markov source, or equivalently the noise-free capacity of constrained sequences. In the presence of the noise, the computation of capacity of Markovian inputs sending over a noisy channel has long been an open problem[111, 94, 106]. Zehavi and Wolf [111], and independently, Shamai and Kofman [98] considered the noisy channel with run-length limited sequences and derived a set of analytical lower and upper bounds on the capacity. Arnold *et al.* [5], Sharma and Singh [100], Pfister *et al.* [86] proposed a Monte Carlo method for computing the exact mutual information of a finite-state machine channel with Markovian inputs. In order to maximize mutual information of Markovian inputs transmitted over a finite-state machine channel, a Blahut-Arimoto Algorithm for Markovian inputs was proposed in [63]. Vontobel *et al.* [106] extend this to a generalised Blahut-Arimoto Algorithm (GBAA) considering local convergence properties. The global convergence of the GBAA requires the concavity of the mutual information between input and output; and the concavity of certain conditional entropy. Han [46] proposed a randomised algorithm to compute the capacity of a finite-state channel which only requires the concavity of mutual information. Han also showed in [47], that the concavity conjuncture of mutual information in finite-state machine channel does not hold in some cases. Hence, despite the good performance of these algorithms in many practical examples, the convergence of these algorithms, in general, is not guaranteed.

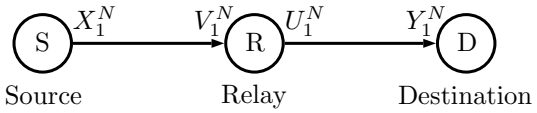

Figure 4.4 A two-hop relay channel.

In this chapter, we consider a two-hop relay channel (see Figure 4.4), where the relay is

- Markovian constrained, i.e., the symbols sent by the relay are generated by a finite-state machine;
- half-duplex.

4.2 System Model

A two-hop half-duplex relay channel, as depicted in Figure 4.4, is denoted by $(\mathcal{X} \times \mathcal{U}, p(v, y|x, u), \mathcal{V} \times \mathcal{Y})$. The channel is defined by four alphabet sets: $\mathcal{X}, \mathcal{U}, \mathcal{V}, \mathcal{Y}$ and a conditional probability function $p(v, y|x, u)$ on $\mathcal{V} \times \mathcal{Y}$ for given inputs $X_n \in \mathcal{X}$ and $U_n \in \mathcal{U}$. For the n-th transmission cycle, the source sends a symbol $X_n \in \mathcal{X}$ to the relay and the relay sends a symbol $U_n \in \mathcal{U}$ to the destination over the channel $p(v, y|x, u)$. The alphabets $\mathcal{X}, \mathcal{U}, \mathcal{V}, \mathcal{Y}$ are assumed to be finite and the relay channel is memoryless in the sense that (v_n, y_n) depends on only current transmitted symbols (x_n, u_n).

We emphasize that the difference between our setting and an ordinary two-hop relay channel is that the relay is not free to choose every letter in $\mathcal{U}$ during each transmission; instead, the relay is Markovian constrained i.e., sequences transmitted by relay are governed by a finite-state machine. We further consider the relay to be half-duplex.

A $(2^{NR}, N)$ code for the two-way relay channel consists of:

- a set of messages $\{1, 2, ..., 2^{NR}\}$,
- an encoder that maps each message w into a codeword $x_1^N(w)$ of length N,
- a relay with a Markovian encoder that generates a symbol $u_n(v_1^{n-1}, u_{n-1})$ depending on the past sequence v_1^{n-1} and the previously emitted symbol

u_{n-1}, except for $u_2(v_1)$ which is solely dependent on the first received symbol v_1,

- a decoding function that maps each received sequence y_1^N into an estimate $\hat{w}(y_1^N)$.

A rate R is achievable if there exists a sequence of $(2^{NR}, N)$ codes with $P_e^{(N)} = \Pr\{\hat{W} \neq W\} \to 0$, as $N \to \infty$. Channel capacity C is defined as the supremum over the set of achievable rates.

4.2.1 Markovian Constraint

We consider a relay node that transmits constrained sequences U_1^N generated by a finite-state machine to the destination. Denote S_n the machine's state at time n, with realization $s_n \in \mathcal{S} \triangleq \{1, ..., M\}$. We assume that the sequence of states S_n forms a time-invariant Markov chain which is characterized by its initial state S_0 and a probability transition matrix $\mathbf{P}$ with the entries P_{ij} being the state transition probabilities of the Markov process denoted by

$$\mathsf{P}_{ij} = \Pr(S_{n+1} = j | S_n = i) \quad \forall i \in \mathcal{S},\ j \in \mathcal{S}. \tag{4.1}$$

For each row, we have $\sum_{j=1}^{M} \mathsf{P}_{ij} = 1$. We define $\mathcal{T}$ to be the set of all *valid* transition from state i to state j, i.e., if a Markov state sequence can be taken from state i to state j with non-zero probability, then it is a valid transition and $(i, j) \in \mathcal{T}$. In addition, we assume that the Markov chain is finite-state, irreducible, and aperiodic so that the chain has a unique stationary distribution $\boldsymbol{\mu} = [\mu_1, \mu_2, ..., \mu_M]$, and any initial distribution tends to the stationary distribution as $n \to \infty$. This stationary distribution is called the steady-state distribution satisfying the stationary condition

$$\sum_{i:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij} = \mu_j\ , \tag{4.2}$$

and $\sum_{j=1}^{M} \mu_j = 1$.

The symbol U_n transmitted by relay is produced by a valid transition from state S_{n-1} to S_n. Once the initial state S_0 is determined, there is a one-to-one correspondence between the input sequences U_1^N and the state sequences S_1^N

for the relay-to-destination channel; therefore, the mutual information rate [41] for relay-to-destination channel may be expressed as

$$\begin{aligned}\mathcal{I}(\boldsymbol{U};\boldsymbol{Y}) &= \lim_{N\to\infty}\frac{1}{N}I(U_1^N;Y_1^N) = \lim_{N\to\infty}\frac{1}{N}I(S_1^N;Y_1^N|S_0)\\ &\overset{(a)}{=} \lim_{N\to\infty}\frac{1}{N}I(S_1^N;Y_1^N) = \mathcal{I}(\boldsymbol{S};\boldsymbol{Y})\ ,\end{aligned} \tag{4.3}$$

where (a) follows from the fact that the choice of initial state does not affect mutual information [98].

Remark: In this work, we focus on binary constrained sequences sent by a relay node in which the inputs for relay-to-destination channel U_n take values on the alphabet $\mathcal{U} = \{0, 1\}$, however our results also extend to the case of a larger alphabet size.

4.2.2 Half-Duplex Constraint

We assume that the relay is half-duplex. In particular, the source is allowed to transmit symbols X_n when the relay is sending symbol $U_n = 0$, i.e., the relay is in reception mode. Otherwise, when the relay is in transmission mode, i.e., $U_n \neq 0$, the relay ignores what the source transmits and the source-to-relay channel is idle. This can be modeled as

$$V_n = \begin{cases} V_n' & \text{if } U_n = 0\\ \text{idle} & \text{if } U_n \neq 0\end{cases} \tag{4.4}$$

where V_n' takes value on $\mathcal{V}$. The probabilities of the relay in reception and transmission modes are denoted by $\Pr(U_n = 0)$ and $\Pr(U_n \neq 0)$, respectively. Let $\mathcal{Q}$ be the set of all transitions from state i to state j that are labeled with 0, i.e., $(i, j) \in \mathcal{Q}$. Under our assumption on the Markov chain, we have

$$\Pr(S_{n-1} = i) = \mu_i \quad \forall (i, j) \in \mathcal{Q} \tag{4.5}$$

$$\Pr(S_n = j|S_{n-1} = i) = \mathsf{P}_{ij} \quad \forall (i, j) \in \mathcal{Q}\ . \tag{4.6}$$

By law of total probability, the probability of the relay being in the reception mode can be written as

$$\Pr(U_n = 0) = \sum_{i,j:(i,j)\in\mathcal{Q}} \mu_i \mathsf{P}_{ij} \ . \tag{4.7}$$

4.3 Cut-set Bound

We study the capacity of a two-hop relay channel with a Markovian constrained relay.

Lemma 3 (Cut-Set Bound)**.** *For a two-hop relay channel with a Markovian constrained relay, the capacity C is upper bounded by*

$$C \leq \lim_{N\to\infty} \sup_{p(x_1^N, u_1^N)} \min \left\{ \begin{array}{l} \dfrac{1}{N}\displaystyle\sum_{n=1}^{N} I(X_n; V_n|U_n) \\ \dfrac{1}{N} I(U_1^N; Y_1^N) \end{array} \right\} . \tag{4.8}$$

Proof. Given any $(2^{NR}, N)$ code for the relay channel with a Markovian constrained relay, the p.m.f on the joint ensemble $W, \boldsymbol{X}, \boldsymbol{U}, \boldsymbol{V}, \boldsymbol{Y}$ is given by

$$\begin{aligned} p(w, \mathbf{x}, \mathbf{u}, \mathbf{v}, \mathbf{y}) &= p(w)p(\mathbf{x}|w)p(u_2|v_1) \\ &\cdot \prod_{n=3}^{N} p(u_n|v_1, ..., v_{n-1}, u_{n-1}) \prod_{n=1}^{N} p(v_n, y_n|x_n, u_n) \ , \end{aligned} \tag{4.9}$$

where the input for relay-to-destination channel U_n is a function of U_{n-1} and all past V_1^{n-1}. For a two-hop relay channel with a Markovian constraint relay, we have

$$C = \lim_{N\to\infty} \sup_{p(x_1^N, u_1^N)} I(W; \boldsymbol{Y})$$

where

$$I(W; \boldsymbol{Y}) \leq \min \left\{ \begin{array}{l} I(U_1^N; Y_1^N) \\ \displaystyle\sum_{n=1}^{N} I(X_n; V_n|U_n) \end{array} \right\} . \tag{4.10}$$

The first inequality is straightforward, consider

$$\begin{aligned} I(W;\boldsymbol{Y}) &= H(Y_1^N) - H(Y_1^N|W) \\ &\overset{(a)}{\leq} H(Y_1^N) - H(Y_1^N|W,U_1^N) \\ &\leq H(Y_1^N) - H(Y_1^N|U_1^N) \\ &= I(U_1^N;Y_1^N)\ , \end{aligned} \tag{4.11}$$

where (a) follows from the fact that conditioning reduces entropy.

To establish the second inequality, consider

$$\begin{aligned} I(W;\boldsymbol{Y}) &\leq I(W;\boldsymbol{V},\boldsymbol{Y}) \\ &= \sum_{n=1}^{N} I(W;V_n,Y_n|V_1^{n-1},Y_1^{n-1}) \\ &\overset{(b)}{=} \sum_{n=1}^{N} I(W;V_n,Y_n|V_1^{n-1},Y_1^{n-1},U_n) \\ &= \sum_{n=1}^{N} (H(V_n,Y_n|V_1^{n-1},Y_1^{n-1},U_n) \\ &\quad - H(V_n,Y_n|W,V_1^{n-1},Y_1^{n-1},U_n)) \\ &\leq \sum_{n=1}^{N} (H(V_n,Y_n|V_1^{n-1},Y^{n-1},U_n) \\ &\quad - H(V_n,Y_n|W,V_1^{n-1},Y^{n-1},X_n,U_n)) \\ &\overset{(c)}{=} \sum_{n=1}^{N} (H(V_n,Y_n|U_n) - H(V_n,Y_n|X_n,U_n)) \\ &= \sum_{n=1}^{N} (H(V_n|U_n) - H(V_n|X_n,U_n)) \\ &= \sum_{n=1}^{N} I(X_n;V_n|U_n)\ , \end{aligned} \tag{4.12}$$

where (b) is valid since U_n is a function of past relay outputs V_1^{n-1} and past input U_{n-1}. The past input U_{n-1} is again a function of V_1^{n-2} and U_{n-2} until U_2 which is a function solely depending on V_1. Knowing V_1^{n-1}, we can generate the relay input U_n without the knowledge of W. Equality (c) is valid due to the discrete memorylessness of the relay channel, i.e., (V_n,Y_n) and (W,V_1^{n-1},Y_1^{n-1}) are conditionally independent given (X_n,U_n). □

Lemma 4. *The upper bound in (4.8) can be further expanded to*

$$C \leq \lim_{N\to\infty} \sup_{p(x|u=0),\mathsf{P}_{ij}} \min \left\{ \begin{array}{l} \displaystyle\sum_{i,j:(i,j)\in\mathcal{Q}} \mu_i \mathsf{P}_{ij} I(X;V|U=0) \\ \displaystyle\frac{1}{N} I(U_1^N; Y_1^N) \end{array} \right\}. \tag{4.13}$$

Proof. Consider the first inequality in (4.8), where

$$\begin{aligned} &\frac{1}{N}\sum_{n=1}^{N} I(X_n;V_n|U_n) \\ &= \frac{1}{N}\sum_{n=1}^{N} \Big(\Pr(U_n=0) I(X_n;V_n|U_n=0) \\ &\qquad + \Pr(U_n\neq 0) I(X_n;V_n|U_n\neq 0) \Big) \\ &\overset{(a)}{=} \sum_{i,j:(i,j)\in\mathcal{Q}} \mu_i \mathsf{P}_{ij} \cdot \frac{1}{N}\sum_{n=1}^{N} I(X_n;V_n|U_n=0) \end{aligned} \tag{4.14}$$

where (a) follows (4.7) and $I(X_n;V_n|U_n\neq 0)=0$ when the relay is in transmission mode. We then follow the standard time sharing argument by introducing a new random variable Z to be uniformly distributed over $[1:N]$ and set $X \triangleq X_Z$, $V \triangleq V_Z$, $U \triangleq U_Z$ and $Y \triangleq Y_Z$ since $Z \to (X,U) \to (V,Y)$, (4.14) becomes

$$\begin{aligned} &\sum_{i,j:(i,j)\in\mathcal{Q}} \mu_i \mathsf{P}_{ij} \cdot I(X;V|U=0,Z) \\ &\leq \sum_{i,j:(i,j)\in\mathcal{Q}} \mu_i \mathsf{P}_{ij} \cdot I(X;V|U=0) \end{aligned} \tag{4.15}$$

Substituting the above into (4.8) yields (4.13). □

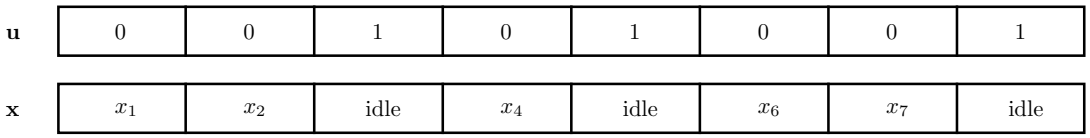

Figure 4.5 Timing Scheme: the source node is able to send messages during transmission of these zero symbols.

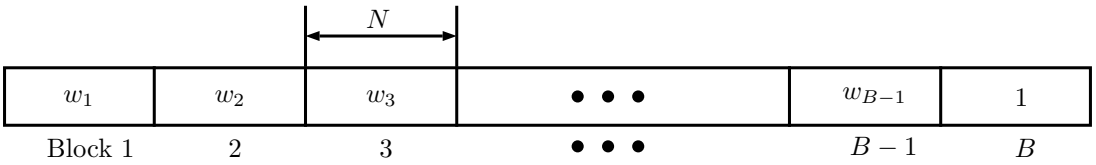

Figure 4.6 Block Markov coding: transmitting $B-1$ blocks in B block transmission.

4.4 Capacity achieving scheme

Theorem 3. *For a two-hop half-duplex relay channel with a Markovian constrained relay, the capacity is given by*

$$C = \lim_{N\to\infty} \max_{p(x|u=0),\mathsf{P}_{ij}} \min \left\{ \begin{array}{l} \sum\limits_{i,j:(i,j)\in\mathcal{Q}} \mu_i \mathsf{P}_{ij} I(X;V|U=0) \\ \dfrac{1}{N} I(U_1^N; Y_1^N) \end{array} \right\}. \tag{4.16}$$

The capacity is achieved, since

1. in the case of two-hop half-duplex relay channel, the zero symbols ($U_n = 0$) from the relay are part of the codeword that conveys messages to the destination; meanwhile, the source node is able to send messages during transmission of these zero symbols as shown in Figure 4.5.

2. the block Markov coding scheme is conducted as shown in Figure 4.6. Considering B blocks of transmission, each of N symbols, a sequence of $B-1$ messages $w_b \in [1, 2^{NR}], b = 1, 2, ..., B-1$ was sent from the source to the relay, and the relay then decodes the message at the end of each block and sends it to the destination with one block delay. Hence, $B-1$ messages will be transmitted over B N-transmission blocks. As $B \to \infty$, for fixed N, the rate $R(B-1)/B$ is arbitrarily close to R [25].

By block Markov coding, the source knows what the relay transmits. This guarantees that the positions of zero symbols ($U_n = 0$) from the relay are revealed to the source node and their proportion in each block b is given by (4.7). In the following, we give the complete proof of Theorem 3.

Proof. Codebook generation. For each block $b \in [1 : B]$, first generate 2^{NR} sequences $u_1^N(w_{b-1})$ of length N according to $p(u_1)\prod_{n=2}^{N} p(u_n|u_{n-1})$, $w_b \in [1 : 2^{NR}]$. Then generate 2^{NR} sequences $x_1^N(w_b|w_{b-1})$ of length N according to $\prod_{n=1}^{N} p(x_n|u_n(w_{b-1}))$, $w_b \in [1 : 2^{NR}]$. Note that since we are only allowed to transmit symbols from the source to the relay while the relay is in reception mode, i.e., $u_n = 0$, it is then equivalent to generate x_1^N according to $\prod_{\substack{n\in[1:N]\\ u_n=0}} p(x_n|u_n = 0)\prod_{\substack{n\in[1:N]\\ u_n\neq 0}} p(x_n|u_n \neq 0)$. The codeword x_1^N consists of two parts, $\mathbf{x}'$ and $\mathbf{x}''$. The first part $\mathbf{x}'$ is generated according to $\prod_{\substack{n\in[1:N]\\ u_n=0}} p(x_n|u_n = 0)$, which contains $N \cdot \Pr(U_n = 0)$ symbols to be transmitted when the relay is in reception mode. The rest $N \cdot \Pr(U_n \neq 0)$ symbols $\mathbf{x}''$ are generated according to $\prod_{\substack{n\in[1:N]\\ u_n\neq 0}} p(x_n|u_n \neq 0)$ which are the idle symbols in codeword x_1^N known by the relay. Hence $x_1^N(w_b|w_{b-1})$ is of length N but only contains $N \cdot \Pr(U_n = 0)$ symbols where each symbol is independently generated according to $p(x|u = 0)$. This defines the random codebook

$$\mathcal{C}_b = \Big\{(\mathbf{x}'(w_b|w_{b-1}), \mathbf{u}(w_{b-1}))\Big\}, \tag{4.17}$$

where $w_{b-1}, w_b \in [1 : 2^{NR}]$ for each $b \in [1 : B]$. The entire codebook is revealed to the source, the relay and the destination prior to communication.

Encoding: To send the message index $w \in [1 : 2^{NR}]$, the source sends the codeword $\mathbf{x}'(w_b|w_{b-1})$ from $\mathcal{C}_b$.

Relay encoding: Let $\mathcal{A}_\epsilon^{(|\mathcal{R}|)}$ defined in [26] denote the set of jointly typical sequences $(\mathbf{x}', \mathbf{v}') \in \mathcal{X}^{|\mathcal{R}|} \times \mathcal{V}^{|\mathcal{R}|}$ under the distribution $p(x, v|u = 0)$, where $|\mathcal{R}| = N \cdot \Pr(U_n = 0)$. At the end of block b, the relay finds the unique $\tilde{w}_b$ such that $(\mathbf{x}', \mathbf{v}') \in \mathcal{A}_\epsilon^{(|\mathcal{R}|)}$. If there are no such codeword or more than one codeword, then an error is declared. In block $b + 1$, the relay transmits $\mathbf{u}(\tilde{w}_b)$ from $\mathcal{C}_{b+1}$.

Decoding: Let $\mathcal{A}_\epsilon^{(N)}$ denote the set of jointly typical sequence $(\mathbf{u}, \mathbf{y}) \in \mathcal{U}^N \times \mathcal{Y}^N$. At the end of block $b + 1$, the destination determines the unique $\hat{w}_b$ such

that $(\mathbf{u}(\hat{w}_b), \mathbf{y}(b+1)) \in \mathcal{A}_\epsilon^{(N)}$. If there are no such codeword or more than one codeword, then an error is declared.

Analysis of probability of error: We analyse the probability of error averaged over codes. Assume without loss of generality that $W_b = 1$. Let $\tilde{W}_b$ be the relay message estimate at the end of block b. The decoder makes an error only if one of the following events occur:

$$\begin{aligned}
\tilde{\mathcal{E}}(b) &= \left\{\tilde{W}_b \neq 1\right\}, \\
\mathcal{E}_1(b) &= \left\{(\boldsymbol{U}(\tilde{W}_b), \boldsymbol{Y}(b+1) \notin \mathcal{A}_\epsilon^{(N)}\right\}, \\
\mathcal{E}_2(b) &= \left\{(\boldsymbol{U}(w_b'), \boldsymbol{Y}(b+1)) \in \mathcal{A}_\epsilon^{(N)} \text{ for some } w_b' \neq \tilde{W}_b\right\},
\end{aligned}$$

Thus the probability of error is upper bounded by

$$\begin{aligned}
\Pr(\mathcal{E}(b)) &= \Pr\{\hat{W}_b \neq 1\} \\
&\leq \Pr(\tilde{\mathcal{E}}(b) \cup \mathcal{E}_1(b) \cup \mathcal{E}_2(b)) \\
&\leq \Pr(\tilde{\mathcal{E}}(b)) + \Pr(\mathcal{E}_1(b)) + \Pr(\mathcal{E}_2(b)),
\end{aligned}$$

By the law of large numbers, $\Pr(\mathcal{E}_1(b))$ tends to zero as $N \to \infty$. The Shannon-McMillan-Breiman theorem extends the Asymptotic Equipartition Property (AEP) to stationary ergodic inputs over memoryless channel, then the joint typicality decoding can be adopted for jointly ergodic processes $(\boldsymbol{U}, \boldsymbol{Y})$ [26]. Therefore, $\Pr(\mathcal{E}_2(b))$ tends to zero as $N \to \infty$ if $R < \frac{1}{N}\mathcal{I}(\boldsymbol{U}; \boldsymbol{Y})$. To upper bound the first term $\Pr(\tilde{\mathcal{E}}(b))$, define

$$\begin{aligned}
\tilde{\mathcal{E}}_1(b) &= \left\{(\boldsymbol{X}'(1|\tilde{W}_{b-1}), \boldsymbol{V}'(b)) \notin \mathcal{A}_\epsilon^{(|\mathcal{R}|)}\right\}, \\
\tilde{\mathcal{E}}_2(b) &= \left\{(\boldsymbol{X}'(w_b'|\tilde{W}_{b-1}), \boldsymbol{V}'(b)) \in \mathcal{A}_\epsilon^{(|\mathcal{R}|)} \text{ for some } w_b' \neq 1\right\},
\end{aligned}$$

Then

$$\begin{aligned}
\Pr(\tilde{\mathcal{E}}(b)) &\leq \Pr(\tilde{\mathcal{E}}(b-1) \cup \tilde{\mathcal{E}}_1(b) \cup \tilde{\mathcal{E}}_2(b)) \\
&\leq \Pr(\tilde{\mathcal{E}}(b-1)) + \Pr(\tilde{\mathcal{E}}_1(b) \cap \tilde{\mathcal{E}}^{\mathsf{c}}(b-1)) + \Pr(\tilde{\mathcal{E}}_2(b)).
\end{aligned}$$

By the law of large numbers, $\Pr(\tilde{\mathcal{E}}_1(b) \cap \tilde{\mathcal{E}}^{\mathsf{c}}(b-1))$ tends to zero as $N \to \infty$. By joint typicality decoding, $\Pr(\tilde{\mathcal{E}}_2(b))$ tends to zero as $N \to \infty$ if $R < \sum_{i,j:(i,j)\in\mathcal{Q}} \mu_i \mathsf{P}_{ij} \cdot I(X;V|U=0)$. Note that $\tilde{W}_0 = 1$ by definition. Hence by induction $\Pr(\tilde{\mathcal{E}}(b))$ tends to zero as $N \to \infty$ for every $b \in [1 : B-1]$. By Lemma 4, we have the converse. This completes the proof of achievability. □

4.5 Conclusion

In this chapter, we provide introduction to two cornerstones on which the second part of the thesis builds: the relay channels and finite-state models. The capacity theorems for a two-hop half-duplex relay channel with a Markovian constrained relay are studied. It first focuses on deriving the cut-set bound, i.e., an upper bound on the capacity. Then this chapter presents the timing strategy which satisfies the half-duplex constraint and it is shown to achieve this bound. This leads to the general capacity formula for a two-hop half-duplex relay channel with a Markovian constrained relay.

Chapter 5

Achievable Rates and Capacity Bounds

In this chapter, we consider the case where the relay-to-destination link is noise-free, the optimal state transition probabilities that give rise to the capacity is determined. This result links the relay channel to Shannon's entropy maximisation by introducing a ***relay adjacency matrix****. For the case where both source-to-relay and relay-to-destination links are noisy, lower bounds on the achievable information rates for various constrained sequences are computed. We conjecture that our numerical bounds are tight. The numerical computed capacities and optimised information rates are significantly higher than the rate achieved by the traditional predetermined time-sharing scheme.*

In this chapter, a special type of Runlength limited (RLL) constraint, namely, the hold time constraint, is introduced in section 5.1. In section 5.2, we derived the capacity achieving input for the case that the relay-to-destination link is noiseless. We further computed a tight lower bound on the capacity via GBAA in section 5.3. Numerical results are discussed in section 5.4.

5.1 Introdcution

MARKOVIAN source based information theoretic models have been extensively studied in the literature for point-to-point communications over

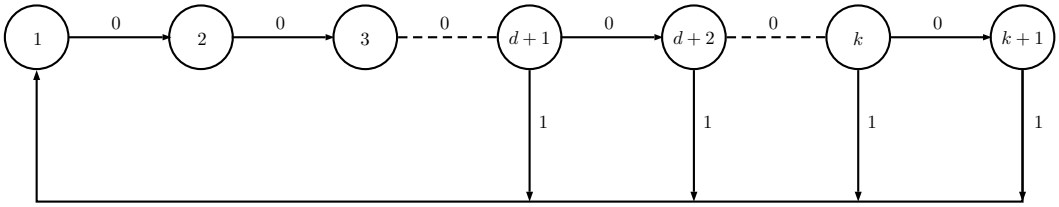

Figure 5.1 State transition diagram for a binary RLL($[d, k], \mathcal{L}_1$) constrained sequence.

noisy and noiseless channels (see references [99, 98, 94, 111, 5, 100, 86, 63, 106, 47]).

We aim at finding capacity of a two-hop half-duplex relay channel, where the relay uses constrained sequences, e.g., RLL sequences, due to

- joint energy and information transmission requirements [40], and/or
- existence of switching noise [92] [19].

We further consider binary RLL sequences and adopt the definition in [9] as shown below.

Definition 1 (RLL($\mathcal{L}_0, \mathcal{L}_1$) sequences [9])**.** *A binary runlength limited constrained sequence RLL $(\mathcal{L}_0, \mathcal{L}_1)$ has $\mathcal{L}_0$ and $\mathcal{L}_1$ as the sets of admissible runlengths of binary symbol 0 and 1, respectively.*

Note that the conventional runlength limitation in [94] can be included in this RLL definition as a special case as introduced below.

Definition 2 (RLL($[d, k], \mathcal{L}_1$) sequences [94])**.** *A binary runlength limited constrained sequence RLL$([d, k], \mathcal{L}_1)$ with its state transition diagram in Fig. 5.1 satisfies the following conditions simultaneously:*

1. *$[d, k]$-constraint – the runs of 0's between successive 1's have length at least d and at most k,*
2. *$\mathcal{L}_1$-constraint – the runs of 1's are dependent on $\mathcal{L}_0 = [d, k]$, and can be:*
   (a) For $d = 0$, $k = 1$, $\mathcal{L}_1 = [1, \infty)$;
   (b) For $d = 0$, $k > 1$, $\mathcal{L}_1 = [0, \infty)$;
   (c) For $d > 0$, $\mathcal{L}_1 = [0, 1]$.

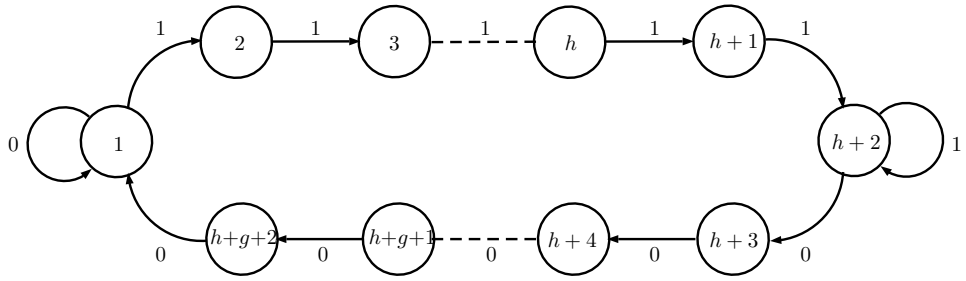

Figure 5.2 State transition diagram for a binary RLL$(\{0\} \cup [1+g, \infty), \{0\} \cup [1+h, \infty))$ constrained sequence.

As mentioned above, the first motivation of considering a relay with constrained sequences is due to the use of joint energy and information transfer in a low-power wireless relay network [40], where the relay uses RLL$([d, k], \mathcal{L}_1)$ sequences with symbol 1 indicating one unit energy transfer and 0 no energy transfer.

The other motivation is due to the existence of switching noise, an induced ISI when switching a digital signal, caused by the relay's switching between reception mode and transition mode [19]. Switching noise can be avoided by better hardware design or introducing a short guard time after switching occurs [90]. Introducing a short guard time after a transition from reception-to-transition mode and transition-to-reception mode results in RLL$(\{0\} \cup [1+g, \infty), \{0\} \cup [1+h, \infty))$ constrained sequences, defined as:

Definition 3 (RLL$\{0\} \cup [1+g, \infty), \{0\} \cup [1+h, \infty))$ constrained sequences)**.** *A binary runlength limited constrained sequence RLL*$(\{0\} \cup [1+g, \infty), \{0\} \cup [1+h, \infty))$ *constrained sequence with its state transition diagram in Figure 5.2 must satisfy the following two conditions simultaneously:*

1. $\{0\} \cup [1+g, \infty)$*-constraint – if no switching then the runlength of* 0*'s between successive* 1*'s is 0, otherwise the runlength of* 0*'s after a* 10 *must be at least* g*,*

2. $\{0\} \cup [1+h, \infty)$*-constraint – if no switching then the runlength of* 1*'s between successive* 0*'s is 0, otherwise the runlength of* 1*'s after a* 01 *must be at least* h*.*

This can be easily applied to our two-hop half-duplex relay channel, where symbol 0 represents relay's reception mode, and 1 its transmission mode. Mo-

tivated by the above applications, we provide the general capacity analysis on a two-hop relay channel with constrained sequences and give examples when the relay has runlength limited constraints. When the relay-to-destination link is noise-free, we derive the capacity achieving Markovian input. When both source-to-relay and relay-to-destination links are noisy, tight lower bounds on information rates can be computed and optimised using the generalised Blahut-Arimoto algorithm (GBAA) [106] modified for this channel. Our numerical results show that significant information rate gains are possible, when compared to that of a predetermined time-sharing strategy.

This chapter is based on the research results reported in [HH18] by the author.

5.2 Capacity with noise-free second link

Corollary 1. *Consider a two-hop relay channel, where the relay-to-destination link is noise-free and the relay is Markovian constraint. Then (4.16) in Theorem 3 becomes*

$$C_{\mathsf{nf}} = \max_{\mathsf{P}_{ij}} \min \left\{ \begin{array}{l} \max\limits_{p(x|u=0)} \sum\limits_{i,j:(i,j)\in\mathcal{Q}} \mu_i \mathsf{P}_{ij} \cdot I(X;V|U=0) \\ \\ \sum\limits_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij} \log \dfrac{1}{\mathsf{P}_{ij}} \end{array} \right\}. \tag{5.1}$$

Proof. For noise-free relay-to-destination link, given the stationary Markovian constraint in (4.2), the second term in (4.16) becomes

$$\lim_{N\to\infty} \frac{1}{N} I(U_1^N; Y_1^N) = H(S_2|S_1) = \sum_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij} \log \frac{1}{\mathsf{P}_{ij}},$$

which can be optimised over P_{ij} for $(i,j) \in \mathcal{T}$.

In (4.16), we observe that, for all $(i,j) \in \mathcal{Q}$, the source is allowed to transmit information with a constant $I(X;V|U=0)$ when $U=0$, where $I(X;V|U=0)$ is solely associated with $p(x|u=0)$. Hence, we conclude the first term in (5.1), which is maximized over $p(x|u=0)$ as well as P_{ij} for $(i,j) \in \mathcal{Q}$. □

Below we provide the generalised maximum mutual information between source and relay.

Definition 4 (Relay Mutual Information Matrix I)**.** *The relay mutual information matrix* I *has entries*

$$\mathsf{I}_{ij} \triangleq \max_{p(x|i,j)} I(X;V|S_{n-1}=i,S_n=j) \tag{5.2}$$

for $(i,j)\in\mathcal{T}$*. In the special case of half-duplex relaying,* $(i,j)\in\mathcal{T}$ *and* $(i,j)\notin\mathcal{Q}$ *then* $\mathsf{I}_{ij}\triangleq 0$*. For invalid transitions* $(i,j)\notin\mathcal{T}$*, we define* $\mathsf{I}_{ij}\triangleq -\infty$*.*

Corollary 2. *Using the above definition, (5.1) in Corollary 1 becomes*

$$C_{\mathsf{nf}} = \max_{\mathsf{P}_{ij}} \min \left\{ \begin{array}{l} \displaystyle\sum_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij}\cdot \mathsf{I}_{ij} \\ \\ \displaystyle\sum_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij}\log\frac{1}{\mathsf{P}_{ij}} \end{array} \right\}. \tag{5.3}$$

For a noise-free point-to-point channel, the computation of the maximum entropy of Markov source can be related to the adjacency matrix associated with the corresponding finite-state machine (see [99]). Motivated by this method, for a relay channel, we define *relay adjacency matrix*, which leads to the maxentropic state transition probabilities.

Definition 5 (Relay Adjacency Matrix $\mathbf{A}(\zeta)$)**.** *The relay adjacency matrix* $\mathbf{A}(\zeta)$ *is given by its entry* $\mathsf{A}_{ij}(\zeta)$ *defined as follows*

$$\mathsf{A}_{ij}(\zeta) = \begin{cases} 2^{\frac{1-\zeta}{\zeta}\cdot \mathsf{I}_{ij}} & \text{if } (i,j)\in\mathcal{T} \\ 0 & \text{otherwise} \end{cases} \tag{5.4}$$

for some $\zeta\in[0,+\infty)$*.*

Remark: When $\zeta=1$, the relay adjacency matrix $\mathbf{A}(1)$ is the standard adjacency matrix in [99].

Next we state the maxentropic state transition probabilities for the Markovian constrained relay in a two-hop half-duplex relay channel. Consider $\mathbf{A}(\zeta)$ in Definition 5 and let $\boldsymbol{l}(\zeta)=[l_1(\zeta),l_2(\zeta),...,l_M(\zeta)]$ and $\boldsymbol{r}^{\mathsf{T}}(\zeta)=[r_1(\zeta),r_2(\zeta),...,r_M(\zeta)]^{\mathsf{T}}$ be the left and right eigenvectors, respectively, corresponding to a real eigenvalue $\lambda(\zeta)$ of the relay adjacency matrix $\mathbf{A}(\zeta)$. Let $c(\zeta)=\frac{1}{\sum_{i=1}^{M} l_i(\zeta)\cdot r_i(\zeta)}$ be a normalisation constant for $\boldsymbol{l}(\zeta)$ and $\boldsymbol{r}(\zeta)$.

Theorem 4 (Maxentropic State Transition Probabilities)**.** *For a two-hop half-duplex relay channel with a Markovian constrained relay, the optimal (maxentropic) state transition probabilities* P^*_{ij} *and stationary distributions* μ^*_i *are of the following forms*

$$\mathsf{P}^*_{ij} = \frac{r_j(\zeta)}{r_i(\zeta)} \cdot \frac{\mathsf{A}_{ij}(\zeta)}{\lambda(\zeta)} \quad \forall i,j : (i,j) \in \mathcal{T} \,, \tag{5.5}$$

$$\mu^*_i = c(\zeta) \cdot l_i(\zeta) \cdot r_i(\zeta) \quad \forall i : (i,j) \in \mathcal{T} \,, \tag{5.6}$$

for some $\zeta \in [0, +\infty)$.

Proof. Let $\mathsf{Q}_{ij} = \mu_i \mathsf{P}_{ij}$ which satisfies $\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij} = 1$ and (5.3) becomes

$$C_{\mathsf{nf}} = \max_{\mathsf{Q}_{ij}} \min \left\{ \begin{array}{l} \displaystyle\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij} \cdot \mathsf{I}_{ij} \\ \displaystyle\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij} \log \left(\frac{\sum_{j':(j',i)\in\mathcal{T}} \mathsf{Q}_{j'i}}{\mathsf{Q}_{ij}} \right) \end{array} \right\} . \tag{5.7}$$

The rest of the proof follows directly from solving (5.7) with respect to Q_{ij}. Notice that the first term in (5.7) is a linear function in Q_{ij} and the second term is concave in Q_{ij} which can be proved using log-sum inequality [26]. Since the concave function is always bounded by 0 and $\log \lambda_{\max}(1)$, there are two cases to be considered.

Case 1: The capacity is limited by that of relay-to-destination link, i.e., the linear function is above the concave function while the concave function reaches its maximum. Hence, we have

$$\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij} \left[\mathsf{I}_{ij} - \log \left(\frac{\sum_{j':(j',i)\in\mathcal{T}} \mathsf{Q}_{j'i}}{\mathsf{Q}_{ij}} \right) \right] > 0 \,. \tag{5.8}$$

In this case, we only need to consider the capacity of the second link. The unique optimal probability assignment P^*_{ij} that gives rise to the capacity is given by [99]

$$\mathsf{P}^*_{ij} = \frac{r_j(1)}{r_i(1)} \cdot \frac{\mathsf{A}_{ij}(1)}{\lambda_{\max}(1)} \tag{5.9}$$

where $\zeta = 1$ and $\lambda_{\max}(1)$ is the maximum eigenvalue of the standard adjacency matrix $\mathbf{A}(1)$ and the noise-free capacity is $\log \lambda_{\max}(1)$.

Case 2: The capacity of source-to-relay link is less than or equal to the capacity of relay-to-destination link, i.e.,

$$\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij}\left[\mathsf{I}_{ij} - \log\left(\frac{\sum_{j':(j',i)\in\mathcal{T}} \mathsf{Q}_{j'i}}{\mathsf{Q}_{ij}}\right)\right] \leq 0\,. \tag{5.10}$$

Since in (5.7), the first term is a linear function in Q_{ij} and the second term is a concave function in Q_{ij}, hence the maximum is achieved when they are equal. This leads to the following optimization problem with linear cost function and convex inequality constraint:

$$\begin{aligned}
&\text{maximize} \sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij}\cdot\mathsf{I}_{ij} && (5.11)\\
&\text{subject to}\ \ \mathsf{Q}_{ij} \geq 0\,, && (C.1)\\
&\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij} - 1 = 0\,, && (C.2)\\
&\sum_{j'':(j'',i)\in\mathcal{T}} \mathsf{Q}_{j''i} - \sum_{j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij} = 0 \quad \forall\ i, && (C.3)\\
&\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij}\left[\mathsf{I}_{ij} - \log\left(\frac{\sum_{j':(j'i)\in\mathcal{T}} \mathsf{Q}_{j'i}}{\mathsf{Q}_{ij}}\right) - \mathsf{T}_{ij}\right] \leq 0 && (C.4)
\end{aligned}$$

where $\mathsf{T}_{ij} = 0$, for noise-free link in all valid transition $(i,j) \in \mathcal{T}$ (while, for noisy case, T_{ij} is defined in Definition 6, which can be any real number associated with the valid transition $(i,j) \in \mathcal{T}$). The rest of this proof is a straightforward routine by introducing Lagrange multipliers (η, γ_i, ζ) to the constraints (C.2), (C.3), and (C.4) correspondingly. We now consider the following Lagrangian

$$\begin{aligned}
\mathcal{L} = &\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij}\mathsf{I}_{ij} - \eta\Big(\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij} - 1\Big)\\
&- \sum_{i:(i,j)\in\mathcal{T}} \gamma_i\Big(\sum_{j'':(j'',i)\in\mathcal{T}} \mathsf{Q}_{j''i} - \sum_{j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij}\Big)\\
&- \zeta\Big(\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij}\left[\mathsf{I}_{ij} - \log\left(\frac{\sum_{j':(j',i)\in\mathcal{T}} \mathsf{Q}_{j'i}}{\mathsf{Q}_{ij}}\right) - \mathsf{T}_{ij}\right]\Big).
\end{aligned} \tag{5.12}$$

We have to solve the following equations

$$\frac{\partial \mathcal{L}}{\partial \mathsf{Q}_{ij}} = 0 \tag{5.13}$$

and yields

$$\frac{\partial \mathcal{L}}{\partial \mathsf{Q}_{ij}} = \mathsf{I}_{ij} - \eta - \gamma_j + \gamma_i - \zeta \Big[\mathsf{I}_{ij} + \log \mathsf{P}_{ij} - \mathsf{T}_{ij} \Big] = 0 \,. \tag{5.14}$$

We get

$$\log \mathsf{P}_{ij} = \mathsf{T}_{ij} + \frac{1-\zeta}{\zeta} \mathsf{I}_{ij} - \frac{\eta}{\zeta} - \frac{\gamma_j}{\zeta} + \frac{\gamma_i}{\zeta} \,. \tag{5.15}$$

Notice that $\mathsf{P}_{ij} = 2^{\{\mathsf{T}_{ij} + \frac{1-\zeta}{\zeta}\mathsf{I}_{ij} - \frac{\eta}{\zeta} - \frac{\gamma_j}{\zeta} + \frac{\gamma_i}{\zeta}\}} \geq 0$, hence the constraint (C.1) is always satisfied. Let A_{ij} be the entries for the relay adjacency matrix $\mathbf{A}$

$$\mathsf{A}_{ij}(\zeta) = \begin{cases} 2^{\mathsf{T}_{ij}} \cdot 2^{\frac{1-\zeta}{\zeta} \cdot \mathsf{I}_{ij}} & \forall (i,j) \in \mathcal{T} \\ 0 & \text{otherwise} \end{cases} \,. \tag{5.16}$$

Further, let $\boldsymbol{r}^{\mathsf{T}}(\zeta) = [r_1(\zeta), r_2(\zeta), ..., r_M(\zeta)]^{\mathsf{T}}$ be the vector with entries $r_i(\zeta) = 2^{-\frac{\gamma_i}{\zeta}}$ and $\lambda(\zeta) = 2^{\frac{\eta}{\zeta}}$, (5.15) yields

$$\mathsf{P}_{ij} = \frac{r_j(\zeta)}{r_i(\zeta)} \cdot \frac{\mathsf{A}_{ij}(\zeta)}{\lambda(\zeta)} \,. \tag{5.17}$$

From constraint (C.2), we have

$$\begin{aligned} \sum_{j:(i,j)\in\mathcal{T}} \frac{r_j(\zeta)}{r_i(\zeta)} \cdot \frac{\mathsf{A}_{ij}}{\lambda(\zeta)} &= 1 \\ \sum_{j:(i,j)\in\mathcal{T}} \mathsf{A}_{ij}(\zeta) \cdot r_j(\zeta) &= \lambda(\zeta) \cdot r_i(\zeta) \\ \mathbf{A}(\zeta) \cdot \boldsymbol{r}(\zeta) &= \lambda(\zeta) \cdot \boldsymbol{r}(\zeta) \,. \end{aligned} \tag{5.18}$$

where $\boldsymbol{r}(\zeta)$ is the eigenvector of the matrix $\mathbf{A}(\zeta)$ with eigenvalue λ. Similarly, for constraint (C.3), we obtain

$$\sum_{i:(i,j)\in\mathcal{T}} \mu_i \frac{r_j(\zeta)}{r_i(\zeta)} \cdot \frac{\mathsf{A}_{ij}(\zeta)}{\lambda(\zeta)} = \mu_j$$
$$\sum_{i:(i,j)\in\mathcal{T}} \frac{\mu_i}{r_i(\zeta)} \cdot \mathsf{A}_{ij}(\zeta) = \frac{\mu_j}{r_j(\zeta)} \cdot \lambda(\zeta) \; . \tag{5.19}$$

Let $\boldsymbol{l}(\zeta) = [l_1(\zeta), l_2(\zeta), ..., l_M(\zeta)]$ be the left eigenvector of $\mathbf{A}(\zeta)$ with entries $l_i(\zeta) = \mu_i/(c(\zeta) \cdot r_i(\zeta))$ such that

$$\boldsymbol{l}(\zeta) \cdot \mathbf{A}(\zeta) = \boldsymbol{l}(\zeta) \cdot \lambda(\zeta) \; . \tag{5.20}$$

Since the summation of μ_i has to be 1, we have

$$\mu_i = c(\zeta) \cdot l_i(\zeta) \cdot r_i(\zeta) \tag{5.21}$$

where $c = 1/(\sum_{i:(i,j)\in\mathcal{T}} l_i(\zeta) \cdot r_i(\zeta))$ is the normalization constant. Finally, substituting (5.17) and (5.21) into constraint (C.4), yields

$$\sum_{i:(i,j)\in\mathcal{T}} c(\zeta) \cdot l_i(\zeta) \cdot r_i(\zeta) \sum_{j:(i,j)\in\mathcal{T}} \left(\frac{r_j(\zeta)}{r_i(\zeta)} \cdot \frac{\mathsf{A}_{ij}(\zeta)}{\lambda(\zeta)} \right.$$
$$\left. \cdot \left[\log\left(\frac{r_j(\zeta)}{r_i(\zeta)} \cdot \frac{\mathsf{A}_{ij}(\zeta)}{\lambda(\zeta)} \right) - \mathsf{T}_{ij} + \mathsf{I}_{ij} \right] \right) = 0 \; , \tag{5.22}$$

and we have

$$\frac{c(\zeta)}{\lambda(\zeta)} \sum_{i,j:(i,j)\in\mathcal{T}} l_i(\zeta) \cdot r_j(\zeta) \cdot \mathsf{A}_{ij}(\zeta) \cdot \mathsf{I}_{ij} = \zeta \cdot \log \lambda(\zeta) \; . \tag{5.23}$$

By the Perron-Forbenius theorem [54], we have $\lambda(\zeta) = \lambda_{\max}(\zeta)$. This completes the proof of Theorem 4. □

It turns out that constraint (C.1) is always satisfied and hence it is not included in the optimisation problem. The solution to this optimisation problem is

$$\mathsf{P}^*_{ij} = \frac{r_j(\zeta)}{r_i(\zeta)} \cdot \frac{\mathsf{A}_{ij}(\zeta)}{\lambda_{\max}(\zeta)} \tag{5.24}$$

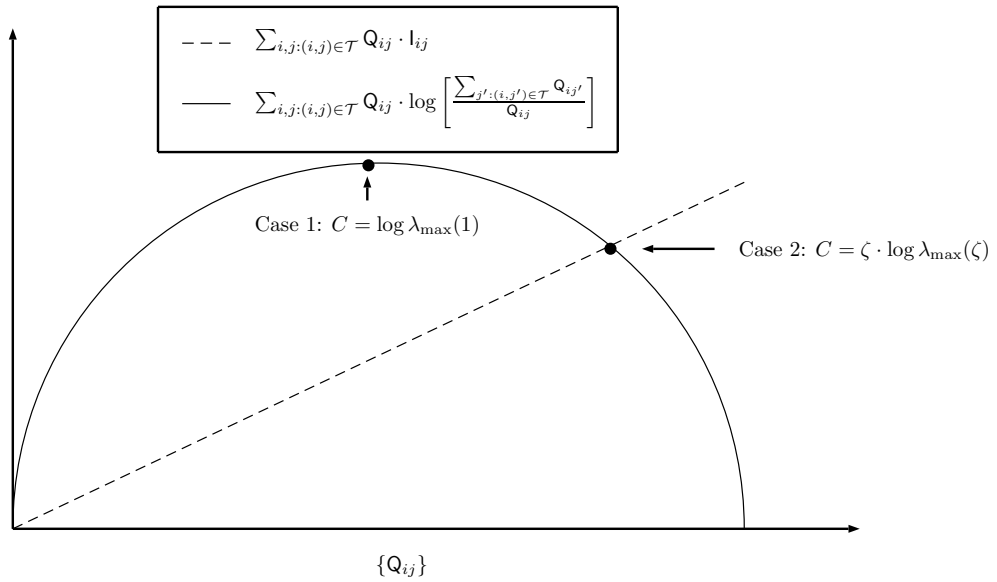

Figure 5.3 Graphical representation of optimization problem (5.11).

where ζ is the Lagrange multiplier for the constraint (C.4) and lies in $[0, +\infty)$. Substituting the solutions μ_i^* and P_{ij}^* into (C.4) yields

$$\frac{c(\zeta)}{\lambda_{\max}(\zeta)} \sum_{i,j:(i,j)\in\mathcal{T}} l_i(\zeta) \cdot r_j(\zeta) \cdot \mathsf{A}_{ij}(\zeta) \cdot \mathsf{I}_{ij} = \zeta \cdot \log \lambda_{\max}(\zeta) \tag{5.25}$$

and ζ can be found by solving the above equation that gives rise to the largest value of $\zeta \cdot \log \lambda_{\max}(\zeta)$.
Remarks:

i) Constraint (5.10) implies that an average information rate requirement is imposed on the relay channel, i.e., $\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij}\mathsf{I}_{ij} \leq H(S_2|S_1)$, which effectively controls the frequency of which sequence to occur.

ii) A closely related topic would be the problem of a Markov source over a costly channel [60, 64], in which, a cost w_{ij} is assigned to every valid transition $(i, j) \in \mathcal{T}$ and the average cost is at most ρ, i.e., $\sum_{i,j(i,j)\in\mathcal{T}} \mathsf{Q}_{ij}\mathsf{w}_{ij} \leq \rho$.

Corollary 3 (Maximal-Eigenvalue Characterization of the noise-free Capacity)**.** *For a two-hop half-duplex relay channel with a Markovian constrained relay and*

a noise-free relay-to-destination link, the capacity is given by

$$C_{\mathsf{nf}} = \zeta \cdot \log \lambda_{\max}(\zeta) \tag{5.26}$$

where $\lambda_{\max}(\zeta)$ *is the largest eigenvalue of* $\mathbf{A}(\zeta)$ *for some fixed* $\zeta \in [0, +\infty)$.

Proof. Corollary 3 follows directly from Theorem 4. We have

$$\begin{aligned}
H(S_2|S_1) &= \sum_{i:(i,j)\in\mathcal{T}} \mu_i^* \sum_{j:(i,j)\in\mathcal{T}} \mathsf{P}_{ij}^* \log \frac{1}{\mathsf{P}_{ij}^*} \\
&= -\sum_{i:(i,j)\in\mathcal{T}} c(\zeta) \cdot l_i(\zeta) \cdot r_i(\zeta) \sum_{j:(i,j)\in\mathcal{T}} \left(\frac{r_j(\zeta)}{r_i(\zeta)} \cdot \frac{\mathsf{A}_{ij}(\zeta)}{\lambda(\zeta)} \right. \\
&\quad \left. \times \left[\log \left(\frac{r_j(\zeta)}{r_i(\zeta)} \cdot \frac{\mathsf{A}_{ij}(\zeta)}{\lambda(\zeta)} \right) - \mathsf{T}_{ij} \right] \right) \\
&= \underbrace{-\frac{c(\zeta)}{\lambda(\zeta)} \sum_{i,j:(i,j)\in\mathcal{T}} l_i(\zeta) \cdot r_j(\zeta) \cdot \mathsf{A}_{ij}(\zeta) \cdot \log r_j(\zeta)}_{\sum_{j:(i,j)\in\mathcal{T}} \mu_j \log r_j(\zeta)} \\
&\quad \underbrace{+\frac{c(\zeta)}{\lambda(\zeta)} \sum_{i,j:(i,j)\in\mathcal{T}} l_i(\zeta) \cdot r_j(\zeta) \cdot \mathsf{A}_{ij}(\zeta) \cdot \log r_i(\zeta)}_{\sum_{i:(i,j)\in\mathcal{T}} \mu_i \cdot \log r_i(\zeta)} \\
&\quad \underbrace{-\frac{c(\zeta)}{\lambda(\zeta)} \sum_{i,j:(i,j)\in\mathcal{T}} l_i(\zeta) \cdot r_j(\zeta) \cdot \mathsf{A}_{ij}(\zeta) \cdot \frac{1-\zeta}{\zeta} \mathsf{I}_{ij}}_{\frac{1-\zeta}{\zeta} \sum_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij} \mathsf{I}_{ij}} \\
&\quad \underbrace{+\frac{c(\zeta)}{\lambda(\zeta)} \sum_{i,j:(i,j)\in\mathcal{T}} l_i(\zeta) \cdot r_j(\zeta) \cdot \mathsf{A}_{ij}(\zeta) \cdot \log \lambda(\zeta)}_{\log \lambda(\zeta)} \\
&= -\frac{1-\zeta}{\zeta} H(S_2|S_1) + \log \lambda(\zeta) \\
&= \zeta \cdot \log \lambda(\zeta) \, .
\end{aligned} \tag{5.27}$$

Thus, for a fixed ζ, the entropy rate is maximized for $\lambda(\zeta)$ being the largest eigenvalue of the relay adjacency matrix $\mathbf{A}(\zeta)$. For an irreducible and non-negative matrix $\mathbf{A}(\zeta)$, the Perron-Forbenius theorem guarantees that there exists an unique real, positive eigenvalue $\lambda_{\max}$ [54]. The right and left eigenvectors $\boldsymbol{r}^{\mathsf{T}}(\zeta)$, $\boldsymbol{l}(\zeta)$ corresponding to the eigenvalue $\lambda_{\max}(\zeta)$ have positive and non-

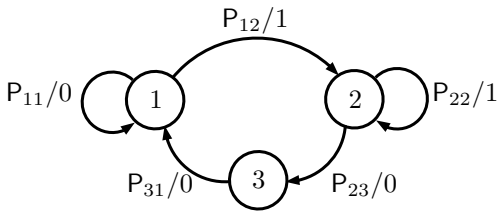

Figure 5.4 State transition diagram with transition probabilities for a RLL($\{0\} \cup [2, \infty), [0, \infty)$) constrained sequence.

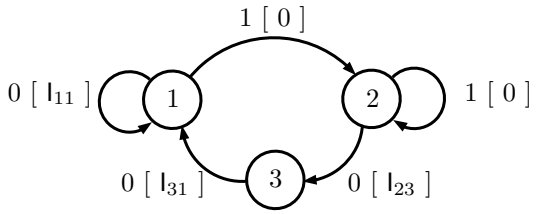

Figure 5.5 State transition diagram labeled with its mutual information in each transition for half-duplex RLL($\{0\} \cup [2, \infty), [0, \infty)$) constraint.

negative entries, respectively. Hence, the maxentropic stationary distributions μ_i^* and state transition probabilities P_{ij}^* will have positive probabilities. □

Remark: The logarithm of the maximal eigenvalue of a noise-free adjacency matrix $\mathbf{A}(1)$, i.e., $\log \lambda_{\max}(1)$, is the maximum achievable noise-free entropy rate of a Markov process according to [99]. Assuming that both links are noiseless, the maximum achievable noise-free entropy rate with conventional time sharing approach is given by $R_{\text{conv}} = \frac{1}{2} \log \lambda_{\max}(1)$ bits/use.

In the following, we will consider an example of a two-hop half-duplex relay, where the relay adopts RLL($\{0\} \cup [2, \infty), [0, \infty)$) constraint with its finite-state machine shown in FIGURE 5.4. We also show the same state transition diagram but labeled with mutual information in each transition in Figure 5.5.

Example 1. *Consider a two-hop half-duplex relay, where the relay has RLL*($\{0\} \cup [2, \infty), [0, \infty)$) *constraint. Assuming that the first link is a binary symmetric channel (BSC) with cross-over probability* $p_1 = 0.1$ *and the second link is noise-*

free, we obtain $\zeta = 0.1609$ *and* $C_{\mathsf{nf}} = 0.4743$*. We also find*

$$\mathbf{A} = \begin{bmatrix} 2.4458 & 1 & 0 \\ 0 & 1 & 2.4458 \\ 2.4458 & 0 & 0 \end{bmatrix},$$
$$\lambda_{\max} = 7.7157\,, \quad \log \lambda_{\max} = 2.9478\,,$$
$$\boldsymbol{l} = [0.8225 \quad 0.4631 \quad 0.3304]\,,$$
$$\boldsymbol{r}^{\mathsf{T}} = [0.6833 \quad 0.5436 \quad 0.4875]\,.$$

The resulting Markov chain has

$$\mathbf{P} = \begin{bmatrix} 0.7134 & 0.2866 & 0 \\ 0 & 0.3602 & 0.6398 \\ 1 & 0 & 0 \end{bmatrix},$$
$$\boldsymbol{\mu} = [0.5765 \quad 0.2582 \quad 0.1652]\,.$$

5.3 Noisy Capacity with both noisy links and its lower bound obtained via GBAA

Now we investigate the noisy capacity by means of GBAA [106] to a two-hop half-duplex relay channel with a Markovian constrained relay. It was shown in [47] that the global convergence of GBAA is in general not theoretically guaranteed. Nevertheless, we may still find a good set of state transition probabilities that give rise to a very tight lower bound on capacity via GBAA.

Consider a two-hop half-duplex relay channel, where the relay-to-destination link is noisy. We are interested in computing its capacity

$$C = \lim_{N\to\infty} \max_{\mathsf{P}_{ij}} \min \left\{ \begin{array}{l} \displaystyle\sum_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij} \cdot \mathsf{I}_{ij} \\ \dfrac{1}{N} I(U_1^N; Y_1^N) \end{array} \right\} \tag{5.28}$$

Following Kavčić's approach [63], by applying chain rule to the second term in (5.28) yields

$$\begin{aligned}\frac{1}{N}I(U_1^N;Y_1^N) &= I(S_1^N;Y_1^N|S_0)\\ &= \frac{1}{N}\sum_{n=1}^{N} I(S_n;Y_1^N|S_{n-1})\\ &= \underbrace{\frac{1}{N}\sum_{n=1}^{N} H(S_n|S_{n-1})}_{-\sum_{i,j:(i,j)\in\mathcal{T}}\mu_i \mathsf{P}_{ij}\log \mathsf{P}_{ij}} - \frac{1}{N}\sum_{n=1}^{N} H(S_n|S_{n-1},Y_1^N) \end{aligned} \tag{5.29}$$

as $N\to\infty$ and the entropy in the second term of (5.29) can be further written as

$$\begin{aligned}-\frac{1}{N}\sum_{n=1}^{N} H(S_n|S_{n-1},Y_1^N) &= \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[\log \Pr(S_n|S_{n-1},Y_1^N)]\\ &= \frac{1}{N}\sum_{n=1}^{N}\sum_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij}\mathbb{E}_{Y_1^N|i,j}[\log \Pr(S_n=j|S_{n-1}=i,Y_1^N)]\end{aligned}$$

where $\mathbb{E}_{Y_1^N|i,j}$ denotes the conditional expectation taken over the observation Y_1^N when $(S_n=j,S_{n-1}=j)$. We define $p_n(i,j|Y_1^N)$ and $p_{n-1}(i,Y_1^N)$ be the a-posteriori probabilities

$$p_n(i,j|Y_1^N) \triangleq \Pr(S_{n-1}=i,S_n=j|Y_1^N)\,, \tag{5.30}$$

$$p_{n-1}(i|Y_1^N) \triangleq \Pr(S_{n-1}=i|Y_1^N)\,, \tag{5.31}$$

and by using Bayes rules, (5.30) yields

$$\begin{aligned}&-\frac{1}{N}\sum_{n=1}^{N} H(S_n|S_{n-1},Y_1^N)\\ &= \sum_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij}\left(\frac{1}{N}\sum_{n=1}^{N}\mathbb{E}_{Y_1^N|i,j}\left[\log\frac{p_n(i,j|Y_1^N)}{p_{n-1}(i|Y_1^N)}\right]\right).\end{aligned} \tag{5.32}$$

To simplify this expression, we define the a-posteriori state transition weight matrix $\mathbf{T}$.

Definition 6 (A-posteriori State Transition Weight Matrix $\mathbf{T}$ [4])**.** *The a-posteriori state transition weight matrix* $\mathbf{T}$ *is given by its entry*

$$\mathsf{T}_{ij} = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} \mathbb{E}\left[\log \frac{p_n(i,j|Y_1^N)^{\frac{p_n(i,j|Y_1^N)}{\mu_i \mathsf{P}_{ij}}}}{p_{n-1}(i|Y_1^N)^{\frac{p_{n-1}(i|Y_1^N)}{\mu_i}}} \right] \tag{5.33}$$

For invalid transitions $(i,j) \notin \mathcal{T}$*, we define* $\mathsf{T}_{ij} \triangleq -\infty$.

This allows us to simplify the general capacity (5.28) into the following compact form

$$C = \max_{\mathsf{P}_{ij}} \min \left\{ \begin{array}{l} \displaystyle\sum_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij} \cdot \mathsf{I}_{ij} \\ \displaystyle\sum_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij} \left[\log \frac{1}{\mathsf{P}_{ij}} + \mathsf{T}_{ij} \right] \end{array} \right\}. \tag{5.34}$$

If we assume that the a-posteriori state transition weight T_{ij} can be computed, then, similar to the noise-free case, in which we define the relay adjacency matrix $\mathbf{A}(\zeta)$, we may also link the noisy capacity to the adjacency matrix with the corresponding finite-state machine by defining the noisy adjacency matrix $\mathbf{G}$ and the noisy-relay adjacency matrix $\tilde{\mathbf{A}}(\zeta)$. Here again, the noisy version of the relay adjacency matrix $\tilde{\mathbf{A}}(\zeta)$ leads to the solution to (5.34).

Definition 7 (Noisy Adjacency Matrix $\mathbf{G}$ [63])**.** *The noisy adjacency matrix* $\mathbf{G}$ *is given by its entry*

$$\mathsf{G}_{ij} = \begin{cases} 2^{\mathsf{T}_{ij}} & \text{if } (i,j) \in \mathcal{T} \\ 0 & \text{otherwise} \end{cases}. \tag{5.35}$$

Definition 8 (Noisy-Relay Adjacency Matrix $\tilde{\mathbf{A}}(\zeta)$)**.** *The noisy-relay adjacency matrix* $\tilde{\mathbf{A}}(\zeta)$ *is given by its entry*

$$\tilde{\mathsf{A}}_{ij}(\zeta) = \mathsf{G}_{ij} \cdot \mathsf{A}_{ij}(\zeta) = \begin{cases} 2^{\mathsf{T}_{ij}} \cdot 2^{\frac{1-\zeta}{\zeta}\cdot \mathsf{I}_{ij}} & \text{if } (i,j) \in \mathcal{T} \\ 0 & \text{otherwise} \end{cases}. \tag{5.36}$$

for some $\zeta \in [0, +\infty)$.

Then the capacity is $C = \zeta \log \tilde{\lambda}_{\max}(\zeta)$ which can be achieved by

$$\mathsf{P}^*_{ij} = \frac{\tilde{r}_j(\zeta)}{\tilde{r}_i(\zeta)} \cdot \frac{\tilde{\mathsf{A}}_{ij}(\zeta)}{\tilde{\lambda}_{\max}(\zeta)} \quad \forall i,j : (i,j) \in \mathcal{T} \tag{5.37}$$

where $\tilde{\lambda}_{\max}(\zeta)$ is the maximum eigenvalue of the noisy-relay adjacency matrix $\tilde{\mathbf{A}}(\zeta)$ and $\tilde{\boldsymbol{r}}^{\mathsf{T}}(\zeta) = [\tilde{r}_1(\zeta), \tilde{r}_2(\zeta), ..., \tilde{r}_M(\zeta)]^{\mathsf{T}}$ is the corresponding right eigenvector.

Remarks:

i) When $\zeta = 1$, the noisy-relay adjacency matrix $\tilde{\mathbf{A}}(1)$ reduces to a noisy adjacency matrix $\mathbf{G}$.

ii) When the relay-to-destination channel is noise-free, the a-posteriori probabilities $p_n(i,j|Y_1^N)$ and $p_{n-1}(i|Y_1^N)$ are either 0 or 1, we have $\mathsf{T}_{ij} = 0$ for all valid transitions $(i,j) \in \mathcal{T}$. The noisy-relay adjacency matrix $\tilde{\mathbf{A}}(\zeta)$ reduces to a relay adjacency matrix $\mathbf{A}(\zeta)$.

iii) The derivation of the maxentropic state transition probabilities (5.37) is omitted since the proof of Theorem 4 already includes T_{ij}.

Note that, although the exact value of T_{ij} is not computable, an estimate of the value T_{ij} can be computed using the sum-product (BCJR, Baum-Welch) algorithm [8] since the a-posteriori probabilities $p_n(i,j|Y_1^N)$ and $p_{n-1}(i|Y_1^N)$ are the outputs of the sum-product algorithm, i.e., [63]

$$\hat{\mathsf{T}}_{ij} = \frac{1}{N} \sum_{n=1}^{N} \log \left[\frac{p_n(i,j|Y_1^N)^{\frac{p_n(i,j|Y_1^N)}{\mu_i \mathsf{P}_{ij}}}}{p_{n-1}(i|Y_1^N)^{\frac{p_n(i|Y_1^N)}{\mu_i}}} \right]. \tag{5.38}$$

For N large, we have (with probability 1) $\lim_{N\to\infty} \hat{\mathsf{T}}_{ij} = \mathsf{T}_{ij}$ and let $\hat{\mathsf{G}}_{ij} = 2^{\hat{\mathsf{T}}_{ij}}$, $\hat{\mathsf{A}}_{ij} = \hat{\mathsf{G}}_{ij} \cdot \mathsf{A}_{ij}$. We are now ready to formulate the GBAA described in [106, 63] to a two-hop relay channel with a Markovian relay node.

Case 1: First consider the case that the capacity is limited by the capacity of relay-to-destination link in (5.34), i.e., the linear function is above the concave function. While the concave function reaches its maximum, we have

the following condition

$$\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij}\left[\mathsf{I}_{ij} - \log\left(\frac{\sum_{j':(i,j')\in\mathcal{T}} \mathsf{Q}_{ij'}}{\mathsf{Q}_{ij}}\right) - \mathsf{T}_{ij}\right] > 0\ . \tag{5.39}$$

In this case, we only need to compute the noisy capacity of relay-to-destination link. This then becomes the problem of finding the optimal Markovian input over a point-to-point noisy channel considered in [63], in which GBAA was first introduced to solve this problem.

Algorithm 2: GENERALISED BLAHUT-ARIMOTO ALGORITHM FOR CASE 1 [63]

Initialisation: Choose an arbitrary distribution P_{ij} that satisfies the following two conditions:

1. If $(i,j) \in \mathcal{T}$ then $0 < \mathsf{P}_{ij} < 1$, otherwise $\mathsf{P}_{ij} = 0$; and
2. For each i, require that $\sum_{j:(i,j)\in\mathcal{T}} \mathsf{P}_{ij} = 1$.

Repeat until convergence

Step 1: For N large, generate u_1^N according to P_{ij} and pass them through the noisy channel according to $p(y|u)$ to get y_1^N.

Step 2: While keeping all P_{ij} fixed, for each $(i,j) \in \mathcal{T}$, run the sum product algorithm and compute the estimate $\hat{\mathsf{T}}_{ij}$.

Step 3: Compute the standard noisy adjacency matrix $\hat{\mathbf{A}}(1)$ as

$$\hat{\mathsf{A}}_{ij}(1) = \begin{cases} 2^{\hat{\mathsf{T}}_{ij}} & \text{if } (i,j) \in \mathcal{T} \\ 0 & \text{otherwise} \end{cases}$$

and find its maximal eigenvalue $\hat{\lambda}_{\max}(1)$ and its corresponding right and left eigenvectors $\hat{\boldsymbol{r}}^{\mathsf{T}}(1)$ and $\hat{\boldsymbol{l}}(1)$, respectively.

Step 4: For all $(i,j) \in \mathcal{T}$, compute the transition probability matrix P_{ij} and the corresponding state distribution μ_i as

$$\mathsf{P}_{ij} = \frac{\hat{r}_j(1)}{\hat{r}_i(1)} \cdot \frac{\hat{\mathsf{A}}_{ij}(1)}{\hat{\lambda}_{\max}(1)} \quad \forall i,j : (i,j) \in \mathcal{T} \,,$$
$$\mu_i = \hat{c}(1) \cdot \hat{r}(1) \cdot \hat{l}(1) \quad \forall i : (i,j) \in \mathcal{T} \,.$$

end

At the end of the execution of Algorithm 2, if condition (5.39) is not satisfied, Case 2 is considered.

Case 2: The capacity of source-to-relay link is less than or equal to the capacity of relay-to-destination link, i.e.,

$$\sum_{i,j:(i,j)\in\mathcal{T}} \mathsf{Q}_{ij} \left[\mathsf{I}_{ij} - \log \left(\frac{\sum_{j':(i,j')\in\mathcal{T}} \mathsf{Q}_{ij'}}{\mathsf{Q}_{ij}} \right) - \mathsf{T}_{ij} \right] \leq 0 \tag{5.40}$$

Algorithm 3: GENERALISED BLAHUT-ARIMOTO ALGORITHM FOR CASE 2

Initialisation: Choose an arbitrary distribution P_{ij} that satisfies the following two conditions:

1. If $(i,j) \in \mathcal{T}$ then $0 < \mathsf{P}_{ij} < 1$, otherwise $\mathsf{P}_{ij} = 0$; and
2. For each i, require that $\sum_{j:(i,j)\in\mathcal{T}} \mathsf{P}_{ij} = 1$.

Repeat until convergence

Step 1: For N large, generate u_1^N according to P_{ij} and pass them through the noisy channel according to $p(y|u)$ to get y_1^N.

Step 2: While keeping all P_{ij} fixed, for each $(i,j) \in \mathcal{T}$, run the sum product algorithm and compute the estimate $\hat{\mathsf{T}}_{ij}$.

Step 3: Compute the noisy-relay adjacency matrix $\hat{\mathbf{A}}(\zeta)$ as

$$\hat{\mathsf{A}}_{ij}(\zeta) = \begin{cases} 2^{\hat{\mathsf{T}}_{ij}} \cdot 2^{\frac{1-\zeta}{\zeta}\mathsf{I}_{ij}} & \text{if } (i,j) \in \mathcal{T} \\ 0 & \text{otherwise} \end{cases}$$

where ζ is the solution to the following equation

$$\frac{\hat{c}(\zeta)}{\hat{\lambda}_{\max}(\zeta)} \sum_{i,j:(i,j)\in\mathcal{T}} \hat{l}_i \cdot \hat{r}_j \cdot \hat{\mathsf{A}}_{ij}(\zeta) \cdot \mathsf{I}_{ij} = \zeta \log \hat{\lambda}_{\max}(\zeta) \ .$$

Find its maximal eigenvalue $\hat{\lambda}_{\max}(\zeta)$ and its corresponding right and left eigenvectors $\hat{\boldsymbol{r}}^{\mathsf{T}}(\zeta)$ and $\hat{\boldsymbol{l}}(\zeta)$, respectively.

Step 4: For all $(i,j) \in \mathcal{T}$, compute the transition probability matrix P_{ij} and the corresponding state distribution μ_i as

$$\mathsf{P}_{ij} = \frac{\hat{r}_j(\zeta)}{\hat{r}_i(\zeta)} \cdot \frac{\hat{\mathsf{A}}_{ij}(\zeta)}{\hat{\lambda}_{\max}(\zeta)} \quad \forall i,j : (i,j) \in \mathcal{T} \ ,$$
$$\mu_i = \hat{c}(\zeta) \cdot \hat{r}(\zeta) \cdot \hat{l}(\zeta) \quad \forall i : (i,j) \in \mathcal{T} \ .$$

end

5.3.1 Zehavi-Wolf Lower Bound

We further provide a lower bound on the capacity, namely, the Zahavi-wolf lower bound.

Lemma 5 (Zehavi-Wolf Lower Bound [111])**.** *The capacity of a memoryless channel under the constraint that the source is stationary and Markovian is lower bounded by*

$$C \geq \sup_{\mathsf{P}_{ij}} I(S_2; Y_2 | S_1) \tag{5.41}$$

where the supremum is taken over all possible probability assignments P_{ij}*.*

Here, we consider binary inputs for both BSC and AWGN channel. The input probability assignment P_{ij} is chosen to be the optimal assignment for noise-free relay-to-destination capacity.

Thus, for the two-hop half-duplex relay channel with stationary Markovian input at the relay, the capacity is lower bounded by

$$C \geq C_{lb} = \max_{p(x|u=0)} \min \left\{ \begin{array}{l} \sum\limits_{i,j:(i,j)\in\mathcal{Q}} \mu_i \mathsf{P}_{ij} \cdot \mathsf{I}_{ij} \\ I(S_2; Y_2|S_1) \end{array} \right\}. \tag{5.42}$$

For both source-to-relay and relay-to-destination links are both Binary symmetric channel (BSC), the lower bound yields

$$C_{lb1} = \max_{p(x|u=0)} \min \left\{ \begin{array}{l} \sum\limits_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij} \cdot \mathsf{I}_{ij} \\ \sum\limits_{j=0}^{M} \mu_j H(\epsilon \cdot p_j) - H(\epsilon) \end{array} \right\} \tag{5.43}$$

and both links are additive white Gaussian noise channels, the lower bound yields

$$C_{lb2} = \max_{p(x|u=0)} \min \left\{ \begin{array}{l} \sum\limits_{i,j:(i,j)\in\mathcal{T}} \mu_i \mathsf{P}_{ij} \cdot \mathsf{I}_{ij} \\ 2\sum\limits_{j=0}^{M} \mu_j H(p(y|1) \cdot p_j + p(y|0) \cdot (1-p_j)) - \frac{1}{2}\log(2\pi e\sigma^2) \end{array} \right\} \tag{5.44}$$

where p_j is the conditional probability of producing a 1 being in state j.

5.4 Numerical Results

In this section, we present numerical results on capacity of a two-hop half-duplex relay channel under different source-to-relay and relay-to-destination links. We assume that the relay has RLL$([d,k],\mathcal{L}_1)$ and RLL$(\{0\}\cup[1+g,\infty),\{0\}\cup[1+h,\infty))$ constraint, respectively.

5.4.1 Noise-free both links

Let us first consider the case that both source-to-relay and relay-to-destination links are noise-free. Hence, we have the a-posteriori state transition weight

Table 5.1 Noise-free capacity versus RLL($[d, k], \mathcal{L}_1$) constraint parameters d and k

$k\backslash d$	0	1	2	3	4	5	6
1	.4360						
2	.5839	.4057					
3	.6557	.5515	.2878				
4	.6962	.6174	.4057	.2232			
5	.7209	.6509	.4650	.3218	.1823		
6	.7369	.6690	.4979	.3746	.2669	.1542	
7	.7475	.6793	.5174	.4057	.3142	.2281	.1335
8	.7547	.6853	. 5293	.4251	.3432	.2709	.1993
9	.7597	.6888	. 5369	.4376	.3620	.2979	.2382
10	.7633	.6909	.5418	.4460	.3746	.3158	.2633
11	.7658	.6922	.5450	.4516	.3833	.3282	.2804
12	.7676	.6930	.5471	.4555	.3894	.3369	.2924
13	.7690	.6935	.5485	.4583	.3937	.3432	.3011
14	.7700	.6938	.5495	.4602	.3968	.3478	.3074
15	.7707	.6939	.5501	.4615	.3991	.3513	.3122
∞	.7729	.6942	.5515	.4650	.4057	.3620	.3282

matrix T with its entries

$$\mathsf{T}_{ij} = \begin{cases} 0 & \text{if } (i,j) \in \mathcal{T} \\ -\infty & \text{otherwise} \end{cases} \tag{5.45}$$

and the relay mutual information matrix I with its entries

$$\mathsf{I}_{ij} = \begin{cases} 1 & \text{if } (i,j) \in \mathcal{Q} \\ 0 & \text{if } (i,j) \in \mathcal{Q}^{\mathrm{c}} \cap \mathcal{T} \\ -\infty & \text{otherwise} \end{cases} . \tag{5.46}$$

Under such setting, we compute and list the capacities based on Theorem 4 and Corollary 3, when the relay adopts different RLL($[d, k], \mathcal{L}_1$) sequences in TABLE 5.1 and RLL($\{0\} \cup [1+g, \infty), \{0\} \cup [1+h, \infty)$) sequences in TABLE 5.2. *Remark:* In the case of unconstrained cases, we have $\zeta = 0.3614$ and $\lambda_{\max} = 4.4035$, yielding the capacity $C = 0.7729$ bits/use. The capacity value 0.7729 coincides with the unconstrained capacity over a noiseless two-hop half-duplex relay channel in [115, 104, 75], i.e., the largest root of the equation $H(x) = x$.

Table 5.2 Noise-free capacity versus RLL($\{0\} \cup [1+g, \infty), \{0\} \cup [1+h, \infty)$) with different g, h

$h \backslash g$	0	1	2	3	4	5	6	7	8
0	.7729	.7418	.6881	.6125	.5515	.5037	.4650	.4329	.4057
1	.6667	.6436	.6067	.5515	.5037	.4650	.4329	.4057	.3823
2	.5946	.5755	.5467	.5037	.4650	.4329	.4057	.3823	.3620
3	.5405	.5240	.5000	.4650	.4329	.4057	.3823	.3620	.3441
4	.4976	.4830	.4623	.4329	.4057	.3823	.3620	.3441	.3282
5	.4624	.4492	.4309	.4057	.3823	.3612	.3441	.3282	.3139
6	.4328	.4207	.4044	.3823	.3611	.3441	.3282	.3139	.3011
7	.4074	.3963	.3814	.3612	.3441	.3282	.3139	.3011	.2894
8	.3853	.3750	.3614	.3441	.3282	.3139	.3011	.2894	.2788
9	.3659	.3563	.3438	.3282	.3139	.3011	.2894	.2788	.2690
10	.3486	.3397	.3280	.3139	.3011	.2894	.2788	.2690	.2600

5.4.2 Noisy Links: Two BSCs with the same cross-over probability p

In this subsection, we consider a two-hop half-duplex relay channel with a Markovian constraint relay and two BSCs links of the same cross-over probability p. We have the a-posteriori state transition weight matrix $\mathbf{T}$ defined in Definition 6 and

$$\mathsf{I}_{ij} = \begin{cases} 1 - H(p) & \text{if } (i,j) \in \mathcal{Q} \\ 0 & \text{if } (i,j) \in \mathcal{Q}^{\mathsf{c}} \cap \mathcal{T} \\ -\infty & \text{otherwise} \end{cases} . \tag{5.47}$$

Under this setting, we compute tight lower bounds on the constrained capacities via Algorithm 2 and Algorithm 3 sequentially, i.e., at the end of the execution of Algorithm 2, if condition (5.39) is not satisfied, then perform Algorithm 3. We compare our results with

- the upper bounds on the noisy capacity, i.e., the capacity C_{nf} of the two-hop relay channel with a noise-free relay-to-destination link, and a BSC source-to-relay link with cross-over probability p,

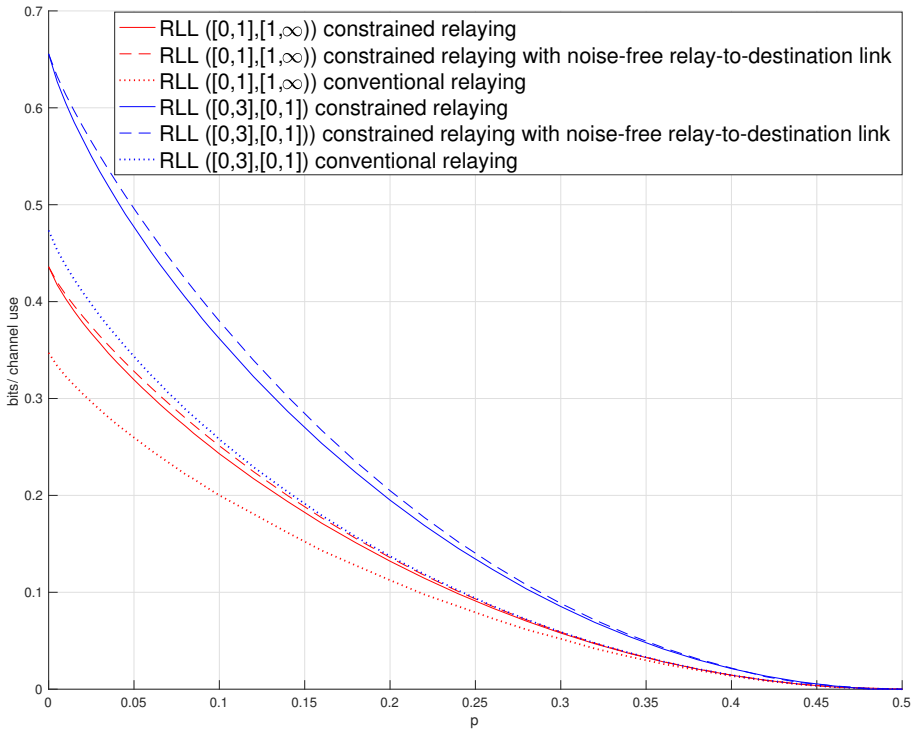

Figure 5.6 Information rates of a BSC two-hop half-duplex relay channel with a relay under RLL$([0,1],[1,\infty))$ and $([0,3],[0,\infty))$ constraints.

- the maximal achievable rate by time-sharing scheme [65] is then given by

$$R_{\text{ts}} = \frac{1}{2}\log\tilde{\lambda}_{\max}(1)\ , \tag{5.48}$$

  which can be computed via Algorithm 2.

In Figure 5.6, we illustrates such comparisons when the relay adopts the RLL$([0,1],[1,\infty))$ and $([0,3],[0,\infty))$ constrained sequences, respectively. We observe that our results are very close to the upper bounds and perform better than the conventional time-sharing schemes. In particular, when both links have $p = 0$, our scheme achieves 0.4360 bits/use and 0.6557 bits/use for RLL$([0,1],[1,\infty))$ and $([0,3],[0,\infty))$ constraints, respectively, while the conventional schemes achieves 0.3471 bits/use and 0.4734 bits/use, respectively. Similar results are obtained for RLL$(\{0\}\cup[2,\infty),[0,\infty))$ and $(\{0\}\cup[3,\infty),\{0\}\cup[2,\infty))$ constraint, as illustrated in FIGURE 5.7.

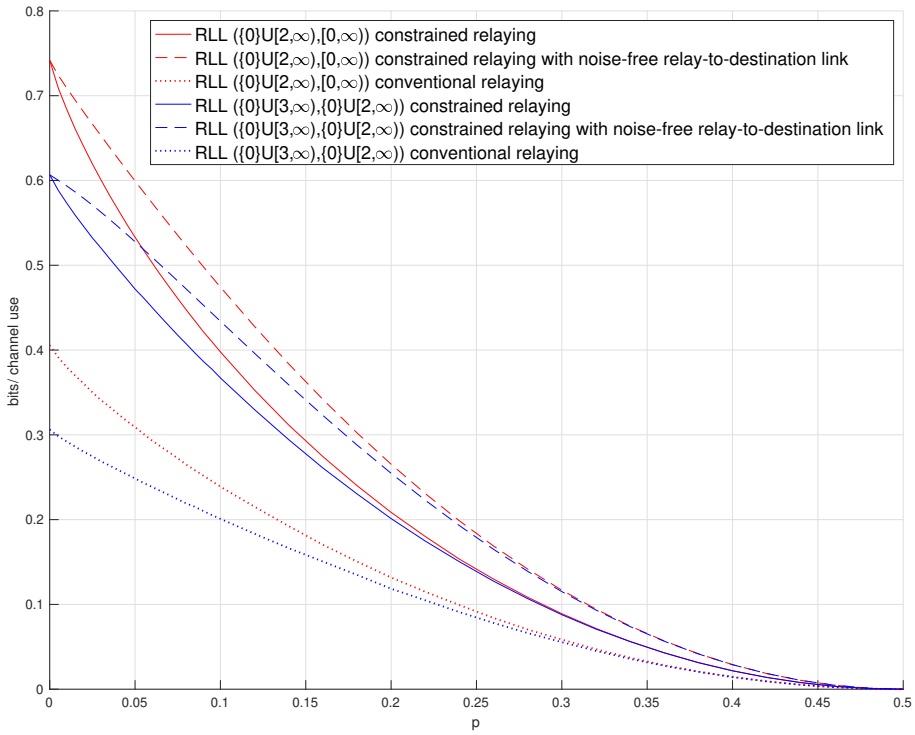

Figure 5.7 Information rates of a BSC two-hop half-duplex relay channel with a relay under RLL$(\{0\} \cup [2, \infty), [0, \infty))$ and $\{0\} \cup [3, \infty), \{0\} \cup [2, \infty))$ constraints.

We conduct the same comparison as that in Figure 5.6, except that the relay uses a RLL$([1, 3], [0, 1])$ sequence, instead. We report our comparisons in Figure 5.8. we computed the optimized information rate for the case that the relay node is RLL$([1, 3], [0, 1])$ constrained. The vertical black dash-dot line at $p = 0.045$ indicates that for cross-over probability $p \leq 0.045$, the optimized information rate is computed via Algorithm 2, i.e., there is always enough information fed into the relay and the relay may transmit information at its full rate. For the case that $p > 0.045$, since condition (5.39) is not satisfied, i.e., the relay may not transmit at its full rate, Algorithm 3 is performed to find an equilibrium point such that the source is feeding information to the relay at the same rate that the relay is transmitting.

In Figure 5.9, we compare our optimized information rate for various RLL$([d, k], \mathcal{L}_1)$ constraint against the unconstrained optimized information rate

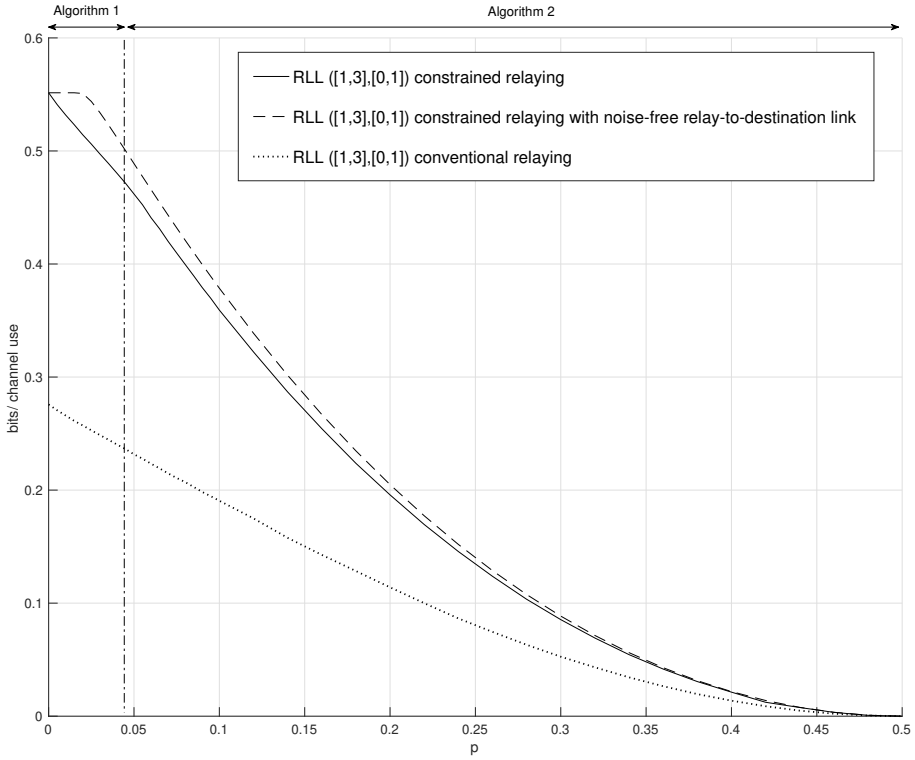

Figure 5.8 Information rates of a BSC two-hop half-duplex relay channel with a relay under RLL$([1,3],[0,1])$ constraint.

in [115]. For RLL$([0,k],\mathcal{L}_1)$ constraint, where $k \in 1,3,7$, we observe that the larger values of k lead to higher optimised information rate. In particular, when $k=7$, our scheme approach to the unconstrained optimised information rate. This agrees with our intuition that as the relay becomes less constrained (i.e., as $k \to \infty$), the problem transforms into optimising mutual information rate for a BSC two-hop half-duplex relay channel with an unconstrained relay.

5.5 Conclusion

In this chapter, we studied the capacity of a two-hop half-duplex channel (source-relay-destination) with a Markovian constrained relay. For this channel, we show that when the zero symbol transmitted by the relay also conveys information; during its transmission, the source is able to transmit information

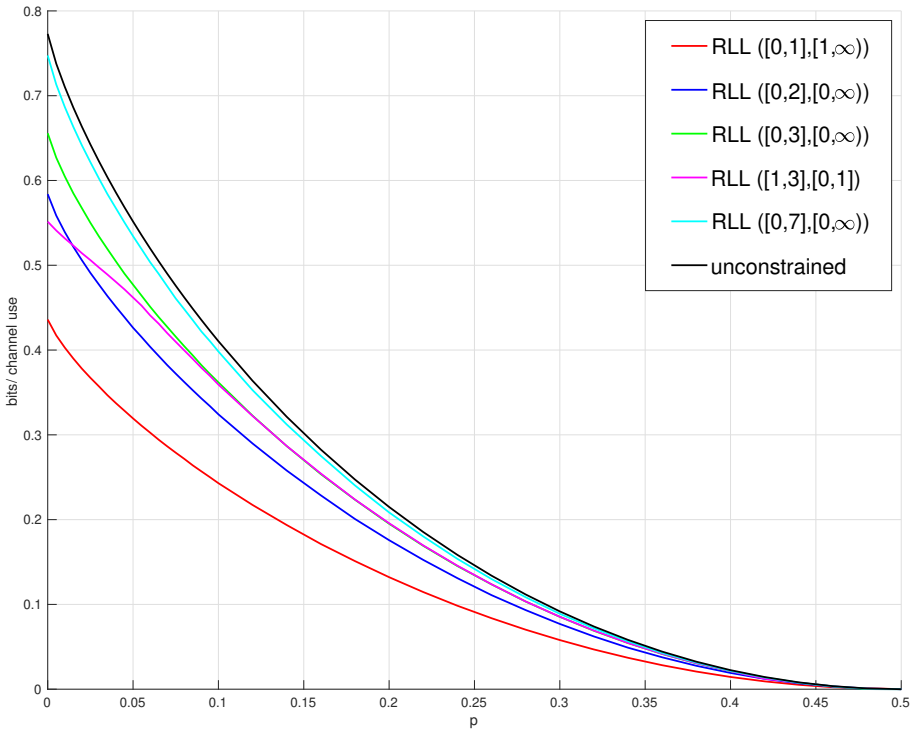

Figure 5.9 Unconstrained optimized information rate [115] vs. optimized information rates of a BSC two-hop half-duplex relay channel with a relay under RLL$([d, k], \mathcal{L}_1)$ constraint.

to the relay, significant rate gains are possible. For the case with a noise-free relay-to-destination link, capacity achieving input is derived. When both source-to-relay and relay-to-destination links are noisy, the constrained mutual information rate are optimised using the GBAA algorithm modified for two-hop relay channels. These results may be instrumental in deriving capacity when constrained sequences are introduced to a more complex network.

Chapter 6

Conclusion and Future Work

6.1 Conlusions

This research focuses on information theoretical studies on non-Gaussian noise channels and Markovian constrained relay channels. In the first part of this thesis, the capacities over non-Gaussian noise channels have been studied, in which the non-Gaussian noise is modelled as Gaussian mixture model. We introduced the envelope Gaussian mixture noise model which models the amplitude distribution of complex Gaussian mixture noise. Its parameters estimation is performed via EM algorithm and Quasi-Newtion method. The capacity achieving-input for a Gaussian mixture noise channel is shown to be of discrete nature. In the second part of this thesis, the capacity of a two-hop Markovian constrained relay channel has been studied where the relay is subject to half-duplex and Markovian constraint. The maxentropic state transition probabilities for relay transmitters are studied for the case that the relay-to-destination link is noiseless. For the case that both source-to-relay and relay-to-destination links are noisy, generalised Blahut-Arimoto Algorithm is performed to compute a tight lower bound on the capacity.

6.2 Summary of Contributions

The following subsections summarise the detail contributions of the research.

Objective 1: New statistic models that better describe the non-Gaussian noise behaviour and relay constraints

A *new, simple and exact* closed form probability density function of the envelope of the Gaussian mixture distributions (i.e. the envelope of independent in-phase and quadrature components of complex non-Gaussian noise) is obtained, namely, the envelope Gaussian mixture model. The proposed model has less pronounced multi-modal behaviour than the envelope of two-term Rayleigh mixture model and it has a linearly decaying tail which is observed in the histogram of power line communication noises. Moreover, the envelope Gaussian mixture model turns into a single Rayleigh distribution when the variances of each Gaussian component has the same variance. Hence, the envelope Gaussian mixture fits naturally into the distribution family.

We further proposed the hold time constraint model for a Markovian constrained relay due to the existence of switching noise, an induced ISI when switching a digital signal, caused by the relay's switching between reception mode and transition mode.

Objective 2: Parameters estimation of the proposed models

A common question after a new model being proposed is how to estimate the parameters given some observations. We adopt the well-known two-step iterative method called the EM Algorithm that finds the maximum likelihood or maximum a posteriori estimates of parameters in statistical models in which observations are treated as "incomplete data". A simulation of the proposed the estimator via the EM Algorithm is performed. The performance of the EM Algorithm is compared against the performance of the Quasi-Newton method. Our simulation shows that both EM Algorithm and Quasi-Newton performs well in terms of maximising the log likelihood function, however the EM Algorithm has relatively slower convergence rate depending on the models and data

size. Quasi-Newton, on the other hand, takes fewer iterations but is more complicated to implement.

Objective 4: Obtain capacity results

In the first part of this thesis, the capacity of the Gaussian mixture noise channel and its capacity-achieving input have been studied. In particular, the symmetric and asymmetric Gaussian mixture noise cases are considered. It is shown that, under average and peak power constraints, the capacity-achieving input is discrete with finitely many mass points. Furthermore, some properties of the capacity-achieving distribution are proved and demonstrated by simulations.

In the second part of this thesis, the capacity theorems for a two-hop half-duplex relay channel with a Markovian constrained relay are studied. It first focuses on deriving the cut-set bound, i.e., an upper bound on the capacity. Then this chapter presents the timing strategy which satisfies the half-duplex constraint and it is shown to achieve this bound. This leads to the general capacity formula for a two-hop half-duplex relay channel with a Markovian constrained relay.

Objective 5: Explore computation methods

In the first part of this thesis, the optimal input for the Gaussian mixture noise channel is numerically computed, i.e., the optimal mass point locations and their probabilities, by using the interior point method by starting from random initial mass point locations and probabilities.

In the second part of this thesis, the relay adjacency matrix is introduced such that the capacity can be easily computed in the case that relay-to-destination link is noiseless. The relay adjacency matrix can be considered as an extend version of adjacency matrix by Shannon for the two-hop relay channel, which is commonly used to compute the largest eigenvalue and corresponding largest eigenvector of non-negative primitive matrices. Moreover, for the case that the relay-to-destination link is noisy, noisy-relay adjacency matrix is introduced. Algorithms are presented to compute the noisy-relay adjacency matrix and compute the optimised information rate which serves as a natural lower bound

on the capacity. The bounds are shown to be very tight while comparing against upper bounds.

6.3 Future Work

We regret if – after having worked for quite a while on these topics – we have to leave the reader with more questions remaining than there were originally. The good thing though is that the following future research projects may be conducted:

Capacity of full-duplex Markovian constrained relay with self-interference

While deriving the capacity of the two-hop half-duplex relay channel with a Markovian constrained relay, we introduced the relay mutual information matrix and the relay adjacency matrix in Chapter 5. Both of them can be naturally extended to the full-duplex case, however it requires taking into account the self-interference power and also it is worth noting that the self-interference is caused by the transmission event of the relay node itself.

Capacity of half-duplex/full-duplex Markovian constrained relay with more than one source node

Since a two-hop Markovian constrained relay channel is the most simple network, it is a natural idea to consider constrained sequences in a more complex network, for example, when more than one source node are introduced, the timing scheme introduced in Chapter 4 maybe used to achieve the capacity. If not, new coding scheme shall be investigated.

Capacity of half-duplex/full-duplex Markovian constrained relay in a line network

Another option to further extend our study is to consider relay in a line network in which there is one source node, one destination node and more than one relay nodes. Assuming that one of the relay node is default and hence the

broken relay is subject to a Markovian constraint. This may requires a new information-theoretical analysis.

Bibliography

[1] Abou-Faycal, I. C., Trott, M. D., and Shamai, S. (2001). The capacity of discrete-time memoryless rayleigh-fading channels. *IEEE Transactions on Information Theory*, 47(4):1290–1301.

[2] Ahlfors, L. (1979). *Complex analysis: an introduction to the theory of analytic functions of one complex variable*. International series in pure and applied mathematics. McGraw-Hill.

[3] Arimoto, S. (1972). An algorithm for computing the capacity of arbitrary discrete memoryless channels. *IEEE Transactions on Information Theory*, 18(1):14–20.

[4] Arnold, D. M. (2003). *Computing information rates of finite-state models with application to magnetic recording*, volume 10. ETH Zurich.

[5] Arnold, D. M., Loeliger, H. A., Vontobel, P. O., Kavcic, A., and Zeng, W. (2006). Simulation-based computation of information rates for channels with memory. *IEEE Transactions on Information Theory*, 52(8):3498–3508.

[6] Aryafar, E., Khojastepour, M. A., Sundaresan, K., Rangarajan, S., and Chiang, M. (2012). Midu: Enabling mimo full duplex. In *Proceedings of the 18th Annual International Conference on Mobile Computing and Networking*, Mobicom '12, pages 257–268, New York, NY, USA. ACM.

[7] Azarian, K., Gamal, H. E., and Schniter, P. (2005). On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels. *IEEE Transactions on Information Theory*, 51(12):4152–4172.

[8] Bahl, L., Cocke, J., Jelinek, F., and Raviv, J. (1974). Optimal decoding of linear codes for minimizing symbol error rate (corresp.). *IEEE Transactions on Information Theory*, 20(2):284–287.

[9] Barbero, A. I., Rosnes, E., Yang, G., and Ytrehus, O. (2014). Near-field passive rfid communication: Channel model and code design. *IEEE Transactions on Communications*, 62(5):1716–1727.

[10] Benyoucef, D. (2003). A new statistical model of the noise power density spectrum for powerline communication. In *Proc. Int. Symp. Power-Line Commun. and its Applicat. (ISPLC)*, pages 136–141.

[11] Bert, L. D., Caldera, P., Schwingshackl, D., and Tonello, A. M. (2011). On noise modeling for power line communications. In *2011 IEEE International Symposium on Power Line Communications and Its Applications*, pages 283–288.

[12] Bharadia, D., McMilin, E., and Katti, S. (2013). Full duplex radios. *SIGCOMM Comput. Commun. Rev.*, 43(4):375–386.

[13] Bishop, C., Bishop, C. M., et al. (1995). *Neural networks for pattern recognition*. Oxford university press.

[14] Blahut, R. (1972). Computation of channel capacity and rate-distortion functions. *IEEE Transactions on Information Theory*, 18(4):460–473.

[15] Bliss, D. W., Parker, P. A., and Margetts, A. R. (2007). Simultaneous transmission and reception for improved wireless network performance. In *2007 IEEE/SP 14th Workshop on Statistical Signal Processing*, pages 478–482.

[16] Bolcskei, H., Nabar, R. U., Oyman, O., and Paulraj, A. J. (2006). Capacity scaling laws in mimo relay networks. *IEEE Transactions on Wireless Communications*, 5(6):1433–1444.

[17] Burr, A. G. and Brown, P. A. (1998). Hf broadcast interference on low voltage mains distribution networks. In *Proc. Int. Symp. Power-Line Commun. and its Applicat. (ISPLC)*, pages 253–262.

[18] Cai, Y., Mutlu, O., Haratsch, E. F., and Mai, K. (2013). Program interference in mlc nand flash memory: Characterization, modeling, and mitigation. In *2013 IEEE 31st International Conference on Computer Design (ICCD)*, pages 123–130.

[19] Calderin, L. (2017). *Flexible Integrated Architectures for Frequency Division Duplex Communication*. PhD thesis, UC Berkeley.

[20] Canete, F. J., Cortes, J. A., Diez, L., Entrambasaguas, J. T., and Carmona, J. L. (2005). Fundamentals of the cyclic short-time variation of indoor power-line channels. In *International Symposium on Power Line Communications and Its Applications, 2005.*, pages 157–161.

[21] Chan, M. H. L. and Donaldson, R. W. (1989). Amplitude, width, and interarrival distributions for noise impulses on intrabuilding power line communication networks. *IEEE Transactions on Electromagnetic Compatibility*, 31(3):320–323.

[22] Chan, T. H., Hranilovic, S., and Kschischang, F. R. (2005). Capacity-achieving probability measure for conditionally gaussian channels with bounded inputs. *IEEE Transactions on Information Theory*, 51(6):2073–2088.

[23] Choi, J. I., Jain, M., Srinivasan, K., Levis, P., and Katti, S. (2010). Achieving single channel, full duplex wireless communication. In *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, MobiCom '10, pages 1–12, New York, NY, USA. ACM.

[24] Cover, T. and Gamal, A. E. (1976). Information transmission through a channel with relay. *IEEE Transactions on Information Theory*, 25:B76–7.

[25] Cover, T. and Gamal, A. E. (1979). Capacity theorems for the relay channel. *IEEE Transactions on Information Theory*, 25(5):572–584.

[26] Cover, T. M. and Thomas, J. A. (2006). *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience.

[27] Cramer, H. (2004). *Random Variables and Probability Distributions*. Cambridge Tracts in Mathematics. Cambridge University Press.

[28] Das, A. (2000). Capacity-achieving distributions for non-gaussian additive noise channels. In *2000 IEEE International Symposium on Information Theory (Cat. No.00CH37060)*, pages 432–.

[29] Degardin, V., Lienard, M., Zeddam, A., Gauthier, F., and Degauquel, P. (2002). Classification and characterization of impulsive noise on indoor powerline used for data communications. *IEEE Transactions on Consumer Electronics*, 48(4):913–918.

[30] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

[31] Ding, Z., Krikidis, I., Rong, B., Thompson, J. S., Wang, C., and Yang, S. (2012). On combating the half-duplex constraint in modern cooperative networks: protocols and techniques. *IEEE Wireless Communications*, 19(6):20–27.

[32] Dong, G., Pan, Y., Xie, N., Varanasi, C., and Zhang, T. (2012). Estimating information-theoretical nand flash memory storage capacity and its implication to memory system design space exploration. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 20(9):1705–1714.

[33] Dostert, K., Waldeck, T., and Zimmermann, M. (1997). Fundamental properties of the low voltage power distribution grid. In *Proc. Int. Symp. Power-Line Commun. and its Applicat. (ISPLC)*, pages 45–50.

[34] Duarte, M., Dick, C., and Sabharwal, A. (2012). Experiment-driven characterization of full-duplex wireless systems. *IEEE Transactions on Wireless Communications*, 11(12):4296–4307.

[35] Duarte, M. and Sabharwal, A. (2010). Full-duplex wireless communications using off-the-shelf radios: Feasibility and first results. In *2010 Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers*, pages 1558–1562.

[36] El Gamal, A. and Zahedi, S. (2005). Capacity of a class of relay channels with orthogonal components. *IEEE Transactions on Information Theory*, 51(5):1815–1817.

[37] Everett, E., Duarte, M., Dick, C., and Sabharwal, A. (2011). Empowering full-duplex wireless communication by exploiting directional diversity. In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 2002–2006.

[38] Fahs, J., Ajeeb, N., and Abou-Faycal, I. (2012). The capacity of average power constrained additive non-gaussian noise channels. In *2012 19th International Conference on Telecommunications (ICT)*, pages 1–6.

[39] Ferreira, H., Lampe, L., Newbury, J., and Swart, T. (2010). *Power Line Communications: Theory and Applications for Narrowband and Broadband Communications over Power Lines*.

[40] Fouladgar, A. M., Simeone, O., and Erkip, E. (2014). Constrained codes for joint energy and information transfer. *IEEE Transactions on Communications*, 62(6):2121–2131.

[41] Gallager, R. (1968). *Information Theory and Relaible Communication)*. John Wiley and Sons, INC, New York, NY, USA.

[42] Gorham. Smith, J. (1969). On the information capacity of peak and average power constrained gaussian channels.

[43] Gotz, M., Rapp, M., and Dostert, K. (2004). Power line channel characteristics and their effect on communication system design. *IEEE Communications Magazine*, 42(4):78–86.

[44] Guo, D., Shamai, S., and Verdu, S. (2005). Mutual information and minimum mean-square error in gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282.

[45] Gursoy, M. C., Poor, H. V., and Verdu, S. (2005). The noncoherent rician fading channel-part i: structure of the capacity-achieving input. *IEEE Transactions on Wireless Communications*, 4(5):2193–2206.

[46] Han, G. (2013). A randomized approach to the capacity of finite-state channels. In *2013 IEEE International Symposium on Information Theory*, pages 2109–2113.

[47] Han, G. and Marcus, B. (2009). Concavity of mutual information rate for input-restricted finite-state memoryless channels at high snr. In *2009 IEEE International Symposium on Information Theory*, pages 1654–1658.

[48] Haneda, K., Kahra, E., Wyne, S., Icheln, C., and Vainikainen, P. (2010). Measurement of loop-back interference channels for outdoor-to-indoor full-duplex radio relays. In *Proceedings of the Fourth European Conference on Antennas and Propagation*, pages 1–5.

[49] Hasna, M. O. and Alouini, M. . (2003). Outage probability of multihop transmission over nakagami fading channels. *IEEE Communications Letters*, 7(5):216–218.

[50] Hasna, M. O. and Alouini, M. . (2004a). Optimal power allocation for relayed transmissions over rayleigh-fading channels. *IEEE Transactions on Wireless Communications*, 3(6):1999–2004.

[51] Hasna, M. O. and Alouini, M. . (2004b). A performance study of dual-hop transmissions with fixed gain relays. *IEEE Transactions on Wireless Communications*, 3(6):1963–1968.

[52] Hirayama, Y., Okada, H., Yamazato, T., and Katayama, M. (2003). Noise analysis on wide-band plc with high sampling rate and long observation time. In *Proc. Int. Symp. Power-Line Commun. and its Applicat. (ISPLC)*, pages 142–147.

[53] Hooijen, O. G. (1998). On the channel capacity of the residential power circuit used as a digital communications medium. *IEEE Communications Letters*, 2(10):267–268.

[54] Horn, R. A. and Johnson, C. R., editors (1986). *Matrix Analysis*. Cambridge University Press, New York, NY, USA.

[55] Hoymann, C., Chen, W., Montojo, J., Golitschek, A., Koutsimanis, C., and Shen, X. (2012). Relaying operation in 3gpp lte: challenges and solutions. *IEEE Communications Magazine*, 50(2):156–162.

[56] Huang, L., Hong, Y., and Viterbo, E. (2014). On parameter estimation of the envelope gaussian mixture model. In *2014 Australian Communications Theory Workshop (AusCTW)*, pages 27–32.

[57] Huang, L., Hong, Y., and Viterbo, E. (2015). On the capacity of the gaussian mixture channel under average and peak power constraints. In *2015 22nd International Conference on Telecommunications (ICT)*, pages 146–150.

[58] Jain, M., Choi, J. I., Kim, T., Bharadia, D., Seth, S., Srinivasan, K., Levis, P., Katti, S., and Sinha, P. (2011). Practical, real-time, full duplex wireless. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*, MobiCom '11, pages 301–312, New York, NY, USA. ACM.

[59] Jamshidian, M. and Jennrich, R. I. (1997). Acceleration of the em algorithm by using quasi-newton methods. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(3):569–587.

[60] Justesen, J. and Høholdt, T. (1984). Maxentropic markov chains (corresp.). *IEEE Transactions on Information Theory*, 30(4):665–667.

[61] Katayama, M., Yamazato, T., and Okada, H. (2006). A mathematical model of noise in narrowband power line communication systems. *IEEE Journal on Selected Areas in Communications*, 24(7):1267–1276.

[62] Katz, M. and Shamai, S. (2004). On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent awgn channels. *IEEE Transactions on Information Theory*, 50(10):2257–2270.

[63] Kavčić, A. (2001). On the capacity of markov sources over noisy channels. In *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, volume 5, pages 2997–3001 vol.5.

[64] Khayrallah, A. S. and Neuhoff, D. L. (1996). Coding for channels with cost constraints. *IEEE Transactions on Information Theory*, 42(3):854–867.

[65] Khojastepour, M. A., Sabharwal, A., and Aazhang, B. (2003). On capacity of gaussian 'cheap' relay channel. In *GLOBECOM '03. IEEE Global Telecommunications Conference (IEEE Cat. No.03CH37489)*, volume 3, pages 1776–1780 vol.3.

[66] Kim, J., Lee, D., and Sung, W. (2012). Performance of rate 0.96 (68254, 65536) eg-ldpc code for nand flash memory error correction. In *2012 IEEE International Conference on Communications (ICC)*, pages 7029–7033.

[67] Kim, Y. H. (2007). Coding techniques for primitive relay channels. In *in Proc. Forty-Fifth Annual Allerton Conf. Commun., Contr. Comput*. Citeseer.

[68] Kramer, G. (2004). Models and theory for relay channels with receive constraints. In *in 42nd Annual Allerton Conf. on Commun., Control, and Computing*.

[69] Kramer, G., Gastpar, M., and Gupta, P. (2005). Cooperative strategies and capacity theorems for relay networks. *IEEE Transactions on Information Theory*, 51(9):3037–3063.

[70] Laneman, J. N., Tse, D. N. C., and Wornell, G. W. (2004). Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *IEEE Transactions on Information Theory*, 50(12):3062–3080.

[71] Laneman, J. N. and Wornell, G. W. (2000). Energy-efficient antenna sharing and relaying for wireless networks. In *2000 IEEE Wireless Communications and Networking Conference. Conference Record (Cat. No.00TH8540)*, volume 1, pages 7–12 vol.1.

[72] Laneman, J. N. and Wornell, G. W. (2002). Distributed space-time coded protocols for exploiting cooperative diversity in wireless networks. In *Global Telecommunications Conference, 2002. GLOBECOM '02. IEEE*, volume 1, pages 77–81 vol.1.

[73] Liu, D., Flint, E., Gaucher, B., and Kwark, Y. (1999). Wide band ac power line characterization. *IEEE Transactions on Consumer Electronics*, 45(4):1087–1097.

[74] Loa, K., Wu, C., Sheu, S., Yuan, Y., Chion, M., Huo, D., and Xu, L. (2010). Imt-advanced relay standards [wimax/lte update]. *IEEE Communications Magazine*, 48(8):40–48.

[75] Lutz, T., Hausl, C., and Kotter, R. (2012). Bits through deterministic relay cascades with half-duplex constraint. *IEEE Transactions on Information Theory*, 58(1):369–381.

[76] Marina, N., Kavcic, A., and Gaarder, N. T. (2008). Capacity theorems for relay channels with isi. In *2008 IEEE International Symposium on Information Theory*, pages 479–483.

[77] Meng, H., Guan, Y. L., and Chen, S. (2005). Modeling and analysis of noise effects on broadband power-line communications. *IEEE Transactions on Power delivery*, 20(2):630–637.

[78] Middleton, D. (1972). Statistical-physical models of urban radio-noise environments - part i: Foundations. *IEEE Transactions on Electromagnetic Compatibility*, EMC-14(2):38–56 –.

[79] Middleton, D. (1979). Canonical non-gaussian noise models: Their implications for measurement and for prediction of receiver performance. *IEEE Transactions on Electromagnetic Compatibility*, EMC-21(3):209–220.

[80] Middleton, D. (1983). Canonical and quasi-canonical probability models of class a interference. *IEEE Transactions on Electromagnetic Compatibility*, EMC-25(2):76–106.

[81] Muroga, S. (1953). On the capacity of a discrete channel. i mathematical expression of capacity of a channel which is disturbed by noise in its every one symbol and expressible in one state diagram. *Journal of the Physical Society of Japan*, 8(4):484–494.

[82] Nabar, R. U., Bolcskei, H., and Kneubuhler, F. W. (2004). Fading relay channels: performance limits and space-time signal design. *IEEE Journal on Selected Areas in Communications*, 22(6):1099–1109.

[83] Nosratinia, A., Hunter, T. E., and Hedayat, A. (2004). Cooperative communication in wireless networks. *IEEE Communications Magazine*, 42(10):74–80.

[84] Pabst, R., Walke, B. H., Schultz, D. C., Herhold, P., Yanikomeroglu, H., Mukherjee, S., Viswanathan, H., Lott, M., Zirwas, W., Dohler, M., Aghvami, H., Falconer, D. D., and Fettweis, G. P. (2004). Relay-based deployment concepts for wireless and mobile broadband radio. *IEEE Communications Magazine*, 42(9):80–89.

[85] Papoulis, A. (1965). Probability, random variables, and stochastic processes. Technical report.

[86] Pfister, H. D., Soriaga, J. B., and Siegel, P. H. (2001). On the achievable information rates of finite state isi channels. In *Global Telecommunications Conference, 2001. GLOBECOM '01. IEEE*, volume 5, pages 2992–2996 vol.5.

[87] Radunovic, B., Gunawardena, D., Key, P., Proutiere, A., Singh, N., Balan, V., and Dejean, G. (2010). Rethinking indoor wireless mesh design: Low power, low frequency, full-duplex. In *2010 Fifth IEEE Workshop on Wireless Mesh Networks*, pages 1–6.

[88] Riihonen, T., Balakrishnan, A., Haneda, K., Wyne, S., Werner, S., and Wichman, R. (2011). Optimal eigenbeamforming for suppressing self-interference in full-duplex mimo relays. In *2011 45th Annual Conference on Information Sciences and Systems*, pages 1–6.

[89] Riihonen, T., Werner, S., and Wichman, R. (2009). Optimized gain control for single-frequency relaying with loop interference. *IEEE Transactions on Wireless Communications*, 8(6):2801–2806.

[90] S. Kiaei, D. Allstot, K. H. and Verghese, N. K. (1998). Noise considerations for mixed-signal rf ic transceivers. *Wireless Networks*, 4(1):41–53.

[91] Sahai, A., Patel, G., and Sabharwal, A. (2011). Pushing the limits of full-duplex: Design and real-time implementation.

[92] Salman, E. (2009). *Switching Noise and Timing Characteristics in Nanoscale Integrated Circuits*. PhD thesis, University of Rochester.

[93] Sari, F., Sari, N., and Mili, L. (2004). Modelling of sea clutter with gaussian mixtures and estimation of the clutter parameter. In *Proceedings of the IEEE 12th Signal Processing and Communications Applications Conference, 2004.*, pages 53–56.

[94] Schouhamer-Immink, K. (1991). *Coding Techniques for Digital Recorders*. Prentice Hall Professional Technical Reference.

[95] Seabra, J. C., Ciompi, F., Pujol, O., Mauri, J., Radeva, P., and Sanches, J. (2011). Rayleigh mixture model for plaque characterization in intravascular ultrasound. *IEEE Transactions on Biomedical Engineering*, 58(5):1314–1324.

[96] Sendonaris, A., Erkip, E., and Aazhang, B. (2003). User cooperation diversity. part i. system description. *IEEE Transactions on Communications*, 51(11):1927–1938.

[97] Shamai, S. and Bar-David, I. (1995). The capacity of average and peak-power-limited quadrature gaussian channels. *IEEE Transactions on Information Theory*, 41(4):1060–1071.

[98] Shamai, S. and Kofman, Y. (1990). On the capacity of binary and gaussian channels with run-length-limited inputs. *IEEE Transactions on Communications*, 38(5):584–594.

[99] Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55.

[100] Sharma, V. and Singh, S. K. (2001). Entropy and channel capacity in the regenerative setup with applications to markov channels. In *Proceedings. 2001 IEEE International Symposium on Information Theory (IEEE Cat. No.01CH37252)*, pages 283–.

[101] Smith, J. G. (1971). The information capacity of amplitude- and variance-constrained sclar gaussian channels. *Information and Control*, 18(3):203 – 219.

[102] Tchamkerten, A. (2004). On the discreteness of capacity-achieving distributions. *IEEE Transactions on Information Theory*, 50(11):2773–2778.

[103] van Der Meulen, E. C. (1971). Three-terminal communication channels. *Advances in Applied Probability*, 3(1):120–154.

[104] Vanroose, P. and van der Meulen, E. C. (1992). Uniquely decodable codes for deterministic relay channels. *IEEE Transactions on Information Theory*, 38(4):1203–1212.

[105] Verdu, S. and Han, T. S. (2006). A general formula for channel capacity. *IEEE Trans. Inf. Theor.*, 40(4):1147–1157.

[106] Vontobel, P. O., Kavcic, A., Arnold, D. M., and Loeliger, H. A. (2008). A generalization of the blahut-arimoto algorithm to finite-state channels. *IEEE Transactions on Information Theory*, 54(5):1887–1918.

[107] Watanabe, M. and Yamaguchi, K. (2003). *The EM algorithm and related statistical models*. CRC Press.

[108] Wu, C. J. (1983). On the convergence properties of the em algorithm. *The Annals of statistics*, pages 95–103.

[109] Yang, Y., Hu, H., Xu, J., and Mao, G. (2009). Relay technologies for wimax and lte-advanced mobile systems. *IEEE Communications Magazine*, 47(10):100–105.

[110] Zabin, S. M. and Poor, H. V. (1991). Efficient estimation of class a noise parameters via the em algorithm. *IEEE Transactions on Information Theory*, 37(1):60–72.

[111] Zehavi, E. and Wolf, J. K. (1988). On runlength codes. *IEEE Transactions on Information Theory*, 34(1):45–54.

[112] Zhang, L. and Guo, D. (2011). Capacity of gaussian channels with duty cycle and power constraints. In *2011 IEEE International Symposium on Information Theory Proceedings*, pages 513–517.

[113] Zimmermann, M. and Dostert, K. (2000). The low voltage power distribution network as last mile access network - signal propagation and noise scenario in the hf-range. *Arch. f. Elektron. u. Übertragungstech., Int. j. of electron. and commun. 54 (2000) H. 1 S. 13-22.*

[114] Zimmermann, M. and Dostert, K. (2002). Analysis and modeling of impulsive noise in broad-band powerline communications. *IEEE Transactions on Electromagnetic Compatibility*, 44(1):249–258.

[115] Zlatanov, N., Jamali, V., and Schober, R. (2015). On the capacity of the two-hop half-duplex relay channel. In *2015 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7.

[116] Zlatanov, N., Sippel, E., Jamali, V., and Schober, R. (2017). Capacity of the gaussian two-hop full-duplex relay channel with residual self-interference. *IEEE Transactions on Communications*, 65(3):1005–1021.