University of Greenwich

Doctoral Thesis

Time Varying Channel Models for 5G Mobile Communication Systems

Author:
Benshuai Xu

Supervisor:
Prof. Predrag Rapajic

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

in the

Mobile and Wireless Communications
School of Engineering

August 2014

DECLARATION

I, Benshuai Xu, certify that this work titled, 'Time Varying Channel Models for 5G Mobile Communication Systems ', has not been accepted in substance for any degree, and is not concurrently being submitted for any degree other than that of Doctor of Philosophy being studied at the University of Greenwich. I also declare that this work is the result of my own investigations except where otherwise identified by references and that I have not plagiarised the work of others

Signed (student): _____

Signed (supervisor): _____

Date: _____

LIST OF PUBLICATIONS

1. B. Xu, Z. Krusevac, P. Rapajic, and Y. Chen. Maximum mutual information rate for the uniformly symmetric variable noise FSMC without channel state information. In *Proc. IEEE 2012 Int. Symp. on Inform. Theory and its Applicat.*, page 41-45, Oct. 2012.

2. B. Xu, P. Rapajic, and Y. Chen. Deficiency of the Gilbert-Elliot channel in modeling time varying channels. In *Proc. 2013 IEEE Wireless Commun. and Netw. Conf.*, page 2609 - 2614, Apr. 2013.

3. B. Xu, Z. Krusevac, P. Rapajic, and Y. Chen. Maximum Mutual Information Rate of A Time Variable Finite State Markov Channel Model for 5G Mobile Communication Systems. Conference paper ready for submission.

4. B. Xu, Z. Krusevac, P. Rapajic, and Y. Chen. A complementary model of the Gilbert-Elliot model for time variable communication channels. Conference paper ready for submission.

5. B. Xu, Z. Krusevac, P. Rapajic, and Y. Chen. Maximum Mutual Information Rate of A Time Variable Finite State Markov Channel Model for 5G Mobile Communication Systems. Journal paper ready for submission.

6. B. Xu, Z. Krusevac, P. Rapajic, and Y. Chen. A Finite State Markov Model of Time Variable Channels for 5G Mobile Communication Systems. Journal paper ready for submission.

ABSTRACT

Researchers all over the world are looking for ways of continuing the evolution of mobile communication technology to its fifth generation (5G). Providing high data rate information transfer to highly mobile users over time varying communication channels remains a shared obstacle. In this thesis, we contribute to these global efforts by providing further fundamental understanding of time varying channels in 5G mobile communication systems and overcome the obstacle.

First, we reopen the door of research in the field of time varying communication channels. The door has almost been closed before by a well-accepted conclusion related to the types of channels. It was 'proven' that mutual information rate of the uniformly symmetric variable noise finite state Markov channel (USVN-FSMC) was maximized by input signals of maximum information entropy. The result means time varying channels and time invariable channels are identical, regarding information rate maximization over input signal probability distribution. We provide evidence that assumptions for the results are not valid for time varying channels and replace them with more practical ones. We confirm, via input signals of non-uniform independent distribution and first order Markov chain, that the mutual information rate of the USVN-FSMC is maximized by input signals with information redundancy.

Second, we provide a solution which dramatically reduces the waste of communication resources in estimating channel state information of time varying mobile communication channels. The orthodox method in dealing with time varying channels is that, the channel is "cut" to pieces in time domain to look like a sequence of time invariable channels for the purpose of state estimation. By doing this the capacity loss is staggering for n-times higher carrier frequency channels and n-dimensional multiple input and multiple output channels, eliminating almost entirely the capacity gain of these two most promising capacity-increasing techniques for 5G. We define the simplest finite state Markov model for time varying channels to explain the essential difference between information processing of time varying channels and time invariable channels. We prove that the full information capacity of the model can be achieved by the differential type encoding/decoding scheme without employing any conventional channel state estimator.

ACKNOWLEDGEMENTS

I am heartily thankful to my supervisor Professor Predrag Rapajic, from whom I learn how to acquire new knowledge.

I am also heartily thankful to my second supervisor Professor Yifan Chen for his independent and encouraging comments on my work.

I would also like to thank my girlfriend Yinting Ye for her unconditional support to my research.

CONTENTS

FIGURES

Abbreviations

2G	**2** **G**eneration
3G	**2** **G**eneration
4G	**4** **G**eneration
5G	**5** **G**eneration
BSC	**B**inary **S**ymmetric **C**hannel
FSMC	**F**inite **S**tate **M**arkov **C**hannel
USVN-FSMC	**U**niformly **S**ymmetric **V**ariable **N**oise **F**inite **S**tate **M**arkov **C**hannel
RIM	**R**esearch **I**n **M**otion Limited
IOS	**I**phone **O**peration **S**ystem
GSM	**G**lobal **P**ositioning **S**ystem
WiMax	**W**orldwide **I**nteroperability for **M**icrowave **A**ccess
LTE	**L**ong **T**erm **E**volution
WiFi	**W**ireless **F**idelity
IEEE	**I**nstitute of **E**lectrical and **E**lectronics **E**ngineers
ETSI	**E**uropean **T**elecommunications **S**tandards **I**nstitute
IMT	**I**nternational **M**obile **T**elecommunications
EU	**E**uropean **U**nion
METIS	**M**obile and **W**ireless **C**ommunications **E**nablers for the **T**wenty-twenty (2000) **I**nformation **S**ociety
NSN	**N**okia **S**iemens **N**etworks
UK	**U**nited **K**ingdom
5GIC	**5G** **I**nnovation **C**entre
NYU	**N**ew **Y**ork **U**niveristy
MIMO	**M**ultiple **I**nput **M**ultiple **O**utput

AWGN	**A**ddtive **W**hite **G**aussian **N**oise
i.i.d.	**I**ndependent and **I**dentical **D**istributed

Symbols

$\mathcal{Y}$	Channel outputs of the simplified model of digital communication systems
y_n	Channel output sequence at time slot n
y'_n	The nth channel output symbol
y^n	Channel output sequence from time slot 0 time slot n
$\mathcal{G}$	Channel state characteristics of the simplified model of digital communication systems
s_n	Channel state at time slot n
$\mathcal{X}$	Channel inputs of the simplified model of digital communication systems
x_n	Channel input sequence at time slot n
x'_n	The nth channel input symbol
x^n	Channel input sequence from time slot 0 time slot n
$\mathcal{Z}$	Additive white Gaussian noise of the simplified model of digital communication systems
$\mathbf{C}$	Information capacity of the USVN-FSMC
$\mathbf{C}_{CSI}$	Information capacity of the Gilbert-Elliot channel assuming perfect channel state information
$\mathbf{C}_{noCSI}$	Information capacity of the Gilbert-Elliot channel without channel state information
$\mathbf{C}_{memoryless}$	Information capacity of the Gilbert-Elliot channel with the state memory equal to 0
s_n	Channel state at time slot n

τ	The initial channel state probability vector of the USVN-FSMC
$\mathbf{P}$	Channel state transition matrix of the USVN-FSMC
P_{mk}	the (m, k)th entry of the channel state transition matrix
π_n	Estimated channel state distribution of the USVN-FSMC conditioned on past inputs and outputs
ρ_n	Estimated channel state distribution of the USVN-FSMC conditioned on past outputs
p_G	The crossover probability of the good state of the Gilbert-Elliot channel
p_B	The crossover probability of the bad state of the Gilbert-Elliot channel
g	The transition probability from the bad state to the good state of the Gilbert-Elliot channel
b	The transition probability from the good state to the bad state of the Gilbert-Elliot channel
$\mathbf{D}$	A diagonal matrix
I	Mutual information
$I(x_n; y_n)$	Mutual information of x_n and y_n
$I(x_n; y_n\|\pi_n)$	Mutual information of x_n and y_n given the channel state estimation result
R	Mutual information rate of the USNV-FSMC under conventional assumptions
R'	Mutual information rate of the USNV-FSMC under new assumptions
$\Pr(y\|x)$	Channel transition probability
H	Information entropy
f_d	Doppler shift in Hertz
$s_T(t)$	Transmitted signals at time t
$s_R(t)$	Receiver signals at time t
$\hat{s}_R(t)$	Recovered transmitter signals at time t
P_r	The total power of the receiver carrier signal of the

	Clarke's model
$\Upsilon(f)$	The power spectral density of the receiver carrier signal
η_n	The channel variation process of the time varying binary symmetric channel model
v_n	The channel noise process of the time varying binary symmetric channel model
b_n	The encoded input signal of the differential encoder and the differential decoder
d_n	The decoded input signal of the differential encoder and the differential decoder
s_k^e	The channel state of the equivalent model of the time varying binary symmetric channel at time slot n
Q_0^e	The initial state distribution of the equivalent model of the time varying binary symmetric channel
$p_{(m)(n)}$	The channel state law of the equivalent model of the time varying binary symmetric channel
$q_{(m)(n)}$	The transition probability of the equivalent model of the time varying binary symmetric channel
z^e	The error function of of the equivalent model of the time varying binary symmetric channel

Chapter 1

Introduction

1.1 A Non-Trivial Obstacle to 5G Mobile Communication Technology

It is the demand from the smart phones market to provided high data rate information transmission to highly mobile users in the next ten years. The data rate of 5G needs to be 1000 times faster than that of 4G. The objective is as difficult as it sounds. We provide evidences that conventional techniques of increasing the data rate cannot achieve this objective.

1.1.1 The Requirements of 5G: High User Mobility and High Data Rate

The world mobile communication market of smart phones, by the time of writing this thesis, runs in trillions of dollars. The market is predicted to double by year 2016 to 4.7 trillion dollars [1, 2]. The highly developing market brings about severe pressure to mobile communication technology.

The market comprises of three essential technical parts: platforms (smart phones), smart phone applications and mobile network connecting the other two parts. The computation performance of the smart phones has been improved significantly without compromising their mobility in the past ten years [3–5]. The smart phone applications, which were originally developed for information retrieval, have also been driven into much broader categories. These developments increase the data volume transmitted over the mobile network dramatically. Therefore, much higher mobile communication data rates are already in need without compromising any user mobility.

1) The development of mobile devices in size: The computer has kept getting smaller and lighter since the first programmable computer was created in 1936 by Konrad Zuse. The first person computer, which size and computing capacity are designed to individuals, was created in 1962. The concept of portable computer, i.e. laptop, was proposed in 1972 and realised in 1975 by IBM. A computer was decreased in size to a mobile phone in 2007 by Steven Jobs. Since then, the communication between computers and that between mobile phones have united and the data volumn transferred over mobile communication networks explored.

2) The development of smart phones in number: According to a report 2013 [3], global broadband mobile subscriptions have reached around 1.7 billion and are predicted to reach 7 billion in 2018 (approximately 85 percent of the world's population) [3]. The majority of mobile broadband devices are, and will continue to be smart phones. It is also predicted by the number of smart phones will grow to 4.5 billion in 2018 from 1.2 billion at the end of 2012 [3].

3) The development of the smart phones in terms of technology: Smart phones have also undergone a major technical development in the past ten years. From Symbian [6], window mobile [7] and RIM [8] to Android [9] and IOS [10], the smart phone operating systems are getting more and more sophisticated. The latest

version of iphone or samsung galaxy can actually be seen as a high-performance computer.

4) The development of smart phone applications: The technical development of smart phone operation systems leads to a fast expansion of smart phone application markets. According to the apple press information on January 7, 2013, customers have downloaded over 40 billion applications, with nearly 20 billion in 2012 alone [11]. The application store of another popular smart phone operation system, android, hits 25 billion downloads on September 26, 2012 [5]. These applications cover a larger range of categories including mobile phone games, factory automation, GPS and location-based services, banking, order tracking, and ticket purchases. By the year of 2020, 50 billion new things will connect to mobile networks.

These developments on smart phone and smartphone application have explicitly defined the next generation mobile communication network (5G). Compared to 4G (e.g. WiMax, LTE), the data rate has to be increased by a 1000 times without compromising any user mobility [12].

1.1.2 The Limitation of The Communication Theory to Provide High Rate Data Transmission to Highly Mobile Users

The requirements of 5G mobile communication technology have been obvious. But the problem is that the Wi-Fi (IEEE 802.11), 2G (ETSI-GSM), 3G (IMT-2000), and 4G (IMT-Advanced) technologies together have reached the channel limit defined by the Shannon information theory. Performances of these communication technologies depend on the user mobility. 2G, i.e., GSM, provides the lowest data rate for users with the highest mobility. Wi-Fi provides the highest data rate to

users with the lowest mobility. The other technologies **compromise** between data rates and user mobility. However, the smart phone communications require the 5G to increase the data rate without compromising any user mobility.

Research projects on 5G have been started in EU (Projects: METIS and NSN network), UK (Project: 5GIC), South Korea (Project: Giga Korea) and America (NYU). In general, the research projects try to increase the channel information capacity by employing n-dimentional MIMO channels or n-times higher frequency channels. [13–17]. These two most promising techniques for increasing the information capacity are insufficient when users are highly mobile.

Problem for empolying n-times higher frequency channel(1): Theoretically, in time invariant or slow varying channels, employing n-times higher frequency means a n-times channel larger bandwidth and a n-times larger channel capacity.

For fast time varying channels, the assumption is no longer valid. In the existing mobile communication systems, training impulses are transmitted periodically for channel state information. The communication channel is assumed unchanged during time intervals between any two consecutive training impulses. An impulse response at the receiver therefore contains necessary information of channel characteristics to recover information symbols transmitted during the relevant interval [18]. For example, in GSM, about 24 bits out of a 100-bits package are used for training purpose. User motions in mobile communications result in Doppler shifts, which lead to time varying phase shifts to all carrier signal components. These phase shifts is seen as random in the literature and cannot be removed in carrier recovery and have secondary effect on the channel characteristics. The time interval during which the channel is seen as unchanged becomes shorter, therefore more frequent training impulses are required to update the knowledge of these channel characteristics.

The Doppler shift increases linearly with the carrier frequency[18]. By contrast, the variation rate of channel characteristics increases linearly with the carrier frequency[18]. When n-times higher frequency channels are employed, the extra training impulses can cancel out the information capacity improvement made by extra bandwidth resource.

Problem (2): The Shannon information theory suggests that the achievable channel capacity increases logarithmically with the transmit power [19]. By contrast, the information capacity of a MIMO channel increases linearly with the number of transmitter antennas. However, any two antennas in MIMO communication systems have to be separated at least by a half of the wavelength of transmitted signals to show different multi-path fading effects [20]. A denser antenna array therefore requires smaller wavelengths, which only come from higher-frequency signals. Problem (1) arises thereafter.

1.2 Our Understanding and Our Proposed Solutions to Achieve 5G

We provide evidence that the real obstacle to 5G is the lack of adequate understanding of the fundamental time varying communication channel.

The mobile communication channel is a typical time varying communication channel. The research field to time varying communication channels has almost been closed for twenty years due to some incomplete results. This thesis is trying to reopen it and develops a further fundamental understanding for the type of channels. The understanding complements the conventional communication theory and leads to a solution of achieving high rate data transmission to highly mobile users.

1.2.1 Incompleteness of Existing Analysis of Time-Varying Channels

Time varying communication channels are often modeled by Finite State Markov Channel (FSMC), among which the uniformly symmetric, variable noise FSMC (USVN-FSMC) is of particular importance [21–24]. The Gilbert-Elliot channel is the simplest USVN-FSMC. It is shown in [21, 22] that the mutual information rate of FSMC is a continuous function of the input distribution. The mutual information rate of the USVN-FSMC is maximized by the channel input of maximum entropy, i.e., uniform and i.i.d.. This result of mutual information rate maximization coincidences with that of time invariant channels. For the reason, it is well accepted that time invariable channels and time varying channels are not fundamentally different. A capacity achieving decoder, named by the decision-feedback decoder, is proposed as a proof for the results in [21, 22]. The decision-feedback decoder is comprised of a recursive state estimator and a conventional maximum-likelihood decoder.

In chapter 3, we show that results in [21, 22] are based in assumptions, which are valid for time invariant channels and slow time varying channels. The assumption is: maximizing the mutual information rate in each state can maximize the mutual information rate of the whole channel. For the reason, the analysis and the decision-feedback decoder employ input signal of maximum information entropy for each channel state.

We show that the local maximization does not lead to global maximization. According to our simulation results, the information redundancy in input signals from the previous channel time slot can decrease the uncertainty of the state distribution in the current channel time slot. Higher maximum mutual information is therefore achievable in the current channel time slot.

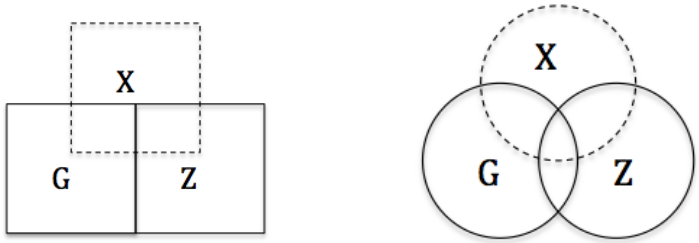
For the reason, in the first channel time slot where the state distribution is assumed, we can choose not to maximize the mutual information. Instead, we employ input signals with information redundancy. The information redundancy decreases the uncertainty of the state distribution of the second channel time slot, in which a higher maximum mutual information is achievable. In the second channel time slot, we again choose not maximize the mutual information. Information redundancy in input signals is used again to decrease the uncertainty of the state distribution in the third time slot. By repeating the process, there is always certain amount of information resource **invested** in channel state estimation. The investment causes some information loss in the beginning, we prove that it pays off in the long term: the mutual information rate of the USVN-FSMC with the decision-feedback decoder employed is maximized by input signals with information redundancy.

1.2.2 The Information Capacity Analysis of the Time Variable Binary Symmetric Channel

The second objective of this thesis is to propose a new simplest model for time varying communication channel. An optimum decoding scheme, which achieve the full information capacity of this new model, is also found. The ultimate purpose of this thesis is to provide a solution to achieve high rate data transmission over time varying mobile communication channels. We provide evidence that the current simplest USVN-FSMC model (i.e. the Gilbert-Elliot channel)of the time varying mobile communication channel is incomplete. Current decoding schemes for time varying communication channels are originally designed for time invariant channels. We can show that it is very difficult to achieve the full information capacity of time varying channels with these decoding schemes.

In mobile communication systems, user motions bring about two new challenges for reliable signal transfer: time varying signal amplitude fading and time varying carrier phase shift[18]. The signal amplitude fading results from the geometric change of the communication environment, while the carrier phase shift is due to the Doppler effect. The Gilbert-Elliot channel model captures only the amplitude fading. However, the channel variation in mobile communications is mainly caused by the Doppler phase shift than on the amplitude variation [25, 26]. In chapter 5 of this thesis, we propose the time-variable-BSC model for the Doppler phase shift, which complements the Gilbert-Elliot channel. We confirm firstly the accuracy of the time-variable-BSC by detailing the mapping between parameters of the time-variable-BSC and the relevant factors of the physical time varying channel. We then prove that the differential encoder and differential decoder achieve the information capacity for the time-variable-BSC.

The differential encoding/decoding scheme does not involve any conventional channel state estimation scheme. It confirms the fundamental difference between information processing of time varying channels and time invariant channels. The Shannon theory assumed that the channel state information had to be known by the receiver before any information transmission can be performed[19]. In what follows, research of channel state estimation and research of signal decoding are done mostly separately, and have developed into two independent research fields, leaded by Kalman filter [27] and Viterbi algorithm [28], respectively. However, although channel state information and additive white Gaussian noise are independent, the channel state estimation and extracting signal in the presence of the Gaussian noise are actually not. Synchronization between these two actions is always required. The relationships are like what is shown in 1.1(b). In some other types of wireless communication where user motions are also highlighted, synchronization has become a shared problem[29–31]. We provide evidence that



(a) (b)

FIGURE 1.1: (a) The relationship between detecting channel states and extracting signals in the presense of additive white Gaussian noise in previous analysis ; (b) The relationship between detecting channel states and extracting signals in the presence of additive white Gaussian noise in this thesis.

the perfect synchronizer is a sufficient condition to achieve the information capacity of time varying communication channels. This can only be done via combining the channel state estimation and signal detection. The differential encoder and the differential decoder are doing just that and achieve the information capacity of the time-variable-BSC.

The report is organized in the following way:

1. In chapter 2, we separate FSMCs of time invariant communication channels and FSMCs of time varying communication channels.

2. In chapter 3, we revisit the existing analysis of mutual information rate of USVN-FSMCs in the literature and provide our analysis of the channel.

3. In chapter 4, we provide a further analysis of the mutual information rate of the USVN-FSMC when the input signal has Markov memory.

4. In chapter 5, we introduce the time-variable-BSC for the mobile communication channel and prove the the information capacity of the channel can be achieved by the differential encoder and differential decoder.

5. Chapter 6 is the conclusion and the future work.

Chapter 2

The Finite State Markov Channels Model of Time Varying Communication Channels

In the literature, time varying channels are defined conceptually by channels whose channel states information are time varying. Time invariable channels are defined by channels whose channel states information are time invariable. Analysis in this thesis requires a clear separation between FSMCs of these two types of channels. This is not easy based on the simple definition of time varying channels and time invariable channels [32–45]. In this chapter, we propose definitions for three types of memory in mobile communication systems. They are called by Markov signal memory, multi-path signal memory and channel variation memory, respectively. We also introduce FSMCs for each memory. FSMCs of time varying channels in this thesis means FSMCs with channel variation memory. FSMCs of time invariable channels means FSMCs with Markov signal memory or multi-path signal memory only.

2.1 The Simplified Wireless Communication System Model

The analysis throughout this thesis is based on the following simplified model of digital communication system,

$$\mathcal{Y} = \mathcal{G}\mathcal{X} + \mathcal{Z}, \tag{2.1}$$

$\mathcal{Y}$ Channel output; $\mathcal{Y}$ could be a variable or a vector. Elements of which are assumed to be numbers in this thesis.

$\mathcal{G}$ Channel state characteristics or channel state information; $\mathcal{G}$ could be a variable, a vector or a matrix. Elements of which are assumed to be numbers in this thesis.

$\mathcal{X}$ Channel input; $\mathcal{X}$ could be a variable or a vector. Elements of which are assumed to be numbers in this thesis.

$\mathcal{Z}$ Additive white Gaussian noise (AWGN). $\mathcal{Z}$ could be a variable or a vector. Elements of which are assumed to be numbers in this thesis.

2.2 Three Types of Memory in The Digital Communication System

In this section, we introduce the Markov signal memory, the multi-path signal memory and the channel variation memory in the system model of (2.1). The channel with each kind of memory can be modeled by a kind of FSMCs. The relationships between three types of FSMCs are also clarified.

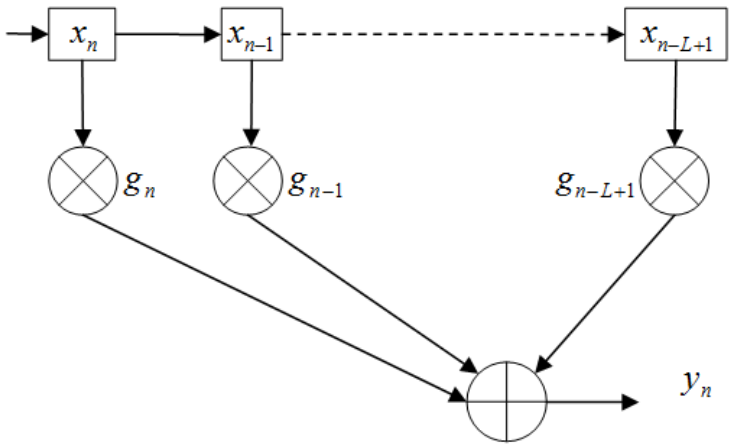


FIGURE 2.1: A discrete model for time invariant communication systems with multi-path delays.

2.2.1 The Markov Signal Memory

The Markov signal memory in this thesis means the mutual dependence of the original transmitted signals. One of the popular example is the Markov channel input[46]. Refering to (2.1),

$$\mathcal{Y} = \mathcal{G}\mathcal{X} + \mathcal{Z}, \tag{2.2}$$

where $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{G}$ and $\mathcal{Z}$ are assumed to be scalers. An L order Markov signal memory means the input sequence x_n in the nth time slot is dependent on the $(x_{n-1},...,x_{n-L})$.

2.3 The Multi-Path Signal Memory

The multi-path channel memory is also called by inter-symbol interference. It refers to the dependence between received signals, which caused by multi-path delays [32, 33, 36, 37, 47–49]. A discrete model of the multi-path communication

system is shown in Fig. 4.1(a). Referring to the system model of (2.1)

$$\mathcal{Y} = \mathcal{G}\mathcal{X} + \mathcal{Z}, \tag{2.3}$$

where $\mathcal{G}$ and $\mathcal{X}$ are assumed to be vectors. $\mathcal{Y}$ and $\mathcal{Z}$ are assumed to be scalars. The multi-path memory is first carried by $\mathcal{Y}$. In order to separate the multi-path signal memory and the Markov signal memory, we express (2.3) by

$$\dot{y}_n = \sum_{l=1}^{L} \dot{g}_l \dot{x}_{n-l+1} + \dot{z}_n, \tag{2.4}$$

where the notations, $\dot{y}_n$, $\dot{x}_n$, $\dot{g}_l$ and $\dot{z}_n$ are only used in this chapter. $\dot{y}_n$ is the the received symbol in the n^{th} time slot. $\dot{y}_n$ depends on the transmitted symbol sequence $\dot{x}_n$ in the nth time slot and some input sequences in previous time slots $(\dot{x}_{n-1}, ..., \dot{x}_{n-L+1})$. All of these transmitted symbols have multi-path components arriving at the receiver in the n time slot; L is the extent of the multi-path interference, and it is also named by the memory order; $\dot{g}_l$ is the lth element of the vector $\mathcal{G}$. It should be noted that perfect knowledge of $\mathcal{G}$ is assumed at the receiver since it is unchanged. $\dot{y}_n$ is dependent on $(\dot{y}_{n-1}, ..., \dot{y}_{n-L+1})$ because any two of them have at least one input symbol in common.

The multi-path channel memory and the signal memory are usually analyzed in the same way. The mutual dependence between the inputs symbols can affect the received signals in the same ways as the multi-path delay does. For example, if the transmitted symbol at time slot n is dependent of those at previous $L-1$ time slots, $\dot{y}_n$ is dependent on $(\dot{y}_{n-1}, ..., \dot{y}_{n-L+1})$ even though it experiences no multi-path delay during transmission.

The multi-path signal memory or Markov signal memory is not fundamentally different from memoryless channels. The information capacity of FSMC of these two types of memory can be achieved by decoders designed for

memoryless channels, such as the maximum-likelihood decoder. The channel with the multi-path signal memory in (2.4) can be modeled by an L-order FSMC, of which the state $s_n = (\dot{x}_n, \dot{x}_{n-1}, ..., \dot{x}_{n-L+1})$[46]. The state transition probability can be expressed by

$$\Pr\left(s_{n+1} \mid s_n\right) = \Pr\left(\dot{x}_{n+1} \mid \dot{x}_n, ..., \dot{x}_{n+1-L}\right). \tag{2.5}$$

The finite state Markov chain can be transformed equivalently into a sequence of discrete memoryless states without losing any information, where $s_n = (\dot{x}_n, \dot{x}_{n-1}, ..., \dot{x}_{n-L+1})$. A simple example of the transformation is shown as follows.

Assuming that the memory order $L = 1$, an input symbol sequence $\dot{x}^n$ can be expressed by

$$\dot{x}^n = (\dot{x}_1, \dot{x}_2, \dot{x}_3, \dot{x}_4, \dot{x}_5, \dot{x}_6, ..., \dot{x}_n). \tag{2.6}$$

When the FSMC of (2.6) has states as follows, $s_1 = \dot{x}_1$, $s_2 = \dot{x}_2$, ..., $s_n = \dot{x}_n$. This Markov chain is equivalent to a sequence of memoryless states, $s_1 = \dot{x}_1\dot{x}_2$, $s_2 = \dot{x}_2\dot{x}_3$, ..., $s_{n-1} = \dot{x}_{n-1}\dot{x}_n$. Refering to Fig. 2.1, the communication system can be understood alternatively that the transmitted signals do not experience multi-path delay, and all input sequences except for those from the first time slot are transmitted repeatedly $L + 1$ times with different signal amplitudes. In the previous example of (2.6), $L = 1$ and each input sequence is repeated 2 times:

At time slot 1, the signal $\dot{x}_1$ and $\dot{x}_2$ are sent.

At time slot 2, the signal $\dot{x}_2$ and $\dot{x}_3$ are sent.

At time slot 3, the signal $\dot{x}_3$ and $\dot{x}_4$ are sent.

At time slot $n - 1$, the signal $\dot{x}_{n-1}$ and $\dot{x}_n$ are sent.

Therefore, conventional decoders for memoryless channel, such as the maximum-likelihood decoder, can perform optimum signal detection for FSMCs with the multi-path signal memory. According to the Shannon theory, maximum mutual information rate of discrete memoryless channels is achieved by uniform, independent and identically distributed (i.i.d.) inputs[50].

2.4 The Channel Variation Memory

The Channel variation memory is mutual dependence between receiver symbols, and the dependence is caused by the variation of the channel state information. Referring to the system model

$$\mathcal{Y} = \mathcal{G}\mathcal{X} + \mathcal{Z}. \tag{2.7}$$

where $\mathcal{Y}$, $\mathcal{G}$, $\mathcal{X}$ and $\mathcal{Z}$ are all scalars We assume that transmitted signals do not experience multi-path delay. We also assume the input signals are i.i.d.. The channel variation memory is firstly carried by $\mathcal{G}$, of which perfect knowledge is no longer assumed at the receiver. In order to clarify the relationship between the multi-path channel memory and the channel variation memory, we express (2.7) by,

$$\dot{y}_n = \ddot{g}_n \dot{x}_n + \dot{z}_n, \tag{2.8}$$

where notations $\dot{y}_n$, $\ddot{g}_n$, $\dot{x}_n$ are used only in this chapter. $\dot{y}_n$ is the output sequence in time slot n and it depends only on the input sequence sent in the same time slot; $\ddot{g}_n$ is the time varying channel state information of time slot n. (2.8) can be also modeled by a FSMC. Assuming that the memory order is L, the channel

state is $s_n = (\ddot{g}_n, \ddot{g}_{n-1}, ..., \ddot{g}_{n-L+1})$, with state transition probability,

$$\Pr\left(s_{n+1} \mid s_n\right) = \Pr\left(\ddot{g}_{n+1} \mid \ddot{g}_n, ..., \ddot{g}_{n+1-L}\right). \tag{2.9}$$

We introduce in the following the relationship between the channel variation memory and the multi-path signal memory. Assuming perfect synchronization between the detection of $\ddot{g}_n$ and $\dot{x}_n$, (2.8) can be expressed by

$$\begin{aligned} \dot{y}_n &= \ddot{g}_n \dot{x}_n + \dot{z}_n \\ &= \dot{g}_n \ddot{x}_n + \dot{z}_n, \end{aligned} \tag{2.10}$$

where $\dot{g}_n$ is unchanged and assumed to be known by the receiver, and $\ddot{x}_n = a_n \dot{x}_n$, of which a_n depends on n; and $\ddot{g}_n = \dot{g}_n \cdot a_n$. (2.10) and (2.4) are equivalent and the channel varying memory can be modeled by the FSMC for multi-path channel memory. This conclusion is important. It shows implicitly that this synchronization between channel state estimation and signal detection is a sufficient condition to achieve the information capacity of the FSMC with channel variation memory. However, to find the perfect synchronization scheme is very difficult. There are always delays between the channel state estimation and signal detection in the presence of noise.

The mobile communication channel is a typical channel with channel variation memory. The lack of adequate understanding of this types FSMCs is one of the main reasons for the difficulty of providing high rate data transmission to users in motion. This thesis therefore focus on reviewing the analysis of the existing FSMCs with channel variation memory. The FSMC is called by the USVN-FSMC. We show the incompleteness of the analysis and provides complementing analysis.

Chapter 3

Mutual Information Rate Analysis of Experiencing Finite State Markov Channels

The chapter focuses on a class of USVN-FSMC where the channel state information is unknown to the transmitter. Results in the literature for the USVN-FSMC show that time variable channels and time invariant channels are the same in terms of maximization of the mutual information rate over channel input probability distribution. We show that the analysis is based on the assumptions for time invariant channels. We propose more practical assumptions for time variable channels. Our analysis of the USVN-FSMC based on the new assumption confirms the fundamental difference between time variable channels and time invariant channels.

3.1 Contrasting The Conventional Assumptions and Our New Assumption for Uniformly Symmetric Variable Noise Finite State Markov Channels

In this section, we show the incompleteness of the conventional assumption for the USVN-FSMC in [21, 22] and propose more practical assumptions in this thesis.

1. Conventional assumption: maximizing the mutual information of each time slot of the channel will maximize the mutual information of whole channel [21, 22];

2. New assumption: maximizing the mutual information of each time slot of the channel might not maximize the mutual information of the whole channel (our assumption).

3.1.1 Uniformly Symmetric Variable Noise Finite State Markov Channels

The channel model considered in this thesis belongs to the class of USVN-FSMCs[22]. The finite channel state space $\mathcal{C} = \{c_1, c_2, ..., c_K\}$ corresponds to K different discrete memoryless channels, respectively. The states have common finite discrete input and output alphabet [22]. The discrete input sequence and output sequence of the channel at time slot n are denoted by x_n and y_n, respectively. The channel state at time slot n is denoted by s_n. The state transition matrix is denoted by $\mathbf{P}$, of which the (m, k)th entry is the probability of transition from state c_m to c_k,

$$P_{mk} = \Pr\left(s_{n+1} = c_k | s_n = c_m\right), \tag{3.1}$$

for $k, m \in 1, ..., K$. The initial state probability vector is denoted by $\boldsymbol{\tau}$ with the kth element being

$$\boldsymbol{\tau}(k) = \Pr(s_0 = c_k). \tag{3.2}$$

Throughout this thesis, we use the following notation:

$$r^n \triangleq (r_1, ..., r_n), \tag{3.3}$$

for $r = x, y$, or s. The initial channel state probability vector $\boldsymbol{\tau}$and the channel state transition matrix $\mathbf{P}$ are assumed to be independent of the channel input. We denote the conditional channel state distribution by two K-dimensional random vectors $\boldsymbol{\pi_n}$ and $\boldsymbol{\rho_n}$. $\boldsymbol{\pi_n}$ is the channel state distribution conditioned on past inputs and outputs, while $\boldsymbol{\rho_n}$ is the channel state distribution conditioned on past outputs only. The kth elements of $\boldsymbol{\pi_n}$ and $\boldsymbol{\rho_n}$ are denoted by

$$\boldsymbol{\pi_n}(k) = \Pr\left(s_n = c_k | x^{n-1}, y^{n-1}\right), \tag{3.4}$$

and

$$\boldsymbol{\rho_n}(k) = \Pr\left(s_n = c_k | y^{n-1}\right), \tag{3.5}$$

respectively.

The numerical example of USVN-FSMC used in the simulation in this chapter is an example of Gilbert-Elliot channel model. The Gilbert-Elliot channel is the two-state USVN-FSMC [21]. The crossover probabilities of "good" and "bad" state of the Gilbert-Elliot channel are denoted by p_G $(0 \leq p_G \leq 0.5)$ and p_B $(0 \leq p_B \leq 0.5)$, respectively, where $p_G < p_B$. The channel state transition probabilities, from the bad state to the good state and from the good state to the bad state, are given by g and b, respectively. The channel memory μ is defined by $\mu = 1 - b - g$.

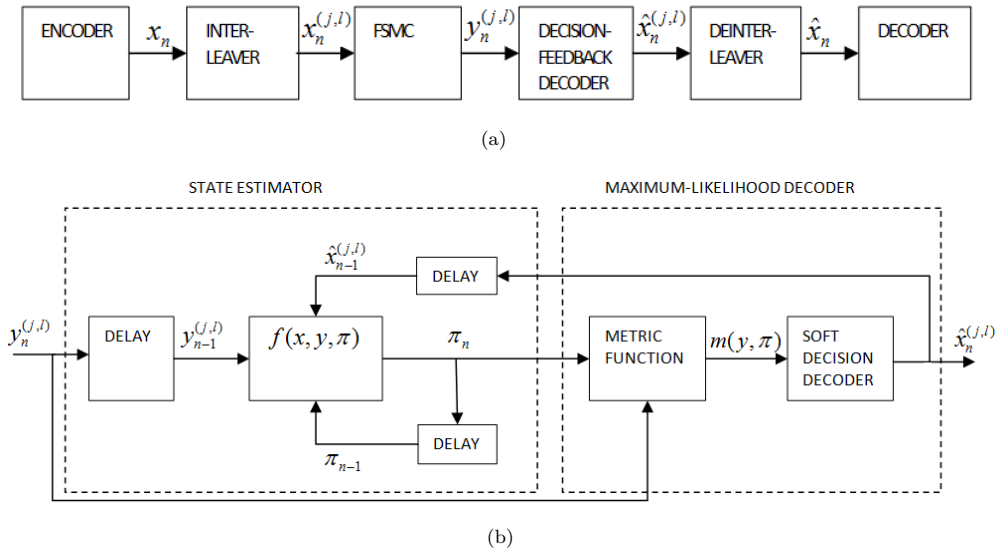


FIGURE 3.1: (a) The system model; (b) the decision-feedback decoder

3.1.2 Decision-Feedback Decoder with The Conventional Assumption

The system model of the USVN-FSMC and the decision feedback decoder are shown in Fig. 3.1(a) and Fig. 3.1(b), respectively[22]. The system is composed of a conventional (block or convolutional) encoder for memoryless channels, a block interleaver, an FSMC, the decision-feedback decoder, and a deinterleaver. The interleaver works as follows: The output of the encoder is stored row by row in a $J \times L$ interleaver, and transmitted over the channel column by column. The deinterleaver performs the reverse operation [22]. Because the effect of the initial channel state dies away, the received symbols within any row of the deinterleaver become mutually independent as J becomes infinite [22]. Each interval, during which a row of output signal is processed, is a channel time slot. The decision feedback decoder includes two parts: the state estimator and the ML decoder. The state estimator uses the following recursive relationship to estimate the channel

state distribution conditioned on the past inputs and outputs[22].

$$\boldsymbol{\pi_{n+1}} = \frac{\boldsymbol{\pi_n}\mathbf{D}\left(x_n, y_n\right)\mathbf{P}}{\boldsymbol{\pi_n}\mathbf{D}\left(x_n, y_n\right)\underline{\mathbf{1}}} \triangleq f\left(x_n, y_n, \boldsymbol{\pi_n}\right) \tag{3.6}$$

where $\mathbf{D}\left(x_n, y_n\right)$ is a $K \times K$ diagonal matrix with the kth diagonal term, $\Pr(y_n = 0|x_n = 0, s_n = c_k)$, and $\underline{\mathbf{1}} = (1, ..., 1)^T$ is a K-dimensional vector. The input to the ML decoder is the channel output y_n, and the estimated state distribution $\boldsymbol{\pi_n}$. Its output is the detected channel input, $\hat{x}_n$, which maximizes $\log \Pr\left(y_n, \boldsymbol{\pi_n}|x_n\right)$. The estimation of the channel input, $\hat{x}_n$, is then fed into the state estimator for the next channel state estimation. For independent input signals, there is a similar recursive estimation formula conditioned on the past output only,

$$\boldsymbol{\rho_{n+1}} = \frac{\boldsymbol{\rho_n}\mathbf{B}\left(y_n\right)\mathbf{P}}{\boldsymbol{\rho_n}\mathbf{B}\left(y_n\right)\underline{\mathbf{1}}} \triangleq \hat{f}\left(y_n, \boldsymbol{\rho_n}\right), \tag{3.7}$$

where $\mathbf{B}\left(y_n\right)$ is a $K \times K$ diagonal matrix with the kth diagonal term $\Pr\left(y_n = 0|s_n = c_k\right)$ and $\mathbf{P}$ is the state transition matrix of the USVN-FSMC[22]. In appendix A, we prove the equivalence between (3.6) and (3.7). In most parts of this thesis, we use (3.7) as the state estimation formula.

Lemma 3.1. *Assuming perfect channel input information, the recursive relationships in(3.6) and (3.7) of the recursive state estimator of the decision-feedback decoder are equivalent.*

Proof: See appendix A for proof.

3.1.2.1 Conventional Assumptions

The decision-feedback decoder with input signals of maximum entropy is capacity achieving for the USVN-FSMC under the following assumptions. The reason of the optimality can be found in the recursive process of the decoder.

- The initial state distribution is assumed.
- The state distribution of other channel time slot can be estimated accurately given the value of past inputs and past outputs.
- Maximum mutual information of any channel time slot is decided by the corresponding state distribution.
- Maximizing the mutual information of each time slot of the channel will maximize the mutual information rate of the whole channel. [21, 22].

3.1.2.2 The Recursive Process of The Decision-Feedback Decoder under The Conventional Assumption

Suppose that a sufficiently large interleaver of size $J \times L$ is implemented. The (possibly) coded signals are arranged row-by-row and transmitted column-by-column. We can decode the received signal row-by-row recursively. The recursion process is as follows:

1. For the first row, $j = 1$. We can treat the first row as L copies of independent channels with π_1 as the state distribution, $\pi_1 = \pi_0 \cdot \mathbf{P}$. The independence holds due to that J is large. The channel transition probability is then given by $\Pr(y|x) = \sum_{k=0,1} p(y|x, s_1)\pi_1(k)$. Maximum mutual information is $\max_{\Pr(x_1)} I(x_1; y_1|\pi_1)$. Obviously, input signals with maximum entropy maximize the mutual information, therefore the achieved mutual information of the recursion is $R_1 = \max_{\Pr(x_1)} I(x_1; y_1|\pi_1)$. Assuming a large enough L and large enough symbol intervals, error probability of $\Pr(\hat{x}_1 = x_1)$ can be made arbitrarily low. Hence we can estimate $\pi_2(k) = \Pr(S_2 = k|x^1, y^1, \pi_1)$ accurately using (3.6).
2. For the second row, $j = 2$. We can treat the second row as L copies of independent channels with π_2 as the state distribution. We can estimate

$\pi_2(k) = \Pr(S_2 = k|x^1, y^1, \pi_1)$ accurately using (3.6). The independence holds due to that J is large. The channel transition probability is then given by $\Pr(y|x) = \sum_{k=0,1} p(y|x, s_2)\pi_2(k)$. Maximum mutual information is $\max_{\Pr(x_2)} I(x_2; y_2|\pi_2)$. Obviously, input signals with maximum entropy maximizes the mutual information, therefore the achieved mutual information of the recursion is $R_2 = \max_{\Pr(x_2)} I(x_2; y_2|\pi_2)$. Assuming a large enough L and large enough symbol intervals, error probability of $\Pr(\hat{x}_2 = x_2)$ can be made arbitrarily low. Hence we can estimate $\pi_3(k) = \Pr(S_3 = k|x^2, y^2, \pi_2)$ accurately using (3.6).

3. For the row, $j = j + 1$. We can treat the second row as L copies of independent channels with π_{j+1} as the state distribution. The independence holds due to that J is large. The channel transition probability is then given by $\Pr(y|x) = \sum_{k=0,1} p(y|x, s_{j+1})\pi_{j+1}(k)$. Maximum mutual information is $\max_{\Pr(x_{j+1})} I(x_{j+1}; y_{j+1}|\pi_{j+1})$. Obviously, input signals with maximum entropy maximizes the mutual information, therefore the achieved mutual information of the recursion is $R_{j+1} = \max_{\Pr(x_{j+1})} I(x_{j+1}; y_{j+1}|\pi_{j+1})$. Assuming a large enough L and large enough symbol intervals, error probability of $\Pr(\hat{x}_{j+1} = x_{j+1})$ can be made arbitrarily low.

4. Step 3 repeats and ends at $j = J$.

The maximum mutual information rate achieved by the decision-feedback decoder with input signals of maximum entropy is

$$R = \lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{J} R_j = \lim_{J\to\infty} \frac{j=1}{J} \sum_{1}^{J} \max I(X_j; Y_j|\pi_j). \tag{3.8}$$

From the recursive process, the mutual information of each channel time slot is maximised. Based on assumption in [21, 22] that mutual information maximization of each channel time slot is independent, it is not difficult to prove that R converge to the information capacity of the USVN-FSMC.

3.1.3 Decision-Feedback Decoder with The New Assumption

In this subsection, we show the hidden problem of the conventional assumptions in [21, 22] and propose a more practical assumption. Under this new assumption, the maximum mutual information rate of the USVN-FSMC with the decision-feedback decoder employed is achieved by input signals with information redundancy.

3.1.3.1 Incompleteness of The Conventional Assumption

In this subsection, we prove via numerical examples that the estimation result of state distribution of any channel time slot depends on the input signal distribution of the previous channel time slot. The simulation method is detailed in appendix B. The estimation results are shown in Fig. 3.2(a), 3.2(b), 3.2(c) and 3.2(d). By this, we prove the dependency between the state estimation result and the information redundancy in input signals.

Fig. 3.2(a) shows the tracking ability of the decision-feedback decoder as the channel input approaches maximum entropy, $\Pr(x_n = 0) = 0.5$. Apparently, the estimator fails to indicate the channel state. This is because the estimator determines the channel state by measuring how much the channel input distribution is modified when it is filtered by the channel. However, for the USVN-FSMC, the maximum-entropy channel input distribution will be modified by the same degree, regardless which state the channel is in[22].

Fig. 3.2(b), 3.2(c) and 3.2(d) show the tracking ability of the decision-feedback decoder with channel input probabilities, $\Pr(x_n = 0) = 0.6$, $\Pr(x_n = 0) = 0.7$ and $\Pr(x_n = 0) = 1$, respectively. In general, the channel states can be tracked successfully for all these input distributions. The best result is obtained with $\Pr(x_n = 0) = 1$, and the performance deteriorates as the probability decreases.

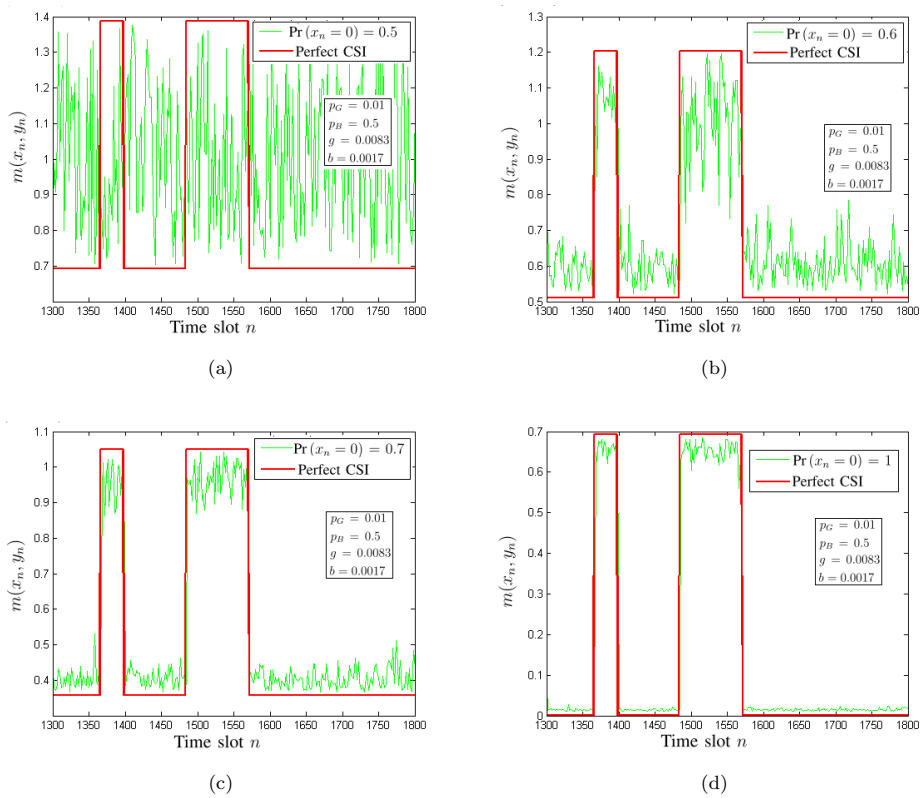


FIGURE 3.2: Tracking ability of the decision-feedback decoder with different channel input probabilities: (a) $\Pr(x_n = 0) = 0.5$, (b) $\Pr(x_n = 0) = 0.6$, (c) $\Pr(x_n = 0) = 0.7$ and (d) $\Pr(x_n = 0) = 1$.

Therefore, we can have the following conclusions:

- As the channel input signals of the previous time slot approach maximum entropy, the channel state of the current time slot of the USVN-FMSC cannot be tracked reliably by the decision-feedback decoder;
- With the channel input signals of the previous time slot with information redundancy, the channel state of the current time slot of the USVN-FMSC can be tracked reliably by the decision-feedback decoder;
- The more redundancy there is in the channel input, the more accurate the estimation is.

3.1.3.2 New Assumptions

The simulation results show the hidden problems of the conventional assumptions. It is clear that maximizing the mutual information of any two consecutive channel time slots depends on each other. Under the case that the mutual information rate of the previous time slot is maximized by input signals with maximum entropy, the estimator in the decision-feedback decoder cannot decrease the uncertainty of state distributions of the current time slot. However, under the case that information redundancy are included in the input signals of the previous time slot, the uncertainty of the channel state distribution of the current time slot will be decreased. It is possible to achieve a higher maximum mutual information of the current channel time slot. Based on the understanding, we propose new assumptions for the USVN-FSMC and the decision-feedback decoder.

- The initial state distribution is assumed (identical).
- The state distribution of other channel time slot can be estimated accurately given the value of past inputs and past outputs (identical).
- Maximum mutual information of any channel time slot is decided by the corresponding state distribution (identical).
- Maximizing the mutual information of each time slot of the channel might not maximize the mutual information of the whole channel(different).

3.1.3.3 The New Recursive Process of The Decision-Feedback Decoder under The New Assumption

Under the new assumption, maximizing the mutual information of each channel time slot independently does not necessary maximize the mutual information rate of the whole channel. Investing certain amount of information resource from input

signals in state estimation, which causes some information loss in the beginning, leads to higher mutual information in future time slots and a higher mutual information rate of the whole channel.

Suppose that a sufficiently large interleaver of size $J \times L$ is implemented. The (possibly) coded signals are arranged row-by-row and transmitted column-by-column. We can decode the received signal row-by-row recursively. The recursion process is as follows:

1. For the first row, $j = 1$. We can treat the first row as L copies of independent channels with π_1 as the state distribution, $\pi_1 = \pi_0 \cdot \mathbf{P}$. The independence holds due to that J is large. The channel transition error probability is then given by $\Pr(y|x) = \sum_{k=0,1} p(y|x, s_1)\pi_1(k)$. Maximum mutual information is $\max_{\Pr(x_1)} I(x_1; y_1|\pi_1)$. Obviously, input signals with maximum entropy maximizes the mutual information. **We choose not to maximize the mutual information rate by employing input signals with information redundancy.** The achieved mutual information of this recursion will be comparatively lower. We denote it here by R'_1, and $R'_1 < R_1$. Assuming a large enough L and large enough symbol intervals, probability of $\Pr(\hat{x}_1 = x_1)$ can be made arbitrarily low.

2. For the second row, $j = 2$. We can estimate the state distribution of the next time slot $\pi'_2(k) = \Pr(S_2 = k|x^1, y^1, \pi_1)$ using (3.6). Because of the redundancy in x_1, π'_2 is with less uncertainty than π_2. We can treat the second row as L copies of independent channels with π'_2 as the state distribution. The independence holds due to that J is large. The channel transition error probability is then given by $\Pr(y|x) = \sum_{k=0,1} p(y|x, s_2)\pi'_2(k)$. Maximum mutual information is $\max_{\Pr(x_2)} I(x_2; y_2|\pi'_2)$. Because π'_2 is with less uncertainty than π_2, $\max_{\Pr(x_2)} I(x_2; y_2|\pi'_2) > \max_{\Pr(x_2)} I(x_2; y_2|\pi_2)$ Obviously, input signals with maximum entropy maximizes the mutual information.

We choose not to maximize the mutual information rate by employing input signals with information redundancy. The achieved mutual information rate denoted by $R'_2 < \max_{\Pr(x_2)} I(x_2; y_2|\pi'_2)$. It should be noted here R'_2 is not necessary smaller than $R_2 = \max_{\Pr(x_2)} I(x_2; y_2|\pi_2)$. Assuming a large enough L and large enought symbol intervals, probability of $\Pr(\hat{x}_2 = x_2)$ can be made arbitrarily low.

3. For row, $j = j + 1$. We can estimate the state distribution of the next time slot $\pi'_{j+1}(k) = \Pr(S_{j+1} = k|x^1, y^1, \pi_j)$ using (3.6). Because of the redundancy in x_j, π'_{j+1} has less uncertainty than π_{j+1}. We can treat the row as L copies of independent channels with π'_{j+1} as the state distribution. The independence holds due to that J is large. The channel transition error probability is then given by $\Pr(y|x) = \sum_{k=0,1} p(y|x, s_{j+1})\pi'_{j+1}(k)$. Maximum mutual information is $\max_{\Pr(x_{j+1})} I(x_{j+1}; y_{j+1}|\pi'_{j+1})$. Because π'_{j+1} is with less uncertainty than π_{j+1}, $\max_{\Pr(x_{j+1})} I(x_{j+1}; y_{j+1}|\pi'_{j+1}) > \max_{\Pr(x_{j+1})} I(x_{j+1}; y_{j+1}|\pi_{j+1})$ Obviously, input signals with maximum entropy maximizes the mutual information. **We choose not to maximize the mutual information rate by employing input signals with information redundancy.** The achieved mutual information rate denoted by $R'_{j+1} < \max_{\Pr(x_{j+1})} I(x_{j+1}; y_{j+1}|\pi'_{j+1})$. It should be noted here R'_{j+1} is not necessary smaller than $R_{j+1} = \max_{\Pr(x_{j+1})} I(x_{j+1}; y_{j+1}|\pi_{j+1})$. Assuming a large enough L and a large enough symbol interval, probability of $\Pr(\hat{x}_{j+1} = x_{j+1})$ can be made arbitrarily low.

4. Step 3 repeats and ends when $j = J$.

The overall achieved mutual information rate is $R' = \lim_{J\to\infty} \frac{1}{J} \sum_{j=1}^{\infty} R'_j$. We prove in the next section of our thesis that, by employing input signals with certain amount of information redundancy $R' > R$.

3.2 Mutual Information Rate Analysis of The Channels with Independent Input Signals Under The New Assumption

In the previous section, the incompleteness of the conventional assumption in [21, 22] for the USVN-FSMC is confirmed and more practical assumptions are proposed. In this section, we provide analysis of the mutual information rate of the USVN-FSMC under the new assumptions and prove that the mutual information rate is maximized by input signals with information redundancy. All the discussions are based on the system model in Fig. 3.1(a), where decision-feedback decoder is employed.

3.2.1 Mutual Information Rate

In this subsection, we show via numerical examples the difference between the mutual information rate of the following three types of USVN-FSMCs in terms of the channel state memory, shown in Fig. 3.3.

1. The USVN-FSMC with infinite channel state memory. This type of USVN-FSMC is also called time invariant USVN-FSMC. Perfect channel state information is usually assumed at the receiver for this type of USVN-FSMC. The mutual information rate can be expressed as $R = \mathcal{I}(Y; X \mid S) = \sum_{k=0}^{k=K} \Pr(c_k)\mathcal{I}(Y; X \mid c_k)$, where Y and X are the channel output and input, respectively, and S is the perfect channel state information. Its mutual information rate is maximized by input signals with maximum entropy. The information capacity of an example Gilbert-Elliot channel assuming perfect channel state information at the receiver is plotted in Fig. 3.3, denoted by C_{CSI}.

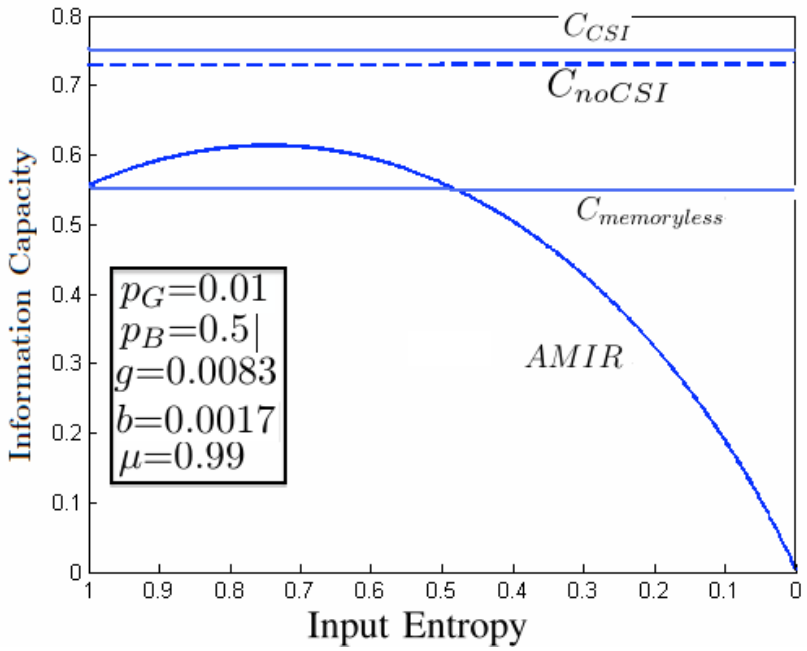


FIGURE 3.3:

1. C_{CSI} is the information capacity of the Gilbert-Elliot channel assuming perfect knowledge of the channel state information ;
2. C_{noCSI} denotes the information capacity of the Gilbert-Elliot without channel state information , of which the value is equal to that in (a) when $n = 5000$;
3. $AMIR$ is the Gilbert-Elliot channel's achievable mutual information rate by the decision-feedback decoder vs. the input entropy;
4. $C_{memoryless}$ is the information capacity of the Gilbert-Elliot channel with the memory equal to 0.

2. The USVN-FSMC with zero channel state memory. This type of USVN-FSMC is also called memoryless USVN-FSMC. Channel state estimation is useless for this type of channel. The channel structure is usually assumed at the receiver, i.e., the initial channel state distribution. The mutual information rate can be expressed as $R = \mathcal{I}(Y; X \mid \rho_0)$. It should be noted that ρ_0 is the state distribution not the perfect state information. The channel transition probability is then given by $\Pr(y|x) = \sum_{k=0,1} p(y|x, s_0)\rho_0(k)$. And its

mutual information rate is maximized by input signals with maximum entropy. The information capacity of an example of memoryless Gilbert-Elliot channel is plotted in Fig. 3.3, denoted by $C_{memoryless}$.

3. The USVN-FSMC with channel state memory larger than zero but not infinite. This type of USVN-FSMC is what we focus on in this thesis. The channel structure is usually assumed in the receiver, i.e., the initial state distribution. For this type of USVN-FSMC, we can decrease the uncertainty of the state distribution by estimating the channel state. This is done at the price of including information redundancy in input signals. The mutual information rate can be expressed as $R = \mathcal{I}\left(Y; X \mid \hat{S}\right)$, where Y and X are the channel output and input, respectively, and $\hat{S}$ is the estimated channel state[21, 22]:

$$\begin{aligned} \mathcal{I}\left(Y; X \mid \hat{S}\right) &= \lim_{J\to\infty} \left(\mathcal{H}\left(y^J \mid \hat{s}^J\right) - \mathcal{H}\left(y^J \mid \hat{s}^J, x^J\right)\right) \\ &= \lim_{J\to\infty} \frac{1}{J} \sum_{n=1}^{J} \left(H\left(y_n \mid \hat{s}_n\right) - H\left(y_n \mid x_n, \hat{s}_n\right)\right), \end{aligned} \tag{3.9}$$

It should be noted here that $\hat{S} = \rho$ is also the estimated state distribution, not the perfect state information. For example, the channel transition probability at n time slot is then given by $\Pr(y|x) = \sum_{k=0,1} p(y|x, s_n)\rho_n(k)$. The mutual information rate of an example Gilbert-Elliot of this type is plotted against the input entropy. The signal detection is performed by the decision-feedback decoder. It is clear that the mutual information rate is not maximized by input signals with maximum entropy.

When input signals with maximum entropy are employed, we are not making use of the state memory. The achieved mutual information rate is the maximum mutual information rate of The USVN-FSMC with zero channel state memory, shown in Fig. 3.3.

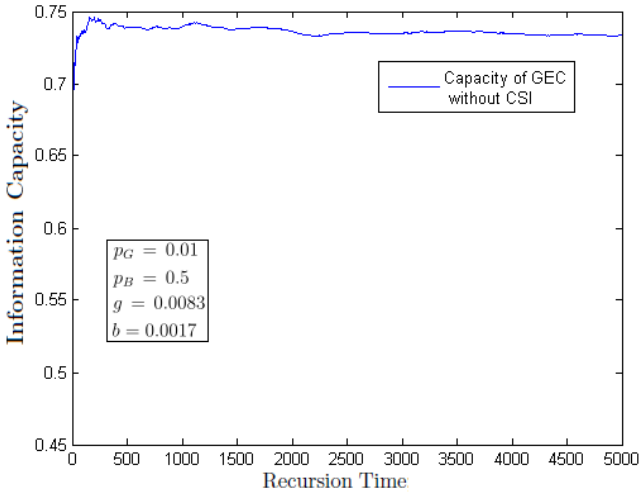


FIGURE 3.4: The information capacity of the Gilbert-Elliot channel obtained by the sum-product algorithm vs. the recursion time; what is shown in the figure is the mean values of 100 simulations [46, 51].

It should be noted that the maximum mutual information rate of the USVN-FSMC obtained by the decision-feedback decoder is not the information capacity of the channel. The C_{noCSI} is the information capacity of the USVN-FSCM obtained by the sum-product algorithm, which is independent of the system model and the decoding scheme [46, 51]. The detail simulation result is shown in Fig.3.4. C_{noCSI} in Fig.3.3 is the value to which the simulation result converges. The sum-product algorithm will be discussed in detail in section 2.3.

3.2.2 Proving The Mutual Information Rate Is Maximized by Input Signals with Information Redundancy

In this subsection, we present a proof that the mutual information rate of the USVN-FSMC with the decision-feedback decoder employed is maximized by input signals with information redundancy. First, it is proven that the state estimator

cannot decrease the uncertainty of the channel state distribution if the channel input is of maximum information entropy. In other words, the estimator does not make use of the channel memory, system model in Fig. 3.1(a) is therefore memoryless. We denote the information capacity of this memoryless USVN-FSMCs as C^{NM}. Second, it is shown that the maximum mutual information rate of the USVN-FSMC assuming no channel state information, $\mathcal{I}_{max}(Y;X \mid \hat{S})$, is larger then C^{NM}. Certain parts of the following proof was done by Zarko Krusevac. My contribution is to link these parts and make them a complete proof.

Lemma 3.2. *For the USVN-FSMC and the channel input of maximum entropy, the recursive formula (3.7) converges toward the vector of stationary state probabilities* $\boldsymbol{\tau} = [Pr(c_1), ..., Pr(c_k)]$. $\boldsymbol{\tau}$ *is the solution of the eigenvector equation* $\mathbf{P}^T\boldsymbol{\tau} = \boldsymbol{\tau}$, *where* $\mathbf{P}^T$ *is the transpose of the channel state transition matrix.*

Proof: Since each $c_k \in \mathcal{C}$ is output symmetric and the marginal input probability $\Pr(x_n)$ is uniform, then $\Pr(y_n|s_n = c_k)$ is also uniform [22], i.e., $\Pr(y_n = y_i|s_n = c_k) = 1/|\mathcal{Y}|$, where $\mathcal{Y}$ is the output alphabet. This is because the crossover probability is the same for both $x_n = 0$ and $x_n = 1$, for any amount of $x_n = 1$"cross" to $x_n = 0$, there are the same amount of $x_n = 1$ "cross" to $x_n = 1$. If x_n is uniform before transmission, it should also be uniform after the transmission. Hence, for the USVC-FSMC, with the channel input of maximum entropy, the recursive formula (3.7) becomes

$$\begin{aligned}
\boldsymbol{\rho_{n+1}}(l) &= \frac{\sum_{k=1}^{K}\Pr(y_n = y_i|s_n = c_k)\Pr(s_n = c_k|y^{n-1})P_{kl}}{\sum_{k=1}^{K}\Pr(y_n = y_i|s_n = c_k)\Pr(s_n = c_k|y^{n-1})} \qquad (3.10)\\
&= \frac{\frac{1}{|\mathcal{Y}|}\sum_{k=1}^{K}\Pr(s_n = c_k|y^{n-1})P_{kl}}{\frac{1}{|\mathcal{Y}|}\sum_{k=1}^{K}\Pr(s_n = c_k|y^{n-1})}\\
&= \frac{\sum_{k=1}^{K}\Pr(s_n = c_k|y^{n-1})P_{kl}}{\sum_{k=1}^{K}\Pr(s_n = c_k|y^{n-1})}, \qquad (3.11)
\end{aligned}$$

where $y_i \in \mathcal{Y}$ and P_{kl} is the (k,l)th entry of the transition matrix $\mathbf{P}$. Because $\sum_{k=1}^{K} \Pr\left(s_n = c_k | y^{n-1}\right) = 1$, $\boldsymbol{\rho_{n+1}}\left(l\right)$ can be expressed by

$$\boldsymbol{\rho_{n+1}}\left(l\right) \quad = \quad \sum_{k=1}^{K} \Pr\left(s_n = c_k | y^{n-1}\right) P_{kl}. \tag{3.12}$$

Thereby, the recursive formula (3.7) converges towards

$$\rho \triangleq \lim_{n\to\infty} \boldsymbol{\rho_n} = \boldsymbol{\tau}, \tag{3.13}$$

where $\boldsymbol{\tau} = \left[\Pr\left(c_1\right), ..., \Pr\left(c_k\right)\right]$ is the vector of stationary state probabilities, which is the solution of the eigenvetor equation $\mathbf{P}^T\boldsymbol{\tau} = \boldsymbol{\tau}$. Therefore, the mutual information rate achieved with the channel input of maximum entropy is actually the information capacity of the memoryless USVN-FSMC.

Lemma 3.3. *For a memoryless USVN-FSMC, the following equality holds,*

$$\mathcal{I}^{NM}_{USVN-FSMC}(Y; X \mid \hat{S}) = \mathcal{I}^{NM}_{FSMC}(Y; X), \tag{3.14}$$

over the set of all i.i.d. input distributions $\mathcal{P}(X)$, where $\mathcal{I}^{NM}_{USVN-FSMC}(Y; X \mid \hat{S})$ is the mutual information rate for the memoryless USVN-FSMC assuming no channel state information , and $\mathcal{I}^{NM}_{FSMC}(Y; X)$ is the mutual information rate of the memoryless FSMC assuming no channel state information .

Proof:

$$\begin{aligned} H\left(y^J|x^J\right) &= \sum_{n=1}^{J} H\left(y_n|x_n, x^{n-1}, y^{n-1}\right) \quad (3.15)\\ &= \sum_{n=1}^{J} H\left(y_n|x_n\right) \\ &= \sum_{n=1}^{J} E\left[-\log \sum_{k=0}^{K} \Pr\left(y_n|x_n, s_n = c_k\right) \Pr\left(s_n = c_k\right)\right] \\ &= \sum_{n=1}^{J} E\left[-\log \sum_{k=0}^{K} \Pr\left(y_n|x_n, s_n = c_k\right) \boldsymbol{\rho_n}\left(k\right)\right] \\ &= \sum_{n=1}^{J} H\left(y_n|x_n, \hat{s}_n\right) \quad (3.16) \end{aligned}$$

where the fourth equality follows from the fact that the FSMC is memoryless and, thereby

$$\boldsymbol{\rho_n}\left(k\right) = \Pr\left(s_n = c_k|y^{n-1}\right) = \Pr\left(s_n = c_k\right). \quad (3.17)$$

Similarly,

$$H\left(y^J|\hat{s}^J\right) = \sum_{n=1}^{J} H\left(y_n|\hat{s}_n\right) = \sum_{n=1}^{J} H\left(y_n \mid y^{n-1}\right) = H\left(y^J\right). \quad (3.18)$$

$$\begin{aligned} \mathcal{I}^{NM}_{USVN-FSMC}(Y; X \mid \hat{S}) &= \lim_{J\to\infty}\left[\frac{1}{J}\sum_{n=1}^{J} H\left(y_n|\hat{s}_n\right) - \frac{1}{J}\sum_{n=1}^{J} H\left(y_n|x_x, \hat{s}_n\right)\right] \\ &= \lim_{J\to\infty}\left[\frac{1}{J}\sum_{n=1}^{J} H\left(y_n\right) - \frac{1}{J}\sum_{n=1}^{J} H\left(y_n|x_n\right)\right] \\ &= \mathcal{I}^{NM}_{FSMC}(Y; X), \quad (3.19) \end{aligned}$$

Because $\mathcal{I}^{NM}_{FSMC}(Y; X)$ is a convex function and is maximized with channel inputs of maximum entropy, $\mathcal{I}^{NM}_{USVN-FSMC}(Y; X \mid \hat{S})$ works the same way. Therefore, the

capacity of memoryless USVN-FSMCs can be expressed by,

$$C^{NM} = \max\left\{\mathcal{I}^{NM}_{USVN-FSMC}(Y;X \mid \hat{S})\right\}. \tag{3.20}$$

Lemma 3.4. *For USVC-FSMC assuming channel state information , the following equality holds,*

$$\mathcal{I}^{CSI}_{USVN-FSMC}\left(Y;X \mid \hat{S}\right) = \mathcal{I}^{CSI}_{FSMC}\left(Y;X\right), \tag{3.21}$$

over the set of all i.i.d. input distribution $\mathcal{P}(X)$, where $\mathcal{I}^{CSI}_{USVN-FSMC}\left(Y;X \mid \hat{S}\right)$ is the mutual information rate of the USVC-FSMC assuming channel state information and $\mathcal{I}^{CSI}_{FSMC}(Y;X)$ is the mutual information rate of the FSMC assuming channel state information .

Proof:

$$\begin{aligned}\mathcal{I}^{CSI}_{USVN-FSMC}\left(Y;X \mid \hat{S}\right) &= \lim_{J\to\infty}\mathcal{I}\left(\left(y^J,\hat{s}^J;x^J\right)\right)\\ &= \lim_{J\to\infty}\mathcal{I}\left(\left(y^J;x^J\right)\right)\\ &= \mathcal{I}^{CSI}_{FSMC},\end{aligned} \tag{3.22}$$

where the second equality follows from the fact that the channel state information has been given, and therefore $\hat{s}^J$ can be taken away from the equation.

Because $\mathcal{I}^{CSI}_{FSMC}(Y;X)$ is a convex function over $\mathcal{P}(X)$ and it is maximized with channel inputs with maximum entropy, $I^{CSI}_{USVN-FSMC}\left(Y;X \mid \hat{S}\right)$ has the same behaviour.

In the following, the relationship among the three mutual information rates is investigated, $\mathcal{I}^{CSI}_{USVN-FSMC}\left(Y;X \mid \hat{S}\right)$, $\mathcal{I}\left(Y;X \mid \hat{S}\right)$ and $\mathcal{I}^{NM}_{USVN-FSMC}(Y;X \mid \hat{S})$. It is shown that, as the channel memory μ increases from 0 to 1, $\mathcal{I}\left(Y;X \mid \hat{S}\right)$ converges to $\mathcal{I}^{CSI}_{USVN-FSMC}\left(Y;X \mid \hat{S}\right)$ from $\mathcal{I}^{NM}_{USVN-FSMC}\left(Y;X \mid \hat{S}\right)$.

In order to investigate the relationship between the mutual information rate and the channel process memory, the channel state structure and the transition structure ratios need to be fixed. For instance, for the Gilbert-Elliott channel, it means that the state cross-over probabilities, p_G and p_B, and the good-to-bad ratio are fixed [21]. We use $\mathcal{I}_{\mu}\left(Y;X \mid \hat{S}\right)$ to denote the mutual information rate of the USVC-FSMC under this assumption.

Let μ denote a measure of the persistent channel memory which has maximum value of μ_{max}, such that

$$\lim_{\mu \to \mu_{max}} \mathbf{P} = \mathbf{P}_{fixed}, \tag{3.23}$$

where $\mathbf{P}$ is the channel state transition matrix, and $\mathbf{P}_{fixed}$ is a special case of $\mathbf{P}$ and it has one element in each row equal to 1 (dominant element) and all other elements equal to 0.

Since the recursive estimation formula (3.7) is linear in $\mathbf{P}$, for $\mu \to \mu_{max}$, it monotonically converges towards

$$\begin{aligned} \lim_{\mu \to \mu_{max}} \boldsymbol{\rho_{n+1}}(l) &= \lim_{\mu \to \mu_{max}} \frac{\sum_{k=1}^{K} \Pr(y_n|s_n = c_k) \Pr(s_n = c_k|y^{n-1}) P_{kl}}{\sum_{k=1}^{K} \Pr(y_n|s_n = c_k) \Pr(s_n = c_k|y^{n-1})} \\ &= \frac{\sum_{r \in R} \Pr(y_n|s_n = c_r) \boldsymbol{\rho_n}(r)}{\sum_{k=1}^{K} \Pr(y_n|s_n = c_k) \boldsymbol{\rho_n}(k)}, \end{aligned} \tag{3.24}$$

where $r \in R$ are dominant elements of the l-th column of $\mathbf{P}_{fixed}$ and $P_{rl} = 1$.

Lemma 3.5. *For any i.i.d. input distribution $P(X)$ which is not uniform, there exists a channel state c_v for which the recursion (3.24) increases and monotonically converges to 1, as time n proceeds. For other states, the recursion (3.24) decreases and monotonically converges to 0.*

Proof: See Appendix C for the proof.

Therefore, there is a monotonic convergence $\lim_{\mu\to\mu_{max}} H(\rho) = 0$ for any i.i.d input distribution which is not uniform, and thereby, monotonic convergence $H\left(y^J|x^J,\hat{s}^J\right)$ to $H\left(y^J|x^J\right)$, $H\left(y^J|\hat{s}^J\right)$ to $H\left(y^J\right)$ and $\mathcal{I}_\mu\left(Y;X\mid\hat{S}\right)$ to $\mathcal{I}^{CSI}_{USVN-FSMC}\left(Y;X\mid\hat{S}\right)$ for $\mu\to\mu_{max}$. The monotonic convergence of $\mathcal{I}_\mu\left(Y;X\mid\hat{S}\right)$ to $\mathcal{I}^{CSI}_{USVN-FSMC}\left(Y;X\mid\hat{S}\right)$ is intuitively satisfactory, because for larger memory the expected dwell time in each state is larger and the next state can be better predicted. For the uniform i.i.d. input distribution, Eq. (3.24) becomes $\lim_{\mu\to\mu_{max}} \boldsymbol{\rho_{n+1}}(l) = \boldsymbol{\rho_n}(l)$ and $\mathcal{H}(\rho)$ is a step function, i.e., $H(\rho) = 0$ for $\mu = \mu_{max}$ and $H(\rho) = H(\boldsymbol{\tau})$ elsewhere, where $\boldsymbol{\tau}$ is the initial channel state distribution.

Therefore, according to *Lemma 3* to *Lemma 6*, for any channel input distribution, $\mathcal{I}\left(Y;X\mid\hat{S}\right)$ converges to $\mathcal{I}^{CSI}_{USVN-FSMC}\left(Y;X\mid\hat{S}\right)$ from $\mathcal{I}^{NM}_{USVN-FSMC}\left(Y;X\mid\hat{S}\right)$, for μ increasing from 0 to 1.

Theorem 3.6. *For the USVC-FSMC, there exists a non-maximum-entropy channel input distribution, with which the obtained mutual information rate assuming no channel state information is larger than that obtained with the channel input of maximum entropy.*

Proof:

We have proved that:

1. if input signals are of maximum information entropy, the mutual information rate achieved by the decision-feedback decoder is the information capacity of memoryless channel, $\mathcal{C}^{NM}\left(Y;X\mid\hat{S}\right)$.

2. as the channel memory goes towards infinite, the maximum mutual information rate achieved by the decision-feedback decoder approaches the information capacity of time invariant channels, $\mathcal{C}^{CSI}_{USVN-FSMC} > \mathcal{C}^{NM}\left(Y;X\mid\hat{S}\right)$.

3. therefore, for a channel with any state memory μ, rather than the input signal with maximum information entropy, there must exist at least one

input signal distribution, in which the mutual information rate is larger than $\mathcal{C}^{NM}\left(Y;X \mid \hat{S}\right)$.

We assume that, the channel state structure and the transition structure of the USVN-FSMC are fixed. For any channel memory μ, the mutual information rate of the USVN-FSMC assuming channel state information is larger than that of the memoryless USVN-FSMC. Because the information capacity of the USVN-FSMC assuming channel state information is a continuous function over all input distributions, there exist a set of channel input distributions, denoted by $\mathcal{P}'(X)$, with which, $\mathcal{I}^{CSI}_{USVN-FSMC}\left(Y;X \mid \hat{S}\right) > \mathcal{C}^{NM}\left(Y;X \mid \hat{S}\right)$. We have proven that $\mathcal{I}\left(Y;X \mid \hat{S}\right)$ converge to $\mathcal{I}^{CSI}_{USVN-FSMC}\left(Y;X \mid \hat{S}\right)$ as μ increases. Therefore, for any input distribution in $\mathcal{P}'(X)$, excluding the input distribution that is of maximum entropy, there is a value of μ, with which $\mathcal{I}\left(Y;X \mid \hat{S}\right) > \mathcal{C}^{NM}\left(Y;X \mid \hat{S}\right)$.

3.3 Discussion: The Decision-Feedback Decoder Cannot Achieve The Full Information Capacity of The Gilbert-Elliot Channel

According to the analysis of mutual information rate of USVN-FSMC, it is clear from Fig. 3.3 that the decision-feedback decoder has not achieved the information capacity of the Gilbert-Elliot channel. The problem for the USVN-FSMC has not been solved completely. We can provide evidence that non-optimality is due to the difficulty of synchronizing the state estimation and signal detection in the decision-feedback decoder. Looking for perfect synchronizer for the Gilbert-Elliot channel could be one of our future research works.

3.3.1 The Non-Optimality of The Decision-Feedback Decoder

In this subsection, we introduce the method which can obtain the information capacity of the Gilbert-Elliot channel independent of the decoding schemes. It is called by "coin toss" in [52]. The method is independent of the system model. After that, we show that the information capacity is higher than the maximum mutual information rate obtained by the decision-feedback decoder.

The mutual information rate expression by the decision-feedback decoder including the estimation-caused information loss, $\mathcal{I}(Y;X \mid \hat{S})$, has been introduced in [23], where $\hat{S}$ is the estimated channel state distribution.

$$\begin{aligned}\mathcal{I}\left(Y;X \mid \hat{S}\right) &= \lim_{J\to\infty}\left(\mathcal{H}\left(y^J \mid \hat{s}^J\right) - \mathcal{H}\left(y^J \mid \hat{s}^J, x^J\right)\right) \\ &= \lim_{J\to\infty}\frac{1}{J}\sum_{n=1}^{J}\left(H\left(y_n \mid \hat{s}_n\right) - H\left(y_n \mid x_n, \hat{s}_n\right)\right), \end{aligned} \tag{3.25}$$

where

$$H\left(y_n \mid \hat{s}_n\right) = \sum_{k=1}^{K}\boldsymbol{\rho_n}\left(k\right)H\left(y_n \mid s_n = c_k\right), \tag{3.26}$$

$$\begin{aligned}&H\left(y_n \mid x_n, \hat{s}_n\right) = \\ &\sum_{i=0,1}\Pr\left(x_n = i\right)\sum_{k=1}^{K}\boldsymbol{\rho_n}\left(k\right)H\left(y_n \mid x_n = i, s_n = c_k\right),\end{aligned} \tag{3.27}$$

where $H\left(y_n \mid \hat{s}_n\right)$ denotes the channel output entropy conditioned on the state distribution, and $H\left(y_n \mid x_n, \hat{s}_n\right)$ is the output entropy conditioned on the state distribution and inputs.

The information capacity of the Gilbert-Elliot channel model can be obtained using a system-independent method. The mutual information of the Gilbert-Elliot

channel can be expressed by[53],

$$\mathcal{I}(Y;X) = \mathcal{H}(Y) - \mathcal{H}(Y \mid X), \tag{3.28}$$

where $\mathcal{H}$ denotes the entropy rate. In [46], it was found that, given input sequence x^n and output sequences y^n with n going to infinity, $-\frac{1}{n}\log p(x^n, y^n) + \frac{1}{n}\log p(x^n)$ converged to the entropy rate $\mathcal{H}(Y|X)$, and $-\frac{1}{n}\log p(y^n)$ converged to $\mathcal{H}(\mathcal{Y})$. $\mathcal{I}(Y;X)$ becomes

$$\begin{aligned} &\mathcal{I}(Y;X) \\ &= \frac{1}{n}\log p(x^n, y^n) - \frac{1}{n}\log(x^n) - \frac{1}{n}\log p(y^n), \end{aligned} \tag{3.29}$$

where $p(y^n)$ and $p(x^n, y^n)$ can be computed using the sum-product algorithm provided in [46]. The results show that the information capacity is achieved by uniform and i.i.d channel inputs[21].

The sum-product algorithm was later simplify to a "coin toss" method. In [21, 46, 52], the Gilbert-Elliot channel information capacity is expressed by

$$\begin{aligned} C &= \lim_{n\to\infty} \max \frac{1}{nI(x^n; y^n)} \\ &= \lim_{n\to\infty} \max \frac{(H(y^n) - H(y^n \mid x^n))}{n} \\ &= \lim_{n\to\infty} \max \frac{(H(y^n) - H(z^n)))}{n}. \end{aligned} \tag{3.30}$$

$H(y^n) = n\log(\mathcal{Y}) = n$ because the output alphabet $\mathcal{Y}$ is uniform and i.i.d.. (3.30) is simplified to,

$$C = 1 - \lim_{n\to\infty} \frac{H(z^n)}{n}. \tag{3.31}$$

$H(z^n)$ can be approximated by generating a long sequence of z^n and evaluating $-\log \Pr(z^n)/n$. The probability of $z_n = 1$ can be calculated recursively as

follows[21, 22]:

$$q_n = v\left(z_{n-1}, q_{n-1}\right), \tag{3.32}$$

if $z_{n-1} = 0$,

$$\begin{aligned} v\left(z_{n-1}, q_{n-1}\right) &= p_G + b\left(p_B - p_G\right) \\ &\quad + \mu\left(q_{n-1} - p_G\right)\left[\left(1 - p_B\right) / \left(1 - q_{n-1}\right)\right], \end{aligned} \tag{3.33}$$

if $z_{n-1} = 1$,

$$v\left(z_{n-1}, q_{n-1}\right) = p_G + b\left(p_B - p_G\right) + \mu\left(q_{n-1} - p_G\right)\left(p_B / q_{n-1}\right). \tag{3.34}$$

Because

$$\begin{aligned} -\log \Pr(z^n)/n &= -\frac{1}{n}\sum_{i=1}^{n} \log \Pr(z_i | z^{i-1}) \\ &= -\frac{1}{n}\sum_{i=1}^{n} \left(z_i \log\left(q_i\right) + \left(1 - z_i\right)\log\left(1 - q_i\right)\right). \end{aligned} \tag{3.35}$$

The limit of (3.35) can be obtained by recursion. z_0 is assumed. This recursion begins with $i = 1$ and stops when i is larger enough that the value of (3.35) converges. In the simulation in this thesis, the recursion stops at $i = 5000$. In each step of the recursion, we generate z_i as a $Bernouli(q_i)$ since q_i is known. If the $Bernouli(q_i)$ result is success, $\log(q_i)$ is added, otherwise $\log(1 - q_i)$ is added. This method is called by "coin toss" in [52].

In Fig. 3.3, the information capacity of the Gilbert-Elliot channel obtained by the "coin toss" method is plotted against the recursive times. In Fig. 3.4, we compare the mutual information rate, denoted in the figure by AMIR, obtained by the decision-feedback decoder with the information capacity, of which the value

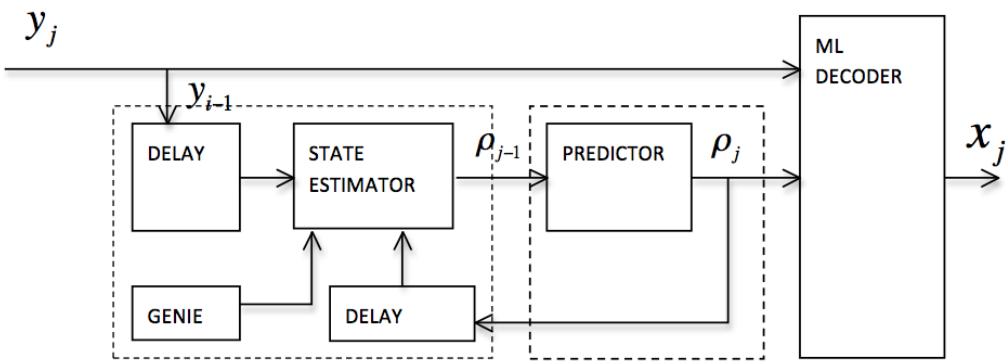


FIGURE 3.5: The decision-feedback decoder with the implicit predictor.

is taken from Fig. 3.3 when recursive time $n = 5000$. Evidently, the decision-feedback decoder is unable to achieve the the information capacity.

3.3.2 A Non-Trivial Reasons for The Non-Optimality of The Decision Feedback Decoder

Many possible reasons contribute to the non-optimality of the decision-feedback decoder. We argue that the delay between the state detection and signal detection is the main reason. We have shown in section 1.3.1.3 that the perfect synchronization between the two detections is a sufficient condition for optimum signal detection of FSMCs with channel variation memory. In this subsection, we show that the decision-feedback decoder tries to solve the problem by guessing the distribution of the next state, which removes the delay of state estimation but results in additional errors.

Refering to formula (3.7) of the decision-feedback decoder

$$\boldsymbol{\rho_{n+1}} = \frac{\boldsymbol{\rho_n}\mathbf{B}\left(y_n\right)\mathbf{P}}{\boldsymbol{\rho_n}\mathbf{B}\left(y_n\right)\underline{\mathbf{1}}} \triangleq f\left(y_n, \boldsymbol{\rho_n}\right), \tag{3.36}$$

where $\mathbf{B}(y_n)$ is a diagonal 2×2 matrix, with the kth ($k = 1, 2$) diagonal term $\Pr(y_n = 0 | s_n = c_k)$ and $\underline{\mathbf{1}} = (1, 1)^T$[22], (3.36) can be devided into two parts:

$$\boldsymbol{\rho_{n+1}} = \left(\frac{\boldsymbol{\rho_n} \mathbf{B}(y_n)}{\boldsymbol{\rho_n} \mathbf{B}(y_n) \underline{\mathbf{1}}} \right) \cdot (\mathbf{P}), \tag{3.37}$$

We show in the following that the state has been actually detected by the formula in the first bracket, while $\mathbf{P}$ in the second bracket predicts the distribution of next state.

$$\begin{aligned}
&\frac{\boldsymbol{\rho_n} \mathbf{B}(y_n)}{\boldsymbol{\rho_n} \mathbf{B}(y_n) \underline{\mathbf{1}}} \\
&= \frac{\begin{bmatrix} \Pr(s_n = c_1) & \Pr(s_n = c_2) \end{bmatrix} \begin{bmatrix} \Pr(y_n | s_n = c_1) & 0 \\ 0 & \Pr(y_n | s_n = c_2) \end{bmatrix}}{\begin{bmatrix} \Pr(s_n = c_1) & \Pr(s_n = c_2) \end{bmatrix} \begin{bmatrix} \Pr(y_n | s_n = c_1) & 0 \\ 0 & \Pr(y_n | s_n = c_2) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}} \\
&= \frac{\begin{bmatrix} \Pr(y_n, s_n = c_1) & \Pr(y_n, s_n = c_2) \end{bmatrix}}{\begin{bmatrix} \Pr(y_n, s_n = c_1) & \Pr(y_n, s_n = c_2) \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix}} \\
&= \begin{bmatrix} \Pr(s_n = c_1 | y_n) & \Pr(s_n = c_2 | y_n) \end{bmatrix}.
\end{aligned} \tag{3.38}$$

Therefore $\frac{\boldsymbol{\rho_n} \mathbf{B}(y_n)}{\boldsymbol{\rho_n} \mathbf{B}(y_n) \underline{\mathbf{1}}}$ has detected the state distribution conditioned on the past outputs. However, what is used for recovering transmitted signal in the ML decoder is the prediction result, shown in Fig. 3.5. Because estimating the channel state in the receiver is always delayed by 1 symbol, which means the detected state information is about the channel state in time slot n when the $n + 1$ symbol arrives at the receiver, this prediction succeeds in removing the delay but obviously additional errors has been introduced by the term $(\mathbf{P})$ in (3.37).

3.4 Chapter Conclusion

In the section, we provide analysis for USVN-FSMC, which complements the existing analysis for time varying channels in the literature. We show that time varying channels are fundamentally different from time invariant channel or slow time varying channels. The exisiting state estimation schemes and decoding schemes are no longer capacity achieving for time varying channels. There could be two non-trival reasons for the non-optimality. First, estimation schemes and the decoding schemes are not powerful enough. Second, models for time varying channels are insufficient to show the essential of the type of channels.

Chapter 4

The Mutual Information Rate Analysis of Uniformly Symmetric Variable Noise Finite State Markov Channels with Markov Input Signals

In this chapter, we test the mutual information rate of USVN-FSMC with an example of Markov signal input. The example we use is a first order symmetric Markov chain. In last chapter, we show via independent input signals that the mutual information rate of USVN-FSMC is maximized by input signals with information redundancy. Our analysis of Markov input signals confirms this result again. More importantly, we find that the Markov input signal can achieve higher mutual information rate than independent input signals with the same information redundancy.

4.1 Channel State Estimation Algorithm for Markov Input Signals and The Estimation Results

In this section, we first show how the uniform inputs with first order memory can be converted to equivalent non-uniform i.i.d. sequence so that (3.7) in chapter 3 can be used. We assume that the alphabet of Markov inputs has two symbols, $\mathcal{X} = \{1, 0\}$, and the transition probabilities between each other are equal, $g_{10} = g_{01}$.

Referring to Fig. 3.1(b), input symbols with memory will be made mutual independent when passing through the interleaver[21–23]. This is done by storing the symbols row by row in a $J \times L$ matrix and transmitting them column by column. An example is shown as follows,

$$x^1, x^2, x^3, x^4, \cdots, x^9 \Longrightarrow \begin{bmatrix} x^1 & x^2 & x^3 \\ x^4 & x^5 & x^6 \\ x^7 & x^8 & x^9 \end{bmatrix} \Longrightarrow x^1, x^4, x^7, x^2, \cdots, x^9.$$

Any two consecutive symbols in the input sequence is dispersed by J in distance and can be regarded as memoryless when J is large. Therefore, when the uniform Markov source is applied to the interleaver directly, it will become uniform and i.i.d.. According to result in section 3.2, the state cannot be tracked reliably as the estimation requires redundancy [23].

However, any two consecutive symbols in the Markov chain are dependent and we can make use of their memory by treating them as one symbol in the interleaver.

An example is shown as follows:

$$x^1, x^2, x^3, x^4, \cdots, x^{18} \Longrightarrow \begin{bmatrix} x^1x^2 & x^3x^4 & x^5x^6 \\ x^7x^8 & x^9x^{10} & x^{11}x^{12} \\ x^{13}x^{14} & x^{15}x^{16} & x^{17}x^{18} \end{bmatrix} \Longrightarrow x^1x^2, x^7x^8, x^{13}x^{14}, x^3x^4, \cdots, x^{17}x^{18}.$$

The inputs is converted to a sequence of new alphabet. In our case, the input will become an i.i.d sequence with new alphabet $\mathcal{X} = \{00, 01, 11, 10\}$, where

$$\begin{aligned} \Pr(x_n = 00) &= 0.5 * (1 - g_{10}), \\ \Pr(x_n = 01) &= 0.5 * g_{10}, \\ \Pr(x_n = 11) &= 0.5 * (1 - g_{10}), \\ \Pr(x_n = 10) &= 0.5 * g_{10}. \end{aligned} \tag{4.1}$$

If $g_{10} = 0.5$, the new sequence is uniform, i.e., $\Pr(x_n = 00) = \Pr(x_n = 01) = \Pr(x_n = 11) = \Pr(x_n = 10) = 0.25$. If $g_{10} \neq 0.5$, the sequence has redundancy and the channel state can be tracked as follows.

Taking a module-2 addition of the two bits of each new symbol, we have

$$\begin{aligned} 00 &\longrightarrow 0 \oplus 0 = 0, \\ 01 &\longrightarrow 0 \oplus 1 = 1, \\ 11 &\longrightarrow 1 \oplus 1 = 0, \\ 10 &\longrightarrow 1 \oplus 0 = 1. \end{aligned} \tag{4.2}$$

Denoting the addition result by $x_n^{'}$, we have

$$\begin{aligned} \Pr\left(x_n^{'} = 0\right) &= 1 - g_{10} \\ \Pr\left(x_n^{'} = 1\right) &= g_{10} \end{aligned} \tag{4.3}$$

The corresponding analytical output probabilities of good state and bad state can be expressed by,

$$\Pr\left(y_n^{'}=0|s_n=c_1\right)=\Pr\left(x_n^{'}=0\right)\cdot\left((1-p_G)^2+(p_G)^2\right)+2\cdot\Pr\left(x_n^{'}=1\right)\cdot(1-p_G)\cdot(p_G)\,, \tag{4.4}$$

$$\Pr\left(y_n^{'}=0|s_n=c_2\right)=\Pr\left(x_n^{'}=0\right)\cdot\left((1-p_B)^2+(p_B)^2\right)+2\cdot\Pr\left(x_n^{'}=1\right)\cdot(1-p_B)\cdot(p_B)\,, \tag{4.5}$$

We denote the statistical average of the output obtained in simulations by $\hat{\Pr}\left(y_n^{'}=0\right)$. According to (3.7), the state can be detected by

$$\boldsymbol{\rho_{n+1}}=\frac{\hat{\boldsymbol{\alpha}}^{'}}{\hat{\beta}^{'}}\cdot\mathbf{P}, \tag{4.6}$$

where $\hat{\boldsymbol{\alpha}}^{'}=\left[|\hat{\Pr}(y_n^{'}=0)-\Pr(y_n^{'}=0|s_n=c_2)|\quad|\hat{\Pr}(y_n^{'}=0)-\Pr(y_n^{'}=0|s_n=c_1)|\right]$,
$\hat{\beta}^{'}=|\hat{\Pr}(y_n^{'}=0)-\Pr(y_n^{'}=0|s_n=c_2)|+|\hat{\Pr}(y_n^{'}=0)-\Pr(y_n^{'}=0|s_n=c_1)|$.

Fig. 4.1(a) shows the tracking ability of estimator with $\mu_s=0$. Obviously, the estimator fails to track the channel state. This is because when the memory is equal to zero, the input is uniform and i.i.d.. According to [21–23], the estimator determines the channel state by measuring how much the channel input distribution is modified after being filtered by the channel. However, the uniform and i.i.d. input distribution will be modified by the same degree, regardless which state the channel is in.

Fig. 4.1(b), 4.1(c) and 4.1(d) show the tracking ability of the estimator with channel input memory, $\mu_s=0.2$, $\mu_s=0.4$ and $\mu_s=0.6$, respectively. For all these inputs, the channel state is tracked successfully. The performance improves as the memory increases. This is because redundancy in the transformed input sequence is an increasing function of the original input's memory, and larger redundancy results in better estimation results [23].

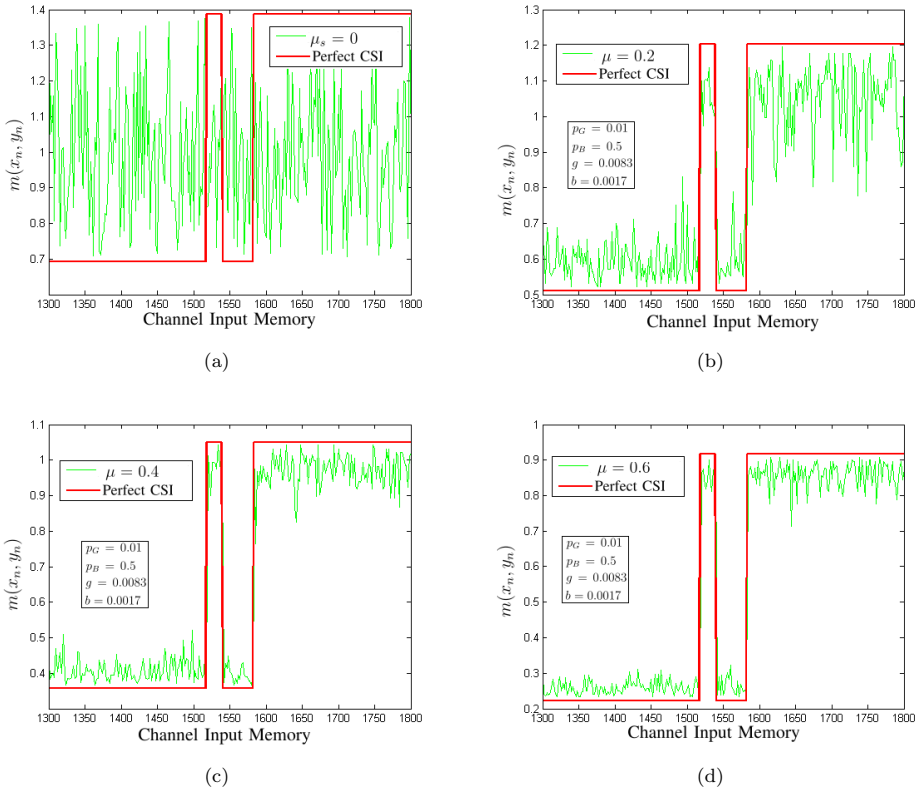
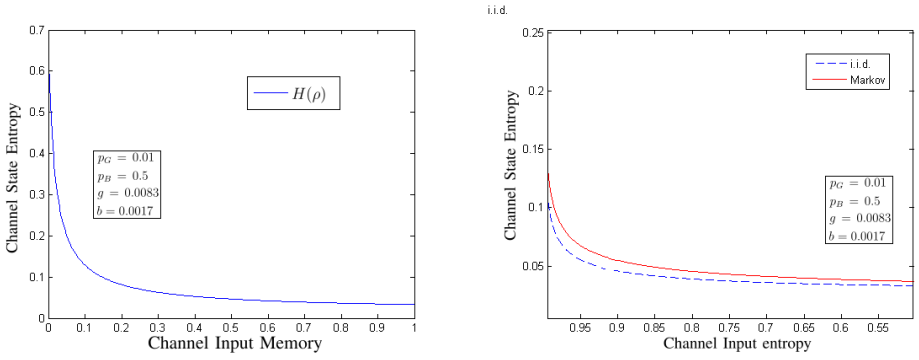


FIGURE 4.1: Tracking ability of the loop-loop estimator with different input memory, μ: (a) $\mu = 0$, (b) $\mu = 0.2$, (c) $\mu = 0.4$ and (d) $\mu = 0.6$.

The results above are also confirmed in Fig. 4.2(a), in which the entropy of the state is plotted against the memory. As the memory increases, the state entropy decreases continuously.

In Fig. 4.2(b), we also compare the estimation ability of the Markov inputs and i.i.d. inputs. The entropy of the estimated state distribution by these two kinds of inputs is plotted against the channel input entropy. With the same input entropy, the state entropy by Markov source is larger. This means that with the Markov source, we are more certain about the channel state. It should be noted that the minimum entropy of the uniform Markov source is 0.5. This is because by taking module-2 addition to the two bits of each input symbol, we enlarge the distance between the two analytical output probability of the two states in (4.6),



(a) (b)

FIGURE 4.2: (a) The entropy of the estimated state distribution vs. the channel input memory; (b)

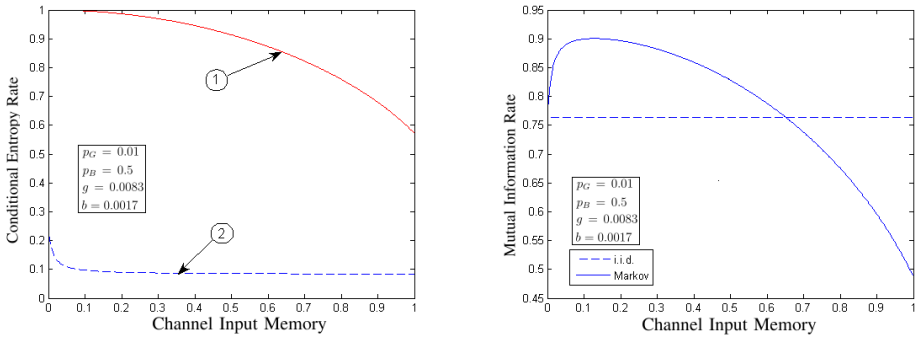
1. The solid line is the entropy of the estimated state distribution by Markov source vs. the channel input entropy.
2. The dash line is the entropy of the estimated state distribution by i.i.d. source vs. the channel input entropy

$\Pr(y_n = 0|s_n = c_1)$ and $\Pr(y_n = 0|s_n = c_2)$, so that it becomes easier to determine the state given the same redundancy.

4.2 Mutual Information Rate Expression of The Channel with Markov Input Signals

In this section, we investigate the mutual information rate of USVN-FSMCs with the Markov input. We show first that compared to uniform and i.i.d. inputs with the same alphabet, larger maximum mutual information rate can be obtained by the Markov input. Second, we show that with the same redundancy, the Markov inputs have better mutual information rate performance than i.i.d. inputs with the same alphabet.

The mutual information rate expression, $R = \mathcal{I}\left(Y; X \mid \hat{S}\right)$, has been introduced in [23], where Y and X are the channel output and input respectively, and $\hat{S}$ is the



(a) (b)

FIGURE 4.3: (a)

1. The solid line is $H\left(y_n \mid \hat{s}_n\right)$ vs. the channel input memory.
2. The dash line is $H\left(y_n \mid x_n, \hat{s}_n\right)$ vs. the channel input memory;

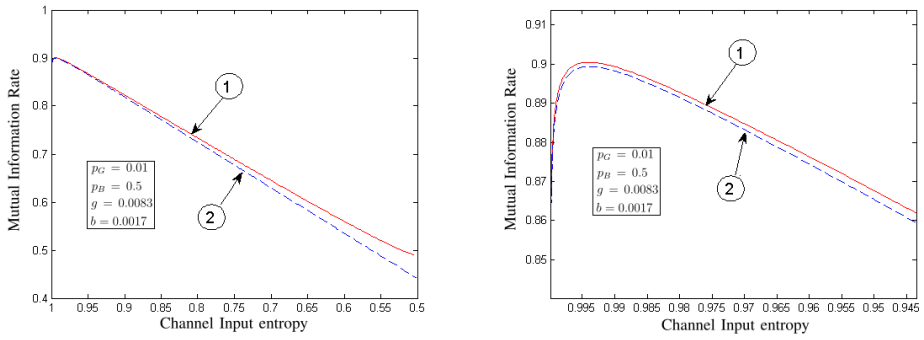
(b)

1. The solid line is the channel mutual information rate by uniform distributed source with memory vs. the channel input memory
2. The dash line is the channel mutual information rate by uniform and i.i.d. source vs. the channel input memory.

estimated channel state distribution. It should be noted that the uniform Markov source is transformed to non-uniform i.i.d. input sequence with larger alphabet by the algorithm in last section. Information about the channel state distribution is obtained by estimation, which is enabled by the redundancy in the transformed input sequence.

$$\begin{aligned} \mathcal{I}\left(Y; X \mid \hat{S}\right) &= \lim_{J\to\infty} \left(\mathcal{H}\left(y^J \mid \hat{s}^J\right) - \mathcal{H}\left(y^J \mid \hat{s}^J, x^J\right)\right) \\ &= \lim_{J\to\infty} \frac{1}{J} \sum_{n=1}^{J} \left(H\left(y_n \mid \hat{s}_n\right) - H\left(y_n \mid x_n, \hat{s}_n\right)\right), \end{aligned} \tag{4.7}$$

where $H\left(\cdot \mid \cdot\right)$ denotes the conditional entropy. $H\left(y_n \mid \hat{s}_n\right)$ denotes the entropy of the channel outputs conditioned on state distribution. $H\left(y_n \mid x_n, \hat{s}_n\right)$ is the entropy of the outputs conditioned on state distribution and inputs.



(a) (b)

FIGURE 4.4: (a)

1. The solid line is the channel mutual information rate by uniform distributed source with memory vs. the channel input entropy.
2. The dash line is the channel mutual information rate by i.i.d. source vs. the channel input entropy.

(b) Enlargement of a part of (a).

4.3 Comparing The Mutual Information Rates Obtained by Markov Source and That by Independent Source

In Fig. 4.3(a), $H(y_n \mid \hat{s}_n)$ and $H(y_n \mid x_n, \hat{s}_n)$ of the channel with uniform inputs with memory are plotted against the channel input memory. They decrease at different rates as the memory increases. Therefore, the mutual information rate is not maximized by inputs with memory which is equal to 0.

In Fig. 4.3(b), the mutual information rates gotten by uniform Markov source and by uniform i.i.d source with the same alphabet are plotted against the channel input memory. Evidently, when the memory is zero, the Markov source changed back to i.i.d. and the two mutual information rates are equal. However, the mutual information rate with Markov source is maximized at a non-zero memory.

In Fig. 4.4(a) and 4.4(b), channel mutual information rates obtained by uniform Markov inputs and i.i.d. inputs with the same alphabet are plotted against input entropy. Obviously, the Markov inputs have achieved better performance. This is because the Markov inputs can track the state more reliably, which indirectly leads to a higher mutual information rate, shown in Fig. 4.2(b).

4.4 Chapter Conclusion

In this chapter, we confirm via a type of Markov input signals that the mutual information rate of USVN-FSMC is maximized with information redundancy. We also prove the possibility that Markov input signals can achieve higher mutual information rate than independent signals. But the advantage is very small and can almost neglected. In other words, there is no much research value in it.

Chapter 5

Information Capacity Analysis of The Time Varying Binary Symmetric Channels

The ultimate purpose of this thesis is to find a solution to achieve high rate data transmission over mobile communication channels. The mobile communication channel is a typical time variable channel. In this chapter, we achieve this objective in two steps. First, we clarify the incompleteness of the current simplest USVN-FSMC model for mobile communication channels. We then introduce a more accurate and simpler model for the time variable mobile communication channels. Second, we prove that the differential encoder and the differential decoder can achieve the information capacity of the new model. In chapter 3 and chapter 4, we revisited the analysis of time variable channels in the literature. We confirm that time invariant channels and time varying channels are fundamentally different. It should be noted here that time invariant channels here also include slow time varying channel. Existing decoding schemes for time varying communication schemes are usually comprised of two parts. The first part is channel state estimator. The second part is signal detector. There are always relative delays

between these two parts. For this reason, existing decoders cannot achieve the full information capacity of time varying communication channels. The differential encoder and the differential decoder combined the two parts and is able achieve the information capacity of the time varying channels.

5.1 The Incompleteness of The Gilbert-Elliot Channel Model

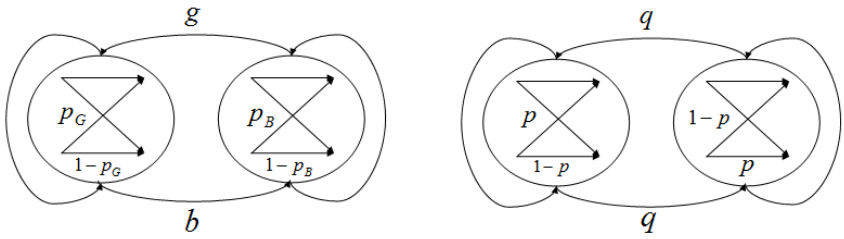
The Gilbert-Elliot model is the simplest FSMC model for time varying communication channels in the literature. In this section, we introduce the physical significance of the Gilbert-Elliot model and show its incompleteness in modeling the time varying channels. A new simplest model for the mobile communication channel is therefore required..

5.1.1 The Discrete Communication System

The mathematical model of the digital communication systems has been expressed in (2.1)

$$\mathcal{Y} = \mathcal{G}\mathcal{X} + \mathcal{Z}. \tag{5.1}$$

In this chapter, we are going to analyze the modeling of the simplest communication channel with channel variation memory. The four factors $\mathcal{Y}$, $\mathcal{G}$, $\mathcal{X}$ and $\mathcal{Z}$ in (5.1) are all scalars, which means that the transmitted symbols are mutually independent and there is no multi-path delay during the transmission. In this paper, we denote the nth transmitted symbol and the nth received symbol by x'_n and y'_n, respectively. And we denote the transmitted symbol sequence and the received



(a) (b)

FIGURE 5.1: (a) The Gilbert-Elliot channel model; (b) The time-varying-BSC.

symbol sequence at time slot n by x_n and y_n, respectively. Since the transmitted symbols are mutually independent, the expectation $E(x'_i \cdot x'_j) = 0$ whenever $i \neq j$.

5.1.2 The Physical Significance of The Gilbert-Elliot Model

The Gilbert-Elliot channel, shown in Fig. 5.1(a), has been used in modeling the time varying receiver signal amplitude fading in time varying mobile communication channel[21–23]. It is comprised of two discrete Binary Symmetric Channels (BSC), denoted by c_1, c_2, respectively. The crossover probabilities at c_1 and c_2 are denoted by p_G and p_B, respectively, and $p_G < p_B$. The state with smaller crossover probability is called as the good state and the other is called as the bad state. We clarify in this subsection how the Gilbert-Elliot channel is used to model the time varying receiver signal amplitude fading in single-path mobile communication system or in multi-path communication system.

First, we assume that there is only one line-of-sight path between the transmitter and the receiver over which the signal is transmitted. The transmitted signal experiences no amplitude attenuation, but the receiver motion can change the communication environment geometrically which may have a different level of additive white Gaussian noise and therefore a different signal to noise ratio. For simplicity, we take the changing of the noise level as a kind of time varying amplitude

fading since the different signal to noise ratio can be caused in another equivalent situation where the signal experiences amplitude fading while the additive white Gaussian noise stays still. The Gilbert-Elliot channel models the channel with strong additive white Gaussian noise by the bad state and models the channel with low additive white Gaussian noise by the good state.

Second, we assume that signals are transmitted to the receiver over multiple propagation paths, but all components of each transmitted symbol arrive at the receiver without any excess delay[18]. Under this assumption, transmitted signals experience certain level of amplitude fading at any time slot. The object motion changes communication environments geometrically, which leads to different amplitude fading level. Similarly, the channel with deep fading is modelled by the bad state and the channel with light fading is modelled by the good state.

Referring back to the scalar model of the time varying communication system in (**??**), the geometric change of the communication environment results in time varying amplitude fading, which leads to an increase or an decrease of the value of $\mathcal{G}$. Therefore, the Gilbert-Elliot channel model can be expressed by

$$\mathcal{Y} = \mathcal{G}\mathcal{X} + \mathcal{Z}, \tag{5.2}$$

where $G \in \{G_1, G_2\}$.

5.1.3 The Limitation of The Gilbert-Elliot Channel Model

In this subsection, we show the incompleteness of the Gilbert-Elliot model in modelling mobile communication channels. The information loss caused by the Doppler shift has not been covered. This incompleteness can be shown via the simplest time varying communication scenario where there is only one transmitter and one receiver, and the signal travels over one single path. We can show that, the

Doppler phase shift cannot be removed in demodulating the carrier signal to base band signal and will cause errors in deciding the transmitted symbol conditioned on demodulated signals. This error has not been covered by the Gilbert-Elliot channel model.

We assume that the receiver is moving at a velocity v. Assuming the wave arriving at the receiver at an angle α to the direction of motion, the Doppler shift in Hertz carried in the carrier signal can be expressed by

$$f_d = \frac{v}{\lambda} \cos \alpha = f_m \cos \alpha, \tag{5.3}$$

where λ is the wavelength and f_m is the maximum doppler shift. The transmitted signal can be expressed by

$$s_T(t) = s(t) \left(\cos(2\pi f_c t) + i \sin(2\pi f_c t) \right), \tag{5.4}$$

The received signal can be expressed by

$$s_R = s_T(t) \left(\cos(2\pi f_c t + 2\pi f_d t) + i \cdot \sin(2\pi f_c t + 2\pi f_d t) \right), \tag{5.5}$$

where s_T is the base band transmitted signal and f_c is the carrier frequency; i is the imaginary unit. The amplitude of the received signal, $|s_R(t)|$, is always equal to that of the transmitted signal, $|s_T(t)|$, shown as follows:

$$\begin{aligned} |s_R(t)| &= |s_T(t)| \sqrt{\cos^2(2\pi f_c t + 2\pi f_d t) + \sin^2(2\pi f_c t + 2\pi f_d t)} \\ &= |s_T(t)|. \end{aligned} \tag{5.6}$$

When this communication channel is modeled by the Gilbert-Elliot channel , the transition probability from the good state to the bad state approaches zero and the channel always stays in the good state. However, we can show that even with

the unweaken received power amplitude, additional bits are required to recover the transmitted symbol. We assume that binary phase shift keying is used here and the signal is recovered using the real part of the received signal[18],

$$R\left\{\left|s_R\left(t\right)\right|\right\} \quad = \quad s_T\left(t\right)\cos\left(2\pi f_c t + 2\pi f_d t\right). \tag{5.7}$$

Multiplying (5.7) by $\cos\left(2\pi f_c t\right)$, we have,

$$\begin{aligned} R\left\{\left|s_R\left(t\right)\right|\right\}\cos\left(2\pi f_c t\right) &= s_T\left(t\right)\cos\left(2\pi f_c t + 2\pi f_d t\right)\cos\left(2\pi f_c t\right) \\ &= \frac{1}{2}s_T\left(t\right)\left(\cos\left(4\pi f_c t + 2\pi f_d t\right) + \cos\left(2\pi f_d t\right)\right). \end{aligned} \tag{5.8}$$

Removing the high frequency part of (5.8) by a low-pass filter[18], we have the recovered signal:

$$\hat{s}_R\left(t\right) \quad = \quad s_T\left(t\right)\cos\left(2\pi f_d t\right). \tag{5.9}$$

The information loss in signal recovery till here is modeled well by the good state of the Gilbert-Elliot channel. However, it is obvious that detection errors might occur due to the Doppler phase shift, $2\pi f_d t$. In slow time varying channels, the Doppler phase shift accumulates very slowly and can be removed or compensated via frequency synchronization. In other words, the Doppler phase shift can still be easily observed using conventional techniques. The Gilbert-Elliot model is still accurate. However, when the receiver is moving in a large enough velocity, for example, in a car. The Doppler phase shift accumulates too fast and becomes very difficult to detect. The consistent effect of the Doppler phase shift is no longer negligible in modelling. Therefore, a model for the relevant information loss is required.

5.2 Modeling The Mobile Communication Channel by The Time-Varying Binary Symmetric Channel

In this section, we model the Doppler phase shift in mobile communications by the time-varying-BSC. We first introduce in detail the Doppler phase shift carried by signals in both single-path time varying channels and multi-path time varying channels. Second, we clarify the physical significance of the time-varying-BSC by matching parameters of the model with the relevant factors of physical channels. Third, we prove that the differential encoder and the differential decoder can synchronise the state detection and signal recovery of the time-varying-BSC, and preserve the information capacity of the model.

5.2.1 Time Varying Binary Symmetric Channel Model

The time-varying-BSC in this thesis belongs to the general set of uniformly symmetric variable noise finite state Markov channels [54–57]. An example of the time-varying-BSC is shown in Fig. 5.1(b). The time-varying-BSC comprises of two BSCs, which are called by non-inverted BSC and inverted BSC. The non-inverted BSC and the inverted BSC have opposite crossover probabilities denoted by p and $1-p$, respectively. And the two BSCs have the same probability of transitioning to each other, denoted by q.

5.2.2 Doppler Phase Shift

In this subsection, we introduce the Doppler phase shift in the received carrier signal in both single-path time varying channels and multi-path time varying channels.

First, we assume the signal only propagates over one line-of-sight path between the transmitter and the receiver. The received carrier signal can be expressed by,

$$s_R(t) = s(t)\left(\cos\left(2\pi f_c t + 2\pi f_d t\right) + i\sin\left(2\pi f_c t + 2\pi f_d t\right)\right), \tag{5.10}$$

where the Doppler shift in Hertz can be expressed by

$$f_d = \frac{v}{\lambda}\cos\alpha = f_m \cos\alpha, \tag{5.11}$$

where v is the motion velocity, λ is the wavelength and f_m is the maximum Doppler shift.

Second, we assume signals are transmitted over multiple paths, in which there is no line-of-sight path between the transmitter and the receiver. We show in the following that Doppler shifts of most carrier signals are within two very small intervals, $[-f_m, -f_m + \rho)$ and $(f_m - \rho, f_m]$, ρ is a small positive real number.

Following the assumption of Clarke's model in [58], antennas of the transmitter are vertically polarized. There are N azimuthal plane waves arriving the receiver with arbitrary carrier phases, arbitrary azimuthal angles of arrival, and equal average amplitude [58]. The power spectrum of the received carrier signal of the Clarke's model has been analyzed in [58]. With the average signal amplitude denoted by U, the total received power can be expressed by

$$P_r = \int_0^{2\pi} U\Pr(\alpha)d\alpha, \tag{5.12}$$

where $U\Pr(\alpha)d\alpha$ is the differential variation of received power with angle. With the power-spectral density of the received signal denoted by $\Upsilon(f)$, the total received power can also be expressed by

$$P_r = \int_{f_c - f_m}^{f_c + f_m} \Upsilon(f)df. \tag{5.13}$$

Combining (5.11)(5.12) and (5.13), we have

$$\Upsilon(f) = \frac{U\left(\Pr\left(\alpha\right) + \Pr\left(-\alpha\right)\right)}{f_m\sqrt{1-\left(\frac{f-f_c}{f_m}\right)^2}}. \tag{5.14}$$

Since α is uniformly distributed over 0 to 2π and $\Pr\left(\alpha\right) = \frac{1}{2\pi}$, the power spectrum becomes

$$\Upsilon(f) = \frac{U}{\pi f_m\sqrt{1-\left(\frac{f-f_c}{f_m}\right)^2}}. \tag{5.15}$$

Equation (5.15) shows infinite power spectral density at $f = f_c \pm f_m$. This means that the Doppler shifts of most arrived waves are within two small ranges, $(-f_m, -f_m + \rho)$ and $(f_m - \rho, f_m)$.

5.2.3 Physical Significance of The Time-Varying Binary Symmetric Channel

In this section, we introduce the time-varying-BSC for the Doppler phase shift. Instant phases of the received carrier components are random and impossible to be modeled. We dig into the phase changes in last two subsections and show their underlying constant changing rate. The time-varying-BSC is actually modeling this changing rate.

Under the single-path assumption, the Doppler shift in the received carrier signal is fixed, f_d or $-f_d$, which depends on the motion direction. We assume the Doppler phase shift starts from 0. Because the Doppler phase shift is equal to $2\pi f_d t$, which is a linear function of t, it rotates on the Cartesian coordinate plan at a constant rate $2\pi f_d$, shown in Fig.5.2. Referring to equation (5.9), the demodulated signal is

$$\hat{s}_R\left(t\right) \;=\; s_T\left(t\right)\cos\left(2\pi f_d t\right). \tag{5.16}$$

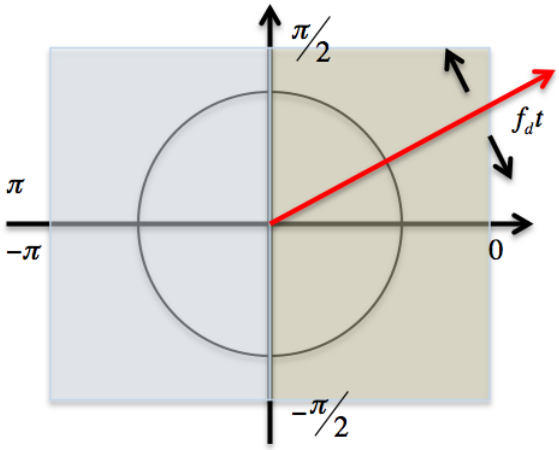


FIGURE 5.2: The Doppler phase shift.

When the phase shift stays in the right side of the Cartesian coordinate plane, $\cos(2\pi f_d t)$ is positive and the signal is in-phase. Assuming binary phase shift keying scheme is used, the transmitted symbol can be recovered without any additional errors. The mathematical system model can be expressed by

$$\mathcal{Y} = \mathcal{G}\mathcal{X} + \mathcal{Z}. \tag{5.17}$$

The channel during the time when the signal stay in-phase can be modeled in the time-varying-BSC by a BSC with the crossover probability, p, shown in Fig. 4.2(b).

When the phase shift stays in the left side of the coordinate plane, the signal would stay out-of-phase. Assuming binary phase shift keying is used, an opposite transmitted symbol will be obtained. The mathematical system model becomes,

$$\begin{aligned} \mathcal{Y} &= \mathcal{G}(-\mathcal{X}) + \mathcal{Z} \\ &= -\mathcal{G}\mathcal{X} + \mathcal{Z}. \end{aligned} \tag{5.18}$$

The channel during this time can be modeled by an inverted BSC, with the probability, $1-p$. Since the phase shift stays in either side of the coordinate plane for

an equal length of time, the transition probabilities, q, between the non-inverted BSC and the inverted BSC are the same, value of which are determined by the motion velocity.

Under the multi-path assumption, the mapping between the physical channel and the time-varying-BSC is not straightforward and it is necessary to clarify following relationships,

1. Excess delay and base band signal: this thesis considers the very narrow base band signal, the bandwidth of which is smaller than ten percent of the carrier frequency. Its symbol duration is long enough that the excess delay encountered by the multi-path components cannot cause inter-symbol interference. For simplicity, we can assume that all multi-path components arrive at the receiver simultaneously.

2. Excess delay and carrier signal: The excess delay is large enough to affect the phase of the carrier signal, which has a very high frequency and a very small wavelength. Usually, the excess delay can be assumed to be a random process and therefore the carrier signal components have random phases, which are uniformly distributed on the interval $(0, 2\pi]$ [18]. It should be noted that we assume that the excess delays encountered by the multi-path components vary at a much slower rate than the Doppler phase shift does. For simplicity, we can assume the excess delays for the multi-path components are stationary for a certain block of time.

3. Doppler phase shift and Carrier signal: The Doppler phase shift is carried by the carrier signal and mixed with the phase shift caused by the excess delay. Both phase shifts continue to exist after the demodulation. However, we can expect the excess-delay-caused phase shift to be eliminated from the demodulated base band signal due to its stationary, but the Doppler phase shift will stay and affect the subsequent detection of the transmitted symbol.

It is fortunate that the Doppler phase shift in the multi-path components is not random and we show in this subsection its effect on symbol detection can be modeled by the time-varying-BSC.

4. Doppler phase shift and base band signal: The Doppler phase shift stays with the modulated base band signal. It should be noted that the Doppler phase shift does not affect the phase of base band signal, which is a combination of random phases of many single-frequency signals and does not have physical meaning. This means the Doppler phase shift does not change the shape of the base band signal. The Doppler phase shift can change the amplitude of the demodulated base band signal and it can be seen as a scalar to the base band signal, which takes values from $[-1, 0) \cup (0, 1]$.

Given the clarifications listed above, the physical significance of the time-varying-BSC under the multi-path propagation assumption can be shown. Based on the analysis in [58, 59], received carrier signals at time t under the multi-path assumption can be expressed by

$$s_R(t) = T_c \cos(2\pi f_c t) - i \cdot T_s \sin(2\pi f_c t), \tag{5.19}$$

where

$$T_c(t) = E_0 \sum_{n=1}^{N} C_n \cos(2\pi f_{d-n} t + \phi), \tag{5.20}$$

and

$$T_s(t) = E_0 \sum_{n=1}^{N} C_n \sin(2\pi f_{d-n} t + \phi). \tag{5.21}$$

where E_0 is the amplitude of local average E-field and C_n is real random variable representing the amplitude of nth wave; f_{d-n} is the Doppler shift of the nth wave. $T_c(t)$ and $T_s(t)$ are both Gaussian random processes at any time t [58, 59].

We assume that the signal is recovered using the real part of (5.19). It can be expected that $\cos(2\pi f_c t)$ is eliminated in demodulation and the transmitted symbol are decided based on $T_c(t)$,

$$T_c(t) = E_0 \sum_{n=1}^{N} C_n \cos(2\pi f_{d-n} t + \phi). \tag{5.22}$$

However, we have shown in section IV-A that Doppler shifts of most arrived waves are within two small ranges, $[-f_m, -f_m + \rho)$ and $(f_m - \rho, f_m]$, where ρ is a very small positive number. Therefore, the Doppler phase shift are also within two small intervals $[-2\pi f_m, -2\pi f_m + 2\pi\rho)$ and $(2\pi f_m - 2\pi\rho, 2\pi f_m]$. In digital communications, the Doppler phase shifts are discretized into equal segments, each of which are represented by one phase shift. It is reasonable to assume that the segment is much larger than the interval and thus all Doppler phase shift are represented by $2\pi f_m t$ or $-2\pi f_m t$. Therefore, $T_c(t)$ can be expressed by,

$$T_c(t) = E_0 \sum_{n=1}^{N} C_n \cos(2\pi f_m t + \phi), \tag{5.23}$$

or

$$T_c(t) = E_0 \sum_{n=1}^{N} C_n \cos(-2\pi f_m t + \phi). \tag{5.24}$$

It can be expected that ϕ of these multi-path components can be removed due to its stationary, as shown in (5.7), (5.8) and (5.9). $T_c(t)$ becomes

$$T_c(t) = E_0 \sum_{n=1}^{N} C_n \cos(-2\pi f_m t). \tag{5.25}$$

Theoretically, the transmitted symbol is decided by the value of $T_c(t)$. When $T_c(t)$ is higher than the threshold 0, the symbol of 1 is transmitted; when $T_c(t)$ is smaller than the threshold 0, the symbol of 0 is transmitted. We also assume the

Doppler phase shift starts from 0. It rotates on the Cartesian coordinate plane at a constant rate $2\pi f_m$ or $-2\pi f_m$. The transmitted symbol can be recovered without any additional errors when the Doppler phase shift is in the right side of the Cartesian plane and $\cos(-2\pi f_m t)$ or $\cos(2\pi f_m t)$ is positive. The mathematical system model can be expressed by

$$\mathcal{Y} = \mathcal{G}\mathcal{X} + \mathcal{Z}. \tag{5.26}$$

The channel during the time is modeled in the time-varying-BSC by the non-inverted BSC.

When the Doppler phase shift stays in the left side of the Cartesian plane, $\cos(-2\pi f_m t)$ or $\cos(2\pi f_m t)$ is negative and the opposite of the transmitted symbol will be obtained if there is no additional detecting technique being used for the Doppler phase shift. The mathematical system model becomes,

$$\mathcal{Y} = -\mathcal{G}\mathcal{X} + \mathcal{Z}. \tag{5.27}$$

The channel during this time can be modeled by the inverted BSC. Since the phase shift rotates at a constant rate, the channel stays in either state for an equal length of time, the transition probabilities between the non-inverted BSC and the inverted BSC are the same.

5.2.4 Differential Encoder and Differential Decoder

In this subsection, we show that the differential encoder and the differential decoder for the time-varying-BSC can synchronize the state estimation and signal detection without any additional information loss. As a result, the encoder and the decoder can achieve the information capacity of the time-varying-BSC. Results in this subsection prove indirectly the hypothesis made in section 3.4.2 that the main

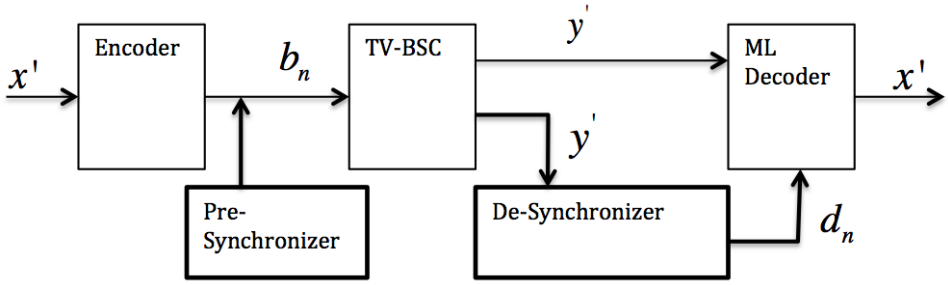


FIGURE 5.3: The channel model of time-varying-BSC with the synchronizer.

difficulty in performing optimum detection for time varying channel model, including the Gilbert-Elliot channel and time-varying-BSC, comes from synchronizing the state detection and signal detection.

5.2.4.1 Synchronizing The Signal Recovery with The State Detection

In chapter 3, we have clarified that the main difficulty to achieving information capacity of time varying channels comes form synchronizing the state detection and signal detection. Before introducing the differential encoder and the differential decoder, we show physically how they synchronize the signal detection and state estimation. Referred to the system model in Fig. 5.3, the decision-feedback decoder for the Gilbert-Elliot channel performs state estimation in the receiver side only. In the differential decoding scheme, part of the work of state estimation is moved to the transmitter. A pre-synchronizer is set up in the transmitter side. Each transmitted symbol is restructured when it is passed through the pre-synchronizer and become sensitive to the varying of the channel state. At the received side, the signal carrying the state information is passed through the de-synchronizer, where the state information is determined and removed from the signal. The transmitted signal is thereafter able to be recovered based on the state information using a conventional decoder, such as the maximum-likelihood

decoder. Since the state information carried by the received signal is instant, no delay occurs between the state estimation and the signal detection.

The advantage of the differential encoder and the differential decoder on synchronization can also be shown mathematically. Referred back to Fig.4.2(b), the time-varying-BSC can be described by [40],

$$y'_n = s_n \oplus x'_n \oplus v_n, \tag{5.28}$$

$$s_n = s_{n-1} \oplus \eta_n \tag{5.29}$$

where y'_n and x'_n denote the n-th bit of the input sequence and the output sequence, respectively; s_n is the channel state information, during which x'_n is transmitted, and η_n is the channel varying process; $\eta_n \in \{0, 1\}$ and $\Pr(\eta_n) = q$; v_n denotes the channel noise process and $\Pr(v_n = 1) = p$.

When the differential encoder and the differential decoder are employed, the transmitted signal is encoded before transmission,

$$b_n = x'_n \oplus b_{n-1}, \tag{5.30}$$

where b_0 is the reference bit. The received signal can be expressed by

$$y'_n = s_n \oplus b_n \oplus v_n \tag{5.31}$$

$$= (s_{n-1} \oplus \eta_n) \oplus (x'_n \oplus b_{n-1}) \oplus v_n. \tag{5.32}$$

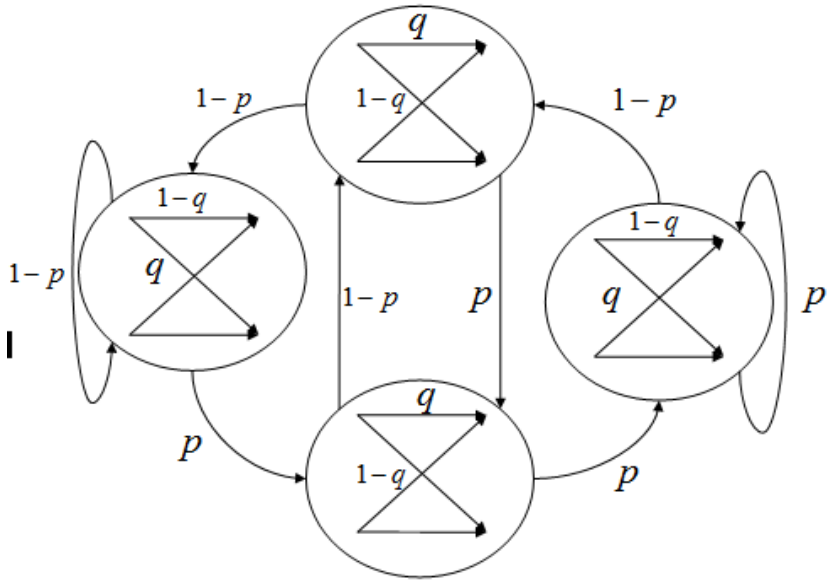


FIGURE 5.4: The new equivalent Markov model of the time-varying-BSC.

When the received signal is passed through the differential decoder, it is decoded by

$$\begin{aligned} d_n &= y'_n \oplus y'_{n-1} \\ &= (s_n \oplus b_n \oplus v_n) \oplus (s_{n-1} \oplus b_{n-1} \oplus v_{n-1}) \\ &= ((s_{n-1} \oplus \eta_n) \oplus (x'_n \oplus b_{n-1}) \oplus v_n) \oplus (s_{n-1} \oplus b_{n-1} \oplus v_{n-1}) \\ &= x'_n \oplus \eta_n \oplus (v_n \oplus v_{n-1}) \,. \end{aligned} \tag{5.33}$$

From (5.33), the instant varying information η_n is carried. What is innovative is that $\eta_n = s_n \oplus s_{n-1}$ is memoryless and the information can be processed using conventional technique in the literature for memoryless channels. Based on instant state information, the transmitted signal can be further recovered.

5.2.4.2 Achieving The Information Capacity of Time-Varying-BSC by The Differential Encoder and The Differential Decoder

In this subsection, we prove that the differential encoder and the differential decoder can achieve the information capacity for the time-varying-BSC. From (5.33), the cascade of the differential encoder, the time-varying-BSC and the differential decoder have formed a new Markov channel. We prove that the encoder and the decoder do not lead to any capacity loss and these two Markov channels have the same information capacity,

The new Markov channel has been defined in [40], shown in Fig. 5.4. The channel state at time slot k is denoted by $s_k^e = c_{i,j}^e = [v_k = i, v_{k-1} = j]$. The initial state distribution is denoted by Q_0^e, of which the $(2i+j)$th element is $\Pr(s_0^e = c_{i,j}^e) = p_i \cdot p_j$ [40].

The transition probability of the new channel is denoted by

$$q_{(2i+j)(2m+n)} = \Pr\left(s_k^e = c_{m,n}^e | s_k^e = c_{i,j}^e\right) = p_m, \tag{5.34}$$

where if $i = n$, $p_m = \Pr(v_k = m) = 1 - p$ for $m = 0$ and $p_m = \Pr(v_k = m) = p$ for $m = 1$; if $i \neq n$, $p_m = 0$.

The equivalent channel state law is denoted by

$$\begin{aligned} p_{(m)(n)}^e &= p_{(m)(2i+j)}^e = \Pr\left(z^e = m | s_k^e = c_{i,j}^e\right) \\ &= \Pr\left(\eta_k = i \oplus j \oplus m\right) = q_{i \oplus j \oplus m} = q_l, \end{aligned} \tag{5.35}$$

where $z^e = x'_n \oplus d_n$ is the error function for the new channel; $q_l = 1 - q$ for $l = 0$ and $q_l = q$ for $l = 1$.

Theorem 5.1. *The differential encoder and the differential decoder are able to achieve the information capacity of the time-varying-BSC.*

Proof We prove the theorem by showing that the new Markov Channel, shown in Fig.5.4 and the time-varying-BSC have equal information capacity.

According to [21], the information capacity of time-varying-BSC and the new Markov channel can be expressed by

$$C = 1 - \lim_{N\to\infty} E\left(H\left(\epsilon_N\right)\right), \tag{5.36}$$

and,

$$C = 1 - \lim_{N\to\infty} E\left(H\left(\epsilon_N^e\right)\right), \tag{5.37}$$

respectively, where ϵ_N and $\epsilon_N^e$is the channel error probability at the Nth use of the time-varying-BSC and the new Markov channel respectively, and $H(\cdot)$ is the binary entropy function; $E(\cdot)$ denotes the expectation.

(5.36) and (5.37) can also be expressed by the following equations due to the stationary stochastic process[21, 53],

$$\begin{aligned} C &= 1 - \lim_{N\to\infty} \frac{1}{N}\left(H\left(z^N\right)\right) \\ &= 1 - \lim_{N\to\infty} \left(-\frac{1}{N}\log_2\left(\Pr\left(z^N\right)\right)\right), \end{aligned} \tag{5.38}$$

and

$$\begin{aligned} C &= 1 - \lim_{N\to\infty} \frac{1}{N}\left(H\left(z^{(e)^N}\right)\right) \\ &= 1 - \lim_{N\to\infty} \left(-\frac{1}{N}\log_2\left(\Pr\left(z^{(e)^N}\right)\right)\right). \end{aligned} \tag{5.39}$$

In the following, we show that the $\Pr(z^N) = \Pr(z^{(e)^N})$.

For the new Markov channel,

$$\begin{aligned}
\Pr\left(z^{(e)^N}\right) &= \sum_{i=0}^{1}\sum_{j=0}^{1}\Pr\left(z^{(e)^N}|s_0^e\right)\Pr\left(s^{(e)^N}|s_0^e=c_{i,j}^e\right)\\
&= \sum_{i=0}^{1}\sum_{j=0}^{1}\Pr\left(z^{(e)^N}|s_0^e==c_{i,j}^e\right)(p_i\cdot p_j)\\
&= \sum_{i=0}^{1}\sum_{j=0}^{1}\left[\sum_{k=0}^{1}\sum_{l=0}^{1}\Pr\left(z^{(e)^N},s_N^e=c_{k,l}^e|s_0^e=c_{i,j}^e\right)\right](p_i\cdot p_j)\\
&= \sum_{i=0}^{1}p_i\sum_{j=0}^{1}p_j\left[\sum_{k=0}^{1}\sum_{l=0}^{1}\Pr\left(z^{(e)^N},s_N^e=c_{k,l}^e|s_0^e=c_{i,j}^e\right)\right]\\
&= \sum_{i=0}^{1}p_i\cdot 1\cdot\left[\sum_{k=0}^{1}\sum_{l=0}^{1}\Pr\left(z^{(e)^N},s_N^e=c_{k,l}^e|s_0^e=c_{i,0}^e\right)\right]\\
&= \sum_{i=0}^{1}p_i\cdot 1\cdot\left[\sum_{k=0}^{1}\underline{\left[\sum_{l=0}^{1}\Pr\left(z^{(e)^N},s_N^e=c_{k,l}^e|s_0^e=c_{i,0}^e\right)\right]}\right]. \qquad (5.40)
\end{aligned}$$

The underlined part in equation (5.40) can be expressed as an backward recursion, of which the first round is shown as follows:

$$\begin{aligned}
\sum_{l=0}^{1}\Pr\left(z^{(e)^N},s_N^e=c_{k,l}^e|s_0^e\right) &= \sum_{l=0}^{1}\Pr\left(z_N^{(e)},z^{(e)^{N-1}},s_N^e=c_{k,l}^e|s_0^e=c_{i,0}^e\right)\\
&= \sum_{l=0}^{1}\sum_{n=0}^{1}\sum_{t=0}^{1}\Pr\left(z^{(e)^{N-1}},s_{N-1}^e=c_{n,t}^e|s_0^e=c_{i,0}^e\right)\\
&\quad \Pr\left(z_N^{(e)}|s_N^e=c_{k,l}^e\right)\Pr\left(c_{k,l}^e|c_{n,t}^e\right)\\
&= \sum_{l=0}^{1}\sum_{n=0}^{1}\sum_{t=0}^{1}\Pr\left(z^{(e)^{N-1}},s_{N-1}^e=c_{n,t}^e|s_0^e=c_{i,0}^e\right)\\
&\quad \Pr\left(z_N^{(e)}|s_N^e=c_{k,l}^e\right)\cdot p_k\\
&= p_k\sum_{l=0}^{1}q_{(z^e|c_{k,l}^e)}\sum_{t=0}^{1}\Pr\left(z^{(e)^{N-1}},\underline{s_{N-1}^e=c_{n,t}^e}|s_0^e=c_{i,0}^e\right)\\
&= p_k\sum_{l=0}^{1}q_{(z^e|c_{k,l}^e)}\sum_{t=0}^{1}\Pr\left(z^{(e)^{N-1}},\underline{s_{N-1}^e=c_{l,t}^e}|s_0^e=c_{i,0}^e\right),
\end{aligned} \qquad (5.41)$$

where the last equality is due to the fact that l has to be equal to n for the recursion to continue. The recursion ends at $\Pr(z^{(e)^N}, s_1^e = c_{k,l}^e | s_0^e)$, (5.40) can then be expressed by

$$\begin{aligned}\Pr\left(z^{(e)^N}\right) &= \underline{\sum_{i=0}^{1} p_i} \cdot \sum_{k=0}^{1} p_k \sum_{l=0}^{1} q_{(z^e|c_{k,l}^e)} \cdot p_l \sum_{l=0}^{1} q_{(z^e|c_{l,t}^e)} \cdots \\ &\quad p_m \sum_{l=0}^{1} q_{(z^e|c_{m,v}^e)} \cdot p_v \cdot \underline{q_{(z^e|c_{v,i}^e)}} \\ &= \sum_{k=0}^{1} p_k \sum_{l=0}^{1} q_{(z^e|c_{k,l}^e)} \cdot p_l \sum_{l=0}^{1} q_{(z^e|c_{l,t}^e)} \cdots \\ &\quad p_m \sum_{v=0}^{1} q_{(z^e|c_{m,v}^e)} \cdot p_v \cdot \underline{\sum_{i=0}^{1} p_i \cdot q_{(z^e|c_{v,i}^e)}}. \end{aligned} \tag{5.42}$$

For the time-varying-BSC, we calculate $\Pr(z^{N+1})$ instead of $\Pr(z^N)$. This is because the new Markov channel has included the differential encoder and the differential decoder, which require one reference bit to be transmitted reliably before

the information transfer. Therefore, $\Pr\left(z^{(e)^N}\right)$ in (5.42) is actually the error probability of the $(N+1)$th information bit of the new Markov channel.

$$\begin{aligned}
\Pr\left(z^{N+1}\right) &= \sum_{i=0}^{1} \Pr\left(z_{N+1}=m, z^N | s_0=i\right) \Pr\left(s_0=i\right) \\
&= 0.5 \cdot \sum_{i=0}^{1} \Pr\left(z_{N+1}=m, z^N | s_0=i\right) \\
&= 0.5 \cdot \sum_{i=0}^{1}\sum_{k=0}^{1} \Pr\left(z_{N+1}=m, z^N, s_N=k | s_0=i\right) \\
&= 0.5 \cdot \sum_{i=0}^{1}\sum_{k=0}^{1} \Pr\left(z^N, s_N=k | s_0=i\right) \\
&\quad \sum_{l=0}^{1} \Pr\left(z_{N+1}=m | s_{N+1}=l\right) \underline{\Pr\left(s_{N+1}=l | s_N=k\right)} \\
&= 0.5 \cdot \sum_{i=0}^{1}\sum_{k=0}^{1} \Pr\left(z^N, s_N=k | s_0=i\right) \\
&\quad \sum_{l=0}^{1} \Pr\left(z_{N+1}=m | s_{N+1}=l\right) \underline{q_{l\oplus k}} \\
&= 0.5 \cdot \sum_{i=0}^{1}\sum_{k=0}^{1} \Pr\left(z^N, s_N=k | s_0=i\right) \sum_{l=0}^{1} p_{m\oplus l} q_{l\oplus k} \\
&= 0.5 \sum_{l=0}^{1} p_{m\oplus l} \cdot \sum_{k=0}^{1} q_{l\oplus k} \sum_{i=0}^{1} \Pr\left(z^N, s_N=k | s_0=i\right). \qquad (5.43)
\end{aligned}$$

The same rule in (5.41) applies.

$$\begin{aligned}
\sum_{i=0}^{1} \Pr\left(z^N, s_N=k | s_0=i\right) &= \sum_{i=0}^{1}\sum_{n=0}^{1} \Pr\left(z_N, z^{N-1}, s_N=k, s_{N-1}=n | s_0=i\right) \\
&= \sum_{i=0}^{1}\sum_{n=0}^{1} \Pr\left(z^{N-1}, s_{N-1}=n | s_0=i\right) \cdot \\
&\quad \underline{\Pr\left(z_N | s_N=k\right)} \Pr\left(s_N=k | s_{N-1}=n\right) \\
&= \underline{p_{(z,k)}} \sum_{i=0}^{1}\sum_{n=0}^{1} \Pr\left(z^{N-1}, s_{N-1}=n | s_0=i\right) \cdot \\
&\quad \Pr\left(s_N=k | s_{N-1}=n\right) \\
&= p_{(z,k)} \sum_{n=0}^{1} q_{k\oplus n} \sum_{i=0}^{1} \Pr\left(z^{N-1}, s_{N-1}=n | s_0=i\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Pr\left(z^{N+1}\right) &= \sum_{i=0}^{1} \Pr\left(z_{N+1}=m, z^N | s_0=i\right) \Pr\left(s_0=i\right) \\
&= 0.5 \sum_{l=0}^{1} p_{m\oplus l} \cdot \sum_{k=0}^{1} q_{l\oplus k} \sum_{i=0}^{1} \Pr\left(z^N, s_N=k | s_0=i\right) \\
&= 0.5 \sum_{l=0}^{1} p_{m\oplus l} \cdot \sum_{k=0}^{1} q_{l\oplus k} \cdot p_{(z,k)} \cdots \sum_{v=0}^{1} q_{r\oplus v} \cdot p_{(z,v)} \cdot \underline{\sum_{i=0}^{1} q_{k\oplus i}} \\
&= 0.5 \sum_{l=0}^{1} p_{m\oplus l} \cdot \sum_{k=0}^{1} q_{l\oplus k} \cdot p_{(z,k)} \cdots \sum_{v=0}^{1} q_{r\oplus v} \cdot p_{(z,v)} \cdot \underline{1}. \qquad (5.44)
\end{aligned}$$

Because of the constellation symmetric that, $p_i = 1-p$ for $i=0$, $p_i = p$ for $i=1$, and $q_i = 1-q$ for $i=0$, $q_i = q$ for $i=1$, (5.44) and (5.42) are the same combinations of the same p_i and q_j multiplication except for the 0.5 in (5.44). Refer back to (5.39)

$$C = 1 - \lim_{N\to\infty} \left(-\frac{1}{N} \log_2 \left(\Pr\left(z^N\right) \right) \right). \qquad (5.45)$$

Since $\Pr(z^{N+1}) = 0.5 \cdot \Pr(z^{(e)^N})$,

$$\begin{aligned}
C &= 1 - \lim_{N+1\to\infty} \left(-\frac{1}{N+1} \log_2 \left(\Pr\left(z^{N+1}\right) \right) \right) \\
&= 1 - \lim_{N+1\to\infty} \left(-\frac{1}{N+1} \log_2 \left(0.5 \cdot \Pr\left(z^{(e)^N}\right) \right) \right) \\
&= 1 - \lim_{N+1\to\infty} \left(-\frac{1}{N+1} \log_2 \left(\cdot \Pr\left(z^{(e)^N}\right) \right) + \frac{1}{N+1} \right) \qquad (5.46) \\
&= 1 - \lim_{N+1\to\infty} \left(-\frac{1}{N+1} \log_2 \left(\cdot \Pr\left(z^{(e)^N}\right) \right) \right) \qquad (5.47) \\
&\Longrightarrow 1 - \lim_{N+1\to\infty} \left(-\frac{1}{N} \log_2 \left(\cdot \Pr\left(z^{(e)^N}\right) \right) \right) = C^e, \qquad (5.48)
\end{aligned}$$

where the third and fourth equalities both follow from the fact that the real number 1 cannot affect the result with N goes to infinity.

5.3 Chapter Conclusion

In this chapter, we clarify that the Doppler phase shift is a major cause for the channel state variation. The Gilbert-Elliot model does not take it into consideration. The model therefore is incomplete for time varying channel. We introduce the time-varying BSC which complements the Gilbert-Elliot model. More importantly, we successfully avoid the issue of synchonizing state estimation and signal detection for time varying channels. And we achieve the information capacity for the new model.

Chapter 6

Conclusions and Future Work

6.1 Conclusion

The thesis serves two aims. The first aim is to differentiate time varying channels and time invariant channels, and to clarify why optimum decoders for the former are not capacity achieving for the latter. The second aim is to provide a solution of achieving high rate data transmission over time varying mobile communication channels. These two aims have both been met.

1. Time varying communication channels and time invariant communication channels are fundamentally different in terms of maximizing the mutual information rate over channel input probability distribution. For time invariant channels, the perfect channel state information is assumed at the receiver. The more information are transmitted, the more information are received. The mutual information rate is maximized by channel inputs of maximum information entropy. However, for time varying channels, we can invest certain amount of information redundancy of channel input signals in channel state estimation. The investment decreases the amount of information transmitted over the channel. But on the other hand, it decreases the

uncertainty of channel state distribution and might lead to higher mutual information rate in the long term. The optimum trade-off between the two effects will lead to maximization of the mutual information rate.

2. The time varying communication channels and time invariant channels are fundamentally different in terms of signal processing. For time invariant channels, estimating channel state and extracting signal in the presence of AWGN (i.e., signal detection) are done independently in succession. Since the channel state information is time invariant. We assume implicitly that channel state estimation and signal detection are already synchronized. But for time varying channels, performing these two actions one after the other independently is not able to achieve the information capacity of the channel. There are always delays between the two actions and extra communication resource are needed for synchronization. We prove the combining channel state estimation and signal detection can save the communication resource for synchronization. And combining these two actions is also a sufficient condition for achieving the information capacity of time varying channels.

6.2 Future Work

The original purpose of this thesis is to correct a well accepted result in the literature for time varying communication channels. However, we unexpectedly end up with attacking one of the most important assumptions of the root communication theory, the Shannon information theory. It has been clear to us that the Shannon information theory are developed for time invariant communication channels only. It has started limiting the development of mobile communication technology, where **time variation** is highlighted. It is the market demand of 5G that asks for

an extension of the Shannon information theory to cover time varying communication channels. In chapter 5 of thesis, we made the first step only. Tremendous research efforts are required in future, serving the following two purposes:

1. To define the information capacity of time varying communication channels and have the information capacity expressed in a closed formula.

2. To exploit the information capacity of time varying channels defined by the new information theory. In other words, the second purpose is to develop 5G mobile communication technology based on the information theory for time varying communication channels.

The future work serving the first purpose is a continuation of the analysis in this thesis. The future work serving the second purpose required cooperative efforts of researcher in mobile communication all over the world. We discuss in this chapter only the future work serving the first purpose. In order to define the information capacity of time varying channels, we need to revisit almost all assumptions and definitions in the Shannon information theory.

6.2.1 Assumptions of The Shannon Wiener Theory for Time Invariable Communication Channels

Shannon Wiener theory of communications starts with definition of communication channel as the system of three fundamental uncertainties: uncertainty of information transmitted $\mathcal{X}$, uncertainty of the observation noise $\mathcal{Z}$, and uncertainty of channel state information $\mathcal{G}$.

$$\mathcal{Y} = \mathcal{G}\mathcal{X} + \mathcal{Z}. \tag{6.1}$$

The Shannon Wiener theory of communications reduces the joint three dimensional communication channel uncertainty model (uncertainty of information transmitted, uncertainty of the observation noise, and uncertainty of channel state) to essentially two communication problems with two uncertainties. We call this reduced model as Shannon Wiener Separation model. The two communication problems with two uncertainties are as follows:

1. Estimation of channel state information $\mathcal{G}$, in the presence of noise $\mathcal{Z}$, training signal being known.
2. Detection of uncertain transmitted information $\mathcal{X}$, in the presence of noise $\mathcal{Z}$, channel state information being measured and certain from the previous measurement.

This approach is correct in the case where channel state information $\mathcal{G}$, does not change in time. The fact that channel state information does not change in time makes existence of channel state information real and observation of channel state information in principle possible (irrespective of the observation noise level).

The Shannon information capacity of the time invariant additive Gaussian communication channel is given as $C = B * log(1 + S/N)$, where B is the channel bandwidth, S/N is the signal to AWGN power ratio [19]. The information rate R, which could be transmitted over the channel with vanishingly small probability of error, must satisfy the inequality $R < C$. The explicit assumptions under which this formula is derived are:

1. Infinite signal time delay (or very long signalling interval).
2. Existence of a very powerful channel coding/decoding scheme.
3. Stationary channel model with bandwidth B defined.
4. Channel characteristics in principle known to the receiver.

6.2.2 Hypothesis for Time Variable Communication Channels

The analysis in chapter 3 proves that, performing the channel state estimation and signal detection separately cannot achieve the information capacity of time varying communication channels. We also confirm that in chapter 5 that, the information capacity of the simplest model of time varying communication channel can be achieved without employing any conventional channel state estimator. According to these understanding, we propose the following hypotheses for time varying communication channels. These hypothesis has been partially proven in chapter 5.

For time varying communication channels $\mathcal{Y} = \mathcal{G}\mathcal{X}+\mathcal{Z}$, the joint three dimensional communication channel uncertainty model is inseparable in the following sense:

1. The information capacity of joint three-dimensional model (as channel information rate with vanishingly small probability of error) does exist and it is different from zero.

2. The information capacity of joint three-dimensional model is no less than the capacity of Shannon Wiener separation model.

3. The absolute time varying channel state information $\mathcal{G}$ cannot be measured accurately (at will) not even in principle. Consequently different receivers with different observation noise have different channel state information maximizing the channel capacity. The channel state information is not absolute. The channel state information is relative – relative to the speed of the channel change and the level of the observation noise.

4. The traditional time invariant channel is a special case of the time varying channel whose channel variation rate is zero.

5. The signalling formats exist at the transmitter, (at least for simple time varying channel models) which eliminates the need for channel state information in order to obtain non-zero information rate at the receiver. For the same channel models, any attempt to establish the measurement of channel state information at the receiver by sending training information at the transmitter, produces loss of information capacity, respectively.

Had these hypothesis been proven, the information capacity of time varying additive Gaussian communication channel might not be as simple as $C = B * log(1 + S/N)$. The explicit assumptions, under which this formula is derived, also need to be revisited.

1. Infinite signal time delay (or very long signalling interval). This assumption is also valid for time varying communication channels. We are also given enough time for processing each transmitted symbol.

2. Existence of a very powerful channel coding/decoding scheme. This assumption is also valid for time varying communication channels. To our knowledge, time invariant communication channels and time varying communication channels are fundamentally different, but not in terms of channel coding/decoding. The turbo coding and other powerful channel codes will be able to deliver data rates close to the information capacity due to constant increase of computational power at reduced cost.

3. Stationary channel model with bandwidth B defined. This assumption is not valid for time varying communication channels. The concept of bandwidth in wireline communication channel is clearly defined. All mentioned wireline channels are very slowly time varying and they could be approximated by time invarying stationary channels. The concepts of autocorrelation, the concepts of spectrum and consequently the concept of bandwidth are rigorously

defined on those channels [60]. However, all these concepts for stationary channels need to be updated.

4. Channel characteristics in principle known to the receiver. We have come to the conclusion that this assumption is not valid for time varying communication channels. And the following question is: what is the new assumption for deriving the information capacity formula for time varying communication channels?

Appendix A

Proving Lemma 1

Proof:

$$\boldsymbol{\pi_{n+1}} = \frac{\boldsymbol{\pi_n}\mathbf{D}\left(x_n, y_n\right)\mathbf{P}}{\boldsymbol{\pi_n}\mathbf{D}\left(x_n, y_n\right)\underline{\mathbf{1}}}, \tag{A.1}$$

where $\boldsymbol{\pi_n} = [\boldsymbol{\pi_n}(1), ..., \boldsymbol{\pi_n}(K)]$.

$$\boldsymbol{\rho_{n+1}} = \frac{\boldsymbol{\rho_n}\mathbf{B}\left(y_n\right)\mathbf{P}}{\boldsymbol{\rho_n}\mathbf{B}\left(y_n\right)\underline{\mathbf{1}}}, \tag{A.2}$$

where $\boldsymbol{\rho_n} = [\boldsymbol{\rho_n}(1), ..., \boldsymbol{\rho_n}(K)]$.

If the input is known by the estimator,

$$\begin{aligned} \boldsymbol{\pi_n}(k) &= \Pr\left(S_n = c_k | x^{n-1}, y^{n-1}\right) \\ &= \Pr\left(S_n = c_k | y^{n-1}\right) = \boldsymbol{\rho_n}(k). \end{aligned} \tag{A.3}$$

$\mathbf{D}\left(x_n, y_n\right)$ is a diagonal $K \times K$ matrix with kth diagonal term, $\Pr\left(y_n = 0 | x_n = 0, S_n = c_k\right)$; and $\mathbf{B}\left(y_n\right)$ is a diagonal $K \times K$ matrix with kth diagonal term $\Pr\left(y_n y_n = 0 | S_n = c_k\right)$. Similarly, $\Pr\left(y_n = 0 | x_n = 0, S_n = c_k\right) = \Pr\left(y_n = 0 | S_n = c_k\right)$, so, $\mathbf{D}\left(x_n, y_n\right) = \mathbf{B}\left(y_n\right)$. Therefore,

$$\boldsymbol{\pi_{n+1}} = \frac{\boldsymbol{\pi_n}\mathbf{D}\left(x_n, y_n\right)\mathbf{P}}{\boldsymbol{\pi_n}\mathbf{D}\left(x_n, y_n\right)\underline{\mathbf{1}}} = \frac{\boldsymbol{\rho_n}\mathbf{B}\left(y_n\right)\mathbf{P}}{\boldsymbol{\rho_n}\mathbf{B}\left(y_n\right)\underline{\mathbf{1}}} = \boldsymbol{\rho_{n+1}}. \tag{A.4}$$

Appendix B

The Estimation Method for The Gilbert-Elliot Channel

Because the Gilbert-Elliot channel has only two states, the estimation recursive formula (3.7) becomes

$$\boldsymbol{\rho_{n+1}} = \frac{\begin{bmatrix}\boldsymbol{\rho_n}(1) & \boldsymbol{\rho_n}(2)\end{bmatrix}\begin{bmatrix}\Pr(y_n=0|s_n=c_1) & 0\\ 0 & \Pr(y_n=0|s_n=c_2)\end{bmatrix}\begin{bmatrix}P_{11} & P_{12}\\ P_{21} & P_{22}\end{bmatrix}}{\begin{bmatrix}\boldsymbol{\rho_n}(1) & \boldsymbol{\rho_n}(2)\end{bmatrix}\begin{bmatrix}\Pr(y_n=0|s_n=c_1) & 0\\ 0 & \Pr(y_n=0|s_n=c_2)\end{bmatrix}\begin{bmatrix}1\\1\end{bmatrix}}, \tag{B.1}$$

where c_1 denotes the good state and c_2 denotes the bad state. (B.1) can be simplified to

$$\boldsymbol{\rho_{n+1}} = \frac{\begin{bmatrix}\Pr(y_n=0, s_n=c_1) & \Pr(y_n=0, s_n=c_2)\end{bmatrix}}{\Pr(y_n=0, s_n=c_1)+\Pr(y_n=0, s_n=c_2)} \cdot \begin{bmatrix}P_{11} & P_{12}\\ P_{21} & P_{22}\end{bmatrix}. \tag{B.2}$$

Because $\Pr(y_n = 0, s_n = c_1) = \Pr(s_n = c_1|y_n = 0) \cdot \Pr(y_n = 0)$ and $\Pr(y_n = 0, s_n = c_2) = \Pr(s_n = c_2|y_n = 0) \cdot \Pr(y_n = 0)$, $\boldsymbol{\rho_{n+1}}$ can be expressed by

$$\boldsymbol{\rho_{n+1}} = \frac{\begin{bmatrix} \Pr(s_n = c_1|y_n = 0) \cdot \Pr(y_n = 0) & \Pr(s_n = c_2|y_n = 0) \cdot \Pr(y_n = 0) \end{bmatrix}}{\Pr(s_n = c_1|y_n = 0) \cdot \Pr(y_n = 0) + \Pr(s_n = c_2|y_n = 0) \cdot \Pr(y_n = 0)} \cdot \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}. \tag{B.3}$$

Eliminating $\Pr(y_n = 0)$ from the denominator and numerator, $\boldsymbol{\rho_{n+1}}$ becomes

$$\boldsymbol{\rho_{n+1}} = \frac{\begin{bmatrix} \Pr(s_n = c_1|y_n = 0) & \Pr(s_n = c_2|y_n = 0) \end{bmatrix}}{\Pr(s_n = c_1|y_n = 0) + \Pr(s_n = c_2|y_n = 0)} \cdot \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}. \tag{B.4}$$

In our simulations, the channel states of the Gilbert-Elliot channel are estimated using equation (B.4). Let $\hat{\Pr}(y_n = 0)$ be the statistical average of the output obtained in simulations. The probabilities $\Pr(s_n = c_1|y_n = 0)$ and $\Pr(s_n = c_2|y_n = 0)$ of the equation (B.4) can be obtained by the following way.

It is assumed that the channel input of the channel, as well as the error probabilities of the good state and the bad state of the Gilbert-Elliot channel, are known by the estimator. The analytical output probabilities of good state and bad state can be obtained by,

$$\Pr(y_n = 0|s_n = c_1) = \Pr(x_n = 0) \cdot (1 - p_G) + (1 - \Pr(x_n = 0)) \cdot p_G, \tag{B.5}$$

$$\Pr(y_n = 0|s_n = c_2) = \Pr(x_n = 0) \cdot (1 - p_B) + (1 - \Pr(x_n = 0)) \cdot p_B. \tag{B.6}$$

The ratio of $\Pr(s_n = c_1|y_n = 0)$ to $\Pr(s_n = c_2|y_n = 0)$ of (B.4) is determined by the distances between $\hat{\Pr}(y_n = 0)$ and these two analytical values, $\Pr(y_n = 0|s_n = c_1)$ and $\Pr(y_n = 0|s_n = c_2)$.

$$\frac{\Pr(s_n = c_1|y_n = 0)}{\Pr(s_n = c_2|y_n = 0)} = \frac{\left|\hat{\Pr}(y_n = 0) - \Pr(y_n = 0|s_n = c_2)\right|}{\left|\hat{\Pr}(y_n = 0) - \Pr(y_n = 0|s_n = c_1)\right|}, \tag{B.7}$$

where "$|*|$" denotes the absolute value of a real number. Substituting (B.7) into (B.4), $\boldsymbol{\rho_{n+1}}$ can be expressed by

$$\boldsymbol{\rho_{n+1}} = \frac{\left[\left|\hat{\Pr}\left(y_n=0\right)-\Pr\left(y_n=0|s_n=c_2\right)\right| \quad \left|\hat{\Pr}\left(y_n=0\right)-\Pr\left(y_n=0|s_n=c_1\right)\right|\right]}{\left|\hat{\Pr}\left(y_n=0\right)-\Pr\left(y_n=0|s_n=c_2\right)\right|+\left|\hat{\Pr}\left(y_n=0\right)-\Pr\left(y_n=0|s_n=c_1\right)\right|} \cdot \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}. \tag{B.8}$$

The metric update function of the ML decoder is able to show the tracking ability of the open-loop estimator. The metric update is based on $\boldsymbol{\rho_n}$, and is expressed as

$$m(x_n,y_n) = -\log\left[\sum_{k=1}^{K}\Pr\left(y_n=0|x_n=0,s_n=c_k\right)\boldsymbol{\rho_n}\left(k\right)\right]-\log\left[\Pr\left(x_n=0\right)\right]. \tag{B.9}$$

With perfect channel state information , the metric update tracks the channel state precisely and has only two values, given by

$$m(x_n,y_n) = -\log\left(p_G\right)-\log\left[\Pr\left(x_n=0\right)\right]. \tag{B.10}$$

and

$$m(x_n,y_n) = -\log\left(p_B\right)-\log\left[\Pr\left(x_n=0\right)\right]. \tag{B.11}$$

for the good and bad states, respectively. When the channel state information is not known, different $\boldsymbol{\rho_n}$ and metric updates will be obtained with different channel input distributions. By comparing these values with the perfect updates, we can analyze the performance of the open-loop state estimator.

Appendix C

Proving Lemma 5

In this section, it is proven that the recursion (3.24) monotonically converges to 1 for one of the channel states, while converges to 0 for other states.

$$\begin{aligned}\lim_{\mu\to\mu_{max}} \boldsymbol{\rho_{n+1}}(l) &= \lim_{\mu\to\mu_{max}} \frac{\sum_{k=1}^{K} \Pr(y_n|s_n=c_k)\Pr(s_n=c_k|y^{n-1})P_{kl}}{\sum_{k=1}^{K} \Pr(y_n|s_n=c_k)\Pr(s_n=c_k|y^{n-1})} \\ &= \frac{\sum_{r\in R} \Pr(y_n|s_n=c_r)\boldsymbol{\rho_n}(r)}{\sum_{k=1}^{K} \Pr(y_n|s_n=c_k)\boldsymbol{\rho_n}(k)},\end{aligned} \tag{C.1}$$

where $r \in R$ are dominant elements of the l-th column of $\mathbf{P}_{fixed}$ and $P_{rl} = 1$. A special case of $\mathbf{P}_{fixed}$ is considered to show the reason of the convergence. The proof of the other types of $\mathbf{P}_{fixed}$ is similar. Assuming that $\mathbf{P}_{fixed}$ is a 3×3 matrix and it has three possible forms:

$$\mathbf{P}_{fixed} = \begin{bmatrix}0&0&1\\0&1&0\\1&0&0\end{bmatrix} \textit{ or } \begin{bmatrix}0&0&1\\0&0&1\\0&1&0\end{bmatrix} \textit{ or } \begin{bmatrix}0&0&1\\0&0&1\\0&0&1\end{bmatrix}, \tag{C.2}$$

For the first form, (C.1) becomes,

$$\begin{aligned}\lim_{\mu\to\mu_{max}} \boldsymbol{\rho_{n+1}}(l) &= \lim_{\mu\to\mu_{max}} \frac{\sum_{k=1}^{K} \Pr(y_n|s_n=c_k)\Pr(s_n=c_k|y^{n-1})P_{kl}}{\sum_{k=1}^{K} \Pr(y_n|s_n=c_k)\Pr(s_n=c_k|y^{n-1})} \\ &= \frac{\sum_{r\in R} \Pr(y_n|s_n=c_r)\boldsymbol{\rho_n}(r)}{\sum_{k=1}^{K} \Pr(y_n|s_n=c_k)\boldsymbol{\rho_n}(k)} \\ &= \frac{\boldsymbol{\rho_n}(r)}{\boldsymbol{\rho_n}(r)+\sum_{k=1,k\neq r}^{K} m(k,r)\boldsymbol{\rho_n}(k)}, \end{aligned} \tag{C.3}$$

where $r \in R$ are the dominant elements, $P_{rl} = 1$, of the l-th column of $\mathbf{P}_{fixed}$ and $m(k,r) = \Pr(y_n|s_n=c_k)/\Pr(y_n|s_n=c_r)$. In this case, there is only one dominant element in each column. For any i.i.d. input distribution which is not uniform, there exists a channel state c_v for which $m(k,v) < 1, k = 1,...,K$. For the channel state c_v, equation (C.3) increases as time n proceeds and monotonically converges to 1. For other states, (C.3) decreases and monotonically converges to 0.

For the second form, as there is no dominant element in the first column of $\mathbf{P}_{fixed}$, $\boldsymbol{\rho_n}(1) \triangleq 0$. The channel becomes a two state channel and

$$\mathbf{P}_{fixed} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \tag{C.4}$$

Similarly, for one of the states, (C.3) converges to 1, while for the other state, the equation converges to 0.

For the third form, as there is no dominant element in either the first column or the second column of $\mathbf{P}_{fixed}$, $\boldsymbol{\rho_n}(1) \triangleq \boldsymbol{\rho_n}(2) \triangleq 0$. Therefore, $\boldsymbol{\rho_n}(3) \triangleq 1$.

Bibliography

[1] Uk is the 'most internet-based major economy'. http://www.bbc.co.uk/news/business-17405016, March 2012.

[2] Web economy in g20 set to double by 2016, google says. http://www.bbc.co.uk/news/business-16753902, January 2012.

[3] Erisson mobility report. http://www.ericsson.com/res/docs/2013/ericsson-mobility-report-june-2013.pdf, June 2013.

[4] App store tops 40 billion downloads with almost half in 2012. http://www.apple.com/pr/library/2013/01/07App-Store-Tops-40-Billion-Downloads-with-Almost-Half-in-2012.html, 2013.

[5] Google play hits 25 billion downloads. http://officialandroid.blogspot.co.uk/2012/09/google-play-hits-25-billion-downloads.html, 2013.

[6] The nokia 9210 communicator heralds the dawn of mobile multimedia. http://press.nokia.com, 2000.

[7] Pocket phone. http://web.archive.org/web/20020124155709/http://www.microsoft.com/mobile/pocketpc/pocketpc2002/default.asp, 2002.

[8] Blackberry 5810 wireless handheld. http://csrc.nist.gov/groups/STM/cmvp/documents/140-1/140sp/140sp408.pdf.

[9] Android, the world's most popular mobile platform. http://developer.android.com/about/index.html, 2013.

[10] what is ois. http://www.apple.com/ios/what-is/, 2013.

[11] App store tops 40 billion downloads with almost half in 2012. http://www.apple.com/pr/library/2013/01/07App-Store-Tops-40-Billion-Downloads-with-Almost-Half-in-2012.html, 2013.

[12] The 5g phone future. *IEEE spectrum*, 2013.

[13] C. Hollanti, J. Lahtonen, and Hsiao-Feng Lu. Maximal orders in the design of dense space-time lattice codes. *IEEE Trans. Inform. Theory*, 54(10):4493–4510, 2008. ISSN 0018-9448. doi: 10.1109/TIT.2008.928998.

[14] E. Torkildson, U. Madhow, and M. Rodwell. Indoor millimeter wave MIMO: Feasibility and performance. *IEEE Trans. Wireless Commun.*, 10(12):4150–4160, 2011. ISSN 1536-1276. doi: 10.1109/TWC.2011.092911.101843.

[15] Wei Ni and Iain B. Collings. A new adaptive small-cell architecture. *IEEE J. Sel. Areas Commun.*, 31(5):829–839, 2013. ISSN 0733-8716. doi: 10.1109/JSAC.2013.130502.

[16] Y. Kishiyama, A. Benjebbour, T. Nakamura, and H. Ishii. Future steps of LTE-A: evolution toward integration of local area and wide area systems. *IEEE Wireless Commun.*, 20(1):12–18, 2013. ISSN 1536-1284. doi: 10.1109/MWC.2013.6472194.

[17] D. Persson, J. Kron, M. Skoglund, and E.G. Larsson. Joint source-channel coding for the mimo broadcast channel. *IEEE Trans. Signal Process.*, 60(4):2085–2090, 2012. ISSN 1053-587X. doi: 10.1109/TSP.2011.2180716.

[18] T.S. Rappaport. *Wireless communications: principles and practice*. Prentice Hall PTR, New Jersey, 2th ed. edition, 2002.

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

[20] Lajos Hanzo, Mohammed El-Hajjar, and Osamah Alamri. Near-capacity wireless transceivers and cooperative communications in the mimo era: evolution of standards, waveform design, and future perspectives. *Proc. IEEE*, 99(8): 1343–1385, 2011.

[21] M. Mushkin and I. Bar-David. Capacity and coding for the gilbert-elliot channels. *IEEE Trans. Inform. Theory*, 35(6):1277–1290, 1989.

[22] A.J. Goldsmith and P.P. Varaiya. Capacity, mutual information, and coding for finite-state markov channels. *IEEE Trans. Inform. Theory*, 42(3):868–886, 1996.

[23] B. Xu, Z. Krusevac, P. Rapajic, and Y Chen. Maximum mutual information rate for the uniformly symmetric variable noise fsmc without channel state information. In *Proc. IEEE 2012 Int. Symp. on Inform. Theory and its Applicat.*, pages 41–45, Oct. 2012.

[24] B. Xu, P. Rapajic, and Y Chen. Deficiency of the Gilbert-Elliot channel in modeling time varying channels. In *Proc. 2013 IEEE Wireless Commun. and Netw. Conf.*, Apr. 2013. Accepted for publication.

[25] Yiqing Zhou, Jiangzhou Wang, and M. Sawahashi. Downlink transmission of broadband ofcdm systems-part ii: effect of doppler shift. *IEEE Trans. Commun.*, 54(6):1097 –1108, Jun. 2006. ISSN 0090-6778. doi: 10.1109/TCOMM.2006.876872.

[26] Say Song Goh, T.N.T. Goodman, and F. Shang. Joint estimation of time delay and doppler shift for band-limited signals. *IEEE Trans. Signal Process.*, 58 (9):4583 –4594, Sept. 2010. ISSN 1053-587X. doi: 10.1109/TSP.2010.2051151.

[27] Rudolph E Kalman and Richard S Bucy. New results in linear filtering and prediction theory. *J. basic Eng.*, 83(3):95–108, 1961.

[28] A.J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Tran. Inform. Theory*, 13(2):260–269, 1967. ISSN 0018-9448. doi: 10.1109/TIT.1967.1054010.

[29] Tao Chen, Jingtai Liu, Guohui Jing, Zheng Liu, and Zhigang Bin. Time synchronization in acoustic localization system based on wireless sensor network. In *Proc. The Sixth World Congr. Intelligent Control and Automation*, volume 1, pages 4590–4594, 2006. doi: 10.1109/WCICA.2006.1713250.

[30] Yuan Zhang, R. Hoshyar, and R. Tafazolli. Channel estimation for mc-cdma with compensation of synchronization errors. In *Proc. IEEE 60th Vehicular Technology Conf.*, volume 2, pages 808–812, 2004. doi: 10.1109/VETECF.2004.1400132.

[31] M. Marx. System issues for time synchronization in real time localization systems with multi path mitigation. In *Proc. 2010 European Wireless Conference*, pages 596–601, 2010. doi: 10.1109/EW.2010.5483467.

[32] Hong Shen Wang and N. Moayeri. Finite-state Markov channel-a useful model for radio communication channels. *IEEE Trans. Veh. Technol.*, 44(1):163 –171, Feb 1995. ISSN 0018-9545. doi: 10.1109/25.350282.

[33] F. Babich and G. Lombardi. A Markov model for the mobile propagation channel. *IEEE Trans. Vehic. Technol.*, 49(1):63–73, 2000.

[34] Parastoo Sadeghi, R Kennedy, P Rapajic, and Ramtin Shams. Finite-state Markov modeling of fading channels-a survey of principles and applications. *IEEE Signal Process. Mag.*, 25(5):57–80, 2008.

[35] Parastoo Sadeghi and Predrag Rapajic. On information rates of time-varying fading channels modeled as finite-state markov channels. *IEEE Trans. Commun.*, 56(8):1268–1278, 2008.

[36] M. Riediger and E. Shwedyk. Communication receivers based on markov models of the fading channel. *IEE Proc. Commun.*, 150(4):275–279, 2003.

[37] C.C. Tan and N.C. Beaulieu. On first-order markov modeling for the rayleigh fading channel. *IEEE Trans. Commun.*, 48(12):2032–2040, 2000.

[38] YL Guan and LF Turner. Generalised FSMC model for radio channels with correlated fading. 146(2):133–137, 1999.

[39] C.D. Iskander and P.T. Mathiopoulos. Fast simulation of diversity nakagami fading channels using finite-state markov models. *IEEE Trans. Broadcast.*, 49(3):269–277, 2003.

[40] Z.B. Krusevac, P.B. Rapajic, and R.A. Kennedy. Basic binary state-space model for time-varying communication channels: uncertainty and information capacity. *IET Commun.*, 1(5):990 –998, oct. 2007. ISSN 1751-8628. doi: 10.1049/iet-com:20060477.

[41] Muriel Medard. The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel. *IEEE Trans. Inform. Theory*, 46(3):933–946, 2000.

[42] L. Li and A.J. Goldsmith. Low-complexity maximum-likelihood detection of coded signals sent over finite-state markov channels. *IEEE Trans. Commun.*, 50(4):524–531, 2002.

[43] C. Komninakis and R.D. Wesel. Joint iterative channel estimation and decoding in flat correlated rayleigh fading. *IEEE J. Sel. Areas Commun.*, 19 (9):1706–1717, 2001.

[44] P. Sadeghi and P. Rapajic. Capacity analysis for finite-state markov mapping of flat-fading channels. *IEEE Trans. Commun.*, 53(5):833–840, 2005.

[45] H. Kong and E. Shwedyk. Sequence detection and channel state estimation over finite state markov channels. *IEEE Trans. Vehic. Technol.*, 48(3):833–839, 1999.

[46] D.M. Arnold, H.-A. Loeliger, P.O. Vontobel, A. Kavcic, and Wei Zeng. Simulation-based computation of information rates for channels with memory. *IEEE Trans., Inform. Theory*, 52(8):3498 –3508, Aug. 2006. ISSN 0018-9448. doi: 10.1109/TIT.2006.878110.

[47] Chengshan Xiao, Jingxian Wu, S.-Y. Leong, Yahong Rosa Zheng, and K.B. Letaief. A discrete-time model for triply selective mimo rayleigh fading channels. *IEEE Trans. Wireless Communications,*, 3(5):1678 – 1688, 2004. ISSN 1536-1276. doi: 10.1109/TWC.2004.833444.

[48] P.B. Rapajic and D. Popescu. Information capacity of a random signature multiple-input multiple-output channel. *IEEE Trans. Commun.,*, 48(8):1245–1248, 2002. ISSN 0090-6778.

[49] G. Forney Jr. Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference. *Information Theory, IEEE Transactions on*, 18(3):363–378, 1972.

[50] C.E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Comput. Commun. Review*, 5(1):3–55, 2001.

[51] M. Rezaeian. Computation of capacity for gilbert-elliott channels, using a statistical method. In *Communications Theory Workshop, 2005. Proceedings. 6th Australian*, pages 56–61. IEEE, 2005.

[52] M. Rezaeian. Computation of capacity for Gilbert-Elliott channels, using a statistical method. In *Proc. 6th Communications Theory Workshop, Australian*, pages 56 –61, Feb. 2005. doi: 10.1109/AUSCTW.2005.1624226.

[53] R.G. Gallager. *Information theory and reliable communication*. Wiley, New York, 1968.

[54] N. Sen, F. Alajaji, and S. Yuksel. Feedback capacity of a class of symmetric finite-state markov channels. *IEEE Trans. Inform. Theory,*, 57(7):4110–4122, 2011. ISSN 0018-9448. doi: 10.1109/TIT.2011.2146350.

[55] C. Pimentel and R.L. Siqueira. Analysis of the go-back- protocol on finite-state markov rician fading channels. *IEEE Trans. Veh. Technol.*, 57(4):2627–2632, 2008. ISSN 0018-9545. doi: 10.1109/TVT.2007.912600.

[56] Chih-Hsien Hou, Jin-Fu Chang, and De-Yi Chen. Sharing of ARQ slots in Gilbert-Elliot channels. *IEEE Trans. Commun.*, 52(12):2070–2072, 2004. ISSN 0090-6778. doi: 10.1109/TCOMM.2004.838680.

[57] Min-Kuan Chang and Shi-Yong Lee. Modeling of single-step power control scheme in finite-state markov channel and its impact on queuing performance. *IEEE Trans. Veh. Technol.*, 58(4):1711–1721, 2009. ISSN 0018-9545. doi: 10.1109/TVT.2008.2004802.

[58] RH Clarke. A statistical theory of mobile-radio reception. *Bell Syst. Tech. J*, 47(6):957–1000, 1968.

[59] SO Rice. Statistical properties of a sine wave plus random noise. *Bell Syst. Tech. J.*, 27:109 –157, Jun. 1948.

[60] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*, volume 9. Technology Press of the Massachusetts Institute of Technology Cambridge:, 1949.