2006

Fundamental limits in Gaussian channels with feedback: confluence of communication, estimation, and control

Jialing Liu
Iowa State University

Fundamental limits in Gaussian channels with feedback: Confluence of communication, estimation, and control

by

Jialing Liu

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Major: Electrical Engineering

Program of Study Committee:
Nicola Elia, Major Professor
Wolfgang Kliemann
Murti V. Salapaka
Zhengdao Wang
Sang Kim
Namrata Vaswani

Iowa State University

Ames, Iowa

2006

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Jialing Liu

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

DEDICATION

To my family

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS

AFSMC	AWGN finite-state Markov channel
AWGN	additive white Gaussian noise
CP	Cover-Pombra
CRB	Cramer-Rao bound
CSI	channel state information
DI	degree of instability
DARE	discrete-time algebraic Riccati equation
DTRCSI	delayed-transmitter-side and receiver-side channel state information
FDLTI	finite-dimensional linear time-invariant
FSMC	finite-state Markov channel
GM	Gauss-Markov
i.i.d.	identically and independently distributed
ISI	inter-symbol interference
LDPC	low-density parity-check
LTI	linear time-invariant
LQG	linear quadratic Gaussian
MEC	minimum-energy control
MIMO	multi-input multi-output
MMSE	minimum mean-squared error
MSE	mean-square error
SISO	single-input single-output
SK	Schalkwijk-Kailath
WDP	writing on dirty paper
WDT	writing on dirty tape

w.r.t.	with respect to

ACKNOWLEDGEMENTS

First and foremost, I thank my major professor Dr. Nicola Elia for his guidance, support, and assistance. I feel honored to be his first student. His highly creative thinking, strong curiosity towards mysteries, aggressive and successful exploration of unknowns, and tireless working impressed me and influenced me greatly. The tremendous amount of interactions with him has made this thesis possible. I consider every discussion, suggestion, and comment invaluable.

I am grateful to Dr. Wolfgang Kliemann, Dr. Murti V. Salapaka, Dr. Zhengdao Wang, Dr. Sang Kim, Dr. Namrata Vaswani for serving on my committee.

I thank Dr. Murti V. Salapaka for his excellent teaching, many discussions and suggestions, and various help. I thank Dr. Wolfgang Kliemann for useful discussions and insightful comments. I thank Dr. Zhengdao Wang for his wonderful teaching, useful discussions, and important suggestions; my interactions with him made it possible for me to solve information theory problems. I thank Dr. Krishna Athreya for his outstanding teaching and helpful discussions; his extremely broad knowledge about Math and Statistics deeply impressed me and led me to thinking stochastically. I am grateful to Dr. Sang Kim, Dr. Aleksandar Dogandžić, Dr. Degang Chen, Dr. Yao Ma, Dr. Namrata Vaswani for many suggestions and discussions. Thanks to them, my academic experience at Iowa State has been very rewarding. I am also grateful to Dr. Sekhar Tatikonda for pleasant collaboration and Dr. Anant Sahai and Dr. Sanjoy Mitter for stimulating discussions.

I would like to thank all my colleagues in Systems and Control Group, Communication and Signal Processing Group, and beyond, for many discussions. To name a few, Wei Mo, Abu Sebastian, Xin Qi, Bingwen Wang, Donatello Materassi, Jeff N. Eisenbeis, Deepak R. Sahoo, and Vikas Yadav. I would like to thank all my friends for their support, encouragement, and help.

My special thanks go to my family, my parents, my parents-in-law, my wife, my Doudou, and my brother, for their unconditional love, support, and happiness during all these years. They always believe in me and have been doing everything they can to help me. This thesis is dedicated to them.

This research was supported by NSF under Grant ECS-0093950.

ABSTRACT

The emerging study of integrating information theory and control systems theory has attracted considerable attention by researchers, mainly motivated by the problems of control under communication constraints, feedback communication, and networked systems. Since in most problems, estimation interacts with communication and control in various ways and cannot be studied isolatedly, it is natural to investigate systems from the perspective of unifying communication, estimation, and control.

This thesis is the first work to advocate such a perspective. To make matters concrete, we focus on communication systems over Gaussian channels with feedback. For some of these channels, their fundamental limits for communication have been studied using information theoretic methods and control-oriented methods but remain open after several decades of research. In this thesis, we address the problems of identifying and achieving the fundamental limits for these Gaussian channels with feedback by applying the unifying perspective.

We establish a general equivalence among feedback communication, estimation, and feedback stabilization over the same Gaussian channels. As a consequence, we see that the information transmission (communication), information processing (estimation), and information utilization (control), seemingly different and usually separately treated, are in fact three sides of the same entity. We then reveal that the fundamental limitations in feedback communication, estimation, and control coincide: The achievable communication rates in the feedback communication problems can be alternatively given by the decay rates of the Cramer-Rao bounds (CRB) in the associated estimation problems or by the Bode sensitivity integrals in the associated control problems. Utilizing the general equivalence, we design optimal feedback communication schemes based on the celebrated Kalman filtering algorithm; these are the first deterministic, optimal feedback communication schemes for these channels (except for the degenerated AWGN case). These schemes also extend the Schalkwijk-Kailath (SK) coding scheme and inherit its useful features, such as reduced coding complexity and improved perfor-

mance. Though for different types of channels, these generalizations are along different lines, they all admit a common interpretation in terms of Kalman filtering of appropriate forms. Thus, we consider that Kalman filtering, the estimation side, acts like the unifier for various problems.

In addition, we show the optimality of the Kalman filtering in the sense of information transmission, a supplement to the optimality of Kalman filtering in the sense of information processing proposed by Mitter and Newton. We also obtain a new formula connecting the mutual information in the feedback communication system and the minimum mean-squared error (MMSE) in the associated estimation problem, a supplement to a fundamental relation between mutual information and MMSE proposed by Guo, Shamai, and Verdu.

To summarize, this thesis demonstrates that the new perspective plays a significant role in gaining new insights and new results in studying Gaussian feedback communication problems. We anticipate that the perspective and the approaches developed in this thesis could be extended to more general scenarios and helpful in building a theoretically and practically sound paradigm that unifies information, estimation, and control.

CHAPTER 1. INTRODUCTION

1.1 Background and motivation

The rapid growth of communication networks opens up new scenarios and possibilities of applications of communication systems in feedback loops and of feedback in communication systems. Examples of practical significance of control over networks include multiple vehicle coordination, air-traffic control, and Micro-Electro-Mechanical Systems (MEMS). In these applications, the control performance is fundamentally limited by the underlying communication networks. The traditional approach of separating the control and the communication problems is inefficient, since it leads either to low utilization of the channels or to control performance degradation due to unnecessary extra coding and decoding delays. Moreover, channels in feedback loops allow us to take advantage of feedback from the decoders to the encoders to drastically reduce the coding complexity and to improve the communication performance.

These considerations have motivated the study of integrated communication and control. This study has two main areas of research. The first focuses on the effect of feedback in communication systems. The simplest case consists of one feedforward channel with a noiseless feedback link from the decoder to encoder, whose study dated back to Shannon [109] and were extensively addressed in [107, 106, 110, 17] and references in the subsequent chapter.

The second considers how the presence of communication channels in the loops affects the control performance and looks for characterizing the requirements imposed on the communication system by the desired performance objective. A great deal of fundamental understanding, highlighted next, has been drawn from characterizing the minimum transmission rate necessary for stabilization for different channel models [132, 133, 4, 113, 87, 32]; very recently however there has been a flurry of papers addressing control performance mostly in the linear quadratic Gaussian (LQG) settings [119, 78, 76]. The channel model is often assumed to be noiseless and without delay but with fixed rate. Due to its simplicity and its strictly positive zero-error capacity, this model eliminates both the need for channel coding and the tradeoff between communication delays and probability of error. This facilitates the analysis of the effect due to finite rate communication and leads to interesting results.

When one considers channels which are not error free, like the binary erasure, binary symmetric, and Gaussian channels, the feedback systems become stochastic systems, and the main finding is that different notions of stochastic stability require different notions of reliable delivery of information through the loop [102, 104]. While the Shannon notion is still appropriate for the "almost sure" stability of the closed loop [117], it is no longer adequate to capture the moment stability [102, 104]. The moment stability requires the *anytime capacity*, a notion of reliable communication stronger than Shannon's notion but weaker than the zero-error one. An important and often overlooked consequence is that while two channels with the same Shannon capacity are equivalent and interchangeable from an information transmission viewpoint, they are not equivalent when used in feedback loops, in the sense that the may not stabilize the moments of the same plants. Unfortunately a mutual information like characterization of anytime capacity is not yet known.

[28] bridges the gap between the two main areas of research by investigating the Gaussian channels in loops. Feedback control and feedback communication over Gaussian channels were shown to be equivalent, and control-oriented approach to study feedback communication was developed. Since the anytime limitations do not arise in Gaussian channels, the analysis is facilitated since it can rely on the classical information theory concepts of Shannon capacity and mutual information (or directed information in feedback settings, as shown in [113]).

Despite the important progresses shown above and summarized in Chapter 2 in more details, the effect of feedback in communication systems is not completely characterized for most of the channels, including the least limiting case of Gaussian channels. Other than the degenerated case of additive white Gaussian noise (AWGN) channel with feedback, a communication scheme that optimally utilizes the feedback is not available to *any* single-input single-output (SISO) Gaussian channels with feedback.

Hence, in this thesis, our efforts will be concentrated on the benchmark problems of identifying and achieving the fundamental limits (i.e. the Shannon capacity) for several types of Gaussian channels with feedback.

1.1.1 Approach

The approach that we will apply is mainly an extension of the control-oriented approach proposed in [28], and it completes the existing picture of communication and control with *estimation.* Estimation interacts with information theory or control theory in various ways and cannot be studied isolatedly. Hence, it is natural to study the feedback communication problem from the unifying perspective encompassing communication, estimation, and control. Note that however, current focuses in many researches of channels in loop are the interplay between communication and control; though estimation is sometimes employed in those researches, its significance has not been completely realized. This thesis is the first work to propose the unifying perspective and to uncover its significance. We remark that this perspective may eventually lead to a paradigm synthesizing communication, estimation, and control, which can be used to address many problems of channels in loops.

More specifically, we will proceed as follows. First, we identify the connections of feedback communication, estimation, and feedback control for Gaussian channels, and second, we utilize such connections to address the optimality and fundamental limitations for the feedback communication problems. This approach is shown to be rather powerful in solving the benchmark problems, and the main findings are listed in the next section.

1.2 Main results

The main results are stated informally as follows.

Result 1 (General equivalence among feedback communication, estimation, and control). *There is a general equivalence among a feedback communication system over a Gaussian channel, an estimation system (i.e. a Kalman filtering system) over the same channel, and a control system (i.e. a minimum-energy control (MEC) system) over the same channel.*

As a consequence, we see that the information transmission (communication), information processing (estimation), and information utilization (control), seemingly different and usually separately treated, are in fact three sides of the same entity.

This result is an extension of those linking feedback communication and feedback control given by Sahai and Mitter [102, 104] and Elia [28]. Here the SISO Gaussian channels include: 1) AWGN channels with feedback, 2) frequency-selective fading Gaussian channels with feedback, where the fading is modeled as a finite-dimensional system, [1] 3) time-selective fading Gaussian channels with feedback and channel state information (CSI), and 4) "writing on dirty paper"-(WDP-) Gaussian channels with feedback and with frequency-selective fading. The precise meaning of the equivalence will be discussed in latter chapters. The feedback used in the communication systems are assumed to be noiseless, a typical and ideal assumption for feedback communication with limitation pointed out in Section 2.2.5.

Since the three types of systems are equivalent, it holds that their fundamental limitations are essentially the same. Note that in the feedback communication system, the fundamental limitation is the achievable rate: There exists a critical value for the signalling rate, above which reliable communication is not achievable and below which reliable communication is achievable. In the (recursive) estimation system, the fundamental limitation is the decay rate of Cramer-Rao bounds (CRB): Whatever estimator one may design, the decay rate of mean-squared error (MSE) cannot be made larger than the decay rate of CRB. In the control system, the fundamental limitation is the Bode sensitivity integral: No matter how one designs the controller, the sensitivity integral cannot be made smaller than a constant determined by how unstable the plant is. Our result states

Result 2 (Agreement of fundamental limitations). *The fundamental limitations in the three systems agree. That is, the achievable rate in the feedback communication system equals half of the decay rate of CRB in the estimation system, and equals the Bode sensitivity integral in the control system.*

This result is an extension of the agreement of fundamental limitations between feedback communication and feedback control given by Elia [28]. The obtained relation between the achievable rate and CRB can be translated into the relation between mutual information and minimum MSE (MMSE), which is a supplement to that obtained by Guo, Shamai, and Verdu in [51]. In loose terms, [51] says that the increasing rate of mutual information w.r.t. the channel parameter SNR is equal to half of the MMSE, whereas here it says that the increasing rate of mutual information w.r.t. time is equal to half of the decreasing rate of the MMSE. Connections between these two results are under current investigation.

[1] By Gaussian channels with memory, researchers normally mean the SISO frequency-selective Gaussian channels, although time-selective Gaussian channels may also have memory; see Section 4.2 for precise definition. These channels are sometimes also referred to as general Gaussian channels (in contrast to the AWGN channels), or even simply as Gaussian channels.

The equivalence given in Result 1 also helps us to construct the optimal feedback communication schemes.

Result 3 (Optimal feedback communication schemes). *The optimality in the three systems coincides, based on which optimal feedback communication schemes can be constructed. The structures of the optimal schemes are given by a simple transform of the Kalman filtering systems, and the parameters of the optimal schemes are given by closed-form expressions or by the solutions to finite-dimensional optimization problems.*

The proposed optimal communication schemes are the first deterministic, optimal communication schemes for these Gaussian channels with feedback (except for the AWGN case). We note that, in our optimal feedback communication schemes, the encoding can be seen as a control problem, and the decoding can be seen as an estimation problem, confirming the claim by Mitter [81].

The presence of Kalman filtering in the optimal feedback communication schemes leads to the information theoretic interpretation of Kalman filtering.

Result 4 (Information theoretic characterization of Kalman filtering). *The Kalman filtering for an unstable process driven by its initial condition, when put in an appropriate form, is optimal in information transmission. The one-step prediction operation in Kalman filtering leads to minimization of the channel input power, and the smoothing operation in Kalman filtering leads to optimal recovery of the transmitted message.*

Another way of saying this is that the optimal feedback communication systems have to implement the Kalman filtering algorithm. The optimality of Kalman filtering in the sense of *information transmission* is a complement to the existing characterization that Kalman filter is optimal in the sense of *information processing* established by Mitter and Newton in [83]. Note that in [83], a Kalman filter for a stable process was studied, whereas here a Kalman filter for an unstable process is studied. This result also completely characterizes the role of Kalman filtering in feedback communication systems, a step further than Yang, Kavcic, and Tatikonda, who identified only the role of Kalman filter as a sufficient statistics generator [136]. It is also interesting to notice that, though the optimal coding schemes are along different directions for different classes of channels, all these schemes can be universally interpreted in terms of

Kalman filtering of appropriate forms. Thus, we consider that Kalman filtering acts as a "unifier" for communication schemes over various channels.

To summarize, we develop a new perspective of unifying information, estimation, and control for the study of Gaussian channels with feedback. The introduction of the new ingredient, namely estimation, plays an important part and provides new insights and new results to the study of feedback communication problems, estimation problems, and control problems.

1.3 Thesis outline

We provide an outline and a summary of contributions for each chapter as follows.

Chapter 1 In this chapter we introduce briefly the motivation of the perspective that unifies communication, estimation, and control; identify the problem we wish to address and describe the main results; and outline the thesis.

Chapter 2 In this chapter we review the relevant literature, mainly including the study of control under communication constraints and the study of feedback information theory.

Chapter 3 In this chapter we study the AWGN channels with feedback. We present a simple motivating observation of the Kalman filtering problem; obtain a Kalman filter based optimal coding scheme; establish the equivalence of information, estimation, and control over this channel; and prove that the fundamental limitations in information, estimation, and control coincide.

The main contribution in this chapter is that, we provide new insights to a problem that has been extensively studied in the past forty years, and the new insight gives us new results, i.e. Results 1 - 4, stated for AWGN channels with feedback. None of these results has been found before. These new insights say that the optimal coding scheme can be obtained from the Kalman filtering system, and the joint study of the feedback coding problem, Kalman filtering problem, and control problem is useful.

We also show that the celebrated Schalkwijk-Kailath (SK) coding scheme is nothing but a simple transform of Kalman filtering, an interesting connection that has never been identified before. Besides, we rediscover the coding scheme studied by Gallager [43], which, unlike other popular coding schemes, is numerically stable but has received little attention in the literature.

Chapter 4 In this chapter we study the frequency-selective Gaussian channels with feedback. We start from finite-horizon analysis. We transform the renowned Cover-Pombra (CP) coding structure (cf. [17]) into a new form, from which we conclude that it has to include a Kalman filter in order to minimize the channel input power. We next rewrite the Kalman filter based coding structure as an estimation system and a control system, and rewrite mutual information in terms of estimation system quantities and control system quantities. We then show that this coding structure in finite-horizon reaches a steady-state as the horizon length increases, and hence the (asymptotic) information rate for the communication system can be represented as half of the decay rate of CRB for the associated estimation system and as the Bode sensitivity integral of the associated control system. Finally, we show that the limiting steady-state feedback communication system of finite-dimension can achieve the feedback capacity, and the construction of the optimal system amounts to solving a finite-dimensional optimization problem.

In this chapter, we obtain Results 1 - 4 for frequency-selective fading Gaussian channels with feedback. The obtained optimal coding scheme based on the celebrated Kalman filtering algorithm is the first solution to such channels, after tens of years of search. Compared with existing results in the literature, our coding scheme is a refinement and extension of the CP coding structure, that is, we show that the CP coding structure essentially contains a Kalman filter; it is an extension of the SK codes and inherits their nice properties of reduced coding complexity and improved performance; and it is also a simplification of the optimal signalling strategy proposed by Yang, Kavcic, and Tatikonda [136]. For the first time, all the major research directions for Gaussian channels with memory are shown to be connected (through Kalman filtering), including the SK codes and their extensions, the CP structures and their extensions, the control-oriented schemes studied by Tatikonda, Mitter, Sahai, Elia, Yang, and Kavcic.

Chapter 5 In this chapter we study the time-selective fading Gaussian channels with CSI and output feedback. We present a feedback communication structure, motivated by an estimation problem or a control problem; specify the parameters for this structure, using closed-form expressions in terms of known channel parameters and the optimal power allocation functions; and prove that this structure with these parameters achieves the feedback capacity and doubly exponential decay of probability of error.

One of our contributions is that, the proposed coding scheme was the *first* capacity-achieving coding scheme over any SISO Gaussian channel (other than the AWGN case) in the literature [74], dated even earlier than the schemes proposed in Chapters 4 and 6. It extends the SK coding scheme for an AWGN channel with feedback along another important direction, namely to *adapt* to the channel variations. Though the channel variations in general lead to difficulties due to the anytime limitation, assuming the knowledge about the channel variations and adapting to them eliminate the anytime limitation in our case. However, due to the delay in the feedback, the multiplexing strategy used in feedforward communication (which is an adaptive strategy) does not apply to the feedback case directly, and we develop the technique of state augmentation to resolve this issue, motivated by the study of the associated control system. Hence, our scheme may be viewed as a non-trivial combination of the SK coding scheme for an AWGN channel with feedback and the feedforward multiplexing coding scheme for a fading channel [47]. These two techniques, namely adaptation and state augmentation, leads to the optimality of the proposed coding scheme, and may be the key techniques to study any feedback channels with known time-variations and delays. We show that the control theoretic equivalence of the optimal coding scheme is a *Markov jump linear system*, which has significant theoretic and practical values and has been a focus in systems and control community; see e.g. [15]. This connection may be useful for studying other fading channels with feedback.

Chapter 6 In this chapter we study the WDP-channels with feedback, which we call *writing on dirty paper with feedback.* In such a setting, there is an interference sequence known to the encoder non-causally but unknown to the decoder. We construct coding schemes to achieve the feedback capacity of AWGN WDP-channels and Gaussian WDP-channels with memory, in which the interference is canceled without incurring any rate loss or extra power overhead, in other words, we achieve *lossless interference cancelation* for these channels. We also demonstrate close connections among communication, estimation, and control.

This is the first study of the WDP problem with feedback, and it demonstrates how feedback can be helpful in WDP problems: Feedback still has the power of greatly reducing coding complexity and coding delay in WDP settings. We show that the optimal dirty paper coding scheme involves a Kalman filtering problem in which there is a process noise known to both the process and the estimator. We develop useful techniques to cope with the presence of interference, and the techniques may be readily applied to more general WDP problems. Since the dirty paper coding study has been considered to be a basic building block in both single-user and multiuser communication problems [38], we envision that our study on the feedback case generates a new avenue for studying many feedback communication problems.

Chapter 7 In this chapter we conclude the thesis and present some interesting directions that will be the subjects of our future research work.

Notations: We represent time indices by subscripts, such as y_t. We denote by y^T the collection $\{y_0, y_1, \cdots, y_T\}$ [2], and $\{y_t\}$ the sequence $\{y_t\}_{t=0}^{\infty}$. We assume that the starting time of all processes is 0, consistent with the convention in dynamical systems but different from the information theory literature. We use $h(x)$ for the differential entropy of the random variable x. For a random vector y^T, we denote its covariance matrix as $K_y^{(T)}$. For a stationary process $\{y_t\}$, we denote its power spectrum as $\mathcal{S}_y(e^{j2\pi\theta})$, and its entropy rate as $\boldsymbol{h}(y)$. We denote $\mathcal{T}_{xy}(z)$ as the transfer function from x to y. We denote "defined to be" as ":=". We use (A, B, C, D) to represent the finite-dimensional linear time-invariant (FDLTI) system

$$\begin{cases} x_{t+1} &= Ax_t + Bu_t \\ y_t &= Cx_t + Du_t. \end{cases} \tag{1.1}$$

In the following chapters, by **feedback** in a communication system, we mean the **output feedback**, namely the feedback of the channel outputs or the feedback of some functions of the channel outputs, unless otherwise specified.

[2]The notation y^T should not be confused with the Tth power of y. The meaning of the superscript notation is clear from the context: y^T is the Tth power of y if y is a scalar, and y^T denotes a vector if a scalar y is not defined but a collection of scalars y_0, y_1,..., y_T is defined.

CHAPTER 2. LITERATURE REVIEW FOR THE JOINT STUDY OF INFORMATION, ESTIMATION, AND CONTROL

In this chapter, we review the existing literature. Due to the rapid growth of the literature, we feel that a survey of results scattered in the literature is necessary and useful. Due to the same reason, our review is by no means complete or comprehensive. In Section 2.1, we review control with limited information. In Section 2.2, we review feedback information theory. We remark that these two are sometimes not differentiable. Some other relevant literature is included in Section 2.3.

2.1 Literature review for control with limited information

Control with limited information has two important features, among others. One is that information needs to be quantized to allow digital communication and processing, the other is that information has to go through channels with uncertainties (e.g. noise, fading, shadowing, etc.). We describe briefly the *quantized control systems* and *control over channels with uncertainties* below.

2.1.1 Quantized control systems

In control systems with digital communication channels, the effect of information quantization is taken into consideration in control systems design, resulting in quantized control systems. In contrast to a conventional control system which assumes that information is transmitted and processed without any cost or limitation, for a quantized control system, the cost or the limitation of information transmission/processing are explicitly considered as follows. The measurements of the plant and/or control inputs have to go through finite-capacity digitalized communication channels and hence are quantized to finite precision; in addition, in many situations, these signals are transmitted and processed only intermittently (namely, they are transmitted and processed only if a change occurs). Therefore, quantized controller usually generates piecewise constant control inputs and are event-driven systems.

Delchamps showed that the closed-loop behavior resulting from quantization of measurements is quite different (and much more complicated) than that resulting from approximation of measurements [19]. In [3] Brockett proposed the minimum attention control which uses piecewise constant control and takes into account of the "attention cost" measuring the overhead for information transmitting and processing. In [132, 133] Wong and Brockett addressed the problems of state estimation and feedback stabilization with finite bandwidth communication constraints. In [1] Borkar and Mitter introduced communication constraints to linear quadratic Gaussian (LQG) control problems. Nair and Evans studied state estimation via a capacity-limited communication channel [86]. These original papers have motivated a lot of research work; see for example [4, 70, 32, 25, 39, 55, 42, 69, 77, 87, 22, 72, 118, 117, 119, 88].

Coarser quantization implies that less information flows between the controller and the plant. Therefore, the minimum quantization density that stabilizes an unstable plant is of interest: It can be used to measure the minimum information needed for stabilization, and it codifies how difficult a system can be controlled. In [32] Elia and Mitter devised the quantizer with minimum density for stabilizing a discrete-time linear time-invariant (LTI) single-input plant that is open-loop unstable. The quantizer design was shown to be an MEC problem, and the minimum density ρ is

$$\rho = \frac{DI-1}{DI+1}, \tag{2.1}$$

depending only on the *degree of instability*

$$DI := \prod_{i=1}^{m} |\lambda_{u,i}|, \tag{2.2}$$

where $\lambda_{u,i}$ are the unstable poles of the system. This implies that if the plant is more unstable, then more information is needed for accomplishing the stabilization task. Later, the minimum-density quantizer design was obtained for multi-input plants [25], for nonlinear plants [71, 72], and for control with performance requirements [24].

Another measure of how much information is needed for stabilization is the minimum bit-rate between the stabilizing quantized controller and the unstable plant. Tatikonda [113] and Nair and Evans [87] provided the minimum bit-rate for stabilizing an unstable linear plant.

Like the previous case, the minimum bit-rate R depends only on the degree of instability, namely

$$R = \log DI. \tag{2.3}$$

Similar results hold for state estimation. Since the bit-rate is directly linked to mutual information and entropy, one may attempt to link this result to the (generalized Kolmogorov) entropy generated by the plant. Nair *et al* showed that such an unstable plant generates entropy at a rate equal to $\log DI$, and hence apparently a channel which can sustain communication rate of at least $\log DI$ is needed [88].

Information quantization makes a dynamical systems *hybrid* in many cases. Such a hybrid nature may cause technical difficulties (such as the discontinuity in the vector fields for continuous-time dynamical systems, see e.g. [4, 71]). Despite of this, sometimes people intensionally introduce information quantization to a system to reduce the communication/computation costs and to address the design problems of hybrid systems or hierarchic systems; see [3, 77, 24, 32, 23, 71]. To summarize, quantized control systems become an interesting device that integrates dynamical systems and control, information theory and communication, and hybrid systems.

2.1.2 Control over probabilistic channels

The above research generally assumes that the channel is digital but noiseless (i.e. deterministic). This noiseless assumption holds true if the channel has a capacity higher than the bit-rate, thanks to the channel coding theorem (though the delay issue needs to be taken care of). In many cases, it is necessary or useful if the noises and uncertainties of the underline channel are considered explicitly. For an example, the optimality for the control over an AWGN channel is easily achieved without quantization or coding, namely sending the un-coded control signal across the noisy channel is optimal [28]; see later chapters for details. This leads to the study of control over probabilistic channels.

Many strategies have been developed to cope with the channel noise and fading; see e.g. [76, 112] for two most recent publications. However, one "universal" strategy exists, as proposed by Elia. This strategy is to, first, extract any channel uncertainties from the *deterministic* mean system, and second, view the uncertainties as those studied in *robust control* [26, 27]. Then the stability of the original system becomes the robust stability of the mean system to those uncertainties. This strategy has been applied to packet-drop networks without side information [26], packet-drop networks with side information but without side information loss [31], packet-drop networks with lossy side information [30], general memoryless fading channels [27], etc. In the scenario of [26], the maximum packet-drop rate that ensures the closed-loop stabilization can be solved analytically, which is again determined by the degree of instability. In other scenarios, a critical quantity resembling the structural singular value in robust control determines the minimum "quality of service" that the channel has to provide to ensure stabilization.

For the problem of estimation over channels with uncertainties, we may also apply the above results by invoking the duality between control and estimation. Other work in this area including [111], which considered the Kalman filtering problem over an erasure channel and characterized the existence of the critical value of packet loss, above which the estimation error covariance becomes unbounded and below which the estimation error covariance is bounded.

2.2 Literature review for feedback information theory

Communication systems with noiseless feedback from the receivers to the transmitters have been studied since Shannon [109]. In [109], Shannon proved that feedback does not increase the capacity for a discrete memoryless channel, namely $C_{fb} = C_{ff}$, where C_{fb} is the feedback capacity and C_{ff} is the feedforward (or forward) capacity. This result was somewhat surprising, since one might expect that the noiseless feedback should benefit the noisy feedforward communication. Such a benefit was later uncovered by Elias, Horstein, etc., who showed that noiseless feedback can improve the performance [33, 52]. More improvement was obtained later by Schalkwijk and Kailath.

2.2.1 The Schalkwijk-Kailath (SK) codes and its extensions

Ten years after Shannon's work, in 1966, substantial improvement was discovered by Schalkwijk and Kailath. While in feedforward communication, researchers were struggling to better address the tradeoff among lower coding complexity, better performance, and higher data rate, in their award-winning papers [107, 106], Schalkwijk and Kailath demonstrated that by utilizing noiseless feedback, a *capacity-achieving* coding scheme for an AWGN channel can be designed with low complexity and very short coding length, under an average power constraint.

Particularly, in [106], it was shown that the following simple signalling strategy is optimal. Given a set of M equally likely messages, uniformly divide the interval $[0, 1]$ into M subintervals, and associate each message to a subinterval center. Pick one center W and transmit it. At time t, the decoder computes a maximum-likelihood estimate $\hat{W}_t$ of W. At time $(t+1)$, the encoder transmit $a(\hat{W}_t - W)$, an *amplified* version of the estimation error, where $a > 1$ is an amplification factor. At the final time, the decoder maps its final estimate to the closest subinterval center as the decoded message. This coding procedure is rather simple, and it leads to *doubly exponentially* decay of the decoding error probability, while achieving the capacity.

The SK codes have been extended to many situations. Gallager reformulated the coding scheme and discussed both the digital and analog (related to rate-distortion and joint source-channel coding) transmission issues [43]. Schalkwijk designed the multi-dimensional signalling for AWGN channels [105]. Omura formulated a stochastic optimal control problem for the SK signalling strategy [92]. Wyner showed that the SK codes generate (singly) exponential decay of error probability if a peak power constraint is used. Butman designed the feedback codes for channels with additive Gaussian noise forming autoregressive processes, and derived tight bound for feedback capacity [5, 6]. More general Gaussian noise channels were also studied, see e.g. [122, 131, 97, 96]. For multi-input multi-output (MIMO) Gaussian channels with feedback (i.e. Gaussian networks with feedback), see e.g. [95, 68, 64].

2.2.2 Computation of feedback capacity and bounds

For discrete memoryless channels and AWGN channels, the feedback capacity equals the feedforward capacity. Dobrushin first showed that channels with memory can have feedback capacity strictly greater than feedforward capacity [21]. Whether there is an improvement or not for a discrete-time additive Gaussian channel with feedback was completely characterized by Yanagi: When the noise is white, $C_{fb} = C_{ff}$; when the noise is blockwise white (independent between blocks), $C_{fb} = C_{ff}$ if the power budget is below some critical value and $C_{fb} > C_{ff}$ otherwise; when the noise is completely colored, $C_{fb} > C_{ff}$ [135]. On the other side, for a colored Gaussian channel, it was pointed out by Pinsker that feedback can at most double the capacity [16], and by Cover and Pombra that feedback can at most improve the capacity by half a bit [17]. Dembo also showed that feedback does not improve capacity for colored Gaussian channels in very high SNR and very low SNR regimes [20]. Numerous upper bounds and lower bounds for Gaussian channels with memory have been obtained in the past decades; see [62] for a recent summary. For Gaussian networks without memory, increase of capacity due to feedback is possible and can be fairly high [95, 68, 53, 54, 63, 64, 65].

Despite of the above progresses, the feedback communication problems were found very challenging: The feedback capacity of *any* Gaussian channel with memory (other than the degenerated case of AWGN channels) could not be computed. As a result, the optimality of the above generalized SK codes for Gaussian channels with memory could not be established. Most of the feedback communication problems remained largely open until very recently, new developments have been obtained mainly based on the unifying perspective of information and control.

2.2.3 Most recent achievements: The interactions with control theory

Researchers had paid little attention on viewing the feedback communication system as a control system (with the notable exception of Omura [92]), but this recently (re-)developed viewpoint has been shown extremely powerful in addressing many long-standing problems in feedback communication.

In [113, 116], Tatikonda and Mitter proposed a unified view of control and information. They thoroughly studied feedback communication systems and the feedback capacity for both discrete and continuous channels. They viewed the feedback communication systems as interconnections of stochastic kernels, which extends Dobrushin's view of communication systems and Willems' view of feedback control systems [21, 99]. They proved that the supremum of *directed information rate* from the channel input to channel output (subject to the power constraint, if any) is the feedback capacity, which is the first input-output characterization of feedback capacity. The directed information rate is defined as

$$\overrightarrow{I}(u \to y) := \lim_{T\to\infty} \frac{1}{T+1} \overrightarrow{I}(u^T \to y^T) := \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} I(u^t; y_t | y^{t-1}), \tag{2.4}$$

provided the limit exists, where $u^t := [u_0, u_1, \cdots, u_t]'$ is the stacked input, and y^t is the stacked output. This differs from the conventional mutual information rate

$$I(u; y) := \lim_{T\to\infty} \frac{1}{T+1} I(u^T; y^T) := \lim_{T\to\infty} \frac{1}{T+1} \sum_{t=0}^{T} I(u^T; y_t | y^{t-1}) \tag{2.5}$$

in that the causal dependence (different from the statistical dependence) of the channel input u_t on channel output y^{t-1} is considered in (2.4). Note that $\overrightarrow{I}(u \to y)$ is indeed "directed" since $\overrightarrow{I}(u \to y) \neq \overrightarrow{I}(y \to u)$ in general. Note also that earlier Massey argued that the supremum of (2.5) does not give us the feedback capacity, and the supremum of (2.4) is an upper bound of the feedback capacity [79]. Tatikonda and Mitter also reformulated the feedback capacity problem as a stochastic control problem, and developed a dynamical programming based solution to compute the finite-horizon feedback capacity (though it still involves rather high computation complexity). To better address the time-delay issue and the feedback issue in many communication problems, they proposed the *sequential rate distortion theory*, motivated in part by control of unstable systems over communication channels (mainly over noiseless digital channels and AWGN channels). This theory tells us how much channel capacity is needed to transmit a *process*, e.g. a video stream, across a channel with specified distortion. It was shown that both the conventional rate-distortion problems and the *successive refinement* problems [35], including the SK coding scheme for transmitting analog sources, are special cases of this theory.

In [102, 104], Sahai and Mitter observed that to communicate over general *noisy* channels delay-sensitive information streams, including non-stationary, non-ergodic sources or even unstable sources, the Shannon capacity becomes inadequate. For instance, if the source to be transmitted over a noisy channel grows exponentially and is known to the transmitter causally, Shannon capacity C of the channel may not be achievable. This is because whenever a code with rate close to C generates an arbitrarily small but nonzero probability of error (which is for sure due to the channel noise), this error cannot be corrected and will causes an unbounded error at the decoder side in later steps. However, the zero-error capacity is too conservative. A fundamental new theory, called the *anytime information theory*, was proposed to address these communication problems. This theory is closely linked to control problems, such as the problem of tracking unstable sources over noisy channels. In particular, Sahai and Mitter demonstrated that moment stabilization (or asymptotic tracking with bounded moments) over noisy channels is equivalent to reliable communication of streams over the same channels. The fundamental limitations in those communication problems are characterized by *anytime capacity*, above which reliable communication of delay-sensitive information streams is possible and below which reliable communication is impossible. In fact, for any rate below the anytime capacity, any error made in previous steps can be eventually corrected. Anytime capacity, corresponding to moment stability of the associated control system, is usually stronger than the Shannon capacity, corresponding to almost sure stability of the associated control system, but it is (much) less demanding than the zero-error capacity. In the AWGN channel case, it is shown that

$$C_{anytime}(\mathcal{P}) = C_{Shannon}(\mathcal{P}) = \frac{1}{2}\log(1+\mathcal{P}), \tag{2.6}$$

whereas the zero-error capacity is zero! It also holds that, for Gaussian channels, moment stability is equivalent to almost sure stability, and hence anytime capacity equals the Shannon capacity. However, in other cases, the anytime capacity has not been found to have a simple characterization such as a mutual information type one, and it is difficult to compute in general. Sahai and Mitter also studied the anytime encoder, anytime decoder, joint source-channel coding, insufficiency of using block codes, etc.

In [28], Elia made an intriguing observation that the celebrated SK coding scheme is merely a rewrite of a special LQG problem (i.e. the MEC problem with Gaussian disturbance), and reliable communication in the former is equivalent to stabilization in the latter. Then he established the general equivalence between reliable feedback communication and feedback stabilization over Gaussian channels with memory [1]. In particular, the problem of feedback communication over a Gaussian channel with memory can be transformed into an MEC problem over the same channel with an open-loop unstable plant, and if the latter is stabilized in closed-loop, the former communicates reliably; additionally, lower bounds of the feedback capacity in the former can be obtained by solving the minimum-energy needed for stabilization in the latter. More precisely, the transmission rate over the channel is shown to equal the degree of instability of the open-loop system, namely

$$R = \log DI, \tag{2.7}$$

and the channel input power is given by

$$P = \|\mathcal{T}_{Zu}(z)\|_2^2 \tag{2.8}$$

[1]One of the differences between this equivalence and that shown by Sahai and Mitter is that, the unstable plant in [28] is driven by its initial condition, and in [102, 104] is driven by an input process.

where $\mathcal{T}_{Zu}$ denotes the transfer function from the Gaussian noise Z to the channel input u, and $\|\cdot\|$ denotes the $\mathcal{H}_2$ norm. Then lower bounds of feedback capacity can be obtained by finding minimum power P while maintaining the closed-loop stability and a given rate R. To solve the minimum power, one can apply many control techniques, such as the Youla parametrization, interpolation conditions, and so on. Elia demonstrated that this approach provides tighter new lower bounds or recover the tightest lower bounds of feedback capacity in existing information theory literature. He also extended the SK codes to achieve these lower bounds, and the coding schemes have an interpretation of tracking unstable sources over Gaussian channels. By investigating the fundamental limitations in the feedback communication problem and the control problem, Elia showed that the achievable rate is alternatively given by the Bode sensitivity integral

$$R = \int_{-\frac{1}{2}}^{\frac{1}{2}} \log S(e^{j2\pi\theta})d\theta \tag{2.9}$$

where $S(z)$ is the sensitivity transfer function of the closed-loop control system. Hence, the fundamental limitation in feedback communication coincides with that in control.

In [137], Yang, Kavcic, and Tatikonda applied the ideas of directed mutual information rate and stochastic control formulation in [113] to compute the feedback capacity for a discrete-input finite-state Markov channel. The optimal input distribution is also characterized. In [136], Yang, Kavcic, and Tatikonda used similar ideas to compute the feedback capacity of a power-constrained Gaussian channel with memory. They uncovered the Markov property of the optimal input distributions for this channel and eventually reduced the finite-horizon stochastic control optimization problem to a manageable size. The optimal input distribution at time t takes the form

$$u_t = d_t'(s_t - \mathbf{E}(s_t|y^{t-1})) + \mathcal{E}_t, \tag{2.10}$$

where $d_t \in \mathbb{R}^m$ is a gain, $s_t \in \mathbb{R}^m$ is the state of the channel at time t (which summarizes the information in previous inputs u^{t-1}), m is the order of the channel, $\mathbf{E}(s_t|y^{t-1})$ represents the sufficient statistics for estimating s_t based on channel outputs y^{t-1} which can be computed using a Kalman filter, $\mathcal{E}_t$ is an independent process representing new information, and its variance $K_{\mathcal{E},t}$ as well as d_t needs to be searched for each t. Using dynamical programming, a solution of computation complexity $O(T+1)$ for computing the finite-horizon feedback capacity C_T is obtained, where $(T+1)$ is the length of the time horizon. Moreover, under a *stationarity conjecture* that the infinite-horizon capacity $C_{fb,\infty}$ equals the stationary capacity (the maximum information rate over all *stationary* input distributions, denoted C_{fb}^{s}), $C_{fb,\infty}$ is given by the solution of a finite dimensional optimization problem. The stationarity conjecture is equivalent to assuming that $d_t = d$ and $K_{\mathcal{E},t} = K_{\mathcal{E}}$ for all t. This is the first computationally efficient [2] method to calculate C_{fb}^{s} or $C_{fb,T}$ for general Gaussian channels with memory.

2.2.4 Other most recent achievements

Most of the new achievements in feedback information theory employed the interactions between information and control. As an notable exception, some recent important results for Gaussian channels with feedback were obtained based on the Cover-Pombra coding structure (called the CP structure). The CP coding structure is a simple linear structure that generates the optimal channel inputs for general Gaussian channels with feedback [17]. Though the structure is rather simple, the number of free parameters to be searched grows fast as the coding length increases and leads to prohibitive computation complexity. By exploiting the special properties of a moving-average Gaussian channel with feedback, Ordentlich discovered the finite rankness of the innovations in the CP structure, which reduces the computation complexity [94]. Shahar-Doron and Feder reformulated the CP structure along this direction, and obtained an SK-based coding scheme to achieve the finite-horizon capacity with reduced computation complexity [108]. Also along this line, Kim studied a first-order moving-average Gaussian channel with feedback, found the closed-form expression for $C_{fb,\infty}$, and obtained an SK-based coding scheme to achieve $C_{fb,\infty}$ [60]. For this channel, the optimal input distribution is indeed stationary, as conjectured by Yang, Kavcic, and Tatikonda in [136]. Moreover, in [61], Kim confirmed the stationary conjecture for stationary Gaussian channels with feedback, based on a careful look at the CP structure, superadditivity of feedback capacity, and the relation between cyclostationary processes and stationary processes.

2.2.5 Limitations of feedback information theory

As we have seen, significant progresses have been made in feedback information theory, especially in the past five years. However, *most* of the results are based on the assumption

[2] Here we do not mean that their optimization problem is convex. In fact the computation complexity for $C_{fb,T}$ is $O(T+1)$, and for $C_{fb,\infty}$ the complexity is determined mainly by the channel order, which does not involve prohibitive computation if the channel order is not too high.

that the feedback link is noiseless, and the literature lacks meaningful results on noisy feedback. In the noiseless feedback case, infinite amount of information can be transmitted across the feedback link without any cost or limitation.

In some situations, the noiseless feedback assumption may be a good approximation. Consider a communication network with base stations and mobile stations. The communication from the mobile station to the base station (the forward transmission) may be very noisy due to the limited resources (e.g. transmission power) available to the mobile station, whereas the communication from the base station to the mobile station (the feedback transmission) may be viewed as noiseless since the base station may have a lot of resources (e.g. transmission power). The study using noiseless feedback says that one can dramatically improve the forward transmission by taking advantage of the feedback transmission, which may be useful in such communication networks. Similar possibilities exist in satellite-ground communication [107], sensor networks with base stations or cluster centers [126], and so on.

Yet, the above described situations in real practice can be much more complicated. For example, in practice one would try to better allocate the bandwidth resources between the feedforward link and the feedback link, and one usually feeds back the decision made by the decoder to the encoder. Additionally, the effect of feedback noise may be very small but never exactly zero. These issues have not been taken into consideration, mainly because they greatly complicate the problem (which is already very difficult) and cannot be studied at the current stage.

The main usefulness of the ideal study based on noiseless feedback of full information about the channel outputs includes the following. First, it gives us the ultimate bound and useful hint of how much feedback (of any kinds) may help us. If the noiseless output feedback leads to substantial improvement in a communication problem in terms of either capacity or coding complexity or performance, it may suggest that we should try to utilize feedback in that problem, even if the feedback link is not perfect. Second, we consider it a necessary step towards the study of the more realistic cases, such as noisy feedback. Third, a feedback communication system with noiseless feedback may be transformed into a control system over one noisy channel using the equivalence shown in [28], which may help solve the associated control problem. Note that a control system with only one noisy channel (either from the plant to the controller or the opposite) is a realistic problem.

Thus, though the noiseless feedback assumption is not quite practical, which makes the so far developed feedback information theory not quite practical, we consider that this simplifying assumption is helpful, these studies have significant theoretic implications, and they may shed important insight on the study of the more practical problems.

Throughout this thesis, we will follow this convention and assume that the channel outputs are fed back **noiselessly** to the encoder.

2.3 Other work

Mitter and Newton revealed the optimality of optimal estimators in the information theoretic sense [82, 83]. They identified the information flow in a Kalman filter, showed that the Kalman filter extracts the right amount of useful information from the observations for estimation purpose, supplies the extracted information to an information store, and dissipates at an optimal rate the old information that becomes no longer useful. Hence, the Kalman filter acts optimally in the information processing sense.

Graham, Baliga, and Kumar observed the convergence of control, communication, and computation [49, 48]. They pointed out that in many applications, the three cannot be treated as separated: "For example, the problem of data fusion in sensor networks is not just an inference problem, or just a computation problem, or just a communication problem. It is a synthesis. When actuation is also involved, as in control, there is a further convergence of theories. Such a systems theory could well be the agenda for the next two decades." It is clear that their main scope largely overlaps with the study of interactions among information, estimation, and control: Though we are not yet explicitly addressing the computation aspect, computation is used in information transmission (e.g. coding and decoding), information processing (e.g. estimation), and information utilization (e.g. decision making). A complete picture of a typical system in both studies comprises the underlying plants, sensing, information transmission, information processing, decision making, another information transmission, and actuation. Kumar and their collaborators have made significant progresses in design and implementation of such systems in their Convergence Lab at the University of Illinois.

Guo, Shamai, and Verdu obtained a new formula that connects the mutual information of the input and output in a Gaussian channel and the MMSE achievable by optimal estimation of the input given the output [51]. In its simplest form, for a scalar AWGN channel

$$y = \sqrt{\mathrm{SNR}}\, u + N \tag{2.11}$$

with SNR being the signal-to-noise ratio (SNR), u being the channel input with *arbitrary* distribution, N being standard Gaussian independent of u, and y being the channel output, it holds that

$$\frac{\mathrm{d}}{\mathrm{dSNR}} I(\mathrm{SNR}) = \frac{1}{2}\mathrm{MMSE}(\mathrm{SNR}), \tag{2.12}$$

where $I(\mathrm{SNR})$ is the mutual information between the input u and output y, and MMSE(SNR) is the MMSE of the input estimate given the output; both $I(\mathrm{SNR})$ and MMSE(SNR) are functions of SNR. In another word, the derivative of the mutual information w.r.t. SNR equals half of the MMSE, *independent of the input distribution.* This intriguing relationship can be extended to various Gaussian channels and draws fundamental connections between information theory and estimation theory. This is closely related to one of the results linking mutual information and MMSE (or CRB) obtained in this thesis.

In [78], Martins and Dahleh studied the performance of control systems with communication constraints. They explicitly showed that how much channel capacity is needed to achieve a desired performance of a control system, and characterized the fundamental limitations of disturbance rejection in terms of a Bode-like formula or information theoretic quantities. Intuitively speaking, they showed that the channel capacity should be no smaller than the rate used for stabilization plus the rate used for disturbance rejection. Elia considered a different setup of the disturbance rejection problem over Gaussian channels [29]. In this case, disturbance rejection by loop shaping does not require extra transmission rate but needs extra power, and the extra power corresponds to how far the sensitivity function is away from the optimal (flat) sensitivity given by $\mathcal{H}_\infty$ loop shaping.

In [75], Liu *et al* studied the collective behavior of a class of dynamical systems under communication constraints, and proved that different amount of information flow inside the system leads to rather different behavior of the system. More specifically, it was shown that, for a group of simple non-mobile agents operating in discrete-time, phase transition emerges from the interactions among agents in the presence of noise: At a noise level higher than some threshold, the system generates symmetric behavior; whereas at a noise level lower than the threshold, the system exhibits spontaneous symmetry breaking. This research belongs to the *cooperation with limited information* framework, also illustrates the interaction between information and dynamical systems, and may help us to understand how information is being utilized in such systems.

We refer to [80, 110, 121, 50, 34, 10, 130, 127, 114, 40, 91, 115, 84, 90] and references therein for other related work.

CHAPTER 3. AWGN CHANNELS WITH FEEDBACK

In this chapter, we study the AWGN channels with feedback. Due to its simplicity, this channel has been extensively investigated. A capacity-achieving coding scheme was first proposed by Schalkwijk and Kailath [107, 106], and variations and different interpretations have been given. In particular, Elia showed that the SK coding scheme is essentially an MEC problem [28]. Nevertheless, this channel is still interesting to us, since its simplicity helps us to easily identify its connections to estimation/control problems. Careful studying of these connections gives us new insights about feedback communication problems, and the insights can be extended to more general situations, as we will see in subsequent chapters.

This chapter is organized as follows. We first present a Kalman filtering problem as a motivating example in Section 3.2, based on which we obtain a feedback communication scheme in Section 3.3. We then prove that this scheme is optimal in the sense of achieving the feedback capacity in Section 3.4. We draw explicit connections to an estimation problem, a tracking-of-unstable-source problem, and a control problem; see Section 3.5. We show that the optimality in these problems coincides, and so do their fundamental limitations. We also illustrate that the SK coding scheme is essentially the Kalman filtering algorithm. In Section 3.6, to overcome the numerical instability problem of the proposed coding scheme, we provide a modified coding scheme suitable for simulation purpose and report the numerical results.

3.1 Channel model and feedback capacity

Consider a discrete-time AWGN channel shown in Fig. 3.1. At time t, $t = 0, 1, \cdots$, this channel is described as

$$y_t = u_t + N_t, \tag{3.1}$$

where u_t is the channel input, N_t is the channel noise and forms an identically and independently distributed (i.i.d.) Gaussian process with zero mean and unit variance, and y_t is the channel output.

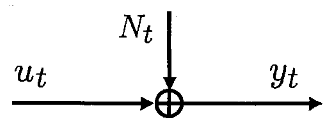


Figure 3.1 An AWGN channel.

We assume that at time t, the encoder can make use of the channel output y^{t-1} or any function of it, by utilizing the noiseless feedback link. Under an average channel input power constraint

$$\lim_{T\to\infty} \frac{1}{T+1} \mathbf{E} u^{T\prime} u^T \leq \mathcal{P} \tag{3.2}$$

with $\mathcal{P} > 0$ being the power budget, it holds that

$$C_{fb}(\mathcal{P}) = C_{ff}(\mathcal{P}) = \frac{1}{2}\log(1+\mathcal{P}), \tag{3.3}$$

where $C_{fb}(\mathcal{P})$ is the feedback capacity and $C_{ff}(\mathcal{P})$ is the feedforward capacity.

3.2 Motivating observations: A simple Kalman filtering problem

To help the reader understand the intuition behind our study, we introduce a simple example before we go into the technical details. A large portion of this research is in fact motivated by the observations made in studying this example.

Consider a standard Kalman filtering problem for a first-order unstable LTI system with noisy measurements:

$$\begin{cases} x_{t+1} &= ax_t \\ r_t &= cx_t \\ \bar{y}_t &= r_t + N_t, \end{cases} \tag{3.4}$$

where x_0 is unknown, $a > 1$ (namely the system is unstable), a and c are known, and N_t is i.i.d. Gaussian with zero mean and unit variance. Though the process $\{\bar{y}_t\}$ is neither stationary, nor even asymptotically stationary, a Kalman filter can be built to guarantee *bounded* error covariance for estimating x_t, and the Kalman filter converges to a time-invariant one, as pointed out in Chapter 14 of [57]. We can obtain the steady-state Kalman filter as

$$\begin{cases} \hat{x}_{t+1} &= a\hat{x}_t + Le_t \\ \hat{r}_t &= c\hat{x}_t \\ e_t &= \bar{y}_t - c\hat{x}_t, \end{cases} \tag{3.5}$$

where

$$L := \frac{a\Sigma c}{1 + c^2\Sigma} \tag{3.6}$$

is called the *Kalman filter gain*; Σ, the asymptotic error covariance for $\hat{x}_t$, is the positive solution to the discrete-time algebraic Riccati equation (DARE)

$$\Sigma = a^2\Sigma - \frac{a^2c^2\Sigma^2}{1 + c^2\Sigma}; \tag{3.7}$$

and e_t is called the *Kalman filter innovation* or simply *innovation.* See Fig. 3.2.

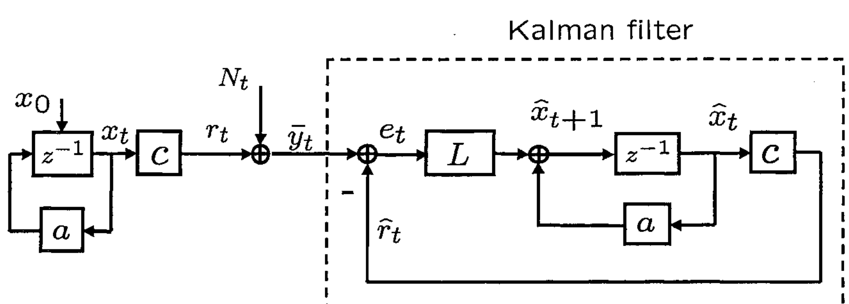


Figure 3.2 A Kalman filtering problem.

It can be easily solved that

$$\begin{aligned} \Sigma &= \frac{a^2 - 1}{c^2} \\ L &= \frac{a^2 - 1}{ac}, \end{aligned} \tag{3.8}$$

and hence the error covariance of estimating r_t is

$$P := \mathbf{E}(r_t - \hat{r}_t)^2 = \mathbf{E}(cx_t - c\hat{x}_t)^2 = c^2\Sigma = a^2 - 1, \tag{3.9}$$

yielding that

$$\log a = \frac{1}{2}\log(1+P), \tag{3.10}$$

which reminds us the capacity formula for an AWGN channel with power budget P.

Note that the Kalman filter is the optimal estimator for an estimation problem. It has been demonstrated by Mitter and Newton that the Kalman filter can be interpreted in an information theoretic sense (but not in the sense of conveying information over some channel), and it is indeed optimal in that sense [83]. Here we would like to ask related but different questions: Does the Kalman filtering problem admit any information transmission interpretation? If yes, does it correspond to the *optimal* communication over an AWGN channel, in light of (3.10)? And where are the encoder, decoder, and message? We answer these questions in the subsequent section.

3.3 Kalman filter based coding scheme

In this section, we propose a coding scheme which achieves the feedback capacity of the AWGN channel. This coding structure, as illustrated in Fig. 3.3, is easily seen to have only a slight change from the Kalman filter shown in Fig. 3.2: Instead of closing the loop after the AWGN N_t (i.e. adding $-\hat{r}_t$ to $\bar{y}_t$), in Fig. 3.3, the loop is closed before the AWGN N_t (i.e. adding $-\hat{r}_t$ to r_t). This does not change anything but the signals between the two adders. In Fig. 3.3, we can identify the encoder, the AWGN channel, and the decoder, that is, we indeed obtain a feedback communication system from the Kalman filtering problem in Fig. 3.2.

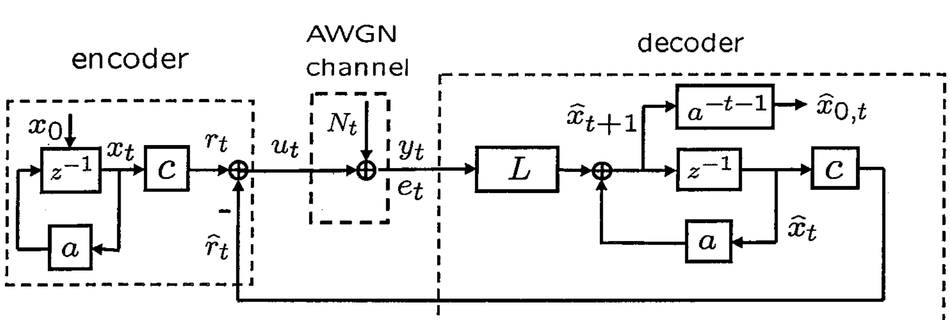


Figure 3.3 A coding structure based on Kalman filtering.

Remark 1. *The rationale behind the optimality of this coding scheme lies in the optimality of the Kalman filtering problem. In Fig. 3.2, if $a > 1$, namely the process to be estimated is unstable, it holds that*

$$x_t = a^t x_0, \tag{3.11}$$

that is, x_t and $r_t := cx_t$ grow exponentially. Compared to the ever growing process $\{r_t\}$, the noise process $\{N_t\}$ is exponentially smaller and smaller and eventually becomes negligible. Therefore, the Kalman filter, the MMSE estimator, *can estimate x_0 with higher and higher precision, which is a fixed point smoothing problem. In fact, let $\hat{x}_{0,t} := a^{-t-1}\hat{x}_t$, then the estimation error $\tilde{x}_{0,t} := x_0 - \hat{x}_{0,t}$ decays exponentially and faster than that given by any other estimator, in the mean-squared sense. This corresponds to that the information about x_0 is being transmitted at the highest rate. Besides, the difference between r_t and $\hat{r}_t$, the optimal one-step prediction of r_t based on $\bar{y}^{t-1}$, is also minimized in the mean-squared sense, which implies that $u_t := r_t - \hat{r}_t$ is minimized in the variance (namely, the power) sense. This leads to the optimality in the feedback communication problem.*

To see that the estimation error $(x_0 - \hat{x}_{0,t})$ decays exponentially in the mean-squared sense, we may reason in an approximate way or a rigorous way. Approximately, it holds that, for t large enough,

$$\mathbf{E}(x_t - \hat{x}_t)^2 \approx \Sigma; \tag{3.12}$$

since $x_t = a^t x_0$ and $\hat{x}_t = a^t \hat{x}_{0,t}$, we then have

$$\mathbf{E}(x_0 - \hat{x}_{0,t})^2 \approx a^{-2t-2}\Sigma, \tag{3.13}$$

that is, the estimation error $\tilde{x}_{0,t}$ decays exponentially at rate $\log a$. *Rigorous computation confirms that this is the right decay exponent of $\tilde{x}_t$, which will be presented shortly in the proof of Theorem 1.*

To summarize the above intuitive explanation, the Kalman filter provides the optimal estimate of the message being transmitted by smoothing operations, as well as the optimal channel input power usage by one-step prediction operations. Combining these two factors, we can obtain an optimal coding scheme based on Kalman filtering.

In the rest of this section, we describe the Kalman filter based coding scheme in detail. We will prove rigorously the optimality of the coding scheme in the next section.

3.3.1 Coding structure and optimal parameters

The encoder and decoder shown in Fig. 3.3 are described in state-space as follows:

$$\text{encoder:} \quad \begin{cases} x_{t+1} &= a x_t \\ r_t &= c x_t \\ u_t &= r_t - \hat{r}_t \end{cases} \tag{3.14}$$

and

$$\text{decoder:} \quad \begin{cases} \hat{x}_{t+1} &= a\hat{x}_t + L y_t \\ \hat{r}_t &= c\hat{x}_t \\ \hat{x}_{0,t} &= a^{-t-1}\hat{x}_{t+1}, \end{cases} \tag{3.15}$$

where

$$\begin{aligned} a &:= \sqrt{1+\mathcal{P}} > 1 \\ c &:= 1 \\ L &:= a - \frac{1}{a}, \end{aligned} \tag{3.16}$$

$\hat{x}_0 := 0$, $\hat{x}_{0,0} := 0$, and x_0 will be determined shortly. Recall that $\mathcal{P} > 0$ is the power budget. It is easy to see that L is chosen according to the Kalman filter gain formula (3.8). We call x_t the *encoder state.*

Interestingly, the encoder may be viewed as a control system, and the decoder may be viewed as an estimation system, as pointed out by Mitter in [81].

3.3.2 Coding processes

The designed communication system can transmit either an analog source or a digital message. We describe the coding processes for both cases below. Let us fix the coding length (or the time horizon) to be $(T+1)$, namely the time spans from time 0 to time T.

3.3.2.1 Transmission of analog source

In this case, we assume that the encoder wishes to convey a Gaussian random variable through the channel and the decoder wishes to learn the random variable, which is a rate-distortion problem or a successive refinement problem (see [113, 102, 45] and reference therein for study of successive refinement and its generalization, the sequential rate-distortion problem).

The coding process is as follows. Assume without loss of generality that the to-be-conveyed message W is distributed as $\mathcal{N}(0, \mathcal{P})$ (if the variance is not $\mathcal{P}$, we can scale W to have the desired variance). To encode, let

$$x_0 := W. \tag{3.17}$$

Then run the system till instant of time T, generating $\hat{x}_{0,t}$ for $t = 0, 1, \cdots, T$. To decode, let $\hat{W}_T := \hat{x}_{0,T}$. The distortion measure is

$$\mathrm{MSE}_{W,T} := \mathbf{E}(W - \hat{W}_T)^2. \tag{3.18}$$

3.3.2.2 Transmission of digital message

To transmit digital messages over the communication system, let us fix $\epsilon > 0$ arbitrarily small. Suppose that we wish to transmit one of a set of

$$M_T := a^{(T+1)(1-\epsilon)} \tag{3.19}$$

messages. We equally partition the interval

$$\left[-\sqrt{\mathcal{P}}\left(1 + \frac{1}{M_T - 1}\right), \sqrt{\mathcal{P}}\left(1 + \frac{1}{M_T - 1}\right)\right] \tag{3.20}$$

into M_T sub-intervals, and map the sub-interval centers to a set of M_T equally likely messages; this is known to both the transmitter and receiver *a priori*.

Suppose now we wish to transmit the message represented by the center W. To encode, define x_0 according to (3.17). Then run the system till instant of time T. To decode, let the decoder estimate $\bar{W}_T$ be

$$\bar{W}_T := \frac{\hat{x}_{0,T}}{1 - a^{-2T-2}}. \tag{3.21}$$

We then map $\bar{W}_T$ into the closest sub-interval center and obtain the decoded message $\hat{W}_T$. We declare an error if $\hat{W}_T \neq W$, and call a (an asymptotic) rate

$$\mathcal{R} := \lim_{T \to \infty} \frac{1}{T+1} \log M_T \tag{3.22}$$

achievable if the probability of error PE_T vanishes as T tends to infinity.

3.4 Coding theorem

The following theorem establishes that, the above described coding scheme is optimal in the sense of information transmission. Therefore, we have indeed obtained optimality in feedback communication based on the optimality in estimation.

Theorem 1. *Let $\{N_t\}$ be AWGN with $N_t \sim \mathcal{N}(0,1)$. Then under the power constraint $\mathbf{E}u^2 \leq \mathcal{P}$,*

i) The coding scheme constructed in Section 3.3.1, following the coding process described in Section 3.3.2, transmits an analog source $W \sim \mathcal{N}(0,\mathcal{P})$ from the encoder to the decoder at the capacity rate

$$C_{fb}(\mathcal{P}) := \frac{1}{2}\log(1+\mathcal{P}), \tag{3.23}$$

with MSE distortion $\mathrm{MSE}_{W,T}$ satisfying the optimal rate-distortion tradeoff function given by

$$C_{fb}(\mathcal{P}) = \frac{1}{2(T+1)}\log\frac{\mathcal{P}}{\mathrm{MSE}_{W,T}} \tag{3.24}$$

for each T.

ii) The coding scheme constructed in Section 3.3.1, following the coding process described in Section 3.3.2, can transmit a digital message from the encoder to the decoder at a rate arbitrarily close to $C_{fb}(\mathcal{P})$, with PE_T decays to zero doubly exponentially.

Proof: We first derive the expressions for u_t and $\hat{x}_{0,t}$. Let $\tilde{x}_t := x_t - \hat{x}_t$. We can express $\tilde{x}_t$ in terms of the initial condition and channel outputs as

$$\tilde{x}_t = a^t W - a^t \sum_{j=0}^{t-1} a^{-j-1} L y_j. \tag{3.25}$$

Hence

$$\sum_{j=0}^{t-1} a^{-j-1} L y_j = W - a^{-t}\tilde{x}_t. \tag{3.26}$$

On the other hand, noticing that $y_t = u_t + N_t$, it also holds that

$$\begin{aligned}
\tilde{x}_t = &\; a\tilde{x}_{t-1} - L(u_{t-1} + N_{t-1}) \\
= &\; (a - Lc)\tilde{x}_{t-1} - LN_{t-1} \\
= &\; a^{-1}\tilde{x}_{t-1} - LN_{t-1} \\
= &\; a^{-t}W - \sum_{j=0}^{t-1} a^{-t+1+j}LN_j.
\end{aligned} \tag{3.27}$$

This leads to that

$$\begin{aligned}
\hat{x}_{0,t-1} = &\; \sum_{j=0}^{t-1} a^{-j-1}Ly_j \\
= &\; W - a^{-t}\tilde{x}_t \\
= &\; (1 - a^{-2t})W + a^{-2t}\sum_{j=0}^{t-1} a^{j+1}LN_j.
\end{aligned} \tag{3.28}$$

Therefore, we have

$$\hat{x}_{0,T} = (1 - a^{-2T-2})W + a^{-2T-2}\sum_{j=0}^{T} a^{j+1}LN_j. \tag{3.29}$$

Also note that

$$u_t = c\tilde{x}_t = a^{-t}W - \sum_{j=0}^{t-1} a^{-t+1+j}LN_j. \tag{3.30}$$

Then we compute the average channel input power, followed by the rate and distortion computation for i), and the rate and probability of error computation for ii). From (3.30), the input power is asymptotically determined only by the term $\sum_{j=0}^{t-1} a^{-t+1+j}LN_j$, which leads to that

$$\mathbf{E}u^2 = L^2 \lim_{t\to\infty}\sum_{j=0}^{t-1} a^{-2t+2+2j} = \frac{(a^2-1)^2}{a^2}\frac{a^2}{a^2-1} = \mathcal{P}. \tag{3.31}$$

i) The MSE distortion is

$$\begin{aligned}
\mathrm{MSE}_{W,T} &= \mathbf{E}\left[a^{-4T-4}W^2 + a^{-4T-4}\sum_{j=0}^{T} a^{2j+2}L^2(N_j)^2\right] \\
&= a^{-4T-4}\mathcal{P} + a^{-4T-4}\mathcal{P}^2\sum_{j=0}^{T} a^{2j} \\
&= a^{-4T-4}\mathcal{P}\left(1+\mathcal{P}\frac{a^{2T+2}-1}{a^2-1}\right) \\
&= a^{-4T-4}\mathcal{P}a^{2T+2} \\
&= \mathcal{P}a^{-2T-2}.
\end{aligned} \tag{3.32}$$

This distortion needs an information rate across the channel to be at least

$$R = \frac{1}{2T+2}\log\frac{\mathcal{P}}{\mathrm{MSE}_{W,T}} = \log a = C_{fb}(\mathcal{P}), \tag{3.33}$$

which is indeed the capacity of the AWGN channel and implies that the optimality is achieved.

ii) By (3.21) and (3.29), we have

$$\bar{W}_T = W + (a^{2T+2}-1)^{-1}\sum_{j=0}^{T} a^{j+1}LN_j, \tag{3.34}$$

that is, $\bar{W}_T$ is an unbiased estimate of W. For each given W, it holds that $\bar{W}_T \sim \mathcal{N}(W, (a^{2T+2}-1)^{-2}\sum_{j=0}^{T} a^{2j+2}L^2)$.

The signalling rate is

$$\mathcal{R} := \lim_{T\to\infty}\frac{1}{T+1}\log M_T = (1-\epsilon)\log a. \tag{3.35}$$

This signalling rate is achievable if the probability of error vanishes. To compute the probability of error, note that for each message W, no error occurs if

$$|\bar{W}_T - W| \le \frac{1}{M_T-1}\sqrt{\mathcal{P}}. \tag{3.36}$$

Then the probability of error satisfies

$$
\begin{aligned}
PE_T &\leq 2Q\left(\frac{\sqrt{\mathcal{P}}/(M_T-1)}{(a^{2T+2}-1)^{-1}L\sqrt{\sum_{j=0}^{T}a^{2j+2}}}\right)\\
&= 2Q\left(\frac{\sqrt{\mathcal{P}}a^{T+1}\sqrt{1-a^{-2T-2}}}{(M_T-1)\sqrt{a^2-1}}\right)\\
&= 2Q\left(a^{(T+1)\epsilon}\frac{\sqrt{1-a^{-2T-2}}}{1-a^{-(T+1)(1-\epsilon)}}\right)\\
&\overset{(a)}{\leq} 2Q\left(a^{(T+1)\epsilon}\right)\\
&\overset{(b)}{\leq} \frac{2}{\sqrt{2\pi}a^{(T+1)\epsilon}}\exp\left(-\frac{1}{2}a^{2(T+1)\epsilon}\right),
\end{aligned}
\tag{3.37}
$$

where inequality (a) follows from

$$
\frac{\sqrt{1-a^{-2T-2}}}{1-a^{-(T+1)(1-\epsilon)}} \geq 1 \tag{3.38}
$$

for any $\epsilon > 0$ and T, and (b) follows from the Chernoff bound

$$
Q(t) \leq \frac{1}{\sqrt{2\pi}t}\exp(-\frac{1}{2}t^2).
$$

Thus, PE_T decreases to zero doubly exponentially. ∎

Remark 2. *We discuss some variants of the above coding scheme. It can be seen that if the message interval shown in (3.20) is chosen to be in some other forms, such as* $[0,1]$ *or* $[10,100]$*, ii) still holds provided that the uniform partition of the interval is fixed and known to both the encoder and decoder before the transmission, since the initial condition does not affect the asymptotic input power and how fast the error* $(\bar{W}-W)$ *decays.*

Another decoding method is to map $\hat{x}_{0,T}$ *directly into the closest sub-interval center and obtain the decoded message* $\hat{W}_T$*. This is asymptotically identical to the above decoding method. The scaling by* $1/(1-a^{-2T-2})$ *is to remove the exponentially vanishing bias in the estimate of* W*; see (3.34) and [43].*

In addition, if we use the time-varying Kalman filter gain instead of the steady-state Kalman filter gain, the above results also hold, since direct computation shows that the time-varying, transient Kalman filtering converges to its steady-state exponentially fast.

Remark 3. *The "error exponent" of the above scheme, defined as*

$$\lim_{T\to\infty} \frac{1}{T+1} \log(\log(\frac{1}{PE_T})), \tag{3.39}$$

equals $2(C_{fb}(\mathcal{P}) - R) = 2\epsilon$. *This is the one computed in [106], and is the* largest *(i.e. the best) error exponent that one can achieve; see [113]. Hence, our scheme is optimal in achieving the capacity and in achieving the best error exponent. Again, this is due to the fact that Kalman filter provides us the optimal estimate of the message, which leads to the fastest decay of probability of error.*

3.5 Connections of information, estimation, and control over an AWGN channel with feedback

In this section, we develop the connections of information, estimation, and control over the AWGN channel with feedback. This provides diversified interpretations why the above coding scheme is optimal. We note that the connections can be generalized to other channels to obtain the optimal coding schemes.

The Kalman filtering problem admits a different state-space representation from the Kalman filter based coding scheme. However, they generate equal signals $\{r_t\}$, $\{e_t\}$, $\{\hat{x}_t\}$, and $\{\hat{x}_{0,t}\}$, and hence are considered to be "equivalent" over any finite time horizon (the precise meaning of equivalence between different systems can be found in Appendix A.2). Below, we rewrite the dynamics of the two systems for convenience.

$$\text{estimation system:} \left\{ \begin{array}{lcl} x_{t+1} & = & a x_t \\ r_t & = & c x_t \\ \bar{y}_t & = & r_t + N_t \\ \hat{x}_{t+1} & = & a\hat{x}_t + L e_t \\ \hat{r}_t & = & C\hat{x}_t \\ e_t & = & \bar{y}_t - \hat{r}_t \\ \hat{x}_{0,t} & = & a^{-t-1}\hat{x}_{t+1} \end{array} \right. \begin{array}{l} \left.\vphantom{\begin{array}{l}x\\x\end{array}}\right\} \text{unknown source} \\ \left.\vphantom{x}\right\} \text{channel} \\ \left.\vphantom{\begin{array}{l}x\\x\\x\\x\end{array}}\right\} \text{Kalman filter} \end{array} \tag{3.40}$$

with $x_0 := W$ and $\hat{x}_0 := 0$.

$$\text{communication system:}\begin{cases} x_{t+1} = ax_t \\ r_t = cx_t \\ u_t = r_t - \hat{r}_t \end{cases}\text{encoder} \quad \begin{cases} y_t = u_t + N_t \end{cases}\text{channel} \quad \begin{cases} e_t = y_t \\ \hat{x}_{t+1} = a\hat{x}_t + Le_t \\ \hat{r}_t = C\hat{x}_t \\ \hat{x}_{0,t} = a^{-t-1}\hat{x}_{t+1} \end{cases}\text{decoder} \tag{3.41}$$

with $x_0 := W$ and $\hat{x}_0 := 0$. Note that this is also a *tracking-of-unstable-source* problem.

Now let

$$\tilde{x}_t := x_t - \hat{x}_t, \tag{3.42}$$

then both systems (3.40) and (3.41) become

$$\text{control system:}\begin{cases} \tilde{x}_{t+1} = (a - Lc)\tilde{x}_t - LN_t = a\tilde{x}_t - Le_t \\ e_t = c\tilde{x}_t + N_t \\ u_t = c\tilde{x}_t, \end{cases} \tag{3.43}$$

where $a > 1$ and $\tilde{x}_0 := W$; see Fig. 3.4 for its block diagram. It is a control system where we want to minimize the power of u by appropriately choosing L, while stabilizing the closed-loop. This is a *minimum energy control* (MEC) problem, which is equivalent to the Kalman filtering problem (see [67]). In essence, the MEC system in Fig. 4.7 is the closed-loop form of the Kalman filtering system and the Kalman filter based coding system, and it is sometimes called the innovations representation of the Kalman filtering system.

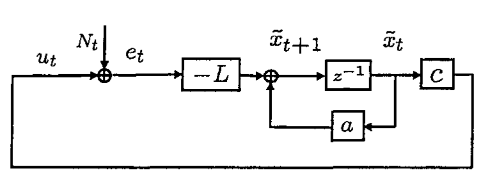


Figure 3.4 The block diagram for the MEC system.

The MEC system has interesting properties as summarized in the following lemma.

Lemma 1. *The optimal L that solves the MEC problem*

$$\min_{L \text{ stabilizing},(3.43)} \quad \lim_{t\to\infty} \frac{1}{t+1}\mathbf{E}u^t u^{t\prime} \tag{3.44}$$

is given by

$$L := \frac{1}{c}\left(a - \frac{1}{a}\right). \tag{3.45}$$

For the optimal L, it holds that

i) The closed-loop pole a_{cl} locates at the reciprocal of the open-loop pole, namely

$$a_{cl} := a - Lc = \frac{1}{a}; \tag{3.46}$$

ii) The transfer function from N to e is an all-pass transfer function

$$\mathcal{T}_{Ne}(z) = \frac{z-a}{z-a^{-1}} \tag{3.47}$$

with a flat power spectrum of magnitude a^2;

iii) $\{e_t\}$ is a white Gaussian process with zero mean and asymptotic variance

$$K_e = a^2. \tag{3.48}$$

Remark 4. *Note that the MEC problem, a special LQG problem, is also a problem of control over noisy channel. Interestingly, the optimal design of the controller L does not require any quantization or coding. Note also that the whiteness of $\{e_t\}$ is consistent with the results in Kalman filtering theory: It has been shown that the innovation processes in Kalman filtering problems are always white [57]; this property will be found especially useful in later chapters.*

Proof: Note that

$$\mathbf{E}(\tilde{x}_{t+1})^2 = (a-Lc)^2\mathbf{E}(\tilde{x}_t)^2 + L^2, \tag{3.49}$$

which leads to that, when the closed-loop is stabilized,

$$\lim_{t\to\infty} \mathbf{E}(\tilde{x}_t)^2 = \frac{L^2}{1-(a-Lc)^2} \tag{3.50}$$

and hence

$$\lim_{t\to\infty} \mathbf{E}(u_t)^2 = \frac{L^2c^2}{1-(a-Lc)^2}. \tag{3.51}$$

Then it is straightforward to see that (3.45) minimizes the average power of u. We can verify i) and ii) directly. The whiteness of $\{e_t\}$ follows from that $\mathcal{T}_{Ne}(z)$ is an all-pass transfer function with a flat spectrum. ∎

3.5.1 Information rate, CRB, and Bode integral

The above shown equivalence among the feedback communication system, estimation system, and control system immediately leads to the following results linking the information rate, CRB, and Bode integral. Define

$$R := \lim_{T\to\infty} \frac{1}{T+1} \overrightarrow{I}(u^T \to y^T) \tag{3.52}$$

to be the (asymptotic) information rate; $\boldsymbol{h}(y)$ to be the entropy rate of process $\{y_t\}$; and $S(e^{j2\pi\theta})$ to be the power spectrum of sensitivity function in (3.43), i.e.,

$$S(e^{j2\pi\theta}) = |\mathcal{T}_{Ny}(e^{j2\pi\theta})|^2. \tag{3.53}$$

Also define

$$\mathrm{MMSE}_{W,T} := \mathbf{E}(W|\bar{y}^T) \tag{3.54}$$

to be the MMSE of estimating W bases on observation $\bar{y}^T$ in the estimation system (3.40);

$$\mathcal{I}_{W,T} := \mathbf{E}\left(\frac{\partial \log p_{W,\bar{y}^T}(W,\bar{y}^T)}{\partial W}\right)^2 \tag{3.55}$$

to be the (Bayesian) Fisher information, where $p_{W,\bar{y}^T}(W,\bar{y}^T)$ is the joint density of W and $\bar{y}^T$; and

$$\mathrm{CRB}_{W,T} := (\mathcal{I}_{W,T})^{-1} \tag{3.56}$$

to be the (Bayesian) CRB [123]. Note that it always holds, as a fundamental limitation in estimation theory, that

$$\mathrm{MSE}_{W,T} \geq \mathrm{CRB}_{W,T}, \tag{3.57}$$

no matter how we design our estimator [123]. This inequality is referred to as the information inequality, Cramer-Rao inequality, or van Trees inequality.

Proposition 1. *Consider the feedback communication system (3.41), estimation system (3.40), and control system (3.43). Assume that $W \sim \mathcal{N}(0, P_W)$ where $0 < P_W < \infty$. It holds that*

$$\begin{aligned} R &:= \lim_{T\to\infty} \frac{1}{T+1} \vec{I}(u^T \to y^T) &&= \lim_{T\to\infty} \frac{1}{T+1} I(W; y^T) \\ &= \boldsymbol{h}(e) - \frac{1}{2}\log 2\pi e &&= \log a \\ &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \log S(e^{j2\pi\theta}) d\theta &&= \lim_{T\to\infty} \frac{\log \mathcal{I}_{W,T}}{2(T+1)} \\ &= -\lim_{T\to\infty} \frac{\log \mathrm{MMSE}_{W,T}}{2(T+1)} &&= -\lim_{T\to\infty} \frac{\log \mathrm{CRB}_{W,T}}{2(T+1)}. \end{aligned} \tag{3.58}$$

Remark 5. *This proposition links the information rate to the entropy rate of the innovations process $\{e_t\}$ (or equivalently $\{y_t\}$), the degree of instability, the Bode sensitivity integral, the increasing rate of the Fisher information, and the decreasing rate of the MMSE and CRB. It implies that the fundamental limitations in information, estimation, and control coincide, a generalization of [28]. It also implies that the increasing rate of directed mutual information is equal to the decreasing rate of MMSE and CRB. The Fisher information, as its name suggests, indeed has an interpretation in terms of mutual information defined in information theory.*

Proof: First, note that

$$\begin{aligned} \vec{I}(u^T \to y^T) &:= \sum_{t=0}^{T} I(u^t; y_t | y^{t-1}) \\ &= \sum_{t=0}^{T} \left(h(y_t|y^{t-1}) - h(y_t|y^{t-1}, u^t) \right) \\ &= h(y^T) - h(N^T) \\ &= h(y^T) - h(y^T|W) \\ &= I(W; y^T). \end{aligned} \tag{3.59}$$

This shows the first equality. The second follows from the definition of entropy rate

$$\boldsymbol{h}(y) = \boldsymbol{h}(e) := \lim_{T\to\infty} \frac{1}{T+1} h(y^T). \tag{3.60}$$

The third is due to that, since $\{e_t\}$ is white,

$$\boldsymbol{h}(e) = \frac{1}{2}\log(2\pi e K_e) = \frac{1}{2}\log 2\pi e + \log a. \tag{3.61}$$

The fourth is because the sensitivity transfer function in this case equals the all-pass transfer function $\mathcal{T}_{Ne}$.

We also note that

$$\begin{aligned} \mathrm{MSE}_{W,T} &= \mathbf{E}\left[a^{-4T-4}W^2 + a^{-4T-4}\sum_{j=0}^{T} a^{2j+2}L^2(N_j)^2\right] \\ &= \mathcal{P}a^{-2T-2}\left(1 - a^{-2T-2}\left(1 - \frac{P_W}{\mathcal{P}}\right)\right), \end{aligned} \tag{3.62}$$

which yields that

$$-\lim_{T\to\infty} \frac{\log \mathrm{MMSE}_{W,T}}{2(T+1)} = \log a. \tag{3.63}$$

Since the Kalman filtering system is linear driven by Gaussian random variables, it holds that

$$\mathrm{MSE}_{W,T} = \mathrm{MMSE}_{W,T} = \mathrm{CRB}_{W,T} = (\mathcal{I}_{W,T})^{-1}. \tag{3.64}$$

Thus we prove the proposition.

■

3.5.2 Connection to the SK coding scheme

The SK coding scheme that achieves the feedback capacity of the AWGN channel is illustrated in Fig. 3.5. In this figure, we can identify the encoder, AWGN channel, decoder, and the feedback link with one-step delay. The coding process is similar to that for the Kalman filter based coding scheme. In the original papers by Schalkwijk and Kailath, the SK coding scheme was obtained essentially as an application of the Robbins-Monro stochastic approximation procedure [107, 106]. Note that there are several versions of the SK scheme; Fig. 3.5 shows a slight variation of the original one proposed in [106].

The Sk coding scheme is another form of the Kalman filter based coding scheme. In other word, the SK coding scheme essentially implements the Kalman filtering algorithm. To see

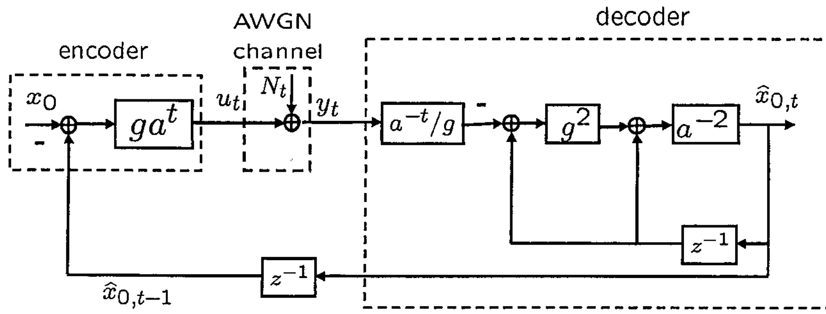


Figure 3.5 The SK coding scheme.

this, note that in the SK scheme, it holds that

$$\begin{aligned} u_t &= ga^t(\hat{x}_{0,t-1} - W) \\ \hat{x}_{0,t} &= \hat{x}_{0,t-1} - a^{-t-2} g y_t; \end{aligned} \tag{3.65}$$

and in the Kalman filter based scheme, it holds that

$$\begin{aligned} u_t &= ca^t(W - \hat{x}_{0,t-1}) \\ \hat{x}_{0,t} &= \hat{x}_{0,t-1} + a^{-t-2} L y_t. \end{aligned} \tag{3.66}$$

Letting

$$\begin{aligned} g &:= \sqrt{a^2 - 1} \\ c &:= -g, \end{aligned} \tag{3.67}$$

both schemes then generate identical channel inputs, outputs, and decoder estimates, respectively, and hence they are equivalent. The optimal choice of g in the SK coding scheme indeed corresponds to the optimal choice of Kalman filter gain.

3.6 Numerical example

In this section, we present numerical examples. We point out the Kalman filter based coding scheme and the original SK coding schemes, all suffer from the problem of numerical instability. Notice that, [107] involves exponentially growing bandwidth, [106] involves an exponentially growing parameter a^t where $a > 1$ and t denotes the time index, and the Kalman filter based coding scheme generates a feedback signal with exponentially growing power.

To overcome this problem and build a coding scheme feasible for simulation [1], we modify our Kalman filter based scheme. Notice that the innovations representation of a Kalman filtering system (i.e. the MEC system) is stable, namely both the unstable source and the unstable Kalman filter are inside the loop that is stabilized. Thus, we obtain a modification based on the innovation representation, as illustrated in Fig. 3.6. It follows the dynamics

$$\begin{cases} \tilde{x}_t &= a\tilde{x}_{t-1} - Ly_{t-1} \\ u_t &= c\tilde{x}_t \\ y_t &= u_t + N_t \\ \hat{x}_{0,t} &= \hat{x}_{0,t-1} + a^{-t-1}Ly_t, \end{cases} \tag{3.68}$$

where $\tilde{x}_{-1} := W/a$ (i.e. $\tilde{x}_0 = W$), $y_{-1} := 0$, and $\hat{x}_{0,-1} = 0$. It can also be derived from (3.41) by letting $\tilde{x}_t := x_t - \hat{x}_t$. Note that the dynamics of $\tilde{x}_t$ is stabilized, so no signals or parameters in (3.68) will be unbounded. In fact, the "control setup" indicated in Fig. 3.6 is the dynamics for the MEC system, stabilized in closed-loop. We remark that the modification coincides with the one studied by Gallager (p. 480, [43]) with minor differences; however, Gallager's scheme has not received enough attention in the literature.

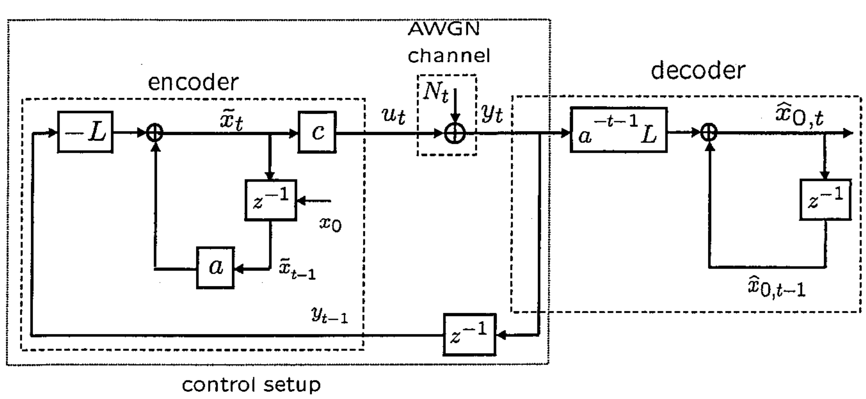


Figure 3.6 The modified feedback coding scheme for an AWGN channel.

We simulate using the modified coding scheme. Let us assume that the power budget is

[1]We claim that our modification is feasible for simulation purpose, since it is numerically stable. However, this modification is not yet feasible for practical purpose, mainly because of the strong assumption on the noiseless feedback. A more practical design is under current investigation.

$\mathcal{P} := 3$, i.e., $C(\mathcal{P}) = 1$ bit per channel use. This leads to that $a = 2$, $L = 1.5$. Fig. 3.7 shows one trial of the coding scheme, in which (a) illustrates how the time average of the channel input power converges to the given power budget, and (b) illustrates the exponential decay of the squared estimation error $(\tilde{x}_t)^2 = (x_t - \hat{x}_t)^2$, as time t increases.

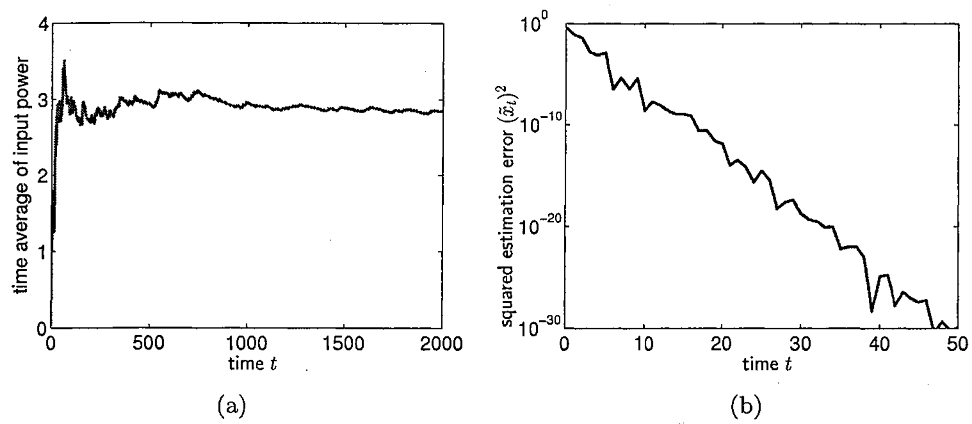


Figure 3.7 (a) Convergence of the average channel input power to the given power budget. (b) Exponential decay of squared estimation error $(\tilde{x}_t)^2$.

In Fig. 3.8, we report the simulated probability of error obtained by running multiple independent trials of the coding scheme. We choose $\epsilon = 0.01$, namely the communication rate $R = 0.99$ bit per channel use, very close to the Shannon capacity. For verification purpose, we also plot the theoretic probability of error, computed using the Gaussian Q-function. We can see that this scheme exhibits very good performance within 120 channel uses for a rate very close to the Shannon capacity.

3.7 Summary

In this section, we study the simple case of AWGN channels with feedback. We show that a Kalman filter based coding scheme achieves the feedback capacity. We draw connections among feedback communication, estimation, and control over this channel, and show that the optimality and fundamental limitations in these problems coincide. We finally verify our study by numerical examples. This chapter lays the foundation for the following chapters. As we will see, the concepts and approaches developed in this chapter carry over to more complicated

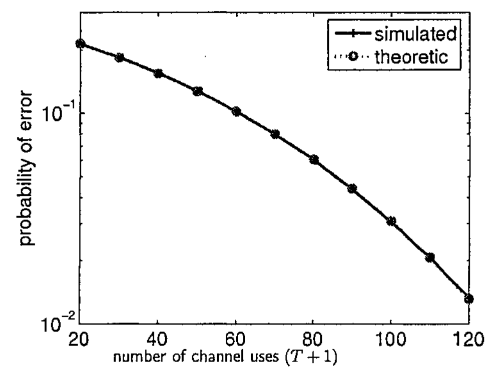


Figure 3.8 Simulated probability of error and theoretic probability of error.

feedback communication systems.

CHAPTER 4. FREQUENCY-SELECTIVE FADING GAUSSIAN CHANNELS WITH FEEDBACK

4.1 Introduction

As we reviewed in Chapter 2, communication systems in which the transmitters have access to noiseless feedback of channel outputs have been widely studied. As one of the most important case, the SISO frequency-selective Gaussian channels with feedback have attracted considerable attention; see [107, 106, 92, 5, 6, 17, 96, 135, 94, 108, 116, 136, 102, 104, 28, 60] and references therein for the capacity computation and coding scheme design for these channels. Note that SISO frequency-selective Gaussian channels are sometimes referred to as the Gaussian channels with memory, or general Gaussian channels, or simply Gaussian channels; for precise description of these channels, see Section 4.2.

We recap some of the major accomplishments for this channel. Schalkwijk and Kailath proposed the SK codes for AWGN channels with feedback, which achieve the asymptotic feedback capacity (i.e. the infinite-horizon feedback capacity, denoted C_∞ [1]) with reduced coding complexity and improved performance [107, 106]. Cover and Pombra presented a rather general coding structure, called the CP structure, to achieve the finite-horizon feedback capacity (denoted C_T, where the horizon spans from time epoch 0 to time epoch T) for Gaussian channels with memory; however, it involves prohibitive computation complexity as the coding length $(T+1)$ increases. Kim obtained the closed-form expression for a first-order moving-average noise channel based on the CP structure [60].

Tatikonda reformulated the problem of computing C_T as a stochastic control optimization problem, and proposed a dynamic programming based solution [116]. This idea was further explored by Yang, Kavcic, and Tatikonda, who uncovered the Markov property of the optimal input distributions for Gaussian channels and eventually reduced the finite-horizon stochas-

[1] We drop the subscript "fb" in feedback capacity notions from now on.

tic control optimization problem to a manageable size [136]. Moreover, under a *stationarity conjecture* that C_∞ equals the stationary capacity (the maximum information rate over all *stationary* input distributions), C_∞ for a general Gaussian channel with memory is given by the solution of a finite dimensional optimization problem. The stationary conjecture has been recently confirmed by Kim [61]. Sahai and Elia identified the connection of feedback communication to the problem of tracking unstable sources over a channel and the problem of feedback stabilization over a channel [102, 28].

As we can see from the literature, it remains an open and longstanding problem to build a coding scheme with reasonable complexity to achieve C_∞ for a Gaussian channel with memory; note that practical codes (rather than random codes) have not been found based on the optimal signalling strategy proposed by Yang, Kavcic, and Tatikonda in [136]. In this chapter, we propose a coding scheme for frequency-selective Gaussian channels with output feedback. This coding scheme achieves C_∞, the asymptotic feedback capacity of the channel; utilizes the Kalman filter algorithm; simplifies the coding processes; and shortens the coding delay. The optimal coding structure is essentially an FDLTI system, which has low design/operation complexity; is also an extension of the SK codes; and leads to a further simplification of the optimal signalling strategy in [136]. The construction of the coding system amounts to solving a finite-dimensional optimization problem. Our solution holds for AWGN channels with inter-symbol interference (ISI) where the ISI is model as a stable and minimum-phase FDLTI system; through the equivalence shown in [116, 136], this channel is equivalent to a colored Gaussian channel with rational noise power spectrums and without ISI. Note that the rationalness assumption is widely used and not too restrictive, since any power spectrum can be arbitrarily approximated by rational ones.

In deriving our optimal coding design in infinite-horizon, we first present finite-horizon analysis (which is closely related to the CP structure) of the feedback communication problem, and then let the horizon length tend to infinity and obtain our optimal coding design which achieves C_∞. In the finite-horizon analysis, we establish the necessity of the Kalman filter: The Kalman filter is not only a device to provide sufficient statistics (which was shown in [136]), but also a device to ensure the power efficiency and to recover the message optimally. This also leads to a refinement of the CP structure, applicable to any Gaussian channels with memory. Additionally, the presence of the Kalman filter in our coding scheme signifies the intrinsic connections among feedback communication, estimation, and control. In particular, we show that the feedback communication problem over a Gaussian channel is essentially an optimal estimation problem, and the achievable rate of the feedback communication system is alternatively given by the decay rate of the CRB for the associated estimation system. Invoking the Bode sensitivity characterization of the achievable rate [28], we conclude that the fundamental limitations in feedback communication, estimation, and control coincide. We then extend the horizon to infinity and characterize the steady-state of the feedback communication problem. We finally show that our optimal scheme achieves C_∞. One main insight gained in this study is that, the perspective of unifying information, estimation, and control, three fundamental concepts, facilitates our development of the optimal feedback communication design.

We also remark that the necessity of the Kalman filter in the optimal coding scheme is not surprising, given various indications of the essential role of Kalman filtering (or MMSE estimators; or MEC, its control theory equivalence; or the sum-product algorithm, its generalization) in optimal communication designs. See e.g. [66, 41, 137, 136, 28, 83]. The study of the Kalman filter in the feedback communication problem along the line of [83] may shed important insights on optimal communication problems and is under current investigation.

This chapter is organized as follows. In Section 4.2, we introduce the channel models. The problem formulation is given in Section 4.3, followed by the problem solution, i.e. the optimal coding scheme and the coding theorem. In Section 4.4, we prove the necessity of the Kalman filter in generating the optimal feedback. In Section 4.5, we provide the connections of the feedback communication problem to an estimation problem and a control problem, and express the achievable rate in terms of estimation theory quantities and control theory quantities. In Section 4.7, we show that our coding scheme is capacity-achieving. Section 4.8 provides a numerical example. Finally we summarize the chapter.

4.2 Channel model

In this section, we briefly describe two Gaussian channel models, namely the colored Gaussian noise channel without ISI and white Gaussian noise channel with ISI.

4.2.1 Colored Gaussian noise channel without ISI

Fig. 4.1 (a) shows a colored Gaussian noise channel without ISI. At time t, this discrete-time channel is described as

$$\tilde{y}_t = u_t + Z_t, \quad \text{for } t = 0, 1, \cdots, \tag{4.1}$$

where u_t is the channel input, Z_t is the channel noise, and $\tilde{y}_t$ is the channel output. We make the following assumptions: The colored noise $\{Z_t\}$ is the output of a finite-dimensional stable and minimum-phase LTI system $\mathcal{Z}(z)$ driven by a white Gaussian process $\{N_t\}$ of zero mean and unit variance, and $\mathcal{Z}(z)$ is at initial rest. For any block size (i.e. coding length) of $(T+1)$, we may equivalently generate Z^T by

$$Z^T = \mathcal{Z}_T N^T, \tag{4.2}$$

where $\mathcal{Z}_T$ is a $(T+1) \times (T+1)$ lower-triangular Toeplitz matrix of the impulse response of $\mathcal{Z}(z)$. We may abuse the notation $\mathcal{Z}$ for both $\mathcal{Z}(z)$ and $\mathcal{Z}_T$ if no confusion arises. As a consequence, $\{Z_t\}$ is asymptotically stationary. [2]

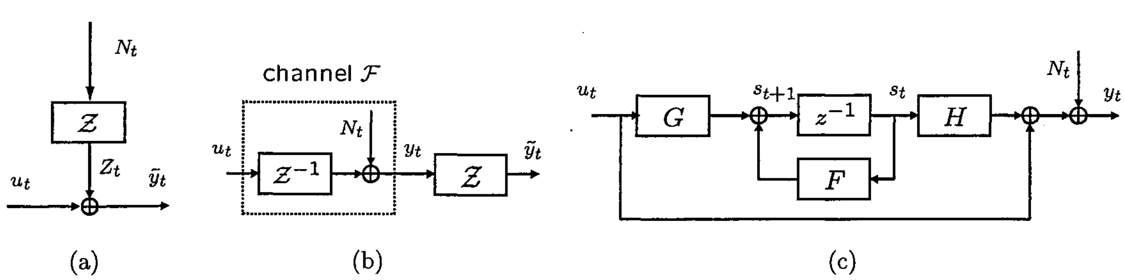


Figure 4.1 (a) A colored Gaussian noise channel without ISI. (b) The equivalent ISI channel with AWGN. (c) State-space realization of channel $\mathcal{F}$.

Note that there is no loss of generality in assuming that $\mathcal{Z}(z)$ is stable and minimum-phase (cf. Chapter 11, [98]), implying that the initial condition of $\mathcal{Z}(z)$ generates no effect on the steady-state. Thus we made the initial rest assumption since our ultimate goal is the

[2]The difference between a stationarity assumption and an *asymptotic* stationarity assumption may result from different starting points of the process: If starting from $t = -\infty$, $\{Z_t\}$ is *stationary*; instead if starting from $t = 0$ as we are assuming here, $\{Z_t\}$ is *asymptotically stationary.* They result in exactly the same steady-state analysis of the feedback communication problem.

steady-state characterization.

4.2.2 White Gaussian channel with ISI

The above colored Gaussian channel induces a new channel, namely a white Gaussian channel with ISI, under a further assumption that $\mathcal{Z}(\infty) \neq 0$ (i.e. $\mathcal{Z}$ is proper but non-strictly proper). More precisely, notice that from (4.1) and (4.2), we have

$$\tilde{y}^T = \mathcal{Z}_T(\mathcal{Z}_T^{-1}u^T + N^T), \tag{4.3}$$

which we identify as a stable and minimum-phase ISI channel with AWGN $\{N_t\}$, see Fig. 4.1 (b). Here $\mathcal{Z}^{-1}(z)$ is also at initial rest. For any fixed u^T and N^T, (a) and (b) generate the same channel output $\tilde{y}^T$. [3] Note that $\mathcal{Z}_T^{-1}$ is the matrix inverse of $\mathcal{Z}_T$, equal to the lower-triangular Toeplitz matrix of impulse response of $\mathcal{Z}^{-1}(z)$.

The initial rest assumption on $\mathcal{Z}^{-1}$ can be imposed in practice equivalently by, first driving the initial condition of the ISI channel to any desired value (known to the receiver) before a transmission, and then removing the response due to that initial condition at the receiver. Such an assumption is also used in [136, 116]. We further assume for simplicity that $\mathcal{Z}(\infty) = 1$; for cases where $g := \mathcal{Z}(\infty) \neq 1$, we can normalize $\mathcal{Z}(z)$ by scaling it by $1/g$. Hence, $\mathcal{Z}_T$ is a lower triangular Toeplitz matrix with diagonal elements all equal to 1 (and thus is invertible).

We can then write the minimal state-space representation of $\mathcal{Z}^{-1}$ as $(F, G, H, 1)$, where $F \in \mathbb{R}^m$ is stable, (F, G) is controllable, (F, H) is observable, and m is the *dimension* or *order* of $\mathcal{Z}^{-1}$. Let us denote the channel from u to y in Fig. 4.1(b) as $\mathcal{F}$, where

$$y^T := \mathcal{Z}_T^{-1}u^T + N^T = \mathcal{Z}_T^{-1}\tilde{y}^T. \tag{4.4}$$

The channel $\mathcal{F}$ is described in state-space as

$$\text{channel } \mathcal{F}: \begin{cases} s_{t+1} &= Fs_t + Gu_t \\ y_t &= Hs_t + u_t + N_t, \end{cases} \tag{4.5}$$

where $s_0 := 0$; see Fig. 4.1 (c). Notice that channel $\mathcal{F}$ is not essentially different than the

[3]More rigorously, the mappings from (u, N) to $\tilde{y}$ are T-equivalent. For a discussion about systems representations and equivalence between different representations, see Appendix A.2.

channel from u to $\tilde{y}$, since $\{y^t\}$ and $\{\tilde{y}^t\}$ causally determine each other.

We concentrate on the case $m \geq 1$; the case that m is 0 (i.e., $\mathcal{F}$ is an AWGN channel) was studied in Chapter 3.

4.3 Problem formulation in steady-state and the solution

Before formulating the steady-state communication problem, we distinguish among the three scenarios: Finite-horizon (i.e. finite coding length), infinite-horizon (i.e. infinite coding length), and steady state. Finite-horizon problems often have time-dependent (i.e. time-varying) and horizon-dependent solutions (similar to finite-horizon Kalman filtering). The horizon-dependence may be removed in the infinite-horizon scenario, and furthermore, the time-dependence may be removed in the steady-state scenario. If we find the (stationary, time-invariant) steady-state solution (which by [61] is also the infinite-horizon solution), we can truncate it and employ the truncation if the practical problem is in finite-horizon but the horizon is large enough. This truncated solution would greatly simplify the implementation while having a performance sufficiently close to finite-horizon optimality.

4.3.1 Problem formulation

For a Gaussian channel with feedback, the channel input may take the form

$$u_t = \gamma_t u^{t-1} + \eta_t y^{t-1} + \zeta_t \tag{4.6}$$

for any $\gamma_t \in \mathbb{R}^{1\times t}$, $\eta_t \in \mathbb{R}^{1\times t}$, and zero-mean Gaussian random variable $\zeta_t \in \mathbb{R}$ which is independent of u^{t-1} and y^{t-1} (cf. [116, 136]). Therefore, the channel inputs are allowed to depend on the channel outputs in a strictly causal manner. Our objective in this chapter is to *design encoder/decoder to achieve the (stationary) asymptotic capacity*, given by

$$C_\infty := C_\infty(\mathcal{P}) := \sup_{\substack{\{u_t\} \text{ stationary,(4.6)} \\ s.t.\ \lim_{T\to\infty} \mathbf{E} u^{T\prime} u^T/(T+1) \leq \mathcal{P}}} \lim_{T\to\infty} \frac{1}{T+1} \overrightarrow{I}(u^T \to y^T) \tag{4.7}$$

where $\mathcal{P} > 0$ is the power budget and $\overrightarrow{I}(u^T \to y^T)$ is the directed information from u^T to y^T (cf. [116]). This capacity has both operational and information meanings. For more details about C_∞, refer to Section 4.7.1.

The problem of solving C_∞ may be equivalently formulated as minimizing the average channel input power while keeping the information rate bounded from below, namely for $\mathcal{R} > 0$,

$$P_\infty(\mathcal{R}) := \inf_{\substack{\{u_t\} \text{ stationary},(4.6) \\ s.t. \lim_{T\to\infty} I(u^T \to y^T)/(T+1) \geq \mathcal{R}}} \lim_{T\to\infty} \frac{1}{T+1} \mathbf{E} u^{T\prime} u^T. \tag{4.8}$$

Therefore $P_\infty(\mathcal{R})$ is the inverse function of $C_\infty(\mathcal{P})$, i.e., $C_\infty(P_\infty(\mathcal{R})) = \mathcal{R}$.

Approach: Our approach to solve the steady-state communication problem is to investigate the finite-horizon problem first, and then let the horizon increase to infinity, which leads to a unified view of infinite-horizon and finite-horizon in terms of Kalman filter based coding scheme. Other approaches not pursued in this thesis are also possible, such as applying the idea in [28] to the optimal signalling strategy in [136], though they generate results not as rich as our approach does.

4.3.2 Coding scheme

The rest of this section presents the solution to the above problem. In this subsection, we introduce an encoder/decoder structure and explain how to choose the parameters to ensure the optimality, and then describe the encoding/decoding process, that is, how we assign the message to be transmitted, and how we recover the message. In the next subsection, we present the coding theorem which states that our encoding/decoding structure with the chosen parameters achieves C_∞. The proof of the theorem will be developed in Sections 4.4 to 4.7.

The encoder/decoder structure

In state-space, the encoder and decoder are described as

$$\text{Encoder:} \begin{cases} x_{t+1} &= Ax_t \\ r_t &= Cx_t \\ u_t &= r_t - \hat{r}_t \end{cases} \tag{4.9}$$

and

$$\text{Decoder:}\begin{cases} \hat{s}_{t+1} &= F\hat{s}_t + L_2 e_t \\ e_t &= y_t - H\hat{s}_t \\ \hat{x}_{t+1} &= A\hat{x}_t + L_1 e_t \\ \hat{r}_t &= C\hat{x}_t \\ \hat{x}_{0,t} &= A^{-t-1}\hat{x}_{t+1}, \end{cases} \tag{4.10}$$

where $\hat{s}_0 := 0$, $\hat{x}_0 := 0$, $A \in \mathbb{R}^{(n+1)\times(n+1)}$, $C \in \mathbb{R}^{1\times(n+1)}$, $L_1 \in \mathbb{R}^{n+1}$, and $L_2 \in \mathbb{R}^m$. We call $(n+1)$ the *encoder dimension*, x_t the *encoder state*, and $\hat{x}_{0,t}$ the *decoder estimate*. See Fig. 4.2 for the block diagram. Observe that $-\hat{r}_t$ is the feedback from the decoder based on the channel output y^{t-1}, and thus u_t depends on y^{t-1} but not y_t. It further follows that $-\hat{r}^t = \mathcal{G}_t^* y^t$ for some strictly lower triangular Toeplitz matrix $\mathcal{G}_t^*$. Here A, C, u_t, etc. depend on n, but we do not specify the dependence explicitly to simplify notations.

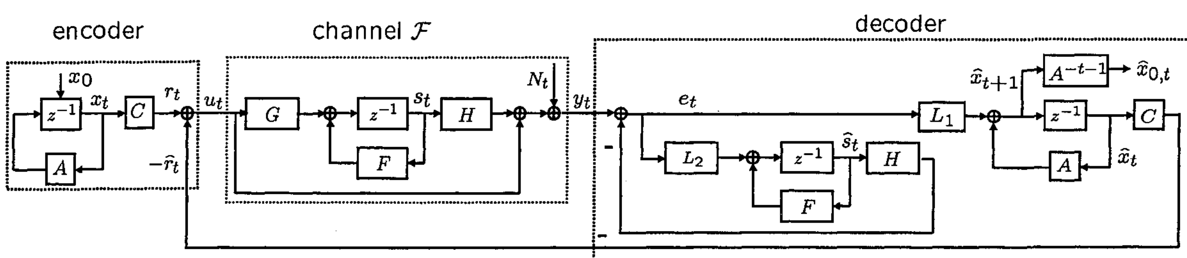


Figure 4.2 The encoder/decoder structure for $\mathcal{F}$.

Optimal choice of parameters

Fix a desired rate $\mathcal{R}$. Let $DI := 2^{\mathcal{R}}$ and $n := m-1$ (recalling that m is the channel dimension), and solve the optimization problem

$$[\boldsymbol{a}_f^{opt}, \Sigma^{opt}] := \underset{\substack{\boldsymbol{a}_f\in\mathbb{R}^n \\ s.t.\ \Sigma=\mathbb{A}\Sigma\mathbb{A}'-\mathbb{A}\Sigma\mathbb{C}'\mathbb{C}\Sigma\mathbb{A}'/(\mathbb{C}\Sigma\mathbb{C}'+1)}}{\arg\inf} \mathbb{D}\Sigma\mathbb{D}', \tag{4.11}$$

where

$$\mathbb{A} := \left[\begin{array}{c|c} A & 0 \\ \hline GC & F \end{array}\right], \mathbb{C} := [C \quad H], \mathbb{D} := [C \quad 0], A := \left[\begin{array}{c|c} 0_{n\times 1} & I_n \\ \hline \pm DI & \boldsymbol{a}_f \end{array}\right], C := \left[\begin{array}{cc} 1 & 0_{1\times n} \end{array}\right]. \tag{4.12}$$

Note that we need to solve (4.11) twice (one for $+DI$ in A and one for $-DI$ in A), and choose the optimal solution as the one with the smaller objective function value. Then we form the optimal A^{opt} based on $\boldsymbol{a}_f^{opt}$, and let (n^*+1) be the number of unstable eigenvalues in A^{opt}, where $n^* \geq 0$.

Now let $n := n^*$, solve (4.11) again, and obtain a new $\boldsymbol{a}_f^{opt}$ and Σ^{opt}. Then form A^{opt}, let $A^* = A^{opt}$, $\Sigma^* = \Sigma^{opt}$, $C^* := [1, 0_{1\times n^*}]$, and form $\mathbb{A}^*$,$\mathbb{C}^*$, and $\mathbb{D}^*$. Let

$$L^* := \begin{bmatrix} L_1^* \\ L_2^* \end{bmatrix} := \frac{\mathbb{A}^*\Sigma^*\mathbb{C}^{*\prime}}{\mathbb{C}^*\Sigma^*\mathbb{C}^{*\prime}+1}. \tag{4.13}$$

As we will show, (A^*, C^*) is observable, and A^* has exactly (n^*+1) unstable eigenvalues.

We assign the encoder/decoder parameters to the scheme built in Fig. 4.2 by letting

$$n := n^*, A := A^*, C := C^*, L_1 := L_1^*, L_2 := L_2^*. \tag{4.14}$$

We then drive the initial condition s_0 of channel $\mathcal{F}$ to zero. Now we are ready to communicate at a rate $\mathcal{R}$ using power $P_\infty(\mathcal{R}) = \mathbb{D}^*\Sigma^*\mathbb{D}^{*\prime}$. [4]

4.3.2.1 Encoding/Decoding process

Transmission of analog source

The designed communication system can transmit either an analog source or a digital message. In the former case, we assume that the encoder wishes to convey a Gaussian random vector through the channel and the decoder wishes to learn the random vector, which is a rate-distortion problem. The coding process is as follows. Assume that the to-be-conveyed message W is distributed as $\mathcal{N}(0, I_{n^*+1})$ (noting that any non-degenerate (n^*+1)-variate Gaussian vector W can be transformed into this form). Assume that the coding length is $(T+1)$. To encode, let $x_0 := W$. Then run the system till time epoch T, obtaining $\hat{x}_{0,t}$, $t = 0, 1, \cdots, T$. To decode, let $\hat{W}_t := \hat{x}_{0,t}$ for $t = 0, 1, \cdots, T$.

The quantities of interest include the squared-error distortion, defined as

$$\mathrm{MSE}(\hat{W}_t) := \mathbf{E}(W - \hat{W}_t)(W - \hat{W}_t)'. \tag{4.15}$$

[4] We see from (4.11) that for *any* channel $\mathcal{F}$, a simple upper bound of the function $P_\infty(\mathcal{R})$ is given by $\min\{(2^{2\mathcal{R}}-1)(\mathcal{Z}(2^{\mathcal{R}}))^2, (2^{2\mathcal{R}}-1)(\mathcal{Z}(-2^{\mathcal{R}}))^2\}$, obtained by using one unstable eigenvalue in A.

It will become clear that $\mathrm{MSE}(\hat{W}_t)$ can be pre-computed before the transmission, and thus the coding length can be determined *a priori* to ensure a desired distortion level.

Transmission of digital message

To transmit digital messages over the communication system, let us first fix $\epsilon > 0$ small enough and the coding length $(T+1)$ large enough. Let

$$\Sigma_x^* := [I_{n^*+1}, 0]\Sigma^*[I_{n^*+1}, 0]'. \tag{4.16}$$

Assume that the matrix $(A^{*\prime})^{-T-1}\Sigma_x^*(A^*)^{-T-1}$ has an eigenvalue decomposition as

$$(A^{*\prime})^{-T-1}\Sigma_x^*(A^*)^{-T-1} = E_T\Lambda_T E_T', \tag{4.17}$$

where $E_T = [e^{(1)}, \cdots, e^{(n^*+1)}]$ is an orthonormal matrix and Λ_T is a positive diagonal matrix. Let $\sigma_{T,i}$ be the square root of the (i,i)th element of Λ_T. Let $\mathcal{B} \in \mathbb{R}^{n^*+1}$ be the unit hypercube spanned by columns of E_T, that is,

$$\mathcal{B} = \left\{ \sum_{i=0}^{n^*} \alpha^{(i)} e^{(i)} \middle| \alpha^{(i)} \in [-\frac{1}{2}, \frac{1}{2}], i = 0, \cdots, n^* \right\}. \tag{4.18}$$

Next we partition the ith side of $\mathcal{B}$ into $(\sigma_{T,i})^{-(1-\epsilon)}$ segments. This induces a partition of $\mathcal{B}$ into M_T sub-hypercubes, where

$$\begin{aligned} M_T &= \prod_{i=0}^{n^*} (\sigma_{T,i})^{-(1-\epsilon)} \\ &= \left[\det\left((A^{*\prime})^{-T-1}\Sigma^*(A^*)^{-T-1}\right)\right]^{-\frac{1-\epsilon}{2}}. \end{aligned} \tag{4.19}$$

We then map the sub-hypercube centers to a set of M_T equally likely messages. The above procedure is known to both the transmitter and receiver *a priori.*

Suppose now we wish to transmit the message represented by the center W. To encode, let $x_0 := W$. Then run the system till time epoch T. To decode, we map $\hat{x}_{0,T}$ into the closest sub-hypercube center and obtain the decoded message $\hat{W}_T$. We declare an error if $\hat{W}_T \neq W$, and call a (an asymptotic) rate

$$R := \lim_{T\to\infty} \frac{1}{T+1} \log M_T \tag{4.20}$$

achievable if the probability of error PE_T vanishes as T tends to infinity. We remark that this coding process is the one used in [28] for Gaussian channels with memory, which was an extension of the SK codes. We also remark that, similar to the analog transmission case, the coding length $(T+1)$ can be pre-determined.

As we have seen, the encoder/decoder design and the encoding/decoding process can be done rather easily. The computation complexity for encoding/decoding grows as $O(T)$. Once again, as in the AWGN case, the encoder may be viewed as a control system, and the decoder may be viewed as an estimation system.

4.3.3 Coding theorem

Theorem 2. *Construct the encoder/decoder shown in Fig. 4.2 using n^*, A^*, C^*, L_1^*, and L_2^*. Then under the power constraint $\mathbf{E}u^2 \leq \mathcal{P}$,*

i) The coding scheme transmits an analog source $W \sim \mathcal{N}(0, I_{n^*+1})$ from the encoder to the decoder at rate $C_\infty(\mathcal{P})$, with MSE distortion $\mathrm{MSE}(\hat{W}_T)$ achieving the optimal asymptotic rate-distortion tradeoff given by

$$C_\infty(\mathcal{P}) = \lim_{T\to\infty} \frac{1}{2(T+1)} \log \frac{1}{\det \mathrm{MSE}(\hat{W}_T)}. \tag{4.21}$$

ii) The coding scheme can transmit digital message from the encoder to the decoder at a rate arbitrarily close to $C_\infty(\mathcal{P})$, with PE_T decays to zero doubly exponentially.

The proof of the theorem will be developed in the subsequent four sections. In Section 4.4, we consider a general coding structure in finite-horizon which may be viewed as a generalization of our optimal coding structure. We show that this general structure essentially contains a Kalman filter. The presence of the Kalman filter links the feedback communication problem to an estimation problem and a control problem, and hence we rewrite the information rate in terms of estimation theory quantities and control theory quantities; see Section 4.5. Sections 4.4 and 4.5 are focused on finite-horizon. In Section 4.6, we extend the horizon to infinity and characterize the steady-state behavior. Then in Section 4.7, we show that our optimal encoder/decoder design is actually the solution to the steady-state communication problem. As we can see, the development highly relies on the interactions among information, estimation, and control.

4.4 Finite horizon: Necessity of Kalman filter for optimal coding

In this section, we consider a finite-horizon coding structure that includes our optimal design in Section 4.3 as a special case. This general structure is useful since: 1) searching over all possible parameters in the general structure achieves C_∞, that is, there is no loss of generality or optimality to focus on this structure only; 2) we can show that to ensure power efficiency (to be explained), the general structure necessarily contains a Kalman filter. The general coding structure is in fact a variation of the CP structure, and hence our Kalman filter characterization leads to a refinement of the CP structure.

4.4.1 Feedback capacity C_T

The following definition of feedback capacity is based on [116].

Definition 1. *The "operational" or "information" finite-horizon feedback capacity C_T, subject to the average channel input power constraint*

$$P_T := \lim_{T\to\infty} \frac{1}{T+1} \mathrm{E} u^{T\prime} u^T \leq \mathcal{P}, \tag{4.22}$$

is

$$C_T(\mathcal{P}) := C_T := \sup \frac{1}{T+1} \vec{I}(u^T \to y^T), \tag{4.23}$$

where $\vec{I}(u^T \to y^T)$ is the directed information from u^T to y^T, and the supremum is over all possible feedback-dependent input distributions satisfying (4.22) and in the form

$$u_t = \gamma_t u^{t-1} + \eta_t y^{t-1} + \zeta_t \tag{4.24}$$

for any $\gamma_t \in \mathbb{R}^{1\times t}$, $\eta_t \in \mathbb{R}^{1\times t}$, and zero-mean Gaussian random variable $\zeta_t \in \mathbb{R}$ independent of u^{t-1} and y^{t-1}.

4.4.2 A general coding structure

Fig. 4.3 illustrates the general coding structure, including the encoder and the *feedback generator*, a portion of the decoder. Below, we fix the time horizon to be $\{0, 1, \cdots, T\}$ and describe the coding structure.

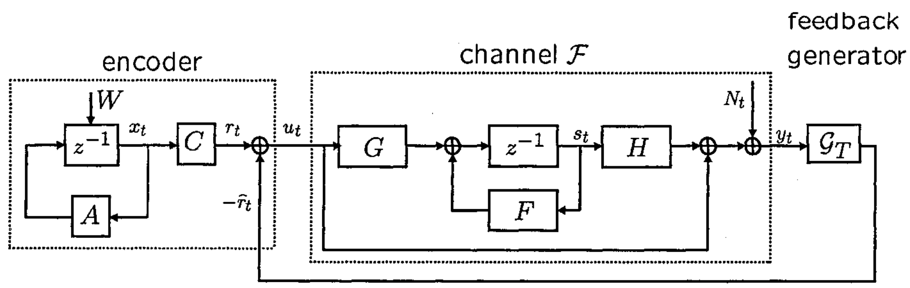


Figure 4.3 A general coding structure for channel $\mathcal{F}$.

Encoder: The encoder follows the dynamics (4.9). We assume that the encoder dimension $(n+1)$ satisfies $0 \le n \le T$, $W \sim \mathcal{N}(0, I_{n+1})$, $A \in \mathbb{R}^{(n+1)\times(n+1)}$, $C \in \mathbb{R}^{1\times(n+1)}$, (A, C) is observable, and none of the eigenvalues of A are on the unit circle or at the locations of the eigenvalues of F. We then let

$$\begin{array}{lllll} \Gamma_n(A,C) & := & \Gamma_n & := & [C', A'C', \cdots, A^{n\prime}C']' \\ \Gamma(A,C) & := & \Gamma & := & [C', A'C', \cdots, A^{T\prime}C']' \\ K_r^{(T)}(A,C) & := & K_r^{(T)} & := & \mathbf{E}r^T r^{T\prime}. \end{array} \tag{4.25}$$

Therefore, Γ_n is the observability matrix for (A, C) and is invertible, Γ has rank $(n+1)$, $r^T = \Gamma W$, and $K_r^{(T)} = \Gamma\Gamma'$.

Feedback generator: The feedback signal $-\hat{r}_t$ is generated through the *feedback generator* $\mathcal{G}_T$, i.e.

$$-\hat{r}^T = \mathcal{G}_T y^T. \tag{4.26}$$

We assume that $\mathcal{G}_T \in \mathbb{R}^{(T+1)\times(T+1)}$ is a strictly lower triangular matrix. Clearly, the optimal encoder/decoder can be viewed as a special case of the general structure. Throughout the chapter, the above assumptions on the encoder/decoder are always assumed unless otherwise specified. For future use purpose, we compute the channel output as

$$y^T = (I - \mathcal{Z}_T^{-1}\mathcal{G}_T)^{-1}(\mathcal{Z}_T^{-1} r^T + N^T). \tag{4.27}$$

Definition 2. *Consider the general coding structure shown in Fig. 4.3. Define*

$$C_{T,n} := C_{T,n}(\mathcal{P}) := \sup_{\substack{A\in\mathbb{R}^{(n+1)\times(n+1)},C,\mathcal{G}_T \\ s.t.\ \mathbf{E}u^{T\prime}u^T/(T+1)\leq\mathcal{P}}} \frac{1}{T+1} I(W;y^T) \tag{4.28}$$

and define its inverse function as $P_{T,n}(\mathcal{R})$.

In other words, $C_{T,n}$ is the finite-horizon information capacity for a *fixed transmitter dimension n*. It holds that $C_{n,n} = C_n$ and hence $\lim_{n\to\infty} C_{n,n} = C_\infty$ (see Lemma 2 and Section 4.7.1). Moreover, as we will show, C_∞ can be achieved using this structure.

4.4.3 Relation between the CP structure for ISI Gaussian channel and the general coding structure

In this subsection, we outline the relation of the CP structure to our general coding structure. In fact, our general coding structure was obtained by studying the CP structure for ISI Gaussian channels. This explains how the general coding structure (and hence the optimal coding scheme) was formulated.

4.4.3.1 CP structure for colored Gaussian noise channel

We briefly review the CP coding structure for the colored Gaussian noise channel specified in Section 4.2.1; see [17, 16] for more details of the CP structure. Let the colored Gaussian noise Z^T have covariance matrix $K_Z^{(T)}$, and

$$u^T := \mathcal{B}_T Z^T + v^T, \tag{4.29}$$

where $\mathcal{B}_T$ is a $(T+1)\times(T+1)$ strictly lower triangular matrix, v^T is Gaussian with covariance $K_v^{(T)} \geq 0$ and is independent of Z^T. [5] This generates channel output

$$\tilde{y}^T = (I + \mathcal{B}_T)Z^T + v^T. \tag{4.30}$$

[5]This v^T is called innovations in [16, 136]; it should not be confused with the Kalman filter innovations in this thesis.

Then the highest rate that the CP structure can achieve in the sense of operational and information is

$$
\begin{aligned}
C_{T,CP}(\mathcal{P}) &= \sup \frac{1}{T+1} I(v^T; \tilde{y}^T) \\
&= \sup \frac{1}{2(T+1)} \log \frac{\det K_{\tilde{y}}^{(T)}}{\det K_Z^{(T)}} \\
&= \sup \frac{1}{2(T+1)} \log \frac{\det((I+\mathcal{B}_T)K_Z^{(T)}(I+\mathcal{B}_T)' + K_v^{(T)})}{\det K_Z^{(T)}},
\end{aligned}
\tag{4.31}
$$

where the supremum is taken over all admissible $K_v^{(T)}$ and $\mathcal{B}_T$ satisfying the power constraint

$$
P_T := \frac{1}{T+1} \mathrm{tr}(\mathcal{B}_T K_Z^{(T)} \mathcal{B}_T' + K_v^{(T)}) \leq \mathcal{P}. \tag{4.32}
$$

Since the operational capacity definitions in [17] and [116] coincide, we have $C_{T,CP}(\mathcal{P}) = C_T(\mathcal{P})$. This may also be seen by observing that, any channel input (4.24) can be rewritten in the form of (4.29), but since (4.24) is sufficient to achieve C_T, we conclude that (4.29) is also sufficient to achieve C_T.

4.4.3.2 CP structure for ISI Gaussian channel

By using the equivalence between the colored Gaussian noise channel and the ISI channel $\mathcal{F}$, we can derive the CP coding structure for $\mathcal{F}$, which is obtained from (4.29) by introducing a new quantity r^T as

$$
r^T := (I+\mathcal{B}_T)^{-1} v^T. \tag{4.33}
$$

By $Z^T = \mathcal{Z}_T N^T$ and $\tilde{y}^T = \mathcal{Z}_T y^T$, we have

$$
\begin{aligned}
u^T &= \mathcal{B}_T \mathcal{Z}_T N^T + (I+\mathcal{B}_T) r^T \\
y^T &= \mathcal{Z}_T^{-1}(I+\mathcal{B}_T)\mathcal{Z}_T N^T + \mathcal{Z}_T^{-1}(I+\mathcal{B}_T) r^T \\
&= \mathcal{Z}_T^{-1}(I+\mathcal{B}_T)(\mathcal{Z}_T N^T + r^T).
\end{aligned}
\tag{4.34}
$$

This implies that, the channel input u^T can be represented as

$$
u^T = (I+\mathcal{B}_T)^{-1} \mathcal{B}_T \mathcal{Z}_T y^T + r^T, \tag{4.35}
$$

which leads to the block diagram in Fig. 4.4.

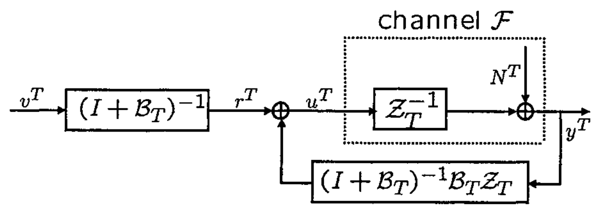


Figure 4.4 The block diagram of the CP structure for ISI Gaussian channel $\mathcal{F}$.

The capacity C_T now takes the form

$$\begin{aligned} C_T(\mathcal{P}) &= \sup \frac{1}{2(T+1)} \log\det K_y^{(T)} \\ &= \sup \frac{1}{2(T+1)} \log\det \left(\mathcal{Z}_T^{-1}(I+\mathcal{B}_T)(\mathcal{Z}_T\mathcal{Z}_T' + K_r^{(T)})(I+\mathcal{B}_T)'\mathcal{Z}_T^{-1\prime}\right) \\ &= \sup \frac{1}{2(T+1)} \log\det(\mathcal{Z}_T\mathcal{Z}_T' + K_r^{(T)}) \end{aligned} \tag{4.36}$$

where the supremum is over the power constraint

$$P_T := \frac{1}{T+1}\mathrm{tr}(\mathcal{B}_T\mathcal{Z}_T\mathcal{Z}_T'\mathcal{B}_T' + (I+\mathcal{B}_T)K_r^{(T)}(I+\mathcal{B}_T)') \leq \mathcal{P}. \tag{4.37}$$

It is easily seen that the capacity in this form is identical to (4.31).

4.4.3.3 Relation of the CP structure with the general coding structure

We can establish correspondence relationship between the CP structure for ISI Gaussian channel $\mathcal{F}$ in Fig. 4.4 and the general coding structure for $\mathcal{F}$ in Fig. 4.2. In fact, the general coding structure for $\mathcal{F}$ in Fig. 4.2 was initially motivated by the CP structure for channel $\mathcal{F}$ in Fig. 4.4.

For any fixed $(K_r^{(T)}, \mathcal{B}_T)$ in the CP structure, define in the general coding structure that

$$\begin{aligned} \mathcal{G}_T &:= (I+\mathcal{B}_T)^{-1}\mathcal{B}_T\mathcal{Z}_T \\ A &:= \Gamma_0^{-1}\left[\begin{array}{c|c} 0 & I_T \\ \hline * & * \end{array}\right]\Gamma_0 \in \mathbb{R}^{(T+1)\times(T+1)} \\ C &:= [1 \;\; 0 \;\; \cdots \;\; 0]\Gamma_0, \end{aligned} \tag{4.38}$$

where $\Gamma_0 := (K_r^{(T)})^{\frac{1}{2}}$, and $*$ can be any number. (Note that the case $K_r^{(T)} \geq 0$ but $K_r^{(T)}$ is not positive definite can be approached by a sequence of positive definite $K_r^{(T)}$, and thus it is sufficient to consider only positive definite $K_r^{(T)}$ in establishing the correspondence relation of the two structures.) Then it is easily verified that $\mathcal{G}_T$ is strictly lower triangular, (A, C) is observable with a nonsingular observability matrix $\Gamma = \Gamma_0$, and A can have eigenvalues not on the the unit circle and not at the locations of F's eigenvalues. Therefore, for any given $(K_r^{(T)}, \mathcal{B}_T)$, we can find an admissible $(A, C, \mathcal{G}_T)$, and it is straightforward to verify that they generate identical channel inputs u^T.

Conversely, for any fixed admissible $(A, C, \mathcal{G}_T)$ with $\in \mathbb{R}^{(n+1)\times(n+1)}$, we can obtain an admissible $(K_r^{(T)}, \mathcal{B}_T)$ as

$$\begin{aligned} \mathcal{B}_T &:= \mathcal{G}_T \mathcal{Z}_T^{-1}(I - \mathcal{G}_T \mathcal{Z}_T^{-1})^{-1} \\ K_r^{(T)} &:= \Gamma(A, C)\Gamma(A, C)', \end{aligned} \tag{4.39}$$

which generates identical channel input u^T as $(A, C, \mathcal{G}_T)$ does.

As a result of the above reasoning, there is a corresponding relation between the CP structure for $\mathcal{F}$ and the general coding structure, and the maximum rate over all admissible $(K_r^{(T)}, \mathcal{B}_T)$ (namely C_T) equals that over all admissible $(A, C, \mathcal{G}_T)$. In other words, we have

Lemma 2.

$$C_T(\mathcal{P}) = C_{T,T}(\mathcal{P}). \tag{4.40}$$

Proof: Note that $C_{T,T}$ is the maximum rate over all admissible $(A, C, \mathcal{G}_T)$ with $\in \mathbb{R}^{(T+1)\times(T+1)}$. ∎

This lemma implies that the general coding structure with an extra constraint $T = n$ becomes the CP structure, that is, in the CP structure, the dimension of A is equal to the horizon length. One advantage of considering the general coding structure is that we can allow $T \neq n$, which makes it possible to increase the horizon length to infinity without increasing the dimension of A, a crucial step towards the Kalman filtering characterization of the feedback communication problem.

4.4.4 The presence of Kalman filter

We first compute the mutual information in the general coding structure.

Proposition 2. *Consider the general coding structure in Fig. 4.3. Fix any* $0 \leq n \leq T$*, and fix any* A, C*, and* $\mathcal{G}_T$*. Then it holds that*

$$\begin{aligned} I(W; y^T) &= I(r^T; y^T) \\ &= \vec{I}(u^T \to y^T) \\ &= \tfrac{1}{2} \log \det K_y^{(T)} \\ &= \tfrac{1}{2} \log \det(I + \mathcal{Z}_T^{-1} K_r^{(T)} \mathcal{Z}_T^{-1'}) \\ &= \tfrac{1}{2} \log \det(I + \mathcal{Z}_T^{-1} \Gamma \Gamma' \mathcal{Z}_T^{-1'}), \end{aligned} \tag{4.41}$$

which is independent *of* $\mathcal{G}_T$.

Proof:

$$\begin{aligned} I(W; y^T) &= h(y^T) - h(y^T | W) \\ &= h(y^T) - h\left((I - \mathcal{Z}_T^{-1} \mathcal{G}_T)^{-1} (\mathcal{Z}_T^{-1} r^T + N^T) | W\right) \\ &\stackrel{(a)}{=} \tfrac{1}{2} \log \det(2\pi e K_y^{(T)}) - h(N^T) \\ &\stackrel{(b)}{=} \vec{I}(u^T \to y^T) \\ &= \tfrac{1}{2} \log \det K_y^{(T)} \\ &= \tfrac{1}{2} \log \det(I + \mathcal{Z}_T^{-1} K_r^{(T)} \mathcal{Z}_T^{-1'}), \end{aligned} \tag{4.42}$$

where (a) is due to $r^T = \Gamma W$, $\det(AB) = \det A \det B$, and $\det(I - \mathcal{Z}_T^{-1} \mathcal{G}_T)^{-1} = 1$; and (b) follows from [28]. ∎

Proposition 2 implies that $I(W; y^T)$ is independent of the feedback generator $\mathcal{G}_T$, and dependent only on $K_r^{(T)}$ or equivalently on (A, C). Thus, fixed (A, C) implies *fixed information rate*, and hence the optimal feedback generator has to be chosen to minimize the average channel input power, which turns out to contain a Kalman filter. Note that the counterpart of this proposition in infinite-horizon was proven in [28]. Now we can define, for a fixed (A, C), the information rate across the channel to be

$$R_T(A, C) := \frac{I(W; y^T)}{T+1}. \tag{4.43}$$

The optimal feedback generator for a given (A, C) is found in the next proposition.

Proposition 3. *Consider the general coding structure in Fig. 4.3. Fix any* $0 \le n \le T$. *Then*

i)

$$P_{T,n}(\mathcal{R}) = \inf_{\substack{A,C,\mathcal{G}_T := \mathcal{G}_T^*(A,C) \\ s.t.\ R_T(A,C) \ge \mathcal{R}}} \frac{1}{T+1} \mathbf{E} u^{T\prime} u^T \tag{4.44}$$

where $\mathcal{G}_T^*(A,C)$ *is the* optimal feedback generator *for a given* (A,C), *defined as*

$$\mathcal{G}_T^*(A,C) := \arg \inf_{(A,C) \text{ fixed}, \mathcal{G}_T} \frac{1}{T+1} \mathbf{E} u^{T\prime} u^T. \tag{4.45}$$

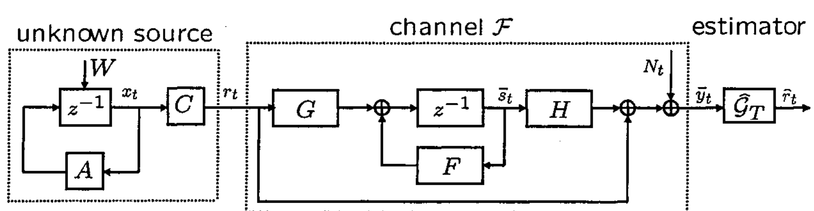
ii) The optimal feedback generator $\mathcal{G}_T^*(A,C)$ *is given by*

$$\mathcal{G}_T^*(A,C) = -\widehat{\mathcal{G}}_T^*(A,C)(I - \mathcal{Z}_T^{-1} \widehat{\mathcal{G}}_T^*(A,C))^{-1}, \tag{4.46}$$

where $\widehat{\mathcal{G}}_T^*(A,C)$ *is the strictly causal MMSE estimator (Kalman filter) of* r^T *given the noisy observation* $\bar{y}^T := \mathcal{Z}_T^{-1} r^T + N^T$, *i.e.,*

$$\widehat{\mathcal{G}}_T^*(A,C) := \arg \inf_{\widehat{\mathcal{G}}_T \in \mathbb{R}^{(T+1)\times(T+1)}} \frac{1}{T+1} \mathbf{E}(r^T - \widehat{\mathcal{G}}_T \bar{y}^T)(r^T - \widehat{\mathcal{G}}_T \bar{y}^T)', \tag{4.47}$$

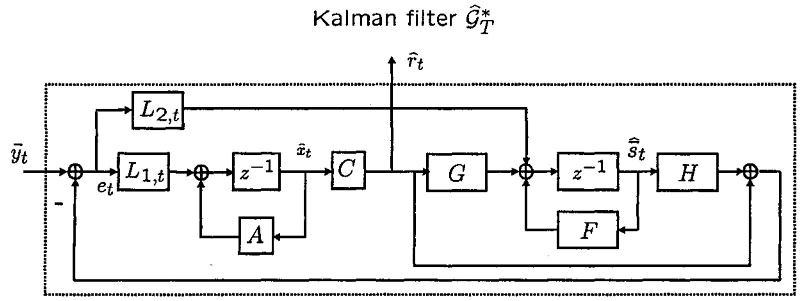
where $\widehat{\mathcal{G}}_T$ *is strictly lower triangular. See Fig. 4.5 for the associated estimation problem, Fig. 4.6 (a) for the Kalman filter* $\widehat{\mathcal{G}}_T^*(A,C)$, *and (b) for the optimal feedback generator* $\mathcal{G}_T^*(A,C)$.



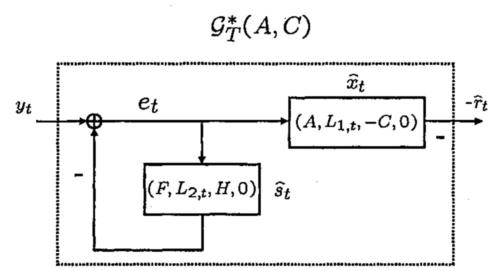
(a)

Figure 4.5 An estimation problem over channel $\mathcal{F}$.

Remark 6. *Proposition 3 reveals that, the minimization of channel input power in a feedback communication problem is equivalent to the minimization of MSE in an estimation problem. This equivalence yields a complete characterization (in terms of the Kalman filter) of optimal feedback generator* $\mathcal{G}_T^*(A,C)$ *for a given* (A,C).



(a)



(b)

Figure 4.6 (a) The Kalman filter $\widehat{\mathcal{G}}_T^*(A, C)$. (b) The Kalman filter based feedback generator $\mathcal{G}_T^*(A, C)$. Here $(A, L_{1,t}, -C, 0)$ with $\hat{x}_t$ denotes a state-space representation with $\hat{x}_t$ being its state at time t, and $\hat{x}_0$ being 0; see (4.57) and (4.60) for $L_{1,t}$ and $L_{2,t}$.

Remark 7. *Proposition 3 i) implies that we may reformulate the problem of $C_{T,n}$ (or $P_{T,n}$) as a two-step problem: In step 1, we fix (A, C), i.e. fixing the rate, and minimize the input power by searching over $\mathcal{G}$; and in step 2, we search over all possible (A, C) subject to the rate constraint. The* role of the feedback generator *$\mathcal{G}$ for any fixed (A, C) is to minimize the input power. Then ii) solves the optimal feedback generator $\mathcal{G}_T^*(A, C)$ by considering the* equivalent *optimal estimation problem in Fig. 4.5 whose solution is the Kalman filter. Notice that the Kalman filter can also give us the optimal estimate of the message W. Hence, the Kalman filter leads to both* power efficiency *and* the best estimate of the message. *We finally note that the necessity of the Kalman filter is not surprising given the previous indications in [106, 6, 102, 116, 83], etc.*

Proof: i) Notice that for any fixed (A, C), $R_T(A, C)$ is fixed. Then from the definition of

$P_{T,n}(\mathcal{R})$, we have

$$\begin{aligned} P_{T,n}(\mathcal{R}) &= \inf_{\substack{A,C,\mathcal{G}_T \\ s.t.\ R_T(A,C)\geq\mathcal{R}}} \frac{1}{T+1}\mathbf{E}u^{T\prime}u^T \\ &= \inf_{\substack{A,C \\ s.t.\ R_T(A,C)\geq\mathcal{R}}} \inf_{(A,C)\text{ fixed},\mathcal{G}_T} \frac{1}{T+1}\mathbf{E}u^{T\prime}u^T. \end{aligned} \tag{4.48}$$

Then i) follows from the definition of $\mathcal{G}_T^*(A,C)$.

ii) Note that for the general coding structure, it holds that

$$u^T = r^T + (-\hat{r}^T) = r^T + \mathcal{G}_T y^T. \tag{4.49}$$

Then, letting

$$\widehat{\mathcal{G}}_T := -\mathcal{G}_T(I - \mathcal{Z}_T^{-1}\mathcal{G}_T)^{-1} \tag{4.50}$$

and $\bar{y}^T := \mathcal{Z}_T^{-1}r^T + N^T$, we have $\mathcal{G}_T y^T = -\widehat{\mathcal{G}}_T\bar{y}^T$. Therefore,

$$\begin{aligned} \mathcal{G}_T^*(A,C) &= \arg\inf_{\mathcal{G}_T} \frac{1}{T+1}\mathbf{E}(r^T + \mathcal{G}_T y^T)(r^T + \mathcal{G}_T y^T)' \\ &= \arg\inf_{\widehat{\mathcal{G}}_T} \frac{1}{T+1}\mathbf{E}(r^T - \widehat{\mathcal{G}}_T\bar{y}^T)(r^T - \widehat{\mathcal{G}}_T\bar{y}^T)'. \end{aligned} \tag{4.51}$$

The last equality implies that the optimal solution $\widehat{\mathcal{G}}_T^*$ is the strictly causal MMSE estimator (with one-step prediction) of r^T given $\bar{y}_T$; notice that $\widehat{\mathcal{G}}_T$ is strictly lower triangular. It is well known that such an estimator can be implemented recursively in state-space as a Kalman filter (cf. [58, 57]). Finally, from the relation between $\mathcal{G}_T$ and $\widehat{\mathcal{G}}_T$, we obtain (4.46). The state-space representation of $\mathcal{G}_T^*(A,C)$ needs only a straightforward computation, as shown in Appendix A.2. ∎

Our study on the general coding structure also refines the CP structure. We can now identify more specific structure of the optimal $(K_v^{(T)}, \mathcal{B}_T)$. Indeed, we conclude that the CP structure needs to have a Kalman filter inside. We may further determine the optimal form of $\mathcal{B}_T$. From (4.39) and (4.46), we have that

$$\mathcal{B}_T^* = -\widehat{\mathcal{G}}_T^*(A,C)\mathcal{Z}_T^{-1}. \tag{4.52}$$

Therefore, to achieve C_T in the CP structure, it is sufficient to search $(K_v^{(T)}, \mathcal{B}_T)$ in the form of

$$\begin{aligned} K_v^{(T)} &:= (I - \widehat{\mathcal{G}}_T^*(A,C)\mathcal{Z}_T^{-1})\Gamma(A,C)\Gamma(A,C)'(I - \widehat{\mathcal{G}}_T^*(A,C)\mathcal{Z}_T^{-1})' \\ \mathcal{B}_T^* &:= -\widehat{\mathcal{G}}_T^*(A,C)\mathcal{Z}_T^{-1}. \end{aligned} \tag{4.53}$$

Additionally, as T tends to infinity, it can be easily shown that $\{v_t\}$ is a stable process in order to achieve C_∞.

We remark that it is possible to derive a dynamic programming based solution ([116]) to compute $C_{T,n}$, and if we further employ the Markov property in [136] and the above Kalman filter based characterization, we would reach a solution with complexity $O(T)$ for computing $C_{T,n}$ and C_T. This may be pursued elsewhere.

4.5 Finite horizon: Feedback rate, CRB, and Bode integral

We have shown that in the general coding structure, to ensure power efficiency for a fixed (A,C), we need to design a Kalman-filter based feedback generator. The Kalman filter immediately links the feedback communication problem to estimation and control problems. In this section, we present a *unified representation* for the general coding structure (with $\mathcal{G}$ being chosen as $\mathcal{G}^*(A,C)$), its estimation theory counterpart, and its control theory counterpart. Then we will establish connections among the information theory quantities, estimation theory quantities, and control theory quantities.

4.5.1 Unified representation of feedback coding system, Kalman filter, and MEC

In this subsection, we will present the dynamics for the estimation problem and the general coding structure, then show that they are governed by one set of equations, which may also be viewed as a control system.

The estimation system

The estimation system in Fig. 4.5 consists of three parts: the unknown source r^T to be estimated or tracked, the channel $\mathcal{F}$ (without output feedback), and the estimator which we choose as the Kalman filter $\widehat{\mathcal{G}}^*$; we assume that (A,C) is fixed and known to the estimator.

The system is described in state-space as

$$\text{estimation system:} \begin{cases} x_{t+1} &= Ax_t \\ r_t &= Cx_t \end{cases}\Big\}\text{unknown source} \\ \begin{aligned} \bar{s}_{t+1} &= F\bar{s}_t + Gr_t \\ \bar{y}_t &= H\bar{s}_t + r_t + N_t \end{aligned}\Big\}\text{channel } \mathcal{F} \\ \left.\begin{aligned} \hat{x}_{t+1} &= A\hat{x}_t + L_{1,t}e_t \\ \hat{r}_t &= C\hat{x}_t \\ \hat{\bar{s}}_{t+1} &= F\hat{\bar{s}}_t + G\hat{r}_t + L_{2,t}e_t \\ e_t &= \bar{y}_t - H\hat{\bar{s}}_t - \hat{r}_t \end{aligned}\right\}\text{Kalman filter } \widehat{\mathcal{G}}^*(A,C) \tag{4.54}$$

with $x_0 := W$, $\bar{s}_0 := 0$, $\hat{\bar{s}}_0 := 0$, and $\hat{x}_0 := 0$. Here $L_{1,t} \in \mathbb{R}^{n+1}$ and $L_{2,t} \in \mathbb{R}^m$ are the time-varying Kalman filter gains specified in (4.59).

The general coding structure with the optimal feedback generator

The optimal feedback generator for a given (A, C) is solved in (4.46), see Fig. 4.6 (b) for its structure. We can then obtain the minimal state-space representation of $\mathcal{G}_T^*(A, C)$, and describe the general coding structure with $\mathcal{G}_T^*(A, C)$ as

$$\text{general coding structure:} \begin{cases} \left.\begin{aligned} x_{t+1} &= Ax_t \\ r_t &= Cx_t \\ u_t &= r_t - \hat{r}_t \end{aligned}\right\}\text{encoder} \\ \left.\begin{aligned} s_{t+1} &= Fs_t + Gu_t \\ y_t &= Hs_t + u_t + N_t \end{aligned}\right\}\text{channel } \mathcal{F} \\ \left.\begin{aligned} \hat{s}_{t+1} &= F\hat{s}_t + L_{2,t}e_t \\ e_t &= y_t - H\hat{s}_t \\ \hat{x}_{t+1} &= A\hat{x}_t + L_{1,t}e_t \\ -\hat{r}_t &= -C\hat{x}_t \end{aligned}\right\}\text{optimal feedback generator} \end{cases} \tag{4.55}$$

with $x_0 := W$, $s_0 := 0$, $\hat{s}_0 := 0$, and $\hat{x}_0 := 0$. See Appendix A.2 for the derivation of the minimal state-space representation of $\mathcal{G}_T^*(A, C)$. It can be easily shown that r_t, $\hat{r}_t$, e_t, x_t, and $\hat{x}_t$ in (4.54) and (4.55) are equal, respectively, and it holds that

$$s_t - \hat{s}_t = \bar{s}_t - \hat{\bar{s}}_t. \tag{4.56}$$

The unified representation

Define

$$
\begin{aligned}
\tilde{x}_t &:= x_t - \hat{x}_t \\
\tilde{s}_t &:= s_t - \hat{s}_t = \bar{s}_t - \hat{\bar{s}}_t \\
\mathbb{X}_t &:= \begin{bmatrix} \tilde{x}_t \\ \tilde{s}_t \end{bmatrix} \\
\mathbb{X}_0 &:= \begin{bmatrix} W \\ 0 \end{bmatrix} \\
\mathbb{A} &:= \left[\begin{array}{c|c} A & 0 \\ \hline GC & F \end{array}\right] \\
\mathbb{C} &:= [C \quad H] \\
\mathbb{D} &:= [C \quad 0] \\
L_t &:= \begin{bmatrix} L_{1,t} \\ L_{2,t} \end{bmatrix}.
\end{aligned}
\tag{4.57}
$$

Note that $\mathbb{X}_t$ is the estimation error for $[x_t', s_t']'$. Substituting (4.57) to (4.54) and (4.55), we obtain that both systems become

$$
\text{control system:} \begin{cases} \mathbb{X}_{t+1} = (\mathbb{A} - L_t\mathbb{C})\mathbb{X}_t - L_t N_t = \mathbb{A}\mathbb{X}_t - L_t e_t \\ e_t = \mathbb{C}\mathbb{X}_t + N_t \\ u_t = \mathbb{D}\mathbb{X}_t; \end{cases}
\tag{4.58}
$$

see Fig. 4.7 for its block diagram. It is a control system where we want to minimize the power of u by appropriately choosing L_t. This is an MEC problem, which is useful for us to characterize the steady-state solution and it is equivalent to the Kalman filtering problem (see [67]).

The signal e_t in (4.58) is called the *Kalman filter innovation* or *innovation* [6], which plays a significant role in Kalman filtering. One fact is that $\{e_t\}$ is a white process, that is, its covariance matrix $K_e^{(T)}$ is a diagonal matrix. Another fact is that e^T and y^T determine each other causally, and we can easily verify that $h(e^T) = h(y^T)$ and $\det K_y^{(T)} = \det K_e^{(T)}$. We remark that (4.58) is the *innovations representation* of the Kalman filter (cf. [57]).

[6] The innovation defined here is different from the innovation defined in [17] or [136].

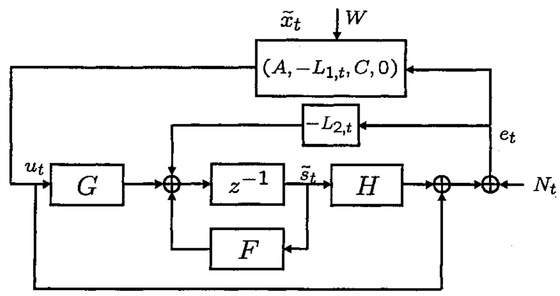


Figure 4.7 The block diagram for the MEC system. Here the block $(A, -L_{1,t}, C, 0)$ with $\tilde{x}_t$ denotes the state-space representation with $\tilde{x}_t$ and W being its state at time t and at time 0.

For each t, the optimal L_t is determined as

$$L_t := \begin{bmatrix} L_{1,t} \\ L_{2,t} \end{bmatrix} := \frac{\mathbb{A}\Sigma_t\mathbb{C}'}{K_{e,t}}, \tag{4.59}$$

where $\Sigma_t := \mathbf{E}\mathbb{X}_t\mathbb{X}_t'$, $K_{e,t} := \mathbf{E}(e_t)^2 = \mathbb{C}\Sigma_t\mathbb{C}' + 1$, and the error covariance matrix Σ_t satisfies the Riccati recursion

$$\Sigma_{t+1} = \mathbb{A}\Sigma_t\mathbb{A}' - \frac{\mathbb{A}\Sigma_t\mathbb{C}'\mathbb{C}\Sigma_t\mathbb{A}'}{\mathbb{C}\Sigma_t\mathbb{C}' + 1} \tag{4.60}$$

with initial condition

$$\Sigma_0 := \begin{bmatrix} I_{n+1} & 0 \\ 0 & 0 \end{bmatrix}, \tag{4.61}$$

This completes the description of the optimal feedback generator for a given (A, C).

The meaning of a unified expression for three different systems (4.54), (4.55), and (4.58) is that the first two are actually two different non-minimal realizations of the third. The input-output mappings from N^T to e^T in the three systems are T-equivalent (see Appendix A.2). Thus we say that the three problems, the optimal estimation problem, the optimal feedback generator problem, and the MEC problem, are *equivalent* in the sense that, if any one of the problems is solved, then the other two are solved. Since the estimation problem and the control problem are well studied, the equivalence facilitates our study of the communication problem. Particularly, the formulation (4.58) yields alternative expressions for the mutual information and average channel input power in the feedback communication problem, as we see in the next subsection.

We further illustrate the relation of the estimation system and the communication system in Fig. 4.8: (b) is obtained from (a) by subtracting $\hat{r}_t$ from the channel input and adding $\mathcal{Z}_T^{-1}\hat{r}_t$ back to the channel output, which does not affect the input, state, and output of $\widehat{\mathcal{G}}_T^*$. It is clearly seen from the block diagram manipulations that the minimization of channel input power in feedback communication problem becomes the minimization of MSE in the estimation problem. This is indeed the extension of how we obtained a Kalman filter based coding scheme from a Kalman filtering system in the AWGN case to the general Gaussian channel case.

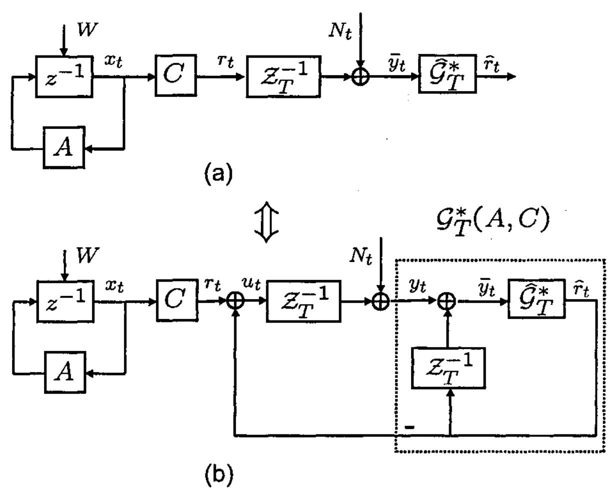


Figure 4.8 Relation between the estimation problem (a) and the communication problem (b).

4.5.2 Mutual information in terms of Fisher information and CRB

Proposition 4. *For any fixed* $0 \leq n \leq T$ *and* (A, C), *it holds that*

i)

$$\begin{aligned} I(W; y^T) &= \frac{1}{2}\log\det K_e^{(T)} = \frac{1}{2}\sum_{t=0}^{T}\log K_{e,t} \\ &= \frac{1}{2}\sum_{t=0}^{T}\log(\mathbb{C}\Sigma_t\mathbb{C}' + 1) \\ &= \frac{1}{2}\log\det \mathrm{MMSE}_{W,T}^{-1} \\ &= \frac{1}{2}\log\det \mathcal{I}_{W,T} \\ &= \frac{1}{2}\log\det \mathrm{CRB}_{W,T}^{-1}; \end{aligned} \tag{4.62}$$

ii)

$$
\begin{aligned}
P_{T,n}(A,C) &= \frac{1}{T+1}\sum_{t=0}^{T}\mathbb{D}\Sigma_t\mathbb{D}' \\
&= \frac{1}{T+1}\text{trace}(\text{CMMSE}_{r,T}) \\
&= \frac{1}{T+1}\sum_{t=0}^{T} CA^t\text{MMSE}_{W,t}A^{t\prime}C',
\end{aligned} \tag{4.63}
$$

where $\text{MMSE}_{W,T}$ *is the minimum MSE of* W*,* $\text{CMMSE}_{r,T}$ *is the causal minimum MSE of* r^T*,* $\mathcal{I}_{W,T}$ *is the Bayesian Fisher information matrix of* W *for the estimation system (4.54), and* $\text{CRB}_{W,T}$ *is the Bayesian CRB of* W *[123].*

Remark 8. *This proposition connects the mutual information to the innovations process and to the Fisher information, (minimum) MSE, and CRB of the associated estimation problem. As a consequence, the finite-horizon feedback capacity* $C_{T,n}$ *is then linked to the smallest possible Bayesian CRB, i.e. the smallest possible estimation error covariance. Thus the fundamental limitation in information theory is linked to the fundamental limitation in estimation theory. As in the AWGN case, we notice that the Fisher information, an estimation quantity, indeed has an information theoretic interpretation as its name suggests. In fact, the connection between mutual information and MMSE holds independent on how the feedback generator is designed, and hence it holds for both feedback communication and feedforward communication.*

Proof: i) First we simply notice that $h(y^T) = h(e^T)$, and $K_{e,t} = \mathbb{C}\Sigma_t\mathbb{C}' + 1$. Next, to find MMSE of W, note that in Fig. 4.5

$$
\bar{y}^T = \mathcal{Z}_T^{-1}\Gamma W + N^T \tag{4.64}
$$

and that $W \sim \mathcal{N}(0,I)$, $N^T \sim \mathcal{N}(0,I)$. Thus, by [58] we have

$$
\text{MMSE}_{W,t} = (I + \Gamma'\mathcal{Z}_T^{-1\prime}\mathcal{Z}_T^{-1}\Gamma)^{-1}, \tag{4.65}
$$

yielding

$$
\begin{aligned}
\det\text{MMSE}_{W,t} &= \det(I + \mathcal{Z}_T^{-1}\Gamma\Gamma'\mathcal{Z}_T^{-1\prime})^{-1} \\
&= \det(I + \mathcal{Z}_T^{-1}K_r^{(T,n)}\mathcal{Z}_T^{-1\prime})^{-1}.
\end{aligned} \tag{4.66}
$$

Besides, from Section 2.4 in [123] we can directly compute the FIM of W to be $(I+\Gamma'\mathcal{Z}_T^{-1\prime}\mathcal{Z}_T^{-1}\Gamma)$. Then i) follows from Proposition 2 and (4.58).

ii) Since $u_t = \mathbb{D}\mathbb{X}_t = C\tilde{x}_t = r_t - \hat{r}_t$ and $\mathbf{E}\tilde{x}_t\tilde{x}_t' = A^t \mathrm{MMSE}_{W,t} A^{t\prime}$, we have $\mathbf{E}(u_t)^2 = \mathbb{D}\Sigma_t\mathbb{D}' = C\mathbf{E}\tilde{x}_t\tilde{x}_t'C' = \mathbf{E}(r_t - \hat{r}_t)^2$, and then ii) follows. ∎

Corollary 1. *For any fixed* $0 \leq n \leq T$ *and* (A, C)*, it holds that*

$$\frac{\mathrm{d}}{\mathrm{d}\mathcal{I}_{W,T}} I(W; y^T) = \frac{1}{2}\mathrm{MMSE}_{W,t}. \tag{4.67}$$

Proof: By

$$I(W; y^T) = \frac{1}{2}\log\det \mathrm{MMSE}_{W,T}^{-1} = \frac{1}{2}\log\det \mathcal{I}_{W,T}, \tag{4.68}$$

we derive

$$\frac{\mathrm{d}}{\mathrm{d}\mathcal{I}_{W,T}} I(W; y^T) = \frac{1}{2}(\mathcal{I}_{W,T})^{-1} = \frac{1}{2}\mathrm{MMSE}_{W,t}, \tag{4.69}$$

based on the matrix differentiation formula proven in [2]. ∎

We can also compute that in the $m = 0$ or $m = 1$ case, the effective SNR equals $(\mathcal{I}_{W,T} - 1)$, which leads to that

$$\frac{\mathrm{d}}{\mathrm{dSNR}} I(W; y^T) = \frac{1}{2}\mathrm{MMSE}_{W,t}, \tag{4.70}$$

the formula linking mutual information and MMSE obtained by Guo, Shamai, and Verdu [51]. Though in our case, the input distribution is Gaussian, more restrictive than theirs assumption of arbitrary input distributions, our formula holds for some colored Gaussian noise channels. We wish to study this problem in full generality; this is subject to ongoing research.

4.5.3 Necessary condition for optimality

Before we turn to the infinite-horizon analysis, we show in this subsection that our general coding structure together with the optimal feedback generator satisfies a "necessary condition for optimality" discussed in [60]. The condition says that, the channel input u_t needs to be orthogonal to the past channel outputs y^{t-1}. This is intuitive since to ensure fastest transmission, the transmitter should not transmit any information that the receiver has obtained, thus the transmitter wants to remove any correlation of y^{t-1} in u_t (to this aim, the transmitter has to access the channel outputs through feedback).

Proposition 5. *In system (4.55), for any* $0 \leq \tau < t$*, it holds that* $\mathbf{E}u_t y_\tau = 0$.

Proof: See Appendix B.1. ∎

4.6 Infinite horizon: Asymptotic behavior of the system

By far we have completed our analysis in finite-horizon. We have shown that the optimal design of encoder and decoder must contain a Kalman filter, and connected the feedback communication problem to an estimation problem and a control problem. Below, we consider the steady-state communication problem, by studying the limiting behavior (T going to infinity) of the finite-horizon solution while fixing the encoder dimension to be $(n+1)$.

4.6.1 Convergence to steady-state

The time-varying Kalman filter in (4.58) converges to a steady-state, namely (4.58) is *stabilized* in closed-loop, u_t, e_t, and y_t will converge to steady-state distributions, and Σ_t, L_t, $\mathcal{G}_t^*(A, C)$, $\widehat{\mathcal{G}}_t^*$, and $K_{e,t}$ will converge to their steady-state values. That is, asymptotically (4.58) becomes an LTI system

$$\text{steady-state:} \begin{cases} \mathbb{X}_{t+1} &= (\mathbb{A} - L\mathbb{C})\mathbb{X}_t - LN_t = \mathbb{A}\mathbb{X}_t - Le_t \\ e_t &= \mathbb{C}\mathbb{X}_t + N_t \\ u_t &= \mathbb{D}\mathbb{X}_t, \end{cases} \tag{4.71}$$

where

$$L := \frac{\mathbb{A}\Sigma\mathbb{C}'}{K_e}, \tag{4.72}$$

$K_e = \mathbb{C}\Sigma\mathbb{C}' + 1$, and Σ is the unique stabilizing solution to the DARE

$$\Sigma = \mathbb{A}\Sigma\mathbb{A}' - \frac{\mathbb{A}\Sigma\mathbb{C}'\mathbb{C}\Sigma\mathbb{A}'}{\mathbb{C}\Sigma\mathbb{C}' + 1}. \tag{4.73}$$

This LTI system is easy to analyze (e.g., it allows transfer function based study) and to implement. For instance, the MEC of an LTI system claims that the transfer function from N to e is an *all-pass* function in the form of

$$\mathcal{T}_{Ne}(z) = \prod_{i=0}^{k} \frac{z - a_i}{z - a_i^{-1}} \tag{4.74}$$

where $a_0, \cdots, a_k$ are the unstable eigenvalues of A or $\mathbb{A}$ (noting that F is stable). Note that this is consistent with the whiteness of innovations process $\{e_t\}$.

The existence of steady-state of the Kalman filter is proven in the following proposition. Notice that (4.58) is a *singular* Kalman filter since it has no process noise; the convergence of such a problem was established in [44].

Proposition 6. *Consider the Riccati recursion (4.60) and the system (4.58).*

i) Starting from the initial condition given in (4.61), the Riccati recursion (4.60) generates a sequence $\{\Sigma_t\}$ *that converges to* Σ_∞ *with rank* $(n+1)$*, the unique stabilizing solution to the DARE (4.73).*

ii) The time-varying system (4.58) converges to the unique steady-state as given in (4.71).

Proof: See Appendix B.2. ■

4.6.2 Steady-state quantities

Now fix (A, C) and let the horizon T in the general coding structure go to infinity. Let $DI(A) := \prod_{i=0}^{k} |a_i|$ be the *degree of instability* of A and $S(e^{j2\pi\theta})$ be the spectrum of the sensitivity function of system (4.71) (cf. [28]). Then the limiting result of Proposition 4 is summarized in the next proposition.

Proposition 7. *Consider the general coding structure in Fig. 4.3. For any* $n \geq 0$ *and* (A, C)*,*

i) The asymptotic information rate is given by

$$\begin{aligned}
R_{\infty,n}(A,C) &:= \lim_{T\to\infty} \frac{1}{T+1} I(W; y^T) \\
&= \boldsymbol{h}(e) - \frac{1}{2}\log 2\pi e \\
&= \log DI(A) \\
&= \int_{-\frac{1}{2}}^{\frac{1}{2}} \log S(e^{j2\pi\theta}) d\theta \\
&= \frac{1}{2}\log(\mathbb{C}\Sigma\mathbb{C}' + 1) \\
&= \lim_{T\to\infty} \frac{\log\det \mathcal{I}_{W,T}}{2(T+1)} \\
&= -\lim_{T\to\infty} \frac{\log\det \mathrm{MSE}_{W,T}}{2(T+1)} \\
&= -\lim_{T\to\infty} \frac{\log\det \mathrm{CRB}_{W,T}}{2(T+1)}.
\end{aligned} \tag{4.75}$$

ii) The average channel input power is given by

$$\begin{aligned} P_{\infty,n}(A,C) &:= \lim_{T\to\infty} \frac{1}{T+1}\mathbf{E}u_T u_T' \\ &= \mathbb{D}\Sigma\mathbb{D}'. \end{aligned} \tag{4.76}$$

Remark 9. *Proposition 7 links the asymptotic information rate across the channel to the entropy rate of the innovations process, to the degree of instability and Bode sensitivity integral ([28]), to the asymptotic increasing rate of the Fisher information, and to the asymptotic decay rate of MSE and of CRB. Recall that the Bode sensitivity integral is the fundamental limitation of the disturbance rejection (control) problem, and the asymptotic decay rate of CRB is the fundamental limitation of the recursive estimation problem. Hence, the fundamental limitations in feedback communication, control, and estimation coincide.*

Remark 10. *Proposition 7 implies that the presence of stable eigenvalues in A does not affect the rate (see also [28]). Stable eigenvalues do not affect $P_{\infty,n}(A,C)$, either, since the initial condition response associated with the stable eigenvalues can be tracked with zero power (i.e. zero MSE). So, we can achieve $C_{\infty,n}$ by a sequence of purely unstable (A,C), and hence the communication problem is related to the tracking of purely unstable source over a communication channel ([102, 28]).*

Proof: Proposition 7 leads to that, the limits of the results in Proposition 4 are well defined. It also holds that

$$\begin{aligned} R_{\infty,n}(A,C) &= \lim_{T\to\infty} \frac{1}{2(T+1)} \sum_{t=0}^{T} \log K_{e,t} \\ &= \lim_{T\to\infty} \frac{1}{2} \log K_{e,t} \\ &= \boldsymbol{h}(e) - \frac{1}{2}\log 2\pi e, \end{aligned} \tag{4.77}$$

where the second equality is due to the Cesaro mean (i.e., if a_k converges to a, then the average of the first k terms converges to a as k goes to infinity), and the last equality follows from the definition of entropy rate of a Gaussian process (cf. [16]).

Now by (4.74), $\{e_t\}$ has a flat power spectrum with magnitude $DI(A)^2$. Then $R_{\infty,n}(A,C) = \log DI(A)$. The Bode integral of sensitivity follows from [28]. The other equalities are the direct applications of the Cesaro mean to the results in Proposition 4. ∎

We finally remark that, if we let the Kalman also perform a smoothing operation to obtain the optimal estimate $\hat{x}_{0,T}$ of x_0, it then holds that $I(W; y^T) = I(x_0; \hat{x}_{0,T})$, which follows from data processing inequality and that $\hat{x}_{0,T}$ is sufficient statistics. This further yields that

$$R_{\infty,n}(A,C) = \lim_{T\to\infty} \frac{1}{T+1} I(W; \hat{W}_T), \tag{4.78}$$

that is, the information rate across the channel is also equal to the end-to-end mutual information rate.

4.7 Infinite horizon: Achievability of asymptotic feedback capacity

In this section, we will prove that $C_{\infty,m-1} = C_\infty$, leading to the optimality of our encoder/decoder design in Section 4.3 in the mutual information sense, and then show that our design achieves C_∞ in the operational sense.

4.7.1 Asymptotic feedback capacities

If the noise in the colored Gaussian channel forms a (an asymptotic) stationary process, then $C_T(\mathcal{P})$ has a finite limit (cf. [60]; the proof utilizes the superadditivity of C_T, similar to the case of forward communication capacities studied in [43]), which also has the operational and information meanings. Precisely, we have

Lemma 3. *For the colored Gaussian noise channel without ISI defined in Section 4.2.1 or for the white Gaussian noise channel with ISI defined Section 4.2.2, the infinite-horizon feedback capacity given by (4.7) is also given by*

$$C_\infty = \lim_{T\to\infty} C_T, \tag{4.79}$$

where C_T is the finite-horizon feedback capacity, and the limit exists and is finite.

Proof: See Appendix B.3. ∎

By Lemma 2, the above implies that

$$\lim_{T\to\infty} C_{T,T} = C_\infty. \tag{4.80}$$

4.7.2 The optimal Gauss-Markov signalling strategy and a simplification

[136] proved that for each input in the form of (4.6), there exists a Gauss-Markov (GM) input that yields the same directed information and same input power. The GM input takes the form

$$u_t = d_t'\tilde{s}_{s,t} + \mathcal{E}_t, \tag{4.81}$$

where $d_t \in \mathbb{R}^m$ is a time-varying gain; $\{\mathcal{E}_t\}$ is a zero-mean white Gaussian process and $\mathcal{E}_t$ is independent on N^{t-1}, u^{t-1}, and y^{t-1}; and $\tilde{s}_{s,t}$ is generated by a Kalman filter

$$\begin{cases} \tilde{s}_{s,t} & := & s_t - \hat{s}_{s,t} \\ \hat{s}_{s,t+1} & = & F\hat{s}_{s,t} + L_{s,t}e_t \\ e_t & = & y_t - H\hat{s}_{s,t}, \end{cases} \tag{4.82}$$

where $\hat{s}_{s,0} = 0$,

$$L_{s,t} := \frac{Q_t\Sigma_{s,t}(H + d_t')' + K_{\mathcal{E}}^{(t)}G}{1 + K_{\mathcal{E}}^{(t)} + (H + d_t')\Sigma_{s,t}(H + d_t')'}, \tag{4.83}$$

$Q_t := F + Gd_t'$, and $\Sigma_{s,t} := \mathbf{E}\tilde{s}_{s,t}\tilde{s}_{s,t}'$ is the estimation error covariance of s_t, satisfying the Riccati recursion

$$\Sigma_{s,t+1} = Q_t\Sigma_{s,t}Q_t' + K_{\mathcal{E}}^{(t)}GG' - \frac{(Q_t\Sigma_{s,t}(H + d_t')' + K_{\mathcal{E}}^{(t)}G)(Q_t\Sigma_{s,t}(H + d_t')' + K_{\mathcal{E}}^{(t)}G)'}{1 + K_{\mathcal{E}}^{(t)} + (H + d_t')\Sigma_{s,t}(H + d_t')'}. \tag{4.84}$$

If one lets $d_t = d$ and $K_{\mathcal{E}}^{(t)} = K_{\mathcal{E}}$ for all t, that is, the input $\{u_t\}$ is a stationary process, then the search over all possible d and $K_{\mathcal{E}}$ solves C_∞, that is,

$$C_\infty(\mathcal{P}) = \max_{d\in\mathbb{R}^m, K_{\mathcal{E}}\in\mathbb{R}} \frac{1}{2}\log(1 + K_{\mathcal{E}} + (H + d')\Sigma_s(H + d')') \tag{4.85}$$

subject to DARE constraint and power constraint

$$\begin{array}{rcl} \Sigma_s & = & Q\Sigma_sQ' + K_{\mathcal{E}}GG' - \frac{(Q\Sigma_s(H+d')'+K_{\mathcal{E}}G)(Q\Sigma_s(H+d')'+K_{\mathcal{E}}G)'}{1+K_{\mathcal{E}}+(H+d')\Sigma_s(H+d')'} \\ \mathcal{P} & = & d'\Sigma_s d + K_{\mathcal{E}}. \end{array} \tag{4.86}$$

We remark that [136] was focused more on the structure of the optimal input distribution and capacity computation, instead of designing a coding scheme; how to encode/decode a message (rather than using a random coding argument) is not clear from [136].

Now we prove that $K_{\mathcal{E}} = 0$, namely $\{\mathcal{E}_t\}$ vanishes in steady-state. [6] This leads to a further simplification of the results in [136].

Proposition 8. *For the GM input (4.81) to achieve C_∞, it must hold that $K_{\mathcal{E}} = 0$.*

Proof: See Appendix B.4. ∎

The vanishing of $\{\mathcal{E}_t\}$ in steady-state helps us to show that, our general coding structure shown in Fig. 4.3 can achieve C_∞, and the encoder dimension needs not be higher than the channel dimension, namely to achieve C_∞ we need A to have at most m unstable eigenvalues, as we will see in the next subsection.

4.7.3 Generality of the general coding structure; finite dimensionality of the optimal solution

In this subsection, we show that the general coding structure is sufficient to achieve mutual information C_∞. In other words, if we search over all admissible parameters $A, C, \mathcal{G}_T$ in the general coding structure, allowing T to increase to infinity and n to increase to $(m-1)$, then we can obtain C_∞. Thus, there is no loss of generality and optimality to consider only the general coding structure with encoder dimension no greater than m.

Definition 3. *Consider the general coding structure in Fig. 4.3. Let*

$$C_{\infty,n} := C_{\infty,n}(\mathcal{P}) := \sup_{A\in\mathbb{R}^{(n+1)\times(n+1)}, C, \mathcal{G}_\infty} \lim_{T\to\infty} \frac{1}{T+1} I(W; y^T) \tag{4.87}$$

subject to

$$P_{\infty,n} := \lim_{T\to\infty} \frac{1}{T+1} \mathbf{E} u^{T\prime} u^T \leq \mathcal{P}. \tag{4.88}$$

In other words, $C_{\infty,n}$ is the infinite-horizon information capacity for a fixed transmitter dimension. Note that $C_{\infty,n}$ exists and is finite. To see this, note Proposition 7, $C_{\infty,n} \leq C_\infty < \infty$, and the fact that

$$C_{\infty,n}(\mathcal{P}) = \sup_{A\in\mathbb{R}^{(n+1)\times(n+1)}, C, \mathcal{G}^*(A,C), (4.88)} R_{\infty,n}(A, C). \tag{4.89}$$

The function $C_{\infty,n}(\mathcal{P})$ also induce $P_{\infty,n}(\mathcal{R})$, the "capacity" in terms of minimum input power subject to an information rate constraint.

[6] $K_{\mathcal{E}} = 0$ was also conjectured and numerically verified by Shaohua Yang (personal communication).

Proposition 9. *Consider the general coding structure in Fig. 4.3.*

i) $C_{\infty,n}$ *is increasing in* n*;*

ii) For channel $\mathcal{F}$ *with order* $m \geq 1$, $C_{\infty,n} = C_\infty$ *for* $n \geq m-1$.

Proof: See Appendix B.5. ■

This proposition suggests that, to achieve C_∞, we may first fix the transmitter dimension as $(n+1)$ and let the dynamical system run to time infinity, obtaining $C_{\infty,n}$, and then increase n to $(m-1)$. The finite dimensionality of the optimal solution is important since it guarantees that we can achieve C_∞ without solving an infinite-dimensional optimization problem.

4.7.4 Achieving asymptotic feedback capacity

In this subsection, we prove that our coding scheme achieves C_∞ in the information sense as well as in the operational sense.

Proposition 10. *For the coding scheme described in Theorem 2,* $R_{\infty,n^*}(A^*, C^*) = C_\infty(\mathcal{P})$ *and* $P_{\infty,n^*}\ (A^*, C^*) = \mathcal{P}$.

Proof: See Appendix B.6. ■

Proposition 11. *The system constructed in Theorem 2 transmits the analog source* $W \sim \mathcal{N}(0, I)$ *at a rate* $C_\infty(\mathcal{P})$*, with MSE distortion* $D(C_\infty(\mathcal{P}))$*, where* $D(\cdot)$ *is the distortion-rate function.*

Proof: See Appendix B.7. ■

Proposition 12. *The system constructed in Theorem 2 transmits a digital message* W *from the transmitter to the receiver at a rate arbitrarily close to* $C_\infty(\mathcal{P})$ *with* PE_T *decays doubly exponentially.*

Proof: See Appendix B.8. ■

Note that, the coding length needed for a pre-specified performance level can be predetermined since $\Sigma^*_{x,T}$ can be solved off-line. Besides, because the probability of error decays doubly exponentially, it leads to much shorter coding length than forward transmission.

4.8 Numerical example

Here we repeat the numerical example studied in [136]. Consider a third-order channel (i.e. $m = 3$) with

$$\mathcal{Z}^{-1} := \frac{1 + 0.5z^{-1} - 0.4z^{-2}}{1 + 0.6z^{-2} - 0.4z^{-3}}. \tag{4.90}$$

In state-space, $\mathcal{Z}^{-1}$ is described as $(F, G, H, 1)$ where

$$\begin{aligned} F &= \begin{bmatrix} 0 & -0.6 & 0.4 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad G = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \\ H &= \begin{bmatrix} 0.5 & -1 & 0.4 \end{bmatrix}. \end{aligned} \tag{4.91}$$

Assume the desired communication rate $\mathcal{R}$ is 1 bit per channel use. We first solve (4.11) with $n = m - 1 = 2$, and find out $n^* = 1$. That is, C_∞ is attained when A has two unstable eigenvalues. Then we solve (4.11) again with $n^* = 1$, and obtain

$$\begin{aligned} A^* &= \begin{bmatrix} 0 & 1 \\ -2 & -0.887 \end{bmatrix} \\ L^* &= \begin{bmatrix} -0.506 & -0.225 & 0.573 & 0.092 & -0.327 \end{bmatrix}'. \end{aligned} \tag{4.92}$$

This yields the optimal power $P_\infty = 0.743$ (or -1.290 dB). Similar computation generates Figure 4.9, the curve of C_∞ against SNR or equivalently $\mathcal{P}$. This curve is identical to that in [136].

We then use the obtained A^*, C^*, and L^* to construct the optimal communication scheme. However, we observe that the optimal communication scheme shown in Fig. 4.2 generates unbounded signals $\{r_t\}$ and $\{\hat{r}_t\}$ due to the instability of A. This is not desirable for the simulation purpose, though the scheme in the form of Fig. 4.2 is convenient for the analysis purpose. Here, we propose a modification of the scheme, see Fig. 4.10. It is easily verify that the system in Fig. 4.10 is T-equivalent to that in Fig. 4.2. As we indicate in Fig. 4.10, the loop including the encoder, the channel, and the feedback link is indeed the control setup, which is stabilized and hence any signal inside is bounded. Note that the encoder now involves $\tilde{x}_{-1}$; we set $\tilde{x}_{-1} := A^{-1}W$, leading to $\tilde{x}_0 := W$, the desired value for $\tilde{x}_0$.

We report the simulation results using the modified communication scheme with the optimal

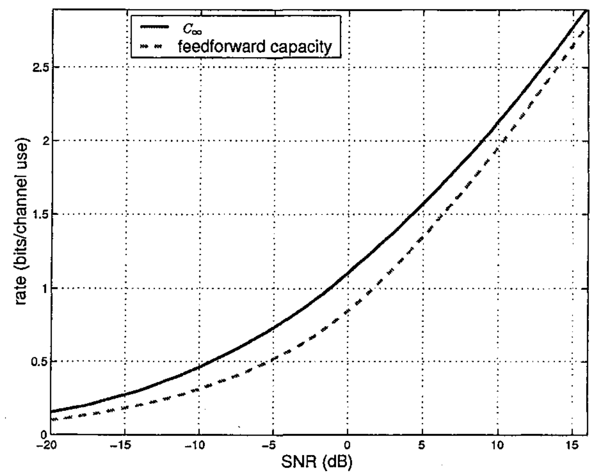


Figure 4.9 The feedback capacity C_∞ and feedforward capacity for channel $\mathcal{F}$ with $\mathcal{Z}^{-1} = (1 + 0.5z^{-1} - 0.4z^{-2})/(1 + 0.6z^{-2} - 0.4z^{-3})$.

parameters given in (4.92). Fig. 4.11 (a) shows the convergence of $\hat{x}_{0,t}$ to x_0, in which $x_0 := [-0.2, -0.7]'$. Fig. 4.11 (a) also shows the time average of the channel input power, which converges to the optimal power $P_\infty = 0.743$. To compute the probability of error, we let $\epsilon = 0.2$, i.e., the signalling rate is equal to $0.8C_\infty$. We demonstrate that this signalling rate is achieved by showing that the simulated probability of error decays to zero, see Fig. 4.11 (b). Fig. 4.11 (b) also plots the theoretic probability of error computed from (B.50), which is almost identical to the simulated curve. In addition, we see that the probability of error decays rather fast within 28 channel uses. The fast decay implies that the proposed scheme allows shorter coding length and shorter coding delay; here coding delay measures the time steps that one has to wait for the message to be decoded at the receiver with small enough error probability.

4.9 Summary

This chapter is a non-trivial generalization of the perspective of unified information, estimation, and control obtained in the AWGN case in Chapter 3. We presented a coding scheme to achieve the asymptotic capacity C_∞ for a Gaussian channel with feedback. The scheme is essentially the Kalman filter algorithm, and its construction involves only a finite dimensional optimization problem. We established connections of feedback communication to estimation

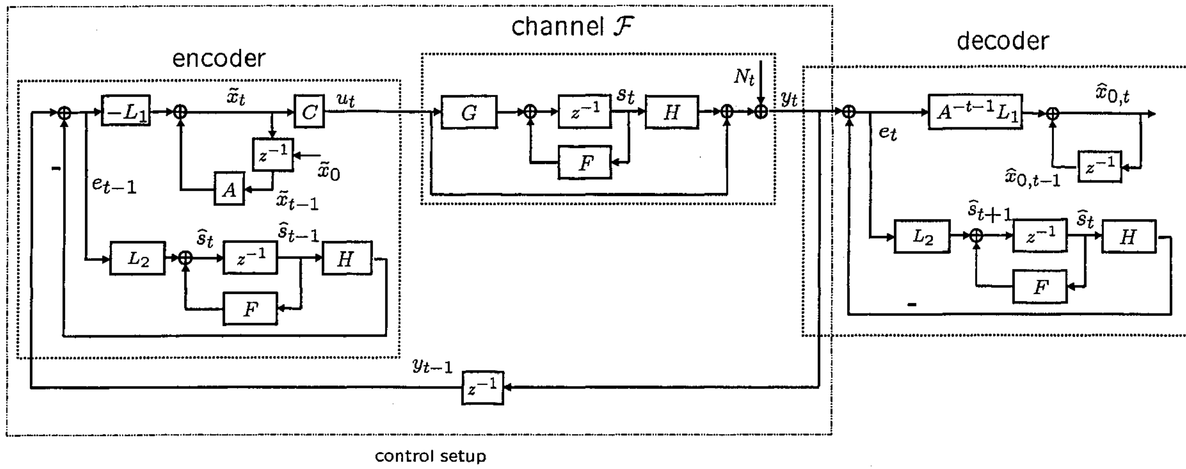
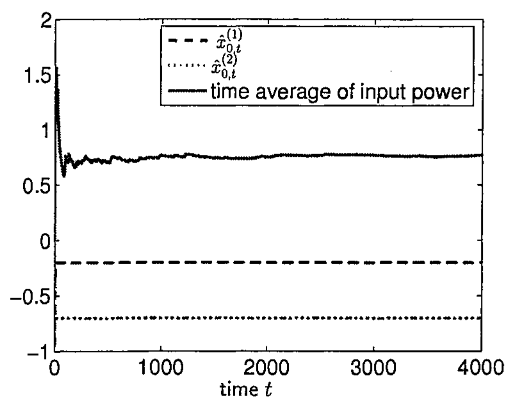
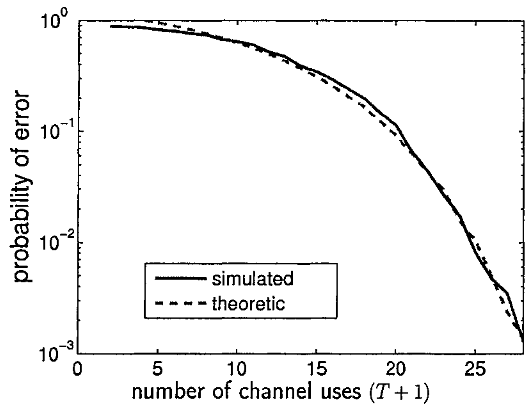


Figure 4.10 The modified feedback communication scheme.

and control. We have seen that concepts in estimation theory and control theory, such as MMSE, CRB, MEC, etc., are useful in studying a feedback communication system. We also verified the results by simulations.



(a)



(b)

Figure 4.11 (a) Convergence of $\hat{x}_{0,t}$ to x_0, and convergence of the average channel input power. (b) Simulated probability of error and theoretic probability of error.

CHAPTER 5. TIME-SELECTIVE FADING GAUSSIAN CHANNELS WITH CHANNEL STATE INFORMATION AND FEEDBACK

5.1 Introduction

In this chapter we study time-selective fading Gaussian channels with feedback. The time-selective fading is usually modeled as Markov chains, and there have been many achievements in the study of Markov channels; see e.g. [46, 47, 7, 125, 89, 137] and references therein. [46] obtained the capacity of a finite-state Markov channel (FSMC) with known channel transition structure but without CSI. [47] solved the capacity of a Markov channel with instantaneous CSI at both the transmitter (or encoder) and receiver (or decoder), or at the receiver only. [7] investigated the capacity of a Markov channel with possibly imprecise or delayed CSI. [125] provided the capacity of an FSMC with CSI delayed at the transmitter side and instantaneous at the receiver side (DTRCSI). It also showed that, for a channel with DTRCSI, the access to the channel output by the transmitter via delayed feedback does not increase the capacity. [89] obtained the delay-constrained capacity for a flat block-fading channel with causal feedback. The above papers are mainly focused on Markov channels without ISI, namely the channel state at each time is independent of the channel inputs up to that time. For ISI channels with output feedback, [137] characterized the capacity and capacity-achieving distribution.

In this chapter, we present a capacity-achieving communication scheme for an FSMC with AWGN, under the assumption of delayed noiseless output feedback and DTRCSI. We consider the case where there is no ISI. Although the access to channel outputs by the transmitter cannot improve capacity in this scenario (as proven in [125]), we show that the proposed scheme utilizing output feedback leads to simpler encoders and decoders, shortens coding delays, and leads to doubly exponential decay of the probability of decoding error, while achieving the feedback capacity.

The proposed communication scheme over an FSMC with output feedback is a *nontriv-*

ial extension of, first, the optimal feedback scheme over an AWGN channel, and second, the optimal forward scheme over an FSMC ([47, 125]). In essence, our optimal feedback communication design for an FSMC consists of a set of *decoupled* subsystems in parallel, and the subsystems are multiplexed to share the forward channel according to the channel state evolution. This introduces to the feedback communication system the ability to *adapt* its operations to the channel variation, which is a crucial step towards obtaining the optimal (in the sense of achieving the Shannon capacity) feedback communication design for time-varying channels. Though the multiplexing idea was widely studied in forward communication ([47, 125]), it has not received sufficient attention in the feedback communication literature. In this chapter, we show that multiplexing according to channel variation will eventually lead to the optimality in feedback communication. However, the presence of (delayed) output feedback considerably complicates the multiplexing design: In forward communication, the subsystems are naturally decoupled from each other (which greatly simplifies the analysis and design), whereas in feedback communication, the decoupling of subsystems needs to be guaranteed through careful design. It turns out that in the feedback case, each subsystem needs to be appropriately designed to depend on "augmented" channel states, and consequently, our transmitter needs to switch among m^2 possible sets of parameters in the case of m channel states with one-step delayed output feedback. Note that this is rather different from the multiplexing design in forward communication, whose transmitter switches among m possible sets of parameters ([47, 125]). In summary, the proposed feedback communication scheme over an FSMC nontrivially combines the Schalkwijk-Kailath feedback design and the forward multiplexing design, and is among the first to achieve the feedback capacity by extending the idea of Schalkwijk and Kailath.

On the other hand, the proposed communication scheme is associated with a Markov jump linear control system over an FSMC: The achievable rate of the former equals the expected open-loop growth rate of the latter if the latter is stabilized in closed-loop, and the optimal channel input power of the former corresponds to the MEC design of the latter. Namely, the optimality in both systems coincide. As a byproduct, we obtain the solution to the MEC over an FSMC. Due to its simplicity, the utilization of control system facilitates our development of the main results. Alternatively, we can also associate the communication problem to a time-varying Kalman filtering problem, in which the time-varying parameters of the to-be-estimated process are known to the estimator. Thus, we see that this study reveals the intrinsic connections among the communication problem over a Markov channel with output feedback, the estimation problem over the same channel, and the control problem of feedback stabilization over the same channel, and it fits into the framework of investigating the interactions among information, estimation, and control.

Our coding design simplifies when the channel states form an i.i.d. process taking a finite number of values, which requires only a simple first-order transmitter and receiver. We also show that the design extends to the case when the channel states are i.i.d. taking an *infinite* number of values, which includes as special cases the widely used Rayleigh, Rician, Nakagami, and Weibull fading channel models.

In addition to leading to the information theoretic results as described above, our study also reveals the confluence of information and control. Particularly, the proposed optimal communication system is associated with a Markov jump linear control system over an FSMC (cf. e.g. [15] for optimal control of Markov jump linear systems): The *achievable communication rate* in the former system equals the expected open-loop growth rate in the latter system if the latter is stabilized in closed-loop, and the *optimal channel input power* of the former corresponds to the minimum-energy control (MEC) design of the latter (cf. [67, 28, 73] for MEC over time-invariant channels). Thus, the optimality in both systems coincides. Due to the simplicity of the associated control system, the utilization of the control based approach facilitates our development of the main results; indeed, the design of the multiplexing communication scheme based on augmented channel states can be derived by studying the MEC of the Markov jump control system. Alternatively, we can also associate the communication problem to a time-varying Kalman filtering problem, in which the time-varying parameters of the to-be-estimated process are known to the estimator. This approach may be extended to more general scenarios. In summary, this study reveals the intrinsic connections among the communication problem over a Markov channel with output feedback, the estimation problem over the same channel, and the control problem of feedback stabilization over the same channel, and it fits into the framework of investigating the interactions among information, estimation, and control.

This chapter is organized as follows. Section 5.2 introduces the channel model and the problem we want to solve. Section 5.3 provides the description of the proposed feedback scheme. This scheme is shown to achieve the capacity in Section 5.4. In Section 5.5 we present a numerical example. Finally we summarize this chapter.

Notations of this chapter: We introduce some special notations in this chapter, to better represent vectors and their components, which are extensively used in this chapter. We use boldface letter $\boldsymbol{x}$ for a vector, and $x^{(i)}$ for the ith element of vector $\boldsymbol{x}$. Note that $(A_n)^m$ is the mth power of A_n, $\boldsymbol{A}_n$ is a vector at time n, and $A_n^{(m)}$ is the mth element of vector $\boldsymbol{A}_n$. We use $a[1], a[2], \cdots$ to represent a collection of fixed numbers.

5.2 Channel Model

Fig. 5.1 (a) shows an FSMC with AWGN, for short, AFSMC. At time t, this discrete-time channel $\mathcal{H}$ is described as

$$\mathcal{H}: \;\; y_t = S_t u_t + N_t, \;\; \text{for } t = 0, 1, 2, \cdots, \tag{5.1}$$

where u_t is the channel input, S_t is the channel state, N_t is the channel noise, and y_t is the channel output. These variables are real-valued. The noise $\{N_t\}$ is independent Gaussian with zero mean and unit variance. The channel state S_t is independent of the channel input u_0^{t-1} and output y_0^{t-1} when conditioned on the previous states. Furthermore, $\{S_t\}$ is a stationary, irreducible, aperiodic, finite-state homogeneous Markov chain and hence is ergodic. The one-step transition probability is

$$p_{ij} := \Pr(S_t = s[j] | S_{t-1} = s[i]), \;\; \text{for } t = 0, 1, 2, \cdots,$$

where $i, j = 1, 2, \cdots, m$; m is the number of states of the Markov chain; and $s[i]$ is a fixed number for each i. Assume that $s[i] \neq s[j]$ if $i \neq j$. Note that $s[i]$ denotes one of the m states of the Markov chain, and also represents the associated channel gain if the channel is in that state. Denote the stationary distribution vector of the Markov chain (which by ergodicity exists and is unique) as $\boldsymbol{\pi} := [\pi[1], \pi[2], \cdots, \pi[m]]$.

Definition 4. *i) Define the set of all possible channel state sequences $\{S_t\}$ as Ω, i.e.,*

$$\Omega := \{\{S_t\}\}. \tag{5.2}$$

ii) Define the set of all possible channel state sequences S^t as Ω_t, i.e.,

$$\Omega_t := \{S^t\}. \tag{5.3}$$

iii) Define the typical set *of sequences $\{S_t\}$ as*

$$\Omega_{TYP} := \left\{ \{S_t\} \middle| \frac{n(j,t)}{t+1} \to \pi[j] \text{ and } \frac{n(j,l,t)}{n(j,t)} \to p_{jl} \text{ as } t \to \infty \right\} \tag{5.4}$$

where for a given channel state sequence $\{S_t\}$, $n(j,l,t)$ is the number of transitions from channel state $s[j]$ to channel state $s[l]$ up to time t; and $n(j,t) := \sum_{l=1}^{m} n(j,l,t)$. [1] *Each sequence in Ω_{TYP} is called a* typical sequence.

By ergodicity of $\{S_t\}$, it holds that $\Pr(\Omega_{TYP}) = 1$. Hereafter, by "with probability one" we mean "for every channel state sequence $\{S_t\} \in \Omega_{TYP}$" or "for every typical sequence".

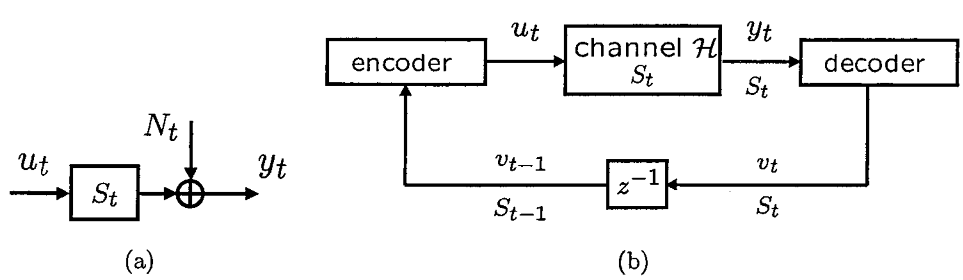


Figure 5.1 (a) An AFSMC $\mathcal{H}$. (b) An AFSMC with DTRCSI and output feedback.

We mainly focus on the case of one-step-delayed transmitter-side and instantaneous receiver-side CSI, denoted DTRCSI. See Appendix C.7 for the multi-step delay case. We also allow the transmitter to have access to the one-step-delayed channel outputs, i.e., the receiver at time t having observed y^t will compute v_t (depending only on y^t) and feed back v_t along with S_t to the transmitter with one step delay. In other words, the channel input u_t can depend on S^{t-1} and v^{t-1}. See Fig. 5.1 (b). Note that the above defined AFSMC has a discrete channel state but continuous channel input, noise, and output (in contrast to the discreteness of FSMC inputs and outputs in [46, 47, 7, 125, 89, 137], with the notable exception of some parts in

[1] *To simplify notations, we do not specify the dependency of $n(j,l,t)$ on the given sequence $\{S_t\}$ here.*

[125]). This channel may be used in modeling the following cases and their generalizations. For one, the channel is subject to both erasure (i.e. discrete channel states) and AWGN (i.e. continuous noise), and the erasures may exhibit either independence or certain time correlation (e.g. forming a 2-state Markov chain). The erasure in this case may be due to some environmental changes (such as buffer overflows) and is likely to exhibit certain time correlation (e.g. forming a 2-state Markov chain). For another, a channel is subject to bursty noises with different noise variances, and the occurrence of bursty noises forms a finite-state Markov chain. The well-known Gilbert-Elliot channel with AWGN falls into this category. We remark that the assumption on the instantaneous, prefect CSI at the receiver side, though often assumed in the literature (see [47, 120, 125], etc.), is not quite realistic (especially in the fast fading case); we employ this assumption in order to simplify the analysis and to gain some conceptual understandings of feedback Markov channel problems. A study taking into consideration of the imperfect CSI will be subject to future work, and our study based on perfect CSI may be found useful in that study.

For an AFSMC with DTRCSI and output feedback, its capacity subject to the *average* channel input power constraint

$$\mathbf{E}_u u^2 \leq \mathcal{P} \tag{5.5}$$

is given by

$$C = \max_{\Gamma(\cdot):\mathbf{E}_{S_t}\Gamma(S_t)\leq\mathcal{P}} \frac{1}{2}\mathbf{E}_{S_t,S_{t+1}} \log(1+(S_{t+1})^2\Gamma(S_t)), \tag{5.6}$$

where S_t is drawn according to the stationary distribution $\boldsymbol{\pi}$, and $\Gamma(\cdot)$ is the power allocation function that maps the channel state S_t to the channel input power $\Gamma(S_t)$. The above capacity formula is obtained by invoking Lemma 2 of [125] (with $d = 1$ and $\sigma_s^2 = 1$ therein). The *optimal* power allocation, denoted $\gamma(\cdot)$, is given by the solution of a set of m equations [2] obtained by applying the Kuhn-Tucker condition (see Appendix B in [125]) and is assumed given throughout this chapter. The objective of the chapter is to design a transmitter and receiver to achieve the capacity given in (5.6) for an AFSMC with DTRCSI and output feedback.

[2]The equations are not linear but the nonlinearity involves fractions only. Hence they can be easily solved numerically.

5.3 The proposed feedback communication scheme

In this section, we propose a communication scheme for an AFSMC with DTRCSI and output feedback. After a brief description of the main idea of our design, we introduce the communication scheme and its dynamics, followed by a choice of parameters for the proposed setup. We then explain the encoding and decoding methods. This scheme will be shown to be capacity-achieving in Section 5.4.

5.3.1 General description of the proposed scheme

We present an informal overview of the proposed scheme before we going into the technical details. In short, the proposed design can be viewed as one that multiplexes a set of subsystems with feedback, each of which uses an augmented channel state. In the degenerated case that the channel is not time-varying, each subsystem simplifies to the one described in previous chapter for AWGN channels.

Suppose that the Markov channel $\mathcal{H}$ has m states. Then our scheme consists of m subsystems (each of which is associated with one channel state) sharing the common channel $\mathcal{H}$. Represent each to-be-transmitted message as an m-dimensional codeword which contains m sub-codewords, and the m sub-codewords uniquely determines the message. Then let each subsystem be associated with one and only one sub-codeword (see Sec. 5.3.8 and Fig. 5.6). At each time epoch, one and only one subsystem is selected to use the forward channel to send its sub-codeword, whereas all other subsystems do not use the channel.

Note that we would like the sub-codewords to be communicated *independently* from each other, which can be ensured if the m subsystems are decoupled. (Recall that the m subsystems in [47, 125] are naturally decoupled.) We can show that the decoupling is possible if our encoder makes use of the augmented channel state, which contains two previous channel states, but *impossible* (except for the degenerated case that $\{S_t\}$ is i.i.d.) if the transmitter only uses the immediate previous channel state. To see this, observe that the transmitter at time t does not have access to S_t, due to the delay in the feedback link. Therefore, the transmitter cannot choose the subsystem associated with S_t to use the forward channel at time t; instead, if the transmitter chooses the subsystem associated with S_{t-1} to use the forward channel at time t, then the channel output y_t depends on both S_{t-1} and S_t. It turns out that the receiver needs to use the augmented channel state for decoupling, and hence that the transmitter needs to use augmented channel state. As we remarked in Section 4.1, this would require a design that is not a trivial extension of the optimal communication design for feedback AWGN channels or of the optimal multiplexing design for forward FSMCs. Note that state augmentation is widely used for control systems with delay. See Sec. 5.3.4 for the details regarding the decoupling of subsystems.

After we ensure the decoupling, the m sub-codewords are thus multiplexed together at the transmitter, and demultiplexed at the receiver, in a mutually independent fashion, namely each sub-codeword is transmitted *without* the interference from other sub-codewords. The original message would be successfully recovered if all m sub-codewords are successfully recovered. Therefore, the average rate of our scheme is the weighted sum of the rates for the subsystems, with the weights being the probabilities that the subsystems are selected to use the forward channel. Beside, the decoupling and the MEC design (see Sec. 5.3.5) ensures that the channel input power at time t depends on $(t-1)$ fade only, and the power of the ith subsystem is designed to converge to $\gamma(s[i])$, which would result in that the average power converges to the weighted sum of of $\gamma(s[i])$, the optimal power (see Appendix C.4). To summarize, each subsystem achieves its Shannon limit by solving the corresponding MEC problem, and the scheme achieves the Markov channel capacity.

The rest of the section is organized as follows. In Section 5.3.2, we describe the communication system, i.e. the dynamics of the encoder and decoder. In 5.3.3, we specify a choice of parameters for the proposed design. We then show that this choice leads to decoupling of sub-systems in Sec. 5.3.4. In Sec. 5.3.5, we show briefly how we derive the proposed scheme from the MEC problem. Finally, we introduce the encoding and decoding methods, probability of decoding errors, and achievable rates.

5.3.2 Coding scheme

Fig. 5.2 shows the proposed communication scheme. In this figure we identify the encoder, the channel $\mathcal{H}$, and the decoder. We call $\boldsymbol{x}_t \in \mathbb{R}^m$ the *encoder state*, with $\boldsymbol{x}_0$ as the initial condition. We call $\hat{\boldsymbol{x}}_{0,t} \in \mathbb{R}^m$ the *decoder estimate*, which is an estimate of $\boldsymbol{x}_0$ at time t. Parameters $A \in \mathbb{R}^{m\times m}$, $\boldsymbol{L} \in \mathbb{R}^m$, and $\boldsymbol{c} \in \mathbb{R}^m$ depend causally on the channel states, and will be chosen to reflect the adaptation of the communication strategy to the channel variation. Since this chapter, we start from the "modified" coding scheme that is numerically stable, rather than from the one obtained directly from the Kalman filtering system. This leads to slight change of some notations.

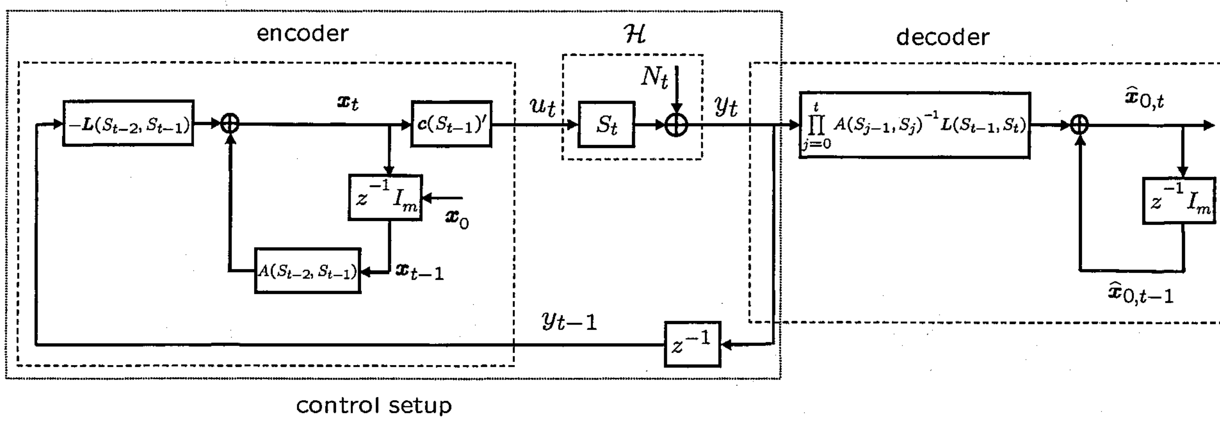


Figure 5.2 The communication scheme and the control setup.

At time t, $t \geq 0$, the system generates signals according to the following dynamics in the listed order:

$$\begin{aligned} \boldsymbol{x}_t &= A(S_{t-2}, S_{t-1})\boldsymbol{x}_{t-1} - \boldsymbol{L}(S_{t-2}, S_{t-1})y_{t-1} \\ u_t &= \boldsymbol{c}'(S_{t-1})\boldsymbol{x}_t \\ y_t &= S_t u_t + N_t \\ \hat{\boldsymbol{x}}_{0,t} &= \hat{\boldsymbol{x}}_{0,t-1} + \prod_{j=0}^{t} A(S_{j-1}, S_j)^{-1}\boldsymbol{L}(S_{t-1}, S_t)y_t, \end{aligned} \tag{5.7}$$

where $\boldsymbol{x}_{-1} := \boldsymbol{x}_0$, $y_{-1} := 0$, and $\hat{\boldsymbol{x}}_{0,-1} := 0$. The above recursions will generate a sequence of receiver estimates $\{\hat{\boldsymbol{x}}_{0,t}\}$ that converges to $\boldsymbol{x}_0$, as we will prove in the next section. We can rewrite the dynamics of the transmitter state $\boldsymbol{x}_t$ as

$$\boldsymbol{x}_t = A_{cl}(S_{t-2}, S_{t-1})\boldsymbol{x}_{t-1} - \boldsymbol{L}(S_{t-2}, S_{t-1})N_{t-1}, \tag{5.8}$$

where

$$A_{cl}(S_{t-2}, S_{t-1}) := A(S_{t-2}, S_{t-1}) - S_{t-1}\boldsymbol{L}(S_{t-2}, S_{t-1})\boldsymbol{c}'(S_{t-2}) \tag{5.9}$$

is the closed-loop matrix for generating $\boldsymbol{x}_t$. Note that (5.8) and (5.9) specifies a control system, referred to as the *control setup*, in which we want to minimize the power of u, namely the channel input power, by designing $\boldsymbol{L}$ for the given A and $\boldsymbol{c}$. This is an MEC problem over a Markov channel, whose solution leads to power efficiency of the proposed scheme and eventually the capacity-achieving property, as we will see in the rest of the chapter. We also remark that the somewhat non-intuitive operations (5.7) in the communication scheme are deduced as a rewrite of the MEC of the control setup (5.8); see Section 5.3.5 for details.

5.3.3 Choice of parameters

Given any $\mathcal{P} > 0$, we choose the parameters in the communication setup as follows. Let $\gamma(\cdot)$ be the optimal power allocation function computed from [125]. Supposing at time $(t-2)$ the channel state $S_{t-2} = s[j]$ for some j, we define

$$\begin{aligned} A(S_{t-2}, S_{t-1}) &:= \operatorname{diag}([1, \cdots, 1, a(S_{t-2}, S_{t-1}), 1, \cdots, 1]) &&\in \mathbb{R}^{m\times m} \\ \boldsymbol{L}(S_{t-2}, S_{t-1}) &:= [0, \cdots, 0, b(S_{t-2}, S_{t-1}), 0, \cdots, 0]' &&\in \mathbb{R}^{m} \\ \boldsymbol{c}(S_{t-2}) &:= [0, \cdots, 0, c(S_{t-2}), 0, \cdots, 0]' &&\in \mathbb{R}^{m}, \end{aligned} \tag{5.10}$$

where $a(S_{t-2}, S_{t-1})$ is the (j,j)th element of $A(S_{t-2}, S_{t-1})$, given by

$$a(S_{t-2}, S_{t-1}) := \sqrt{\gamma(S_{t-2})(S_{t-1})^2 + 1} \quad ; \tag{5.11}$$

$b(S_{t-2}, S_{t-1})$ is the jth element of $\boldsymbol{L}(S_{t-2}, S_{t-1})$, given by

$$\begin{aligned} b(S_{t-2}, S_{t-1}) &:= \frac{\gamma(S_{t-2})S_{t-1}}{\sqrt{\gamma(S_{t-2})(S_{t-1})^2 + 1}} = \frac{\gamma(S_{t-2})S_{t-1}}{a(S_{t-2}, S_{t-1})} \\ &= \frac{1}{S_{t-1}}\left(a(S_{t-2}, S_{t-1}) - \frac{1}{a(S_{t-2}, S_{t-1})}\right), \end{aligned} \tag{5.12}$$

where the last equality is also true for the case $S_{t-1} = 0$ if we treat $0/0 = 0$; and $c(S_{t-2})$ is the jth element of $\boldsymbol{c}(S_{t-2})$, given by

$$c(S_{t-2}) := 1. \tag{5.13}$$

Whenever S_t, $t < 0$, is encountered, it is treated as $s[1]$. Note that $a(S_{t-2}, S_{t-1}) = 1$ implies that $S_{t-1} = 0$ or $\gamma(S_{t-2}) = 0$. Note also that the above choice uses the augmented channel state (S_{t-2}, S_{t-1}) as we have mentioned in Section 5.3.1.

5.3.4 Decoupling of states and MEC

To see that the proposed communication system is decoupled under the above choice of parameters, let us assume $S_{t-2} = s[j]$. We then obtain that $A_{cl}(S_{t-2}, S_{t-1})$ in (5.9) is a diagonal matrix with (i,i)th element being 1 if $i \neq j$, and being

$$a(S_{t-2}, S_{t-1}) + S_{t-1}b(S_{t-2}, S_{t-1})c(S_{t-2}) = a(S_{t-2}, S_{t-1})^{-1} \tag{5.14}$$

if $i = j$. Hence, we have

$$x_t^{(i)} = \begin{cases} a(S_{t-2}, S_{t-1})^{-1}x_{t-1}^{(j)} - b(S_{t-2}, S_{t-1})N_{t-1} & \text{if } i = j \\ x_{t-1}^{(i)} & \text{if } i \neq j; \end{cases} \tag{5.15}$$

or equivalently in matrix form (noticing that $A_{cl}(S_{t-2}, S_{t-1}) = A(S_{t-2}, S_{t-1})^{-1}$)

$$\boldsymbol{x}_t = A(S_{t-2}, S_{t-1})^{-1}\boldsymbol{x}_{t-1} - \boldsymbol{L}(S_{t-2}, S_{t-1})N_{t-1}. \tag{5.16}$$

More explicitly,

$$\begin{bmatrix} x_t^{(1)} \\ \vdots \\ x_t^{(j)} \\ \vdots \\ x_t^{(m)} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & & & & \\ 0 & 0 & \cdots & a(S_{t-2}, S_{t-1})^{-1} & \cdots & 0 \\ \cdots & \cdots & & & & \\ 0 & \cdots & & 0 & 1 & 0 \\ 0 & \cdots & & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{t-1}^{(1)} \\ \vdots \\ x_{t-1}^{(j)} \\ \vdots \\ x_{t-1}^{(m)} \end{bmatrix} + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ b(S_{t-2}, S_{t-1}) \\ 0 \\ \vdots \\ 0 \end{bmatrix} N_{t-1} \tag{5.17}$$

This illustrates that, conditioned on the channel state sequence $\{S_t\}$, the evolution of subsystem of $x_t^{(i)}$ does not involve $x_0^{(l)}$ for any $l \neq i$.

Therefore, associated with each channel state $s[j]$, there is one subsystem with transmitter state $x^{(j)}$, and the sub-codeword $x_0^{(j)}$ is transmitted independently from other sub-codewords throughout, which shows the decoupling of subsystems (see Appendix C.2 for the proof of decoupling); note that the decoupling is ensured by more complicated design than in the forward communication case. See Fig. 5.3 for the timeline of subsystems' operations. Suppose

$S_{t-2} = s[j]$ and $S_{t-1} = s[l]$. Then at time t, the jth transmitter state $x_{t-1}^{(j)}$ is updated to a new value $x_t^{(j)}$ (by using the feedback channel), whereas all other transmitter states hold their previous values. In the meantime, the lth receiver estimate $\hat{x}_{0,t-1}^{(l)}$ is updated to a new value $\hat{x}_{0,t}^{(l)}$ (by using the forward channel), whereas all other decoder estimates hold their previous values. Thus, for each subsystem, typically it goes through this cycle of operations: holding – updating receiver estimate – updating transmitter state – holding. The transition is triggered by the channel state one step before, and any two subsystems do not perform the same updating operation at the same time. That is, our design ensures mutually exclusive updates among subsystems and hence interaction-free evolution for subsystems. This simplifies the encoding/decoding processes.

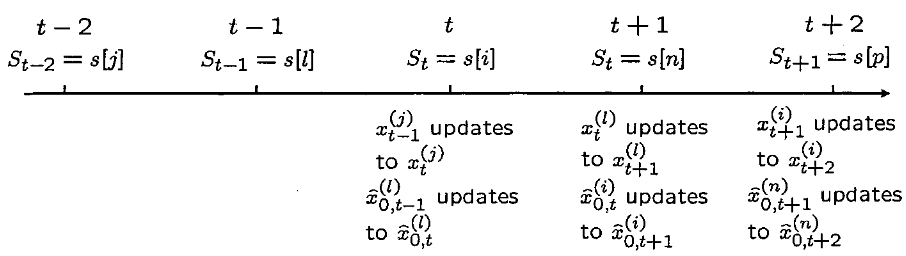


Figure 5.3 Timeline of the operations. The first row lists the time instants, and the second lists the channel states. How the encoder states and the decoder estimates are updated is shown below the time axis. For each time instant, the encoder states and decoder estimates not listed here hold their past values.

As a final remark to this subsection, we note that, in Fig. 5.3, though at time $(t-1)$, the receiver has S_{t-1}, it does not update the subsystem associated with S_{t-1}, namely $\hat{x}_{0,t-1}^{(l)}$ remains as $\hat{x}_{0,t-2}^{(l)}$. This is because u_{t-1} is associated with S_{t-2} (as $\boldsymbol{c}_{t-2}$ chose subsystem associated with S_{t-2} to use the forward channel at time $(t-1)$). This in turn implies that to ensure decoupling, the receiver at time $(t-1)$ should defer the update of the subsystem associated with S_{t-1} until time t.

5.3.5 Minimum-energy control

Equations (5.11)-(5.13) also ensure that, at the time instant when the subsystem for $x^{(j)}$ is activated, the closed-loop eigenvalue of this subsystem locates at the reciprocal of the open-loop eigenvalue (see (5.14), in which the open-loop eigenvalue is $a(S_{t-2}, S_{t-1})$ and the closed-loop one is its reciprocal), which resembles the MEC design for a Gaussian channel without fading [67]. The resulting (minimum) power of u_t, namely the channel input power, will be computed in Proposition 15. In this subsection, we brief discuss the MEC problem.

Consider the MEC of the Markov jump linear system shown in Fig. 5.4 (a), in which $A(S_{t-1}, S_t)$ and $\boldsymbol{c}(S_{t-1})$ are given, and $\boldsymbol{L}$ is the controller to be designed to ensure the closed-loop stability and minimum power of u. The discrete state S^t is known to the controller at time t. In the special case that S_t is constant throughout, this reduces to the MEC of an LTI system. If A and C are given according to (5.10), (5.11), and (5.13), then $\boldsymbol{L}$ according to (5.10) and (5.12) is the *optimal* choice. To see this, note that the subsystems are decoupled, so each subsystem acts like an LTI system. Since each subsystem implements the MEC design, it can be shown that the composite system also minimizes the power of u.

If instead, the choice of A and C is *not* given but a constraint of A is present, i.e., the average growth rate of the open-loop system is fixed, then it can be shown that A and C given in (5.10), (5.11), and (5.13) will lead to the minimum power. Detail computation is omitted here, but the validity will become clear by combining 1) the relation between the control problem and the communication problem shown below and 2) the optimality of the communication problem established in the next section.

In Fig. 5.4 (b) we illustrate how we obtain the optimal communication scheme of Fig. 5.2 from the MEC. By linearity, it holds that

$$\boldsymbol{x}_t = \tilde{\boldsymbol{x}}_t + \hat{\boldsymbol{x}}_t, \tag{5.18}$$

where $\tilde{\boldsymbol{x}}_t$ is the zero-input response (due to initial condition $\boldsymbol{x}_0$), and $\hat{\boldsymbol{x}}_t$ is the zero-state response (due to external input y_0^t); this is shown in the left part of Fig. 5.4 (b). Note that $-\hat{\boldsymbol{x}}_t$ can be generated from y_0^t as indicated in the right part of Fig. 5.4 (b). Since $\boldsymbol{x}_t$ is bounded and $\tilde{\boldsymbol{x}}_t$ grows approximately exponentially, it holds that

$$\tilde{\boldsymbol{x}}_t \approx -\hat{\boldsymbol{x}}_t. \tag{5.19}$$

Therefore, the right part can approximate $\tilde{\boldsymbol{x}}_t$ and hence $\boldsymbol{x}_0$ without actually knowing $\boldsymbol{x}_0$ beforehand. Then some block diagram transformations lead to the proposed communication scheme in Fig. 5.2 which can be used to convey $\boldsymbol{x}_0$.

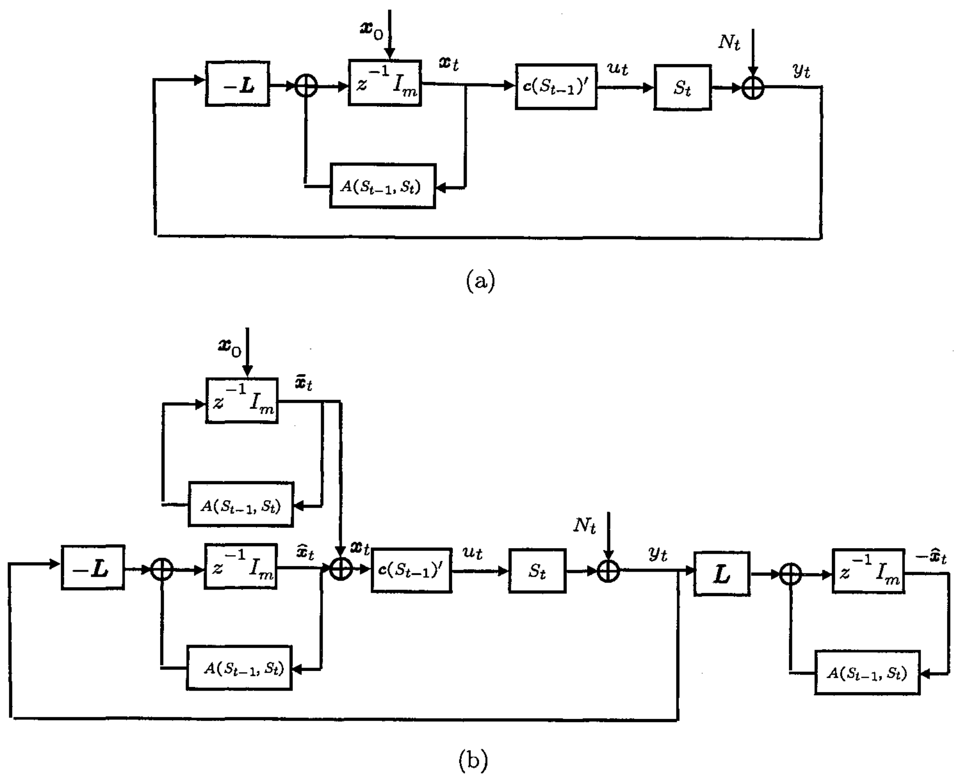


Figure 5.4 (a) MEC of the Markov jump linear system. (b) Intermediate step towards the optimal communication scheme.

5.3.6 Kalman filtering and state augmentation

In this subsection, we present the Kalman filtering system associated with the feedback coding system, and illustrate the state augmentation technique.

Fig. 5.5 (a) shows a time-varying process to be estimated. Let us assume that the $\boldsymbol{c}(S_{t-1})$ vector is forced to use delayed state S_{t-1}. This delay can be "eliminated" by the state augmentation technique, widely used in estimation and control for systems involving delay; see e.g. [10]. Define the augmented state as

$$\bar{S}_t := [S_{t-1}, S_t]', \tag{5.20}$$

then the Markov chain S_t with number of states being m induces a Markov chain $\bar{S}_t$ with number of states being m^2. Let

$$\begin{aligned}\bar{y}_t &= [1 \quad 0]\bar{S}_t + N_t \\ \bar{\boldsymbol{c}}(\bar{S}_t) &= \boldsymbol{c}(S_{t-1}).\end{aligned} \tag{5.21}$$

Then the process involves no delay constraint. We can then easily design the time-varying Kalman filter for this process; note that the Kalman filtering algorithm works for time-varying case as long as the time-varying parameters are known to the Kalman filter causally. It is then relatively simple to figure out the structure of A and $\boldsymbol{c}$ to lead to a decoupled problem, as we demonstrated for the MEC problem. After we obtain the Kalman filtering system, we can easily obtain Kalman filter based coding system and finally the numerically stable coding system described in previous subsections.

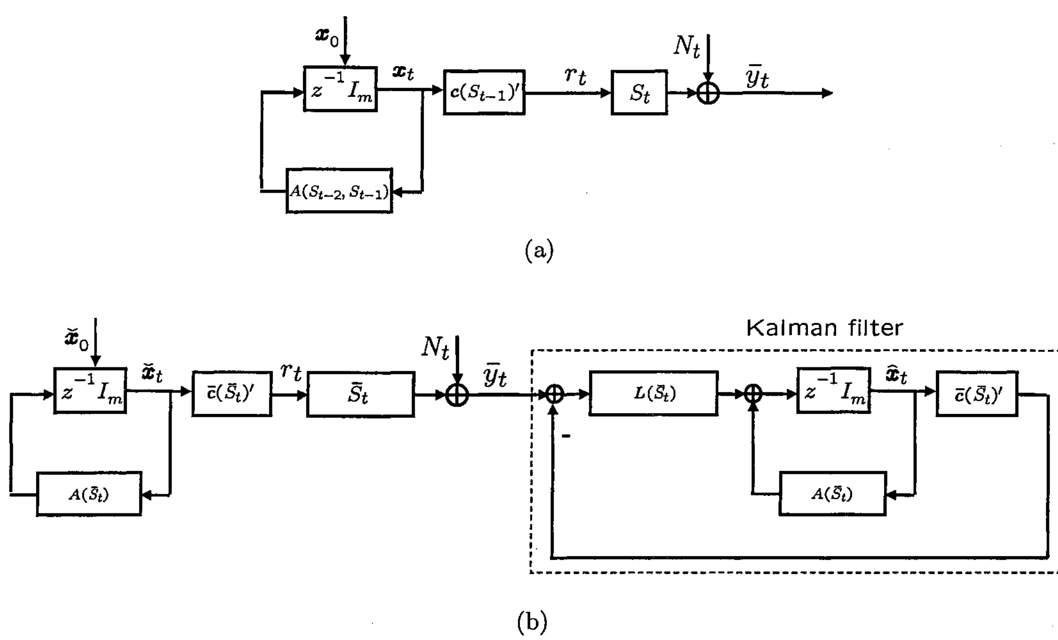


Figure 5.5 (a) A process needed to be estimated. The $\boldsymbol{c}(S_{t-1})$ vector is forced to use delayed state S_{t-1}. (b) A process needed to be estimated and the Kalman filter. After state augmentation, no delay is involved.

5.3.7 Assumptions

In what follows, the choice of parameters according to (5.10)-(5.13) and Assumptions A1) and A2) are assumed unless otherwise specified:

A1) $\gamma(s[j]) > 0$ for each j, i.e., each state $s[j]$ is assigned with a nonzero power;

A2) $s[j] \neq 0$ for each j, i.e., the channel contains no erasure.

We adopt these assumptions only for convenience; our main results hold true if the assumptions do not hold. In fact, when A1 and A2 hold, we have $a(s[j], s[l]) > 1$ and that the mapping from $x_{t-1}^{(j)}$ to $x_t^{(j)}$ in (5.15) is always "strictly contractive", which would lead to a convergence result easily and simplify our development of the main results. However, if A1 does not hold, then whenever the subsystem assigned with zero power is activated, the transmitter does not transmit any information; if A2 does not hold, then whenever the channel state is an erasure, the transmitted signal does not reach the receiver. In either case, the receiver receives only an AWGN and *no* information (apart from the CSI) flows across either the forward channel or the feedback channel. At those moments when no information flows, the transmitter state, receiver state, and the receiver estimate remain as the immediate previous moments, which is different from other moments. More specifically, there is a positive probability that the entire transmitter state $\boldsymbol{x}$ is completely "frozen" at some time t, namely, we may have $A(S_{t-2}, S_{t-1}) = I$ and $\boldsymbol{L}(S_{t-2}, S_{t-1}) = 0$ for some t and hence $\boldsymbol{x}_t = \boldsymbol{x}_{t-1}$, which is not "strictly contractive" at this moment. This would require a couple of extra, minor steps in establishing the (same) main results. For convenience, we would like to develop the main results under A1 and A2 in the main body of the chapter, and defer the description of the extra steps in Appendix C.5.

We also note that A2 can be directly verified from the given channel model, and A1 can also be easily verified by 1) checking the optimal solution of $\gamma(s[i])$ computed by a numerical solver (noting that the decision variables $\Gamma(s[i])$, $i = 1, \cdots, m$, are inside a compact region and a number of numerical tools are available); or 2) applying the "complementary slackness" (if an inequality constraint holds strictly if and only if its multiplier is zero, namely $\Gamma(s[i]) > 0$ if and only if the ith multiplier is zero; cf. [2]) to the optimization problem.

5.3.8 Encoding/decoding method

Define

$$\bar{a}[j] := \prod_{l=1}^{m} a(s[j], s[l])^{\pi[j]p_{jl}}, \text{ for } j = 1, \cdots, m \tag{5.22}$$

and

$$\tilde{a} := \prod_{j=1}^{m} \bar{a}[j]. \tag{5.23}$$

Then it holds that $\bar{a}[j] > 1$ and $\tilde{a} > 1$, since $a(s[j], s[l]) > 1$ for each j and l.

Encoding and decoding

The proposed communication scheme can also transmit digital messages or analog sources. We focus on transmission of digital messages here, and discuss the analog case in Section 5.4.4.

Fix the coding length to be $(T+1)$, i.e., we use the channel from time 0 to time T. We define $\mathcal{B} \in \mathbb{R}^m$ to be the unit hypercube centered at the origin and with each side (denoted $\mathcal{B}^{(j)}$) being the interval $[-\frac{1}{2}, \frac{1}{2}]$. For any fixed $\epsilon > 0$, and for each j, let $\mathcal{B}^{(j)}$ be uniformly partitioned into $\lfloor M_T^{(j)} \rfloor$ subintervals, where

$$M_T^{(j)} := \bar{a}[j]^{(T+1)(1-\epsilon)} \tag{5.24}$$

and $\lfloor M \rfloor$ denotes the largest integer no greater than M. For each T and j, it holds that

$$\lfloor M_T^{(j)} \rfloor = \frac{M_T^{(j)}}{\xi_T^{(j)}} \tag{5.25}$$

for some $\xi_T^{(j)} \in [1, 2)$. Now $\mathcal{B}$ is partitioned into $M_T := \prod_{j=1}^{m} \lfloor M_T^{(j)} \rfloor$ sub-hypercubes. Let the center of each sub-hypercube represent one of a set of M_T equally likely messages. Call the sub-hypercube centers the *codewords*, the sub-interval centers the *sub-codewords*, and the set of codewords the *codebook*.

For encoding, choose a message from the M_T centers, say W, and let $\boldsymbol{x}_0 := W$. Then $\boldsymbol{x}_0$ enters the system (5.7) and generates channel input sequence u^T. For decoding, the decoder generates $\hat{x}_{0,T}^{(j)}$ for each j, and then decides the decoded message $\hat{W}_T$ to be the subinterval center closest to $\hat{\boldsymbol{x}}_{0,T}$. See Fig. 5.6 for a simple example of a codebook.

Once again, we see that the encoding and decoding are fairly simple. In fact, the computation complexity for the encoding and decoding mainly involves a product of $m \times m$ diagonal

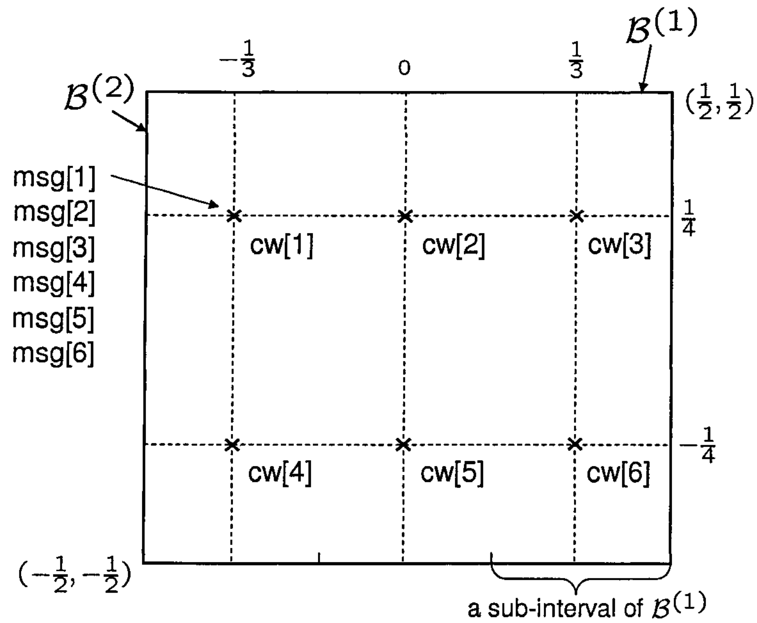


Figure 5.6 An example of codebook. Assume $m = 2$, $\lfloor M_t^{(1)} \rfloor = 3$, and $\lfloor M_t^{(2)} \rfloor = 2$. The six messages are represented by the six codewords. Suppose that message msg[1] is to be conveyed. Then codeword cw[1] is to be transmitted, and the sub-codewords are $x_0^{(1)} = -1/3$ and $x_0^{(2)} = 1/4$. The two sub-codewords are transmitted through two decoupled subsystems. At the receiver side, if *both* sub-codewords are correctly decoded, then the codeword cw[1] and hence message msg[1] can be correctly recovered.

matrices $\prod_{j=0}^{T} A(S_{j-1}, S_j)^{-1}$ and grows linearly in $(T+1)$, where $(T+1)$ is the number of channel uses.

Signalling rate

We define the signalling rate as

$$R := \lim_{T\to\infty} \frac{1}{T+1} \log M_T = \lim_{T\to\infty} \frac{1}{T+1} \sum_{j=1}^{m} \log \lfloor M_T^{(j)} \rfloor, \tag{5.26}$$

if the limit exists.

Probability of error and achievable rate

We declare a *decoding error* if the decoder's decision $\hat{W}_T$ is not equal to the transmitted codeword W. To compute the probability of error PE_T, we first define the probability of error for the jth sub-codeword conditioned on the channel state sequence S^T as

$$PE_{T|S}^{(j)} := \Pr\left(\hat{x}_{0,T}^{(j)} \text{ and } x_0^{(j)} \text{ are in different subintervals of } \mathcal{B}^{(j)}|S^T\right). \tag{5.27}$$

Since conditioned on S^T, $x^{(i)}$ and $x^{(j)}$ evolve independently, we can independently compute $PE_{T|S}^{(j)}$ for each j. We then have that, $PE_{T|S}$, the probability of error for the message W conditioned on the channel state sequence S^T, as

$$PE_{T|S} = 1 - \prod_{j=1}^{m}(1 - PE_{T|S}^{(j)}). \tag{5.28}$$

Consequently, the probability of error for decoding W, which averages $PE_{T|S}$ over all possible channel state sequences S^T, is

$$PE_T := \sum_{S^T \in \Omega_T} PE_{T|S} \Pr(S^T); \tag{5.29}$$

recall that Ω_t is defined as the set of all possible channel state sequences of infinite length. We remark that, though the above definitions are for some fixed $\boldsymbol{x}_0$, since the probability of error for different $\boldsymbol{x}_0$ shares the same asymptotic behavior [3], it is sufficient to study the probability of error for one fixed initial condition $\boldsymbol{x}_0$.

We call a signalling rate R, defined in (5.26), achievable if PE_T decays to zero as T tends to infinity.

5.4 Achieving capacity of AFSMC

In this section, we show that the feedback communication scheme proposed in Section 5.3 along with the parameters given by (5.10)-(5.13) is capacity-achieving, and it leads to doubly exponential decay of error probability. Our main result is summarized in Theorem 3.

Theorem 3. *Suppose $\mathcal{H}$ is any given AFSMC, where the channel state S_t forms an ergodic Markov process and is available instantaneously to the decoder and with one step delay to the encoder. Given any $\mathcal{P} > 0$, let $\gamma(\cdot)$ be the capacity-achieving power allocation that maps the channel state S_t to the channel input power $\gamma(S_t)$ and such that $\mathbf{E}_{S\sim\boldsymbol{\pi}}\gamma(S) \leq \mathcal{P}$ holds. Then,*

[3] In fact for t sufficiently large, $\hat{\boldsymbol{x}}_{0,t}$ has the form $\hat{\boldsymbol{x}}_{0,t} = \boldsymbol{x}_0 + \Delta_t$ (see (5.43)), where Δ_t does not depend on $\boldsymbol{x}_0$. Therefore, asymptotically the decoding error does not depend on $\boldsymbol{x}_0$.

the feedback communication scheme described in Section 5.3, along with the parameters given by (5.10)-(5.13) under Assumptions A1) and A2), transmits at a rate arbitrarily close to the feedback capacity

$$C = \frac{1}{2}\mathbf{E}_{S_t,S_{t+1}} \log(1 + (S_{t+1})^2\gamma(S_t)) = \log\tilde{a} \tag{5.30}$$

under average input power constraint

$$\mathbf{E}_u u^2 \leq \mathcal{P}. \tag{5.31}$$

Moreover, asymptotically $PE_{t|S}$ decays to zero doubly exponentially for any given typical sequence $\{S_t\} \in \Omega_{TYP}$.

Remark 11. Note that for a given channel state sequence $\{S_t\}$, $PE_{t|S}$ may or may not decay doubly exponentially. However, if the given sequence $\{S_t\}$ is typical, then $PE_{t|S}$ decays doubly exponentially. Since typical sequences form a probability one set, we conclude that with probability one, $PE_{t|S}$ decays doubly exponentially; in other words, almost every "sample trajectory" of $PE_{t|S}$ decays doubly exponentially. Similar to the AWGN case, doubly exponential decay is only possible if an *average* channel input power constraint is used (cf. [134]). It is also worth noting that the average probability of error PE_t does not decay doubly exponentially, though $PE_{t|S}$ decays doubly exponentially with probability one. The doubly exponential statement uses channel state sequences $\{S_t\}$ of *infinite* length, whereas the average probability of error PE_t uses channel state sequences S^t (though t can be very large, it is not infinity).

To prove this theorem, we first compute the achievable rate, and then prove the doubly exponential decay, followed by showing that the power constraint (5.31) is satisfied. See Appendix C.1 for the proof of $C = \log\tilde{a}$.

5.4.1 Achievable rate

To show that the proposed scheme can transmit at a rate *arbitrarily* close to capacity C, we need to prove that, for any $\epsilon > 0$ small enough, the rate $(1-\epsilon)C$ is achievable. This development is facilitated by considering the control setup. In fact, for an AFSMC, whenever the control setup in Fig. 5.4 (a) is open-loop unstable but closed-loop stabilized, the communication system in Fig. 5.2 achieves a rate determined by the growth rate of the open-loop, similar to the case of Gaussian channels without time-selective fading [28].

Our choice of parameters in Section 5.3.3 leads to that,the open-loop of the control setup is unstable. The average rate of growth for this unstable system, defined as $\mathbf{E}_{S_t,S_{t+1}} \log A(S_t, S_{t+1})$, is $\log \tilde{a}$ (following the derivation in Appendix C.1). This is exactly the maximum achievable rate for the proposed communication system. Below, we show that the control setup is stabilized in closed-loop, based on which we prove the achievable rate. Note that the stability of the control setup is in the sense of, first, the boundedness and convergence to zero of the first moment $\mathbf{E}\boldsymbol{x}_t$, and second, the boundedness of the variance-covariance matrix $\Sigma_t := \mathbf{E}\boldsymbol{x}_t\boldsymbol{x}_t' - \mathbf{E}\boldsymbol{x}_t\mathbf{E}\boldsymbol{x}_t'$, *both conditioned on a given channel state sequence* $\{S_t\}$. Finally, note that if one fixes the choice of $A(S_t, S_{t+1})$, then the rate is fixed, and thus the choice of $\boldsymbol{L}(S_t, S_{t+1})$ is to ensure the stability of the closed-loop as well as the minimum input power, which is exactly an MEC problem.

We introduce the following notation. Fix a sequence $\{S_t\}$. Recalling that if $S_{\tau-1} = s[j]$ and $S_\tau = s[l]$, we have

$$A(S_{\tau-1}, S_\tau)^{-1} = \operatorname{diag}\left([1, \cdots, 1, a(s[j], s[l])^{-1}, 1, \cdots, 1]\right), \tag{5.32}$$

we can then obtain, for any t,

$$\begin{aligned}\prod_{\tau=0}^{t} A(S_{\tau-1}, S_\tau)^{-1} &= \operatorname{diag}\left(\left[\prod_{l=1}^{m} a(s[1], s[l])^{-n(1,l,t)}, \cdots, \prod_{l=1}^{m} a(s[m], s[l])^{-n(m,l,t)}\right]\right) \\ &:= \operatorname{diag}\left(\left[\phi_t^{(1)}, \cdots, \phi_t^{(m)}\right]\right) := \Phi_t,\end{aligned} \tag{5.33}$$

where $\phi_t^{(j)}$ and Φ_t are defined in an obvious way.

Lemma 4. *Assume the hypotheses of Theorem 3, and fix a channel state sequence $\{S_t\}$ in Ω. Then for the control setup (5.8),*

i) Φ_t is the state transition matrix, namely the response due to initial condition $\boldsymbol{x}_0$ is $\boldsymbol{x}_{t+1} = \Phi_t\boldsymbol{x}_0$. For every $j = 1, \cdots, m$, it holds that $0 < \phi_t^{(j)} \leq 1$ for any t, and that

$$\lim_{t\to\infty} \phi_t^{(j)} \to 0 \text{ if } \{S_t\} \in \Omega_{TYP}; \tag{5.34}$$

ii) For any fixed initial condition $\boldsymbol{x}_0$, it holds that $-|x_0^{(j)}| \leq \mathbf{E}x_t^{(j)} \leq |x_0^{(j)}|$ for any t, and *that*

$$\lim_{t\to\infty} \mathbf{E}\boldsymbol{x}_t \to 0 \text{ if } \{S_t\} \in \Omega_{TYP}; \tag{5.35}$$

iii) For any $\boldsymbol{x}_0$, $\Sigma_t := \mathbf{E}(\boldsymbol{x}_t - \mathbf{E}\boldsymbol{x}_t)(\boldsymbol{x}_t - \mathbf{E}\boldsymbol{x}_t)'$ *is a diagonal matrix, i.e., the components of* $\boldsymbol{x}_t$ *are mutually independent, and there exist* $\underline{c}$ *and* $\bar{c}$ *such that*

$$0 < \underline{c} \le \Sigma_t^{(j)} \le \bar{c} \le \infty \tag{5.36}$$

for any t*, where* $\Sigma_t^{(j)}$ *is the* (j, j)*th element of* Σ_t*.*

Proof: See Appendix C.2. ∎

Now we can use the stability of the control setup to establish the reliable communication across the channel. Note that the stability of the control setup implies that the communication scheme does not involve unbounded signals, as claimed in Section 5.1.

Proposition 13. *Assume the hypotheses of Theorem 3. Then the communication system reliably transmits at rate*

$$R = (1-\epsilon)\log\tilde{a} = (1-\epsilon)C, \tag{5.37}$$

for any given $\epsilon > 0$*.*

Proof: To establish the achievable rate, we first compute the signalling rate, followed by proving that the error probability goes to zero, which implies that the signalling rate is achievable.

Signalling rate

For any $\{S_t\}$ in Ω_{TYP}, it holds that

$$\begin{aligned} R &:= \lim_{T\to\infty} \frac{\sum_{j=1}^m \log\lfloor M_T^{(j)}\rfloor}{T+1} \\ &= \lim_{T\to\infty} \left(\frac{\sum_{j=1}^m \log M_T^{(j)}}{T+1} - \frac{\sum_{j=1}^m \log \xi_T^{(j)}}{T+1} \right) \\ &= \lim_{T\to\infty} \frac{\sum_{j=1}^m \log \bar{a}[j]^{(T+1)(1-\epsilon)}}{T+1} \\ &= (1-\epsilon)\sum_{j=1}^m \log \bar{a}[j] \\ &= (1-\epsilon)\log\tilde{a}. \end{aligned} \tag{5.38}$$

Probability of error

The proof of vanishing probability of error is essentially an asymptotic equi-partition (AEP) based argument (cf. [16]). Let us define

$$\Omega_{t,\mu} := \left\{ S^t \left| \left| \frac{n(j,t)}{t+1} - \pi[j] \right| < \mu \text{ and } \left| \frac{n(j,l,t)}{n(j,t)} - p_{jl} \right| < \mu \right. \right\}, \tag{5.39}$$

where $\mu > 0$. From AEP or the Weak Law of Large Numbers, it holds that

$$\Pr(\Omega_{t,\mu}) \to 1 \tag{5.40}$$

as t tends to infinity. Then Ω_t is partitioned into two subsets: $\Omega_{t,\mu}$ and $\Omega^c_{t,\mu}$. Our goal is to show that the probability of error $PE_{T|S}$ satisfies

$$PE_{T|S} \leq \kappa(T) \tag{5.41}$$

for some vanishing function $\kappa(\cdot) \geq 0$ which is independent of T and S^T, as long as $S^T \in \Omega_{T,\mu}$. This would lead to vanishing probability of error PE_T. More precisely, note that

$$\begin{aligned} PE_T &:= \sum_{S^T \in \Omega_T} PE_{T|S} \Pr(S^T) \\ &= \sum_{S^T \in \Omega_{T,\mu}} PE_{T|S} \Pr(S^T) + \sum_{S^T \in \Omega^c_{T,\mu}} PE_{T|S} \Pr(S^T) \\ &\leq \sum_{S^T \in \Omega_{T,\mu}} \kappa(T) \Pr(S^T) + \sum_{S^T \in \Omega^c_{T,\mu}} \Pr(S^T) \\ &\leq \kappa(T) + \Pr(\Omega^c_{T,\mu}), \end{aligned} \tag{5.42}$$

where we have used that $\kappa \geq 0$ and is independent of S^T, $\sum_{S^T \in \Omega_{T,\mu}} \Pr(S^T) \leq 1$, and $PE_{T|S} \leq 1$. Therefore, as T increases, the averaged probability of error PE_T would decay to zero.

To show that $PE_{T|S}$ decays to zero on $\Omega_{T,\mu}$ as T goes to infinity, we may first investigate the behavior of $PE^{(j)}_{T|S}$ on $\Omega_{T,\mu}$. Fix any channel state sequence S^T and initial condition $\boldsymbol{x}_0$. It is straightforward to compute that

$$\hat{\boldsymbol{x}}_{0,t} = \boldsymbol{x}_0 - \prod_{\tau=0}^{t} A(S_{\tau-1}, S_\tau)^{-1} \boldsymbol{x}_{t+1} = \boldsymbol{x}_0 - \Phi_t \boldsymbol{x}_{t+1}. \tag{5.43}$$

Because the noise is zero mean i.i.d. Gaussian, $\boldsymbol{x}_{t+1}$ is Gaussian with distribution $\mathcal{N}(\mathbf{E}\boldsymbol{x}_{t+1}, \Sigma_{t+1})$, *if conditioned on S^t and $\boldsymbol{x}_0$.* Then notice that $\mathbf{E}\boldsymbol{x}_{t+1} = \Phi_t \boldsymbol{x}_0$. Therefore, $\hat{\boldsymbol{x}}_{0,t}$ is an m-variate Gaussian distributed as

$$\mathcal{N}\left(\boldsymbol{x}_0 - (\Phi_t)^2\boldsymbol{x}_0, (\Phi_t)^2\Sigma_{t+1}\right), \tag{5.44}$$

and particularly, for each j, $\hat{x}_{0,t}^{(j)}$ is a univariate Gaussian distributed as

$$\mathcal{N}\left(x_0^{(j)} - (\phi_t^{(j)})^2 x_0^{(j)}, (\phi_t^{(j)})^2 \Sigma_{t+1}^{(j)}\right). \tag{5.45}$$

We now assume without loss of generality that $x_0^{(j)}$ is the center of the ith subinterval of $\mathcal{B}^{(j)}$. See Fig. 5.7. We can thus derive the following expression of $PE_{T|S}^{(j)}$:

$$\begin{aligned} PE_{T|S}^{(j)} &= Q\left(\frac{0.5/\lfloor M_T^{(j)}\rfloor + x_0^{(j)}(\phi_T^{(j)})^2}{\phi_T^{(j)}\sqrt{\Sigma_{T+1}^{(j)}}}\right) + Q\left(\frac{0.5/\lfloor M_T^{(j)}\rfloor - x_0^{(j)}(\phi_T^{(j)})^2}{\phi_T^{(j)}\sqrt{\Sigma_{T+1}^{(j)}}}\right) \\ &= Q\left(\frac{0.5\xi_T^{(j)}/M_T^{(j)} + x_0^{(j)}(\phi_T^{(j)})^2}{\phi_T^{(j)}\sqrt{\Sigma_{T+1}^{(j)}}}\right) + Q\left(\frac{0.5\xi_T^{(j)}/M_T^{(j)} - x_0^{(j)}(\phi_T^{(j)})^2}{\phi_T^{(j)}\sqrt{\Sigma_{T+1}^{(j)}}}\right) \\ &= Q\left(\frac{0.5\xi_T^{(j)}}{\bar{a}[j]^{(T+1)(1-\epsilon)}\phi_T^{(j)}\sqrt{\Sigma_{T+1}^{(j)}}} + \frac{x_0^{(j)}\phi_T^{(j)}}{\sqrt{\Sigma_{T+1}^{(j)}}}\right) + Q\left(\frac{0.5\xi_T^{(j)}}{\bar{a}[j]^{(T+1)(1-\epsilon)}\phi_T^{(j)}\sqrt{\Sigma_{T+1}^{(j)}}} - \frac{x_0^{(j)}\phi_T^{(j)}}{\sqrt{\Sigma_{T+1}^{(j)}}}\right). \end{aligned} \tag{5.46}$$

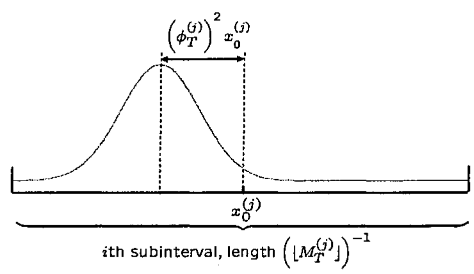


Figure 5.7 The location of $x_0^{(j)}$ in a subinterval of $\mathcal{B}^{(j)}$ and the distribution of $\hat{x}_{0,T}^{(j)}$, with mean $x_0^{(j)} - (\phi_T^{(j)})^2 x_0^{(j)}$ and variance $(\phi_T^{(j)})^2\Sigma_{T+1}^{(j)}$.

Notice that

$$\begin{aligned} \bar{a}[j]^{(T+1)(1-\epsilon)}\phi_T^{(j)} &= \left(\prod_{l=1}^m a(s[j], s[l])^{\pi[j]p_{jl}}\right)^{(T+1)(1-\epsilon)} \prod_{l=1}^m a(s[j], s[l])^{-n(j,l,T)} \\ &= \left(\prod_{l=1}^m a(s[j], s[l])^{d(j,l,T)}\right)^{T+1}, \end{aligned} \tag{5.47}$$

where

$$d(j,l,T) := (1-\epsilon)\pi[j]p_{jl} - \frac{n(j,l,T)}{T+1}. \tag{5.48}$$

Since

$$\frac{n(j,l,T)}{T+1} = \frac{n(j,l,T)}{n(j,T)}\frac{n(j,T)}{T+1} \to \pi[j]p_{jl}, \tag{5.49}$$

it holds that for sufficiently large T and sufficiently small μ, $d(j,l,T)$ is non-positive for any S^T in $\Omega_{T,\mu}$.

Let $a := \min_l a(s[j], s[l]) > 1$ and $\alpha := a^\epsilon > 1$. It then follows from (5.47), (5.48), and $d(j,l,T) \le 0$ that

$$\begin{aligned}
\bar{a}[j]^{(T+1)(1-\epsilon)}\phi_T^{(j)} &\le \left(\prod_{l=1}^{m} a^{d(j,l,T)}\right)^{T+1} \\
&= a^{(T+1)(1-\epsilon)\pi[j]-n(j,T)} \\
&= a^{-\epsilon n(j,T)+(T+1)(1-\epsilon)\left(\pi[j]-\frac{n(j,T)}{T+1}\right)} \\
&= \alpha^{-n(j,T)+(T+1)\left(\frac{1}{\epsilon}-1\right)\left(\pi[j]-\frac{n(j,T)}{T+1}\right)}.
\end{aligned} \tag{5.50}$$

Therefore,

$$\begin{aligned}
Q_{T,1} &:= Q\left(\frac{0.5\xi_T^{(j)}}{\bar{a}[j]^{(T+1)(1-\epsilon)}\phi_T^{(j)}\sqrt{\Sigma_{T+1}^{(j)}}} + \frac{x_0^{(j)}\phi_T^{(j)}}{\sqrt{\Sigma_{T+1}^{(j)}}}\right) \\
&\le Q\left(\frac{\xi_T^{(j)}}{2\sqrt{\Sigma_{T+1}^{(j)}}}\alpha^{n(j,T)+(T+1)\left(\frac{1}{\epsilon}-1\right)\left(\pi[j]-\frac{n(j,T)}{T+1}\right)} + \frac{x_0^{(j)}\phi_T^{(j)}}{\sqrt{\Sigma_{T+1}^{(j)}}}\right) = Q\left(\alpha^{\eta_T^{(j)}+\zeta_T^{(j)}}\right),
\end{aligned}$$

where

$$\eta_T^{(j)} := n(j,T) + (T+1)\left(\frac{1}{\epsilon}-1\right)\left(\pi[j]-\frac{n(j,T)}{T+1}\right) \tag{5.51}$$

and

$$\zeta_T^{(j)} := \log_\alpha\left\{\frac{\xi_T^{(j)} + 2\alpha^{-n(j,T)\left[1+\left(\frac{1}{\epsilon}-1\right)\left(\frac{(T+1)\pi[j]}{n(j,T)}-1\right)\right]}x_0^{(j)}\phi_T^{(j)}}{2\sqrt{\Sigma_{T+1}^{(j)}}}\right\}. \tag{5.52}$$

For μ sufficiently small and T sufficiently large, it holds that

$$\left[1+\left(\frac{1}{\epsilon}-1\right)\left(\frac{(T+1)\pi[j]}{n(j,T)}-1\right)\right] > 0; \tag{5.53}$$

by $\xi_T^{(j)} \in [1,2)$, $x_0^{(j)} \in (-1,1)$, $\phi_T^{(j)} \in (0,1]$, and $\Sigma_{T+1}^{(j)} \in (\underline{c},\bar{c})$ for any j, T, and S^T, one can easily show that ζ_T is uniformly bounded, namely $\zeta_T \in (\underline{\zeta}_1,\bar{\zeta}_1)$ where the bounds are independent of j, T, $\boldsymbol{x}_0$, and S^T. Also for T sufficiently large, it holds that

$$\left(\frac{1}{\epsilon}-1\right)\left|\pi[j]-\frac{n(j,T)}{T+1}\right| < \left(\frac{1}{\epsilon}-1\right)\mu. \tag{5.54}$$

Pick μ sufficiently small such that

$$\pi[j]-\left(\frac{1}{\epsilon}-1\right)\mu > \sigma \tag{5.55}$$

for some $\sigma > 0$ and for all $j = 1, 2, \cdots, m$. Note that such μ and σ exist since m is finite. This would yield that, for T large enough, $\eta_T > (T+1)\sigma$. Consequently, $(\eta_T + \zeta_T)$ goes to infinity and $Q_{T,1}$ vanishes. More precisely,

$$Q_{T,1} < Q(\alpha^{(T+1)\sigma+\underline{\zeta}_1}). \tag{5.56}$$

Similarly,

$$\begin{aligned} Q_{T,2} &:= Q\left(\frac{0.5\xi_T^{(j)}}{\bar{a}[j]^{(T+1)(1-\epsilon)}\phi_T^{(j)}\sqrt{\Sigma_{T+1}^{(j)}}} - \frac{x_0^{(j)}\phi_T^{(j)}}{\sqrt{\Sigma_{T+1}^{(j)}}}\right) \\ &< Q(\alpha^{(T+1)\sigma+\underline{\zeta}_2}). \end{aligned} \tag{5.57}$$

for some $\underline{\zeta}_2$.

Letting $\underline{\zeta} := \max\{\underline{\zeta}_1, \underline{\zeta}_2\}$, we have that

$$PE_{T|S}^{(j)} = Q_{T,1} + Q_{T,2} < 2Q(\alpha^{(T+1)\sigma+\underline{\zeta}}). \tag{5.58}$$

vanishes. Then invoking the union bound

$$PE_{T|S} = 1 - \prod_{j=1}^{m}(1 - PE_{T|S}^{(j)}) \le \sum_{j=1}^{m} PE_{T|S}^{(j)} < 2mQ(\alpha^{(T+1)\sigma+\underline{\zeta}}), \tag{5.59}$$

we deduce that $PE_{T|S}$ would converge to zero on $\Omega_{T,\mu}$ for sufficiently small μ. Thus we prove that PE_T decays to zero, i.e., rate R is achievable. [4] ■

[4] We may also employ a modified decoding after we obtain $\hat{\boldsymbol{x}}_{0,T}$, by letting $\hat{W}_T$ be the subinterval center closest to $(I-(\Phi_T)^2)^{-1}\hat{\boldsymbol{x}}_{0,T}$. This removes the estimate bias and hence the asymptotic behavior analysis of the communication scheme remains the same.

5.4.2 Doubly exponential decay of probability of error

Proposition 14. *Assume the hypotheses of Theorem 3. Then*

i) For sufficiently large T, for any j and $\{S_t\} \in \Omega_{TYP}$,

$$PE_{T|S}^{(j)} \leq \beta_1 \exp\{-\exp(\beta_2(T+1)+o(T))\} \tag{5.60}$$

for some $\beta_1, \beta_2 > 0$, where $o(T)$ denotes the terms with lower order than T, namely $o(T)/T$ vanishes as T tends to infinity;

ii) For any j and $\{S_t\} \in \Omega_{TYP}$, the decay exponent for $PE_{T|S}^{(j)}$ is

$$e^{(j)} := \lim_{T\to\infty} \frac{1}{T+1} \log(\log(\frac{1}{PE_{T|S}^{(j)}})) = 2\epsilon \log(\bar{a}[j]), \tag{5.61}$$

and the decay exponent for $PE_{T|S}$ is

$$e := \lim_{T\to\infty} \frac{1}{T+1} \log(\log(\frac{1}{PE_{T|S}})) = \min_j e^{(j)} = 2\epsilon \log \underline{a}, \tag{5.62}$$

where $\underline{a} := \min_j \bar{a}[j]$.

Remark 12. This proposition claims that, for each sub-message $x_0^{(j)}$, essentially its probability of error decays doubly exponentially, and hence $PE_{T|S}$ decays doubly exponentially with respect to T for large enough T, provided that we exclude a set of $\{S_t\}$ that would occur with zero probability and on which only zero communication rate can be achieved. This notion of decay of error probability is stronger than what we have proven in Proposition 13 regarding the probability of error averaged over all finite-length typical sequences S^T, analogue with that the Strong Law of Large Numbers is stronger than the Weak Law of Large Numbers. The decay exponent of $PE_{T|S}$ is the smallest (slowest) decay exponent of $PE_{T|S}^{(j)}$ among all j. When the channel has no fading, i.e., $m = 1$ and $a = a(s[j], s[l])$ for all j and l, the decay exponent becomes $e = 2\epsilon \log a = 2(C - R)$, and we recover the decay exponent obtained in [107, 106].

Proof: See Appendix C.3. ∎

5.4.3 Power computation

Proposition 15. *Assume the hypotheses of Theorem 3. Then the average channel input power is*

$$\mathrm{E}u^2 = \sum_{i=1}^{m} \pi[i]\gamma(s[i]), \tag{5.63}$$

and hence satisfies power constraint (5.31).

Proof: See Appendix C.4. ∎

Combining the results in Propositions 13-15, we have completed the proof for Theorem 3. Noting that the entropy rate of the channel output is indeed the transmission rate, we also conclude that we obtain the MEC design over an AFSMC with an entropy rate constraint.

5.4.4 Transmission of Gaussian random vector

With some modifications we can transmit a Gaussian random vector over the AFSMC, in parallel with the AWGN channel case. Let us use the same parameters given in (5.10)-(5.13) and $\boldsymbol{x}_0 \sim \mathcal{N}(0, \mathcal{P}I_m)$, and follow the dynamics (5.7). For a given channel state sequence $\{S_t\} \in \Omega_{TYP}$, we obtain the MSE distortion as

$$\mathrm{MSE}(\hat{\boldsymbol{x}}_{0,T}) := \mathbf{E}(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0,T})(\boldsymbol{x}_0 - \hat{\boldsymbol{x}}_{0,T})' = (\Phi_T)^2 \mathbf{E}\boldsymbol{x}_{T+1}\boldsymbol{x}'_{T+1}, \tag{5.64}$$

which, by rate-distortion theory, requires an asymptotic rate to be at least

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{2(T+1)} \log \frac{\mathcal{P}^m}{|\mathrm{MSE}(\hat{\boldsymbol{x}}_{0,T})|} &= \lim_{T\to\infty} \frac{1}{2(T+1)} \log \frac{\mathcal{P}^m}{\prod_{j=1}^{m} (\phi_T^{(j)})^2 |\mathbf{E}\boldsymbol{x}_{T+1}\boldsymbol{x}'_{T+1}|} \\
&\overset{(a)}{=} -\lim_{T\to\infty} \frac{1}{T+1} \prod_{j=1}^{m} \phi_T^{(j)} \\
&\overset{(b)}{=} \log \tilde{a},
\end{aligned} \tag{5.65}$$

where (a) follows from the boundedness of $\mathcal{P}^m$ and $|\mathbf{E}\boldsymbol{x}_{t+1}\boldsymbol{x}'_{t+1}|$, and (b) follows from a derivation similar to Appendix C.3. Since $\log \tilde{a}$ is the capacity, we conclude that $\boldsymbol{x}_0$ is successively refined at the capacity rate of the AFSMC.

5.4.5 Special case: AWGN i.i.d. fading channel

Theorem 3 directly applies to the case that $\{S_t\}$ forms a discrete i.i.d. process. However, a simplified capacity-achieving feedback scheme with a *scalar* transmitter state exists. Assume that the channel state has an i.i.d. distribution given by

$$\Pr(S_t = s[i]) = p[i] \text{ for } t = 0, 1, \cdots, \tag{5.66}$$

where for $i = 1, 2, \cdots, m$, $p[i]$ and $s[i]$ are fixed numbers. Given any power budget $\mathcal{P} > 0$, we choose the parameters in the communication scheme as

$$\begin{array}{llll} A(S_{t-2}, S_{t-1}) & := A(S_{t-1}) & := \sqrt{(S_{t-1})^2\mathcal{P} + 1} & \in \mathbb{R} \\ \boldsymbol{L}(S_{t-2}, S_{t-1}) & := \boldsymbol{L}(S_{t-1}) & := \frac{S_{t-1}\mathcal{P}}{\sqrt{(S_{t-1})^2\mathcal{P}+1}} & \in \mathbb{R} \\ \boldsymbol{c}(S_{t-2}) & := 1 & \in \mathbb{R}. & \end{array} \tag{5.67}$$

Note that A and $\boldsymbol{c}$ in this design do *not* require the augmented channel state (S_{t-2}, S_{t-1}); S_{t-1} is sufficient. We can show that this design leads to a transmission rate arbitrarily close to the capacity (proof skipped for brevity)

$$C_{iid} = \frac{1}{2}\sum_{i=1}^{m} p[i] \log(1 + s[i]^2\mathcal{P}) = \frac{1}{2}\mathbf{E}_S \log(1 + S^2\mathcal{P}). \tag{5.68}$$

Note that no power adaptation is needed in the capacity formula and in the proposed scheme.

Remark 13. We remark that a direct application (without multiplexing) of the ideas in [28] does not achieve the Shannon capacity. We conjecture that such a direct application can at most achieve the anytime capacity, which is less than the Shannon capacity. In *anytime information theory* [102], tracking a constantly growing unstable source through a channel is studied, the information theoretic interpretations are provided, and the fundamental limitations are captured by anytime capacity. In this chapter, by using the multiplexing structure that adapts according to the channel state variation, the decoder no longer tracks a constantly growing unstable source, but a source that grows faster when the channel is in good state, and grows slower when the channel is in bad state, and thus we can achieve the Shannon capacity. That is, this adaptation idea is essential to the optimality, and it significantly generalizes the ideas used in [102, 117] for an erasure channel.

5.4.5.1 AWGN i.i.d. fading channel with infinite channel states

AWGN i.i.d. fading channels with infinite channel states include many channels as the special cases, such as the Rayleigh, Rician, Nakagami, and Weibull fading channels. Here we focus on the scenario of real channel state-spaces; the scenario of complex channel state-spaces can be studied similarly. Assume that the channel state forms an i.i.d. process with density $p_S(s)$, where $s \in \mathbb{R}$. Then the channel capacity is

$$C_{iid,inf} = \frac{1}{2}\mathbf{E}_{S\sim p_S} \log(1 + S^2\mathcal{P}). \tag{5.69}$$

Then we construct a coding scheme with a scalar transmitter state as in the finite state-space case, using the choice of parameters given in (5.67). As we show in Appendix C.6, this scheme achieves the feedback capacity given in (5.69). We point out that the proof makes use of the fact that the transmitter can be designed as a scalar system and hence this proof may not be directly applicable to Markov channels with infinite states.

5.5 Numerical example

Consider a Gilbert-Elliot fading channel with AWGN, i.e. an AFSMC with only two states. We simulate the proposed scheme for this channel. Fig. 5.8 shows the simulated $PE^{(j)}_{T|S}$ and $PE_{T|S}$ for a randomly chosen (typical) $\{S_t\}$, as well as the theoretic $PE_{T|S}$ computed from (5.46) and (5.28). We see that the probability of error decays rather fast within 20 channel uses. However, the decay of probability of error is not very smooth, caused by instantaneous deviations from the typical channel behaviors, though $\{S_t\}$ may be typical in the long run. This may be improved by considering a "turbo mode" of using larger power at those atypical instants, which does not affect the average power constraint (under further investigation); see [103] for the idea of turbo mode.

Fig. 5.9 (a) shows the decay of PE_T, which averages both the doubly exponential decay of error probability for typical sequences and the exponential decay of error probability for atypical sequences. [5] These fast decays imply that the proposed scheme allows shorter coding

[5]This implies that our error probability averaged over all channel state sequences decays exponentially. [103] presented a coding scheme to reduce the decoding errors caused by atypical channel state sequences for "streaming" communication. The same idea may be applied here to lead to doubly exponential decay of the averaged error probability; this is subject to future research.

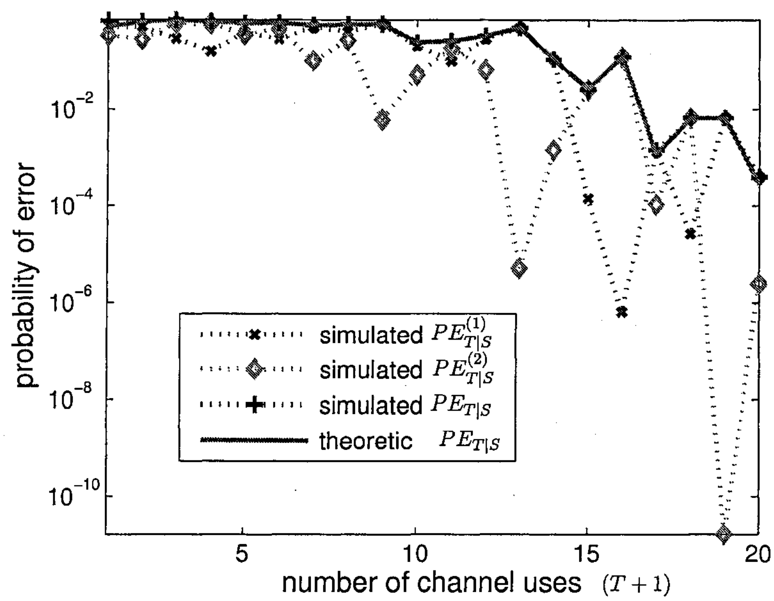
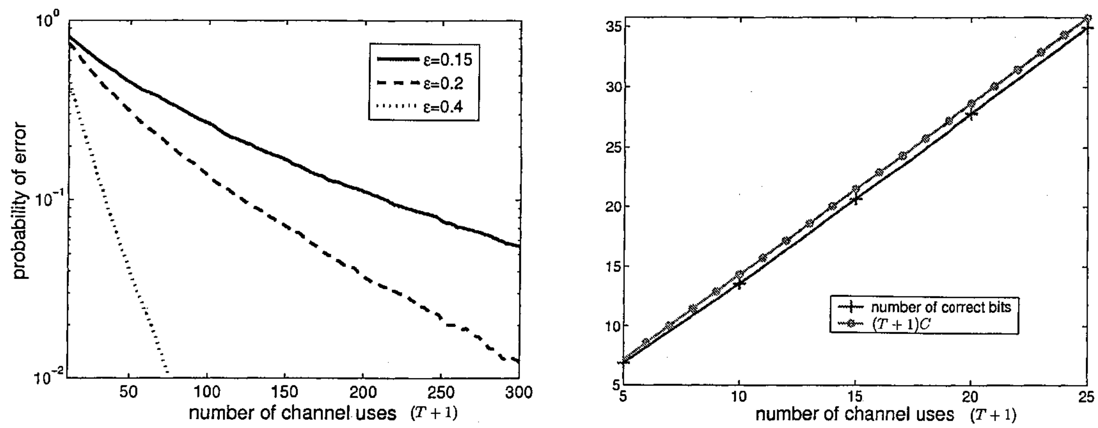


Figure 5.8 Simulated $PE_{T|S}^{(j)}$, simulated $PE_{T|S}$, and theoretic $PE_{T|S}$. $s[1] = 2$, $s[2] = 1$, $p_{11} = 0.65$, $p_{22} = 0.38$, $p = 3$, and $\epsilon = 0.2$ (i.e. $R = 0.8C$).

length and shorter coding delay; here coding delay measures the time steps that one has to wait for the message to be decoded at the decoder with small enough error probability. The short coding delay is also reflected in Fig. 5.9 (b), where we compare the message and the decoded message bit by bit and count how many bits are correctly obtained by the decoder. At time $T = 24$, the channel can transmit 35.8 bits if at each step the capacity C is attained, and the simulation shows that on average 34.9 bits are actually correctly decided.

Therefore, though the availability of output feedback at the encoder does not affect the capacity when DTRCSI is available, we have seen that, output feedback can considerably simplify the coding design and coding process while achieving the capacity (in contrast to the more sophisticated designs to approach the capacity using Turbo codes or LDPC codes, see e.g. [101]), and it leads to better performance in terms of probability of error than the forward coding schemes in the literature. Recall that turbo codes or LDPC codes need long coding lengths of at least several thousands to achieve a decent performance, and in [101], coding length of 8000 was used over a Markov channel. However, since power adaptation has not been employed in those schemes, a fair and more accurate comparison is not yet available.



(a) (b)

Figure 5.9 (a) Theoretic PE_T. (b) The number of bits that has been correctly decided and the number of bits that could be correctly decided if at each step the capacity rate is attained.

5.6 Summary

In this chapter, we proposed a capacity-achieving feedback communication scheme for an AFSMC with precise CSI available to the decoder immediately and to the encoder with delay. This scheme is essentially a multiplexing of the scheme obtained for AWGN channels with feedback switching according to the augmented channel states. The scheme greatly simplifies the complexity in the coding design and coding processes. The error probability decreases to zero doubly exponentially and it shortens the coding length, compared with existing coding schemes over an AFSMC in the literature. We established the equivalence among feedback communication with over over an AFSMC, estimation over the same channel, and feedback stabilization over the same channel. We have seen that the utilization of the estimation/control-theoretic equivalence of the proposed coding scheme facilitates the development. In fact, the idea of introducing multiplexing and augmented channel states was motivated by studying the estimation/control system; particularly the multiplexing idea is motivated by the research of Markov jump linear control systems, and the augmented channel state idea is motivated by the research of estimation/control systems with delay. We remark that these two ingredients may play significant role in studying any feedback communication systems with known time-variations and delays.

CHAPTER 6. WRITING ON DIRTY PAPER WITH FEEDBACK

6.1 Introduction

The study of lossless interference cancelation in a communication system has attracted considerable interest from researchers, since the publication of Costa's celebrated article "Writing on Dirty Paper" [14]. Costa considered a power constrained discrete-time Gaussian channel, in which there are two independent processes that corrupt the channel inputs. One process, a sequence $\{\xi_t\}$ of i.i.d. $\mathcal{N}(0, Q)$ random variables, is completely known to the encoder *non-causally* and is unknown to the decoder; this is referred to as the *interference* (or channel states). The other process, a sequence $\{N_t\}$ of i.i.d. $\mathcal{N}(0, 1)$ random variables (independent of $\{\xi_t\}$), is unknown to neither the encoder nor the decoder. Costa named this model as the *writing on dirty paper*- (WDP-) channel model. Fig. 6.1 illustrates this model, in which W is the message, u is the channel input, ξ is the interference, N is the channel noise, $y := u+\xi+N$ is the channel output, and $\hat{W}$ is the decoded message. If ξ is zero (or is also known to the receiver), then the channel can achieve a rate $\mathcal{R}$ with the channel input power being at least $\mathcal{P}(\mathcal{R}) := 2^{2\mathcal{R}}-1$. If ξ is not zero and is not known to the receiver, then obviously the minimum channel input power $\mathcal{P}_{WDP}(\mathcal{R})$ for achieving rate $\mathcal{R}$ is bounded by

$$\mathcal{P}(\mathcal{R}) \leq \mathcal{P}_{WDP}(\mathcal{R}) \leq (1+Q)\mathcal{P}(\mathcal{R}). \tag{6.1}$$

What is surprising is that the lower bound is achievable, namely there exists a strategy such that we can transmit across the channel *as if the interference did not exist.* In other words, we can achieve *lossless interference cancelation,* by which we mean that the interference is "canceled" without incurring any power increase or rate loss; such an optimal strategy was introduced by Costa in [14]. Note that the strategy of letting the encoder ignore the interference knowledge or letting the decoder try the brutal force way to cancel the interference incurs power increase or rate loss, and hence they are only suboptimal.

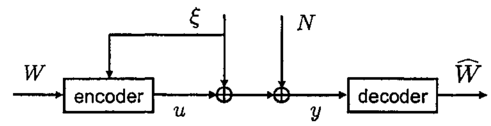


Figure 6.1 AWGN WDP-channel model.

Costa's results have been generalized to various situations; see [128, 129, 36, 12, 11, 37, 138, 13, 139, 8, 38] and references therein. [12, 11] extended the results to the case of ergodic interference and colored Gaussian noise. [36, 37] showed that lossless interference cancelation is possible for arbitrarily varying interference, provided that the encoder and decoder share a common random dither signal. [138] showed that, as long as the interference and the noise are Gaussian (not necessarily memoryless, stationary, or ergodic), the channel has a capacity as if the interference did not exist. Various coding schemes were also provided; see e.g. [36, 37, 38]. On the other hand, if the interference is known to the transmitter only in a *causal* manner (in which case the problem is sometimes referred to as *writing on dirty tape* (WDT)), the problem is much more involved; in fact both the capacity computation problem and the capacity-achieving problem remain to be solved. See [128, 37, 129] for suboptimal coding strategies for the WDT-channels.

These results have found broad applications in information hiding [85], digital watermarking [11], precoding for ISI channels [36], and precoding for broadcast channels [8]. To summarize its significance, the dirty paper coding study has been considered to be a basic building block in both single-user and multiuser communication problems [38].

The above results are focused on the case where the encoder does not receive any feedback from the decoder. In many situations, however, it is possible for the encoder to access the information at the decoder-side in a strictly causal way, namely the encoder has feedback from the decoder. As we have shown, the availability of such feedback usually allows us to considerably simplify the coding scheme, to improve the performance, and to increase the capacity. We note that little is done to extend the dirty paper coding to feedback communication systems.

In this chapter, we consider the WDP-channel where there is a noiseless feedback from the decoder to the encoder with one-step delay. We focus on the scenario of non-causal encoder-side information about the interference (rather than the WDT-scenario). In the case of arbitrarily varying interference and AWGN, we present a Kalman filter based coding scheme to achieve the lossless interference cancelation. We then extend this result to the case of WDP-channels with both AWGN and ISI. The proposed scheme greatly simplifies the encoder/decoder design and encoding/decoding processes, guarantees doubly exponentially decay of probability of error, and is optimal in the sense of achieving the capacity. This scheme also extends the Sk codes for AWGN non-WDP-channels to WDP-channels with both AWGN and ISI.

Our study reveals the intimate connections among information, and control, and estimation over a WDP-channel. We show that the feedback communication over a WDP-channel is essentially equivalent to a Kalman filtering problem, a tracking problem, and an MEC problem. This extends the perspective of integrating information, control, and estimation to WDP-channels with feedback. Such a unifying perspective may be applicable and useful to more general feedback communication problems, such as multiuser WDP-channels and WDT-channels with feedback.

One potential application area of the research under noiseless feedback assumption is the sensor networks (as we mentioned in Section 3.1), in which the forward communication from the sensors to the base station (or the cluster center, if any) may be very noisy due to the limited power of the sensors, whereas the feedback communication from the base station to the sensors may be viewed as noiseless due to the high power of the base station. The interference to a sensor may be the signals sent by its neighboring sensors. Since the signals for neighboring sensors are usually correlated, the interference is partially known to this sensor. Our results say that we can significantly improve the forward transmission by taking advantage of both the feedback transmission and the knowledge about interference, which may be useful in sensor networks.

The rest of the chapter is organized as follows. In Section 6.2, we introduce the optimal coding scheme for a WDP-channel with AWGN and arbitrarily varying interference. In Section 6.3, we study a WDP-channel with arbitrarily varying interference, AWGN, and ISI. In Section 6.4, we provide a numerical example. Finally we summarize this chapter.

6.2 The AWGN case

Consider the power constrained WDP-channel with AWGN shown in Fig. 6.2. Let the average power budget be $\mathcal{P} > 0$. Let $\{\xi_t\}$ be an interference sequence known non-causally to the encoder. This interference sequence can be deterministic or random; we assume that

$$\frac{1}{T+1}\left(\sum_{j=0}^{T}(1+\mathcal{P})^{-\frac{j}{2}}\mathbf{E}\xi_j\right)^2 \to 0 \tag{6.2}$$

as T tends to infinity (noting that $\mathbf{E}\xi_j = \xi_j$ in the deterministic case). Let further $\{N_t\}$ be AWGN with $N_t \sim \mathcal{N}(0,1)$. We first describe the proposed coding scheme and the coding process, then present the coding theorem, followed by the proof, and finally discuss the connections to a minimum-energy control problem and a Kalman filtering problem.

6.2.1 Coding scheme

Fig. 6.2 illustrates the designed coding system, in which we can identify the encoder, decoder, and WDP-channel. Let us fix the time horizon to be $\{0, 1, \cdots, T\}$, namely the number of channel uses is $(T+1)$.

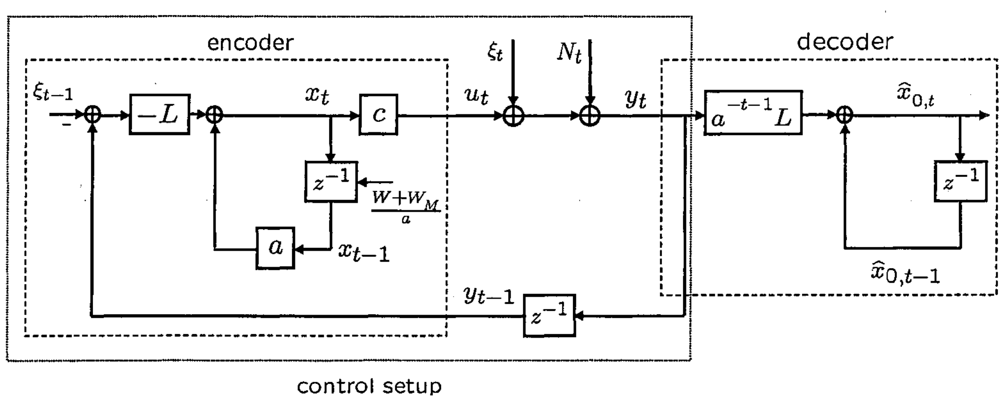


Figure 6.2 The optimal coding scheme for a WDP-channel with AWGN. The dotted box indicates a control system, referred to as the control setup.

The encoder/decoder structures

In state-space, the encoder and decoder are described as

$$\text{encoder:} \quad \begin{cases} x_t &= a x_{t-1} - L(y_{t-1} - \xi_{t-1}) \\ u_t &= c x_t \\ y_t &= u_t + \xi_t + N_t \end{cases} \tag{6.3}$$

and

$$\text{decoder:} \qquad \hat{x}_{0,t} = \hat{x}_{0,t-1} + a^{-t-1} L y_t, \tag{6.4}$$

where

$$\begin{aligned} a &:= \sqrt{1+\mathcal{P}} > 1 \\ c &:= 1 \\ L &:= a - \frac{1}{a}, \end{aligned} \tag{6.5}$$

$\xi_{-1} := 0$, $y_{-1} := 0$, $\hat{x}_{0,-1} := 0$, and x_0 will be determined shortly. We call x_t the *encoder state* and $\hat{x}_{0,t}$ the *decoder state.*

Transmission of analog source

The designed communication system can transmit either an analog source or a digital message. In the former case, the coding process is as follows. Assume without loss of generality that the to-be-conveyed message W is distributed as $\mathcal{N}(0, \mathcal{P})$ (if the variance is not $\mathcal{P}$, we can scale W to have the desired variance). To encode, let

$$x_{-1} := \frac{W + W_M}{a}, \tag{6.6}$$

i.e. $x_0 := W + W_M$, where [1]

$$W_M := -\frac{\sum_{j=0}^{T} a^{-j-1} L \xi_j}{1 - a^{-T-2}}. \tag{6.7}$$

Then run the system till instant of time T, generating $\hat{x}_{0,t}$ for $t = 0, 1, \cdots, T$. To decode, let $\hat{W}_T := \hat{x}_{0,T}$. The distortion measure is

$$\mathrm{MSE}(\hat{W}_T) := \mathbf{E}(W - \hat{W}_T)^2. \tag{6.8}$$

Transmission of digital message

To transmit digital messages over the communication system, let us fix $\epsilon > 0$ arbitrarily

[1]Another choice which is asymptotically equivalent to (6.7) is $W_M := -\sum_{j=0}^{T} a^{-j-1} L \xi_j$.

small. Suppose that we wish to transmit one of a set of

$$M_T := a^{(T+1)(1-\epsilon)} \tag{6.9}$$

messages. We equally partition the interval

$$\left[-\sqrt{\mathcal{P}}\left(1+\frac{1}{M_T-1}\right), \sqrt{\mathcal{P}}\left(1+\frac{1}{M_T-1}\right)\right] \tag{6.10}$$

into M_T sub-intervals, and map the sub-interval centers to a set of M_T equally likely messages; this is known to both the transmitter and receiver *a priori.*

Suppose now we wish to transmit the message represented by the center W. To encode, define x_{-1} according to (6.6). Then run the system till instant of time T. To decode, let the decoder estimate $\bar{W}_T$ be

$$\bar{W}_T := \frac{\hat{x}_{0,T}}{1-a^{-2T-2}}. \tag{6.11}$$

We then map $\bar{W}_T$ into the closest sub-interval center and obtain the decoded message $\hat{W}_T$. [2] We declare an error if $\hat{W}_T \neq W$, and call a (an asymptotic) rate

$$R := \lim_{T\to\infty} \frac{1}{T+1} \log M_T \tag{6.12}$$

achievable if the probability of error PE_T vanishes as T tends to infinity.

Interestingly, the coding scheme can be interpreted as follows. When the codebook, namely the mapping between the messages and initial condition x_0, has taken into account of the non-causal knowledge of the interference $\{\xi_t\}$, the encoder works *strictly causally* in the sense that at time t, the encoder only needs the knowledge of ξ_{t-1}. Whether this could be useful remains to be seen.

6.2.2 Coding theorem

Theorem 4. *Let $\{\xi_t\}$ be an arbitrarily varying interference sequence (deterministic or random) known to the encoder non-causally and satisfying (6.2), and $\{N_t\}$ be AWGN with $N_t \sim \mathcal{N}(0,1)$. Then under the power constraint $\mathbf{E}u^2 \leq \mathcal{P}$,*

[2] Another decoding method is to map $\hat{x}_{0,T}$ directly into the closest sub-interval center and obtain the decoded message $\hat{W}_T$. This is asymptotically identical to the above decoding method. The scaling by $1/(1-a^{-2T-2})$ is to remove the bias in the estimate of W; see (3.34) and [43].

i) The coding scheme constructed in Section 6.2.1 transmits an analog source $W \sim \mathcal{N}(0, \mathcal{P})$ *from the encoder to the decoder at the capacity rate*

$$C(\mathcal{P}) := \frac{1}{2}\log(1+\mathcal{P}), \tag{6.13}$$

with MSE distortion $\mathrm{MSE}(\hat{W}_T)$ *satisfying the optimal rate-distortion tradeoff function given by*

$$C(\mathcal{P}) = \frac{1}{2(T+1)}\log\frac{\mathcal{P}}{\mathrm{MSE}(\hat{W}_T)} \tag{6.14}$$

for each T.

ii) The coding scheme constructed in Section 6.2.1 can transmit digital messages from the encoder to the decoder at a rate arbitrarily close to $C(\mathcal{P})$*, with* PE_T *decays to zero doubly exponentially.*

Remark 14. *Compared with the coding scheme for AWGN non-WDP-channels, the one for AWGN WDP-channels has two main differences: The presence of "process noise" (cf. [57])* ξ_{t-1} *at the encoder, and the presence of initial condition offset* W_M*. In fact, these are the two main techniques needed to generalize the coding schemes designed for non-WDP-channels to WDP-channels. The presence of the process noise* ξ_{t-1} *at the encoder leads to that, the interference* ξ_t *does not affect the encoder state* x_t*, since the process noise cancels the interference before the interference enters the encoder state. This follows that the channel input* u_t *is not affected by the interference since* $u_t = cx_t$*. In addition, by linearity of our coding scheme, the decoder state* $\hat{x}_{0,T}$ *depends affinely on* x_0*,* ξ*, and* N*. The terms associated with* x_0 *and* ξ *are known to the encoder before the transmission because* ξ^T *is known to the encoder non-causally, and therefore the encoder can offset* x_0 *to cancel the term associated with* ξ *in* $\hat{x}_{0,T}$*, namely, the decoded message would not be affected by the interference. In summary, we can achieve lossless cancelation of interference* ξ *by applying these two techniques.*

Proof: See Appendix D.1. ∎

6.2.3 Connections to control problem, tracking problem, and estimation problem

In this subsection, we briefly discuss the equivalent relations of our coding scheme to an MEC problem, a tracking-of-unstable-source problem, and a Kalman filtering problem. These relations are conceptually appealing since they provide another example that information, control, and estimation, three fundamental concepts, can be studied in a unified framework, and if any one of these problems is solved, other problems can be solved.

The dynamics of x_t in (6.3), repeated here as

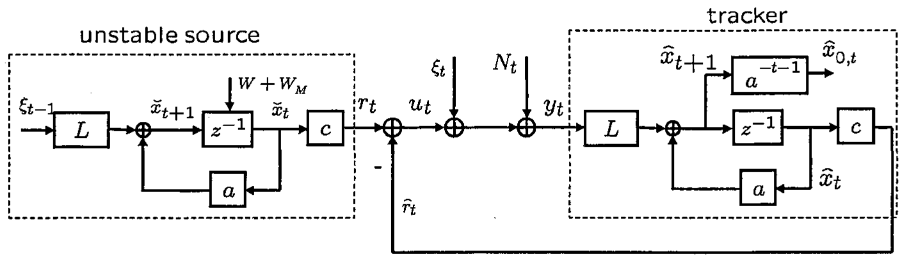
$$\begin{cases} x_t &= a x_{t-1} + L(\xi_{t-1} - y_{t-1}) \\ u_t &= c x_t \\ y_t &= u_t + \xi_t + N_t, \end{cases} \tag{6.15}$$

is indeed a control system, as indicated in Fig. 6.2. The control system is open-loop unstable with its pole at a, and is closed-loop stabilized with its pole at $1/a$, which is an MEC problem and the power of u is minimized over all possible choice of stabilizing L. Therefore, the optimality in feedback communication coincides with the optimality in control, and reliable feedback communication using (6.3) and (6.4) is equivalent to feedback stabilization of (6.15).

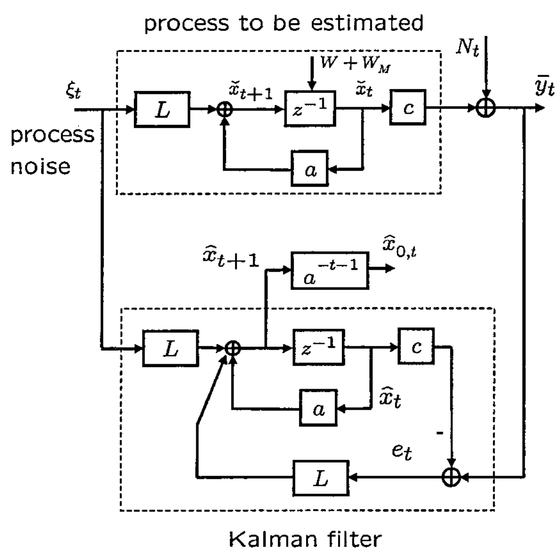
Fig. 6.3 (a) illustrates a system closely related to the coding scheme shown in Fig. 6.2. The dynamics in Fig. 6.3 (a) is

$$\begin{cases} \check{x}_{t+1} &= a\check{x}_t + L\xi_t \\ r_t &= c\check{x}_t \\ u_t &= r_t - \hat{r}_t \\ y_t &= u_t + \xi_t + N_t \\ \hat{x}_{t+1} &= a\hat{x}_t + Ly_t \\ \hat{r}_t &= c\hat{x}_t \\ \hat{x}_{0,t} &= a^{-t-1}\hat{x}_{t+1}, \end{cases} \tag{6.16}$$

where $\check{x}_0 := (W + W_M)$ and $\hat{x}_0 := 0$. To see the relation to the coding scheme, letting $x_t := \check{x}_t - \hat{x}_t$ in (6.16), we indeed obtain the dynamics of (6.3) and (6.4) through straightforward manipulation. Therefore, (6.16) generates the same channel inputs, the same channel outputs, the same decoded message as (6.3) and (6.4) do; namely, the two systems are T-equivalent. However, Fig. 6.3 (a) has an interpretation of tracking unstable source (i.e. $\{r_t\}$) over a communication channel by applying the internal mode principle. It holds that asymptotic tracking of an unstable source over a (WDP- or non-WDP-) channel is equivalent to reliable feedback communication over the same channel.



(a)



(b)

Figure 6.3 (a) Tracking of unstable source. (b) The associated Kalman filtering problem. Note that the "process noise" ξ_t enters both the process to be estimated and the Kalman filter.

We can furthermore arrange Fig. 6.3 (a) to obtain the block diagram shown in Fig. 6.3 (b), which is a Kalman filtering problem (cf. Fig. 1.1 in [57]).The dynamics of the Kalman filtering shown in Fig. 6.3 (b) is

$$\begin{cases} \check{x}_{t+1} &= a\check{x}_t + L\xi_t \\ \bar{y}_t &= c\check{x}_t + N_t \\ \hat{x}_{t+1} &= a\hat{x}_t + L\xi_t + Le_t \\ e_t &= \bar{y}_t - c\hat{x}_t \\ \hat{x}_{0,t} &= a^{-t-1}\hat{x}_{t+1}, \end{cases} \tag{6.17}$$

where $\check{x}_0 := (W + W_M)$ and $\hat{x}_0 := 0$. Note that $\check{x}_t$, $\hat{x}_t$, and e_t in (6.17) are identical to those in (6.16), respectively. That is, the two systems are T-equivalent for any T. The feedback communication problem is then linked to an estimation problem, and their optimalities coincide. We also see from Fig. 6.3 (b) that the interference sequence becomes the process noise common to both the process to be estimated and the Kalman filter. Then a well-known fact about Kalman filtering (cf. [57]) implies that the interference does not affect the estimation error $(\check{x}_t - \hat{x}_t)$ and hence the channel input $u_t = c(\check{x}_t - \hat{x}_t)$. Therefore, if we consider the offset of initial condition as an offset of the "codebook" (i.e. the partition and the assignment of messages), which takes into account of the non-causal knowledge about the interference, our dirty paper coding scheme is merely a reformulation of Kalman filtering algorithm for the case of process noise known causally to both the process and the Kalman filter. The connection of the optimal dirty paper coding scheme to the Kalman filter obtained here further confirms the optimality of the Kalman filter in the information theoretic sense. The role of the Kalman filter in more general communication setups, such as writing on dirty tape with feedback, is under current investigation.

6.3 The ISI Gaussian channel case

This section presents the optimal coding scheme for a WDP-channel with AWGN and ISI. We first describe the channel model. We then introduce the encoder/decoder structures and explain how to choose the parameters to ensure the optimality, and describe the encoding/decoding processes. Finally we present the coding theorem.

6.3.1 Channel model and feedback capacity

The WDP-channel $\mathcal{F}$ with ISI and AWGN is described in state-space as

$$\mathcal{F} : \begin{cases} s_{t+1} &= Fs_t + Gu_t \\ y_t &= Hs_t + u_t + \xi_t + N_t, \end{cases} \tag{6.18}$$

where $s_0 := 0$, $F \in \mathbb{R}^m$ is stable, (F, G) is controllable, (F, H) is observable, m is the *dimension* or *order* of $\mathcal{F}$, and ξ^T is an interference sequence known to the encoder non-causally and unknown to the decoder. See Fig. 6.4 (a) for the block diagram of $\mathcal{F}$. Through the equivalence shown in Section 4.2, the results developed for this channel hold for WDP-channels with colored Gaussian noise with rational power spectrum.

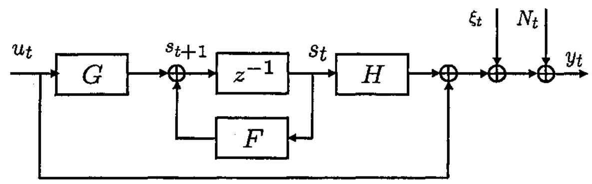


Figure 6.4 State-space representation of the WDP-channel $\mathcal{F}$.

Our focus is to find a (stationary, time-invariant) steady-state coding scheme to achieve the highest possible information rate, in other words, we wish to achieve the (stationary) asymptotic capacity, given by

$$C_\infty := C_\infty(\mathcal{P}) := \sup_{\{u_t\}} \lim_{T\to\infty} \frac{1}{T+1} I(u^T \to y^T) \tag{6.19}$$

where the supremum is over all (asymptotically) stationary input sequences $\{u_t\}$ satisfying the power constraint

$$\lim_{T\to\infty} \frac{1}{T+1} \mathbf{E} u^{T\prime} u^T \leq \mathcal{P} \tag{6.20}$$

and in the form of

$$u_t = \gamma_t u^{t-1} + \beta_t \xi^T + \eta_t y^{t-1} + \zeta_t \tag{6.21}$$

for any $\gamma_t \in \mathbb{R}^{1\times t}$, $\beta_t \in \mathbb{R}^{1\times(T+1)}$, $\eta_t \in \mathbb{R}^{1\times t}$, and zero-mean Gaussian random variable $\zeta_t \in \mathbb{R}$. Here $\mathcal{P} > 0$ is the power budget and $I(u^T \to y^T)$ is the directed information from u^T to y^T. The asymptotic capacity problem admits a finite-dimensional time-invariant solution, which has rather low design/operation complexity.

6.3.2 Coding scheme

The encoder/decoder structures

In state-space, the encoder and decoder are described as

$$\text{Encoder:} \begin{cases} x_t &= Ax_{t-1} - L_1 e_{t-1} \\ u_t &= Cx_t \\ \bar{s}_t &= F\bar{s}_{t-1} - \xi_{t-1} - H\bar{s}_{t-1} \\ e_{t-1} &= y_{t-1} - \xi_{t-1} - H\bar{s}_{t-1} \end{cases} \tag{6.22}$$

and

$$\text{Decoder:} \begin{cases} \hat{s}_{t+1} &= F\hat{s}_t + L_2\hat{e}_t \\ \hat{e}_t &= y_t - H\hat{s}_t \\ \hat{x}_{0,t} &= \hat{x}_{0,t-1} + A^{-t-1}L_1\hat{e}_t, \end{cases} \tag{6.23}$$

where $s_0 := 0$, $\bar{s}_{-1} := 0$, $\xi_{-1} := 0$, $\hat{s}_0 := 0$, $\hat{x}_{0,-1} := 0$, $A \in \mathbb{R}^{(n+1)\times(n+1)}$, $C \in \mathbb{R}^{1\times(n+1)}$, $L_1 \in \mathbb{R}^{n+1}$, and $L_2 \in \mathbb{R}^m$. We call $(n+1)$ the *encoder dimension.* See Fig. 6.5 for the block diagram. Here A, C, u_t, etc. depend on n, but we do not specify the dependence explicitly to simplify notations.

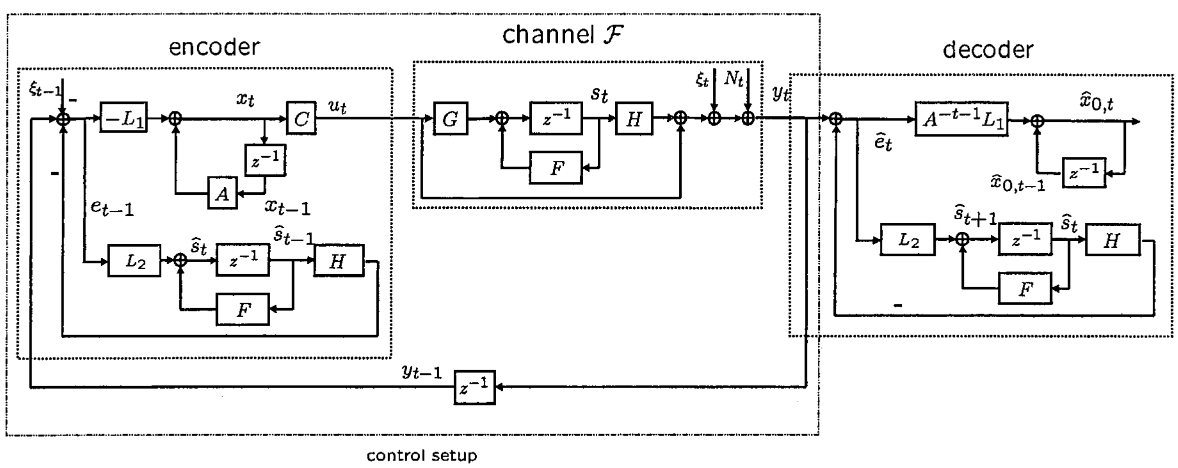


Figure 6.5 The optimal coding scheme for the WDP-channel $\mathcal{F}$.

Optimal choice of parameters

Fix a desired rate $\mathcal{R}$. Let $DI := 2^{\mathcal{R}}$ and $n := m - 1$ (recalling that m is the channel dimension), and solve the optimization problem

$$[\boldsymbol{a}_f^{opt}, \Sigma^{opt}] := \arg \inf_{\substack{\boldsymbol{a}_f \in \mathbb{R}^n \\ s.t.\ \Sigma = \mathbb{A}\Sigma\mathbb{A}' - \mathbb{A}\Sigma\mathbb{C}'\mathbb{C}\Sigma\mathbb{A}'/(\mathbb{C}\Sigma\mathbb{C}'+1)}} \mathbb{D}\Sigma\mathbb{D}', \tag{6.24}$$

where

$$\begin{aligned} \mathbb{A} &:= \left[\begin{array}{c|c} A & 0 \\ \hline GC & F \end{array}\right], \mathbb{C} := \left[\begin{array}{cc} C & H \end{array}\right], \mathbb{D} := \left[\begin{array}{cc} C & 0 \end{array}\right], \\ A &:= \left[\begin{array}{c|c} 0_{n\times 1} & I_n \\ \hline \pm DI & \boldsymbol{a}_f \end{array}\right], C := \left[\begin{array}{cc} 1 & 0_{1\times n} \end{array}\right]. \end{aligned} \tag{6.25}$$

Note that we need to solve (6.24) twice (one for $+DI$ in A and one for $-DI$ in A), and choose the optimal solution as the one with the smaller objective function value. Then we form the optimal A^{opt} based on $\boldsymbol{a}_f^{opt}$, and let (n^*+1) be the number of unstable eigenvalues in A^{opt}, where $n^* \geq 0$.

Now let $n := n^*$, solve (6.24) again, and obtain a new $\boldsymbol{a}_f^{opt}$ and Σ^{opt}. Then form A^{opt}, let $A^* = A^{opt}$, $\Sigma^* = \Sigma^{opt}$, $C^* := [1, 0_{1\times n^*}]$, and form $\mathbb{A}^*$,$\mathbb{C}^*$, and $\mathbb{D}^*$. Let

$$L^* := \begin{bmatrix} L_1^* \\ L_2^* \end{bmatrix} := \frac{\mathbb{A}^*\Sigma^*\mathbb{C}^{*\prime}}{\mathbb{C}^*\Sigma^*\mathbb{C}^{*\prime}+1}. \tag{6.26}$$

It holds that (A^*, C^*) is observable, and A^* has exactly (n^*+1) unstable eigenvalues.

We assign the encoder/decoder parameters to the scheme built in Fig. 6.5 by letting

$$n := n^*, A := A^*, C := C^*, L_1 := L_1^*, L_2 := L_2^*. \tag{6.27}$$

We then drive the initial condition s_0 of channel $\mathcal{F}$ to 0. Now we are ready to communicate at a rate $\mathcal{R}$ using power $\mathbb{D}^*\Sigma^*\mathbb{D}^{*\prime}$, which is the minimum power needed to sustain the pre-specified rate $\mathcal{R}$.

The encoding/decoding processes

Transmission of analog source

Assume that the to-be-conveyed message W is distributed as $\mathcal{N}(0, I_{n^*+1})$ (noting that any non-degenerate (n^*+1)-variate Gaussian vector W can be transformed in this form). Assume that the coding length is $(T+1)$. To encode, let

$$x_0 := W + W_M, \tag{6.28}$$

where

$$W_M := -\sum_{j=0}^{T} A^{-j-1} L_1 \xi_j + \sum_{j=0}^{T}\sum_{i=0}^{j-1} A^{-j-1} L_1 H (F - L_2 H)^{t-1-j} L_2 \xi_j. \tag{6.29}$$

Then run the system till time epoch T. To decode, let $\hat{W}_T := \hat{x}_{0,T}$. The distortion is defined as

$$\mathrm{MSE}(\hat{W}_T) := \mathbf{E}(W - \hat{W}_T)(W - \hat{W}_T)'. \tag{6.30}$$

Transmission of digital message

To transmit digital messages over the communication system, let us first fix $\epsilon > 0$ small enough and the coding length $(T+1)$ large enough. Let

$$\Sigma_x^* := [I_{n^*+1}, 0]\Sigma^*[I_{n^*+1}, 0]'. \tag{6.31}$$

Assume that the matrix $(A^{*\prime})^{-T-1}\Sigma_x^*(A^*)^{-T-1}$ has an eigenvalue decomposition as

$$(A^{*\prime})^{-T-1}\Sigma_x^*(A^*)^{-T-1} = E_T \Lambda_T E_T', \tag{6.32}$$

where $E_T = [e^{(1)}, \cdots, e^{(n^*+1)}]$ is an orthonormal matrix and Λ_T is a positive diagonal matrix. Let $\sigma_{T,i}$ be the square root of the (i,i)th element of Λ_T. Let $\mathcal{B} \in \mathbb{R}^{n^*+1}$ be the hypercube spanned by columns of E_T, that is,

$$\mathcal{B} = \left\{ \sum_{i=0}^{n^*} \alpha^{(i)} e^{(i)} \middle| \alpha^{(i)} \in [-\frac{1}{2}, \frac{1}{2}], i = 0, \cdots, n^* \right\}. \tag{6.33}$$

Next we partition the ith side of $\mathcal{B}$ into $(\sigma_{T,i})^{-(1-\epsilon)}$ segments. This induces a partition of $\mathcal{B}$ into M_T sub-hypercubes, where

$$\begin{aligned} M_T &= \prod_{i=0}^{n^*} (\sigma_{T,i})^{-(1-\epsilon)} \\ &= \left[\det\left((A^{*\prime})^{-T-1}\Sigma^*(A^*)^{-T-1}\right)\right]^{-\frac{1-\epsilon}{2}}. \end{aligned} \tag{6.34}$$

We then map the sub-hypercube centers to a set of M_T equally likely messages. The above procedure is known to both the transmitter and receiver *a priori.*

Suppose now we wish to transmit the message represented by the center W. To encode, define x_0 according to (6.28). Then run the system till time epoch T. To decode, we map $\hat{x}_{0,T}$ into the closest sub-hypercube center and obtain the decoded message $\hat{W}_T$. We declare an error if $\hat{W}_T \neq W$, and call a (an asymptotic) rate

$$R := \lim_{T\to\infty} \frac{1}{T+1} \log M_T \tag{6.35}$$

achievable if the probability of error PE_T vanishes as T tends to infinity.

As we can see, the encoder/decoder design and the encoding/decoding processes are rather simple. The computation complexity involved in coding grows as $O(T+1)$.

6.3.3 Coding theorem

Theorem 5. *Construct the encoder/decoder shown in Fig. 6.5 using n^*, A^*, C^*, L_1^*, L_2^*, and W_M. Then under the average power constraint $\mathbf{E}u^2 \leq \mathcal{P}$,*

i) The coding scheme transmits an analog source $W \sim \mathcal{N}(0, I_{n^*+1})$ from the encoder to the decoder at rate $C_\infty(\mathcal{P})$, with MSE distortion $\mathrm{MSE}(\hat{W}_T)$ achieving the optimal asymptotic rate-distortion tradeoff given by

$$C_\infty(\mathcal{P}) = \lim_{T\to\infty} \frac{1}{2(T+1)} \log \frac{1}{\det \mathrm{MSE}(\hat{W}_T)}. \tag{6.36}$$

ii) The coding scheme can transmit digital messages from the encoder to the decoder at a rate arbitrarily close to $C_\infty(\mathcal{P})$, with PE_T decaying to zero doubly exponentially.

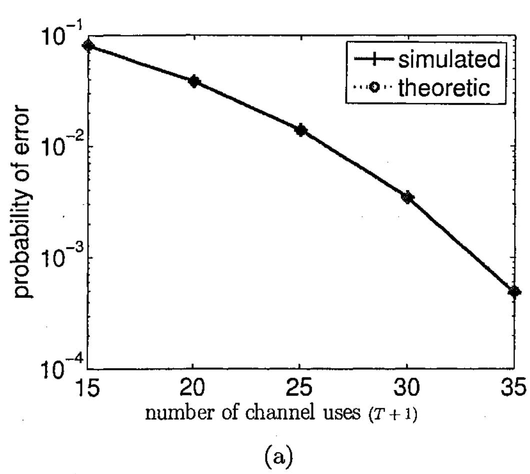
Remark 15. *The main idea of this dirty paper coding scheme is still the two techniques used for the AWGN WDP-channel: The presence of process noise ξ_{t-1} at the encoder, and the presence of initial condition offset W_M which can be viewed as an offset of the "codebook". In addition, it still holds that the feedback communication problem is equivalent to an MEC problem, tracking-of-unstable-source problem, and Kalman filtering problem; details are skipped for brevity.*

Proof: See Appendix D.2. ∎

6.4 Numerical example

In this section, we provide a numerical example of the proposed optimal coding scheme for AWGN WDP-channel. Assume that the power budget is $\mathcal{P} := 3$, that is, we may achieve any rate equal to $C(\mathcal{P}) - \epsilon = 1 - \epsilon$ bit per channel use for any $\epsilon > 0$. Let $\epsilon := 0.05$, i.e. the desired rate is $R = 0.95C(\mathcal{P}) = 0.95$ bit per channel use. Then we can construct the coding scheme according to Section 6.2.1.

Simulation shows that R is indeed achieved since the decoder can distinguish among $M_T := a^{(T+1)(1-\epsilon)}$ messages with the probability of error decaying to zero (in a doubly exponential fashion); see Fig. 6.6 (a). Due to the fast decay of probability of error, the coding length (and hence the coding delay) can be rather short (within 100) to attain a good performance. The shortened coding length also implies that the preview of the interference ξ^T can be short. Additionally, the average channel input power converges to $\mathcal{P}$, see Fig. 6.6 (b). For short coding length, however, the consumed power can be slightly different from the given power budget, since most of our analysis holds asymptotically. Fig. 6.6 (b) also shows the convergence of the decoder estimate $\bar{W}_t$ to the message W.



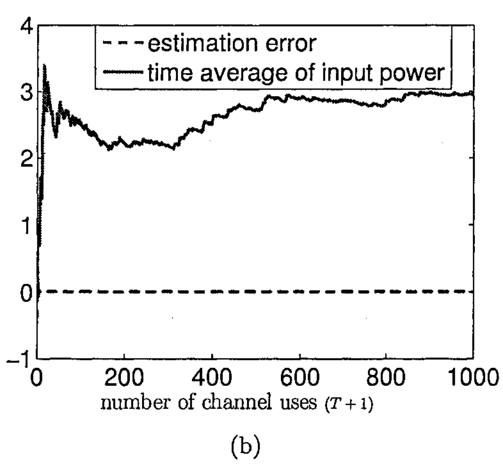


Figure 6.6 (a) Simulated probability of error and theoretic probability of error. (b) Convergence of the average channel input power, and vanishing of estimation error $(\bar{W}_T - W)$.

6.5 Summary

In this chapter, we have proposed capacity-achieving coding schemes for a WDP-channel with AWGN and for a WDP-channel with AWGN and ISI, both having noiseless output feedback. The interference is assumed to be known to the encoder non-causally before the transmission and unknown to the decoder, and can be arbitrarily varying, deterministic or random. We achieved lossless interference cancelation, that is, we canceled the interference without extra power overhead or rate loss. We exhibited the connections among feedback communication, feedback control, and estimation over a WDP-channel. In establishing lossless interference cancelation, we developed techniques which may be readily applied to more general WDP-channels. We also discussed the potential usefulness of our feedback communication schemes to sensor networks, in which we may wish to take advantage of the high-quality feedback link and the knowledge of interference to improve the forward communication.

CHAPTER 7. CONCLUSIONS AND FUTURE DIRECTIONS

7.1 Thesis summary

In the past few years, considerable efforts have been devoted to the subject of interactions between information and control, due to its theoretical and practical importance in the study of control under communication constraints, feedback information theory, networked systems, etc.

This thesis is an addition to this subject. In this thesis, we have explicitly brought *estimation* into the existing picture, and established a new perspective of unifying communication, estimation, and control. We have demonstrated that this perspective arises naturally and is powerful in solving important problems in feedback information theory. In particular, we designed the Kalman filter based coding schemes to achieve the feedback capacity of AWGN channels, frequency-selective fading Gaussian channels, time-selective fading Gaussian channels, and WDP-channels with Gaussian noises. These feedback coding schemes are equivalent to Kalman filtering problems and MEC problems, and their optimality and fundamental limitations coincide. We also obtained a new information theoretic characterization of Kalman filtering, a new formula connecting mutual information and CRB or MMSE, connection between time-selective fading channels and Markov jump linear control, and so on. We anticipate that the perspective, techniques, and results developed in this thesis could be extended to more general scenarios and helpful in establishing a theoretically and practically sound framework that unifies information, estimation, and control.

7.2 Future research directions

Research will continue on the following aspects:

1. Noisy feedback. As we pointed out in Section 2.2.5, the noiseless feedback assumption leads to the major limitation of feedback information theory. It is then necessary to

study the more realistic problem of communication with noisy feedback, after the ideal problem of communication with noiseless feedback has been resolved. We remark that by far we are not aware of any meaningful positive results regarding noisy feedback in the literature. Negative results include that, in the noisy feedback case, the SK coding scheme and its extensions leads to non-vanishing probability of error for a fixed signalling rate, or leads to a vanishing signalling rate if the probability of error decays to zero, as the coding length tends to infinity.

One way out may be to realize that we do not need infinite coding length; on the contrary, due to the fast decay of probability of error, a very short coding length can be used, and that very short coding length may correspond to small enough probability of error *and* high enough signalling rate. This would require a careful comparison of the resultant rate and performance to the feedforward case, to see if there is any improvement using the noisy feedback, and if there is, how much improvement we can get.

Another way is to, before the channel output is processed or sent back, we quantize it with high precision. Then the feedback link can be viewed as digital and noiseless when transmitting the digitalized channel output. The price we need to pay is that now the quantization error is further added to the channel output, namely the effective channel noise becomes the Gaussian noise plus the quantization error, resulting in a new effective non-Gaussian channel with noiseless feedback. It is known that the SK-like schemes work for non-Gaussian noise channels, so it is possible that we obtain a constant signalling rate with vanishing probability of error. The technical difficulty is that now the effective channel noise is correlated with the message, channel input, and noises at other times, which complicates the analysis. If we invoke the equivalence between information and control, this may necessitate a study of control over noisy channels together with quantization, a problem not investigated to the best of our knowledge.

We also conjecture that if the message arrives at the encoder in a streaming fashion, we may have a better chance to fight against the feedback noise. This will also be studied. We expect the noisy feedback problem to be very challenging, but any progress in this regard may have high theoretic and practical significance. Theoretically, the noisy feedback case bridges the noiseless feedback case (related to standard Kalman filtering) and the pure feedfoward communication case (related to the sum-product algorithm, a generalization of the Kalman filtering algorithm). Practically, it leads to feasible feedback communication solutions.

2. Reducing the computation complexity in solving the optimization problem (4.11) for Gaussian channels with memory. The optimization (4.11) has $(m-1)$ free parameters to be searched (where m is the order of the channel) and is not convex. When the channel order is not too high, it can be solved with affordable computation cost using MATLAB. However, when the channel order is very high, say $m = 100$, directly solving (4.11) would be very difficult, if not possible.

   There are several possible ways to reduce the computation complexity for (4.11). First, we may try model reduction for high order channels, which reduces the channel order significantly without affecting much of the input-output behavior of this channel, and then we can solve with lower computation complexity the feedback capacity for the reduced-order channel. The feedback capacity for the reduced-order channel is expected to be close enough to that for the original channel; a rigorous proof will be provided for this statement.

   Second, we may explore the properties of the optimal coding schemes for Gaussian channels with memory. In our simulations, we have observed many interesting properties. For example, we have seen that a first-order encoder is sufficient to achieve the feedback capacity in very high SNR and very low SNR regimes, no matter how high the channel order is. This may be related to the fact that feedback does not improve the capacity in very high SNR and very low SNR regimes [20], namely the channel behaves as if it is memoryless. In many cases, we also observed that the optimal encoder dimension can be significantly smaller than the channel order. Hence, it may be interesting to quantify precisely the optimal encoder dimension, which may require a careful study of the underline Riccati recursions. It is also interesting to study the locations of the unstable eigenvalues in a optimal coding scheme.

   Third, we may try to convexify the optimization problem (4.11). Numerous convexifying techniques exist in the linear matrix inequalities (LMI) study. So far our efforts to convexifying (4.11) has produced little progress; in fact, we have found that this problem is equivalent to the structured static output feedback problem, which remains open (though there is a possibility that some structured static output feedback can be convexified). On the other side, the problem of finite-horizon feedback capacity has been shown to be convex (with $O((T+1)^2)$ free variables, unfortunately) [124]. This implies that (4.11) is also equivalent to an infinite-dimensional convex optimization problem; whether we can reduce it to a finite-dimensional convex problem remains to be seen. We remark that any progress in this regard would also have high impact on the computational methods of controller design.

3. Error exponents. In Gaussian channels with memory and feedback, the probability of error decays doubly exponentially, and we would like to compute the optimal error exponents defined in (3.39) which determines the decay rate of probability of error. Error exponents have been investigated in [113] for feedback communication systems; however, how to compute them is not available in the literature. The connection between MMSE (or CRB) and probability of error established in this thesis may be crucial. Thanks to that connection, it is expected that the decay rate of CRB is linked to the error exponent, and solving the optimal error exponent may boil down to a finite-dimensional optimization problem, similar to the feedback capacity problem. This is under current investigation.

4. MIMO channels with feedback. It is believed that feedback can have important advantages in communication networks. With the problems of SISO Gaussian channels with feedback and WDP-channels with feedback being resolved, we are now ready to attack the problem of MIMO channels with feedback; note that the broadcast channels, a class of MIMO channels, are closely related to the WDP-channels. We anticipate that the connections among information, estimation, and control, extended properly, will play fundamental roles in this research. Besides, it is worth noting that, though in the SISO case, the optimal channel input does not require a feedforward term, the MIMO case usually needs a mixture of feedback and feedforward signalling. Such a feedforward signalling may be similar to the streaming source used in anytime encoder, but it can also be easily incorporated in a Kalman filtering framework as the process noise. Hence, the Kalman filter based approach developed in this thesis may be found useful in studying MIMO problems.

5. Writing on dirty tape with feedback. In a WDT-problem, the interference is known to the encoder *causally* and unknown to the decoder. This problem is much more complicated than the WDP-problem, and in the feedforward case, both the capacity computation problem and capacity-achieving problem remain largely open. However, this problem has important applications in precoding, watermarking, and data hiding. We wish to study the WDT-problem with feedback, due to its importance and the fact that it is the natural extension of the WDP-problem with feedback. This may be related to the streaming communication problem, because the information about the interference arrives at the encoder causally as a stream.

6. Cooperation with limited information. We conjecture that information transmission /processing and feedback are central to generating cooperative behavior. Without information transmission/processing or feedback, there is no cooperation or cooperative behavior. However, in a cooperative system with multiple agents, it is not easy to separate the notions of information and control: The transmission of information can be viewed as feedback among agents; information processing is intertwined with estimation and detection; and the processed information is utilized for decision making and generating control commands. Therefore, it is necessary and advantageous to study information, estimation, and control *jointly* in a cooperative system. We would like to see the broader impact of our study on cooperation problems.

   Our preliminary results have already shown that information exchange considerably affects the cooperative behavior [75]. Roughly speaking, no cooperative behavior occurs if the amount of information exchange is limited below some level, and cooperative behavior emerges otherwise. Noise, as a measure of limited information, was introduced in our model, in contrast to noiseless models in the literature [56]. We emphasize that the introduction of noise is important for studying the interplay between information/feedback and the cooperative behavior, since it allows us, first, to keep track of how information is conveyed, processed, and utilized in such a system, and second, to understand how local information exchange may be used to generate global behaviors.

   Particularly, we may proceed as follows. We wish to obtain an explicit characterization of information flow in our multiple-agent system, based on the information flow concepts developed in [83] or [130]. We may also generalize [75] to multiple hypothesis testing scenario in which richer system behavior is possible; note that [75] concerns about a binary hypothesis testing problem. The recent work on distributed hypothesis testing in [90] has been found relevant. We will test the idea of if the convergence to the right hypothesis has anything to do with stability of information exchange.

7. Other directions. We may also pursue along other directions. For one example, we can study the time-selective fading Gaussian channels whose channel state takes continuous values. This would include the Rayleigh fading channels, Nakagami fading channels, and Rician fading channels as special cases. We envision that our results in Chapter 5 extends to this case, with probably an argument of partitioning the channel state space and a justification of switching two limits, as was done in [47]. For another example, we may propose an alternative dynamical programming based solution to compute the finite-horizon feedback capacity, with the advantage over [136] that the obtained optimizing parameters can be directly used to construct the optimal finite-horizon time-varying coding scheme. We may also attempt to confirm the stationarity conjecture in an alternative way, preferably based on connections among information, estimation, and control. This is of independent interest to the study of many estimation/control problems, dynamic programming problems, and Markov decision problems; if such an alternative method does exist, we may be able to establish that the optimal solutions in these problems are stationary. Possible approaches are listed as follows. First, we may investigate if the double indexed sequence $C_{T,n}$ satisfies the uniform convergence property based on analysis of Riccati equations; if confirmed, then we may be able to prove that $\lim_{n\to\infty} \lim_{n\to\infty} C_{T,n}$ equals the limit of $C_{T,T}$, namely the asymptotic feedback capacity C_∞. Second, we may try to find a solution to the associated Bellman equation, since the existence of such a solution implies the stationarity conjecture; note that the cost function is already known but the cost-to-go function needs to be identified. Third, noting that the dynamical programming problem associated to this feedback capacity problem is singular, we may study a sequence of perturb dynamical problems which are: 1) regular, 2) admit stationary optimal solutions, and 3) approaching the singular problem. This may also prove the stationarity conjecture. Finally, we remark that we are currently investigating the extension of the formula linking mutual informaiton and MMSE obtained by Guo, Shamai,

and Verdu in [51] to more general settings, as indicated in Section 4.5.

APPENDIX A. SYSTEMS REPRESENTATIONS AND EQUIVALENCE

The concept of system representations and the equivalence between different representations are extensively used in this paper. In this subsection, we briefly introduce system representations and the equivalence. For more thorough treatment, see e.g. [93, 9, 18].

A.1 Systems representations

Any discrete-time linear system can be represented as a linear mapping (or a linear operator) from its input space to output space; for example, we can describe a SISO linear system as

$$y^t = \mathcal{M}_t u^t \tag{A.1}$$

for any t, where $\mathcal{M}_t \in \mathbb{R}^{(t+1)\times(t+1)}$ is the matrix representation of the linear operator, $u^t \in \mathbb{R}^{t+1}$ is the stacked input vector consisting of inputs from time 0 to time t, and $y^t \in \mathbb{R}^{t+1}$ is the stacked output vector consisting of outputs from time 0 to time t. For a (strictly) causal SISO LTI system, $\mathcal{M}_t$ is a (strictly) lower-triangular Toeplitz matrix formed by the coefficients of the impulse response. Such a system may also be described as the (reduced) transfer function, whose inverse z-transform is the impulse response; by a (reduced) transfer function we mean that its zeros are not at the same location of any pole.

A causal SISO LTI system can be realized in state-space as

$$\begin{cases} x_{t+1} &= Ax_t + Bu_t \\ y_t &= Cx_t + Du_t, \end{cases} \tag{A.2}$$

where $x_t \in \mathbb{R}^l$ is the state, $u_t \in \mathbb{R}$ is the input, and $y_t \in \mathbb{R}$ is the output. We call l the *dimension* or the *order* of the realization. The state-space representation (A.2) may be denoted as (A, B, C, D). Note that in the study of input-output relations, it is sometimes convenient to assume that the system is relaxed or at initial rest (i.e. zero input leads to zero output), whereas in the study of state-space, we generally allow $x_0 \neq 0$, which is not at initial rest. For MIMO systems, linear time-varying systems, etc., see [9, 18].

The state-space representation of an causal FDLTI system $\mathcal{M}(z)$ is not unique. We call a realization (A, B, C, D) *minimal* if (A, B) is controllable and (A, C) is observable. All minimal realizations of $\mathcal{M}(z)$ have the same dimension, which is the minimum dimension of all possible realizations. All other realizations are called *non-minimal.*

An example

We demonstrate here how we can derive a minimal realization of a system. Consider $\mathcal{G}_T^*(A, C)$ in (4.46) in Section 4.4, which is given by

$$\mathcal{G}_T^*(A, C) = -\widehat{\mathcal{G}}_T^*(I - \mathcal{Z}_T^{-1}\widehat{\mathcal{G}}_T^*)^{-1}, \tag{A.3}$$

where the state-space representations for $\widehat{\mathcal{G}}_T^*(A, C)$ and $\mathcal{Z}_T^{-1}$ are illustrated in Fig. 4.8 (b) and Fig. 4.1 (c). Since (A.3) suggests a feedback connection of $\widehat{\mathcal{G}}^*$ and $\mathcal{Z}^{-1}$ as shown in Fig. A.1, we can write the state-space for $\mathcal{G}^*$ as

$$\left\{ \begin{array}{lcl} \hat{x}_{t+1} & = & A\hat{x}_t + L_{1,t}e_t \\ \hat{r}_t & = & C\hat{x}_t \\ \hat{\bar{s}}_{t+1} & = & F\hat{\bar{s}}_t + G\hat{r}_t + L_{2,t}e_t \\ e_t & = & \bar{y}_t - H\hat{\bar{s}}_t - \hat{r}_t \\ s_{a,t+1} & = & Fs_{a,t} + G\hat{r}_t \\ \bar{y}_t & = & y_t + Hs_{a,t} + \hat{r}_t. \end{array} \right. \tag{A.4}$$

Then let $\hat{s}_t := \hat{\bar{s}}_t - s_{a,t}$, and we have

$$\left\{ \begin{array}{lcl} \hat{x}_{t+1} & = & A\hat{x}_t + L_{1,t}e_t \\ \hat{r}_t & = & C\hat{x}_t \\ \hat{s}_{t+1} & = & F\hat{s}_t + L_{2,t}e_t \\ e_t & = & y_t - H\hat{s}_t. \end{array} \right. \tag{A.5}$$

It is straightforward to check that this dynamics is controllable and observable, and therefore it is a minimum realization of $\mathcal{G}^*$.

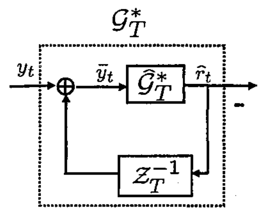


Figure A.1 $\mathcal{G}^*$ is a feedback connection of $\widehat{\mathcal{G}}^*$ and $\mathcal{Z}^{-1}$.

A.2 Equivalence between representations

Definition 5. *i) Two FDLTI systems represented in state-space are said to be* equivalent *if they admit a common transfer function (or a common transfer function matrix) and they are both stabilizable and detectable.*

*ii) Fix $0 \leq T < \infty$. Two linear mappings $\mathcal{M}_{i,T} : \mathbb{R}^{q(T+1)} \to \mathbb{R}^{p(T+1)}$, $i = 1, 2$, both at initial rest, are said to be T-*equivalent *if for any $u^T \in \mathbb{R}^{q(T+1)}$, it holds that*

$$\mathcal{M}_{1,T}(u^T) = \mathcal{M}_{2,T}(u^T). \tag{A.6}$$

We note that i) is defined for FDLTI systems, whereas ii) is for general linear systems. i) implies that, the realizations of a transfer function are not necessarily equivalent. However, if we focus on all realizations that do not "hide" any unstable modes, namely all the unstable modes are either controllable from the input or observable from the output, they are equivalent; the converse is also true. ii) concerns about the *finite-horizon* input-output relations only. Since the states are not specified in ii), it is not readily extended to infinite horizon: Any unstable modes "hidden" from the input and output will grow unboundedly regardless of input and output, which is unwanted.

Examples

As we mentioned in Section 4.2.2, for any u^T and N^T, Fig. 4.1 (a) and (b) generate the same channel output $\tilde{y}^T$. That is, the mappings from (u^T, N^T) to $\tilde{y}^T$ for the two channels are identical, and both are given by

$$\tilde{y}^T = \mathcal{Z}_T(\mathcal{Z}_T^{-1}u^T + N^T). \tag{A.7}$$

Thus, we say the two channels are T-equivalent.

The feedback communication system (4.55), estimation system (4.54), and control system (4.58) are T-equivalent, since for any N^T, they generate the same innovations e^T.

APPENDIX B. PROOFS OF RESULTS IN CHAPTER 4

B.1 Proof of Proposition 5: Necessary condition for optimality

In this subsection, we show that our general coding structure, in the form of (4.58), satisfies the necessary condition for optimality as presented in Proposition 5.

Since $\{y_t\}$ is interchangeable with the innovations process $\{e_t\}$, in the sense that they determine each other causally and linearly, it suffices to show that $\mathbf{E}u_t e_\tau = 0$. Note that

$$u_t = \mathbb{D}\mathbb{X}_t = \mathbb{D}\mathbb{A}\mathbb{X}_{t-1} - \mathbb{D}L_{t-1}e_{t-1}, \tag{B.1}$$

and thus

$$\begin{aligned} \mathbf{E}u_t e_{t-1} &= \mathbf{E}\mathbb{D}\mathbb{A}\mathbb{X}_{t-1}e_{t-1} - \mathbb{D}L_{t-1}K_{e,t-1} \\ &\overset{(a)}{=} \mathbf{E}\mathbb{D}\mathbb{A}\mathbb{X}_{t-1}\mathbb{X}_{t-1}'\mathbb{C}' + \mathbf{E}\mathbb{D}\mathbb{A}\mathbb{X}_{t-1}N_{t-1} - \mathbb{D}\mathbb{A}\Sigma_{t-1}\mathbb{C}' \\ &= \mathbb{D}\mathbb{A}\Sigma_{t-1}\mathbb{C}' + 0 - \mathbb{D}\mathbb{A}\Sigma_{t-1}\mathbb{C}' = 0, \end{aligned} \tag{B.2}$$

where (a) follows from (4.58) and (4.59). Similarly we can prove $\mathbf{E}u_t e_\tau = 0$ for any $\tau < t-1$.

B.2 Proof of Proposition 6: Convergence to steady-state

In this subsection, we show that system (4.58) converges to a steady-state, as given by (4.71). To this aim, we first transform the Riccati recursion into a new coordinate system, then show that it converges to a limit, and finally prove that the limit is the unique stabilizing solution of the DARE. The convergence to the steady-state follows immediately from the convergence of the Riccati recursion.

Consider a coordinate transformation given as

$$\underline{\mathbb{A}} := \Phi\mathbb{A}\Phi^{-1} := \begin{bmatrix} A & 0 \\ 0 & F \end{bmatrix}, \ \underline{\mathbb{C}} := \mathbb{C}\Phi^{-1}, \ \underline{\Sigma}_t := \Phi\underline{\Sigma}_t\Phi', \tag{B.3}$$

where

$$\Phi := \begin{bmatrix} I_{n+1} & 0 \\ -\phi & I_m \end{bmatrix}, \tag{B.4}$$

and ϕ is the unique solution to the Sylvester equation

$$F\phi - \phi A = -GC. \tag{B.5}$$

Note that the existence and uniqueness of ϕ is guaranteed by the assumption on A that $\lambda_i(-A) + \lambda_j(F) \neq 0$ for any i and j (see Section 4.4.2).

This transformation transforms $\mathbb{A}$ into block-diagonal form with the unstable and stable eigenvalues in different blocks, and transforms the initial condition Σ_0 to

$$\underline{\Sigma}_0 := \Phi \begin{bmatrix} I_{n+1} & 0 \\ 0 & 0 \end{bmatrix} \Phi' = \begin{bmatrix} I & -\phi' \\ -\phi & \phi\phi' \end{bmatrix}. \tag{B.6}$$

Therefore, the convergence of (4.60) with initial condition Σ_0 is equivalent to the convergence of

$$\underline{\Sigma}_{t+1} = \underline{\mathbb{A}\Sigma_t\mathbb{A}'} - \frac{\underline{\mathbb{A}\Sigma_t\mathbb{C}'\mathbb{C}\Sigma_t\mathbb{A}'}}{\underline{\mathbb{C}\Sigma_t\mathbb{C}'} + 1} \tag{B.7}$$

with initial condition $\underline{\Sigma}_0$. By [44], $\underline{\Sigma}_t$ would converge if

$$\det\left(\begin{bmatrix} 0 & 0 \\ 0 & I_m \end{bmatrix} - \underline{\Sigma}_0 \begin{bmatrix} I_{n+1} & 0 \\ 0 & X_{22} \end{bmatrix}\right) \neq 0, \tag{B.8}$$

where X_{22} is a positive semi-definite matrix (whose value does not affect our result here). Since

$$\begin{aligned} \det\left(\begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} - \begin{bmatrix} I & -\phi' \\ -\phi & \phi\phi' \end{bmatrix}\begin{bmatrix} I & 0 \\ 0 & X_{22} \end{bmatrix}\right) &= \det\left(\begin{bmatrix} -I & \phi' X_{22} \\ \phi & I - \phi\phi' X_{22} \end{bmatrix}\right) \\ &= \det(-I)\det\left(I - \phi\phi' X_{22} + \phi\phi' X_{22}\right) \\ &\neq 0, \end{aligned} \tag{B.9}$$

we conclude that $\underline{\Sigma}_t$ converges to a limit $\underline{\Sigma}_\infty$.

This limit $\underline{\Sigma}_\infty$ is a positive semi-definite solution to

$$\underline{\Sigma}_\infty = \underline{\mathbb{A}\Sigma_\infty\mathbb{A}'} - \frac{\underline{\mathbb{A}\Sigma_\infty\mathbb{C}'\mathbb{C}\Sigma_\infty\mathbb{A}'}}{\underline{\mathbb{C}\Sigma_\infty\mathbb{C}'} + 1}. \tag{B.10}$$

By [57], (B.10) has a unique stabilizing solution because $(\underline{\mathbb{A}}, \underline{\mathbb{C}})$ is observable and $\underline{\mathbb{A}}$ does not have any eigenvalues on the unit circle. Therefore, $\underline{\Sigma}_\infty$ is this unique stabilizing solution, which can be computed from (B.10) as (see also [44])

$$\underline{\Sigma}_\infty = \begin{bmatrix} \underline{\Sigma}_{11} & 0 \\ 0 & 0 \end{bmatrix} \tag{B.11}$$

where $\underline{\Sigma}_{11}$ is the positive-definite solution to a reduced order DARE

$$\underline{\Sigma}_{11} = A\underline{\Sigma}_{11}A' - \frac{A\underline{\Sigma}_{11}(C+H\phi)'(C+H\phi)\underline{\Sigma}_{11}A'}{(C+H\phi)\underline{\Sigma}_{11}(C+H\phi)'+1}. \tag{B.12}$$

and has rank $(n+1)$ (cf. [44]). Thus, Σ_t converges to

$$\Sigma_\infty = \begin{bmatrix} \underline{\Sigma}_{11} & \underline{\Sigma}_{11}\phi' \\ \phi\underline{\Sigma}_{11} & \phi\underline{\Sigma}_{11}\phi' \end{bmatrix} \tag{B.13}$$

with rank $(n+1)$.

B.3 Proof of Lemma 3: Convergence of C_T

It suffices to prove the lemma for the colored Gaussian noise case, since the other channel model is equivalent to this channel. We first show that the limit exist and is finite, and then show that the limit is indeed C_∞.

The proof of the first claim is similar to [60]. Essentially we show that the sequence $\{(T+1)C_T\}$ is superadditive, namely for any k and l

$$(k+l+2)C_{k+l+1} \geq (k+1)C_k + (l+1)C_l, \tag{B.14}$$

which would follow that

$$\lim_{T\to\infty} C_T = \sup_T C_T \tag{B.15}$$

by Fekete's Lemma [100]. Since $\sup_T C_T$ is finite (by the finiteness of $C_{T,ff}$, the finite-horizon feedforward capacity, and by the uniform boundedness of $C_T - C_{T,ff}$, cf. [17, 116]), the limit of C_T exists and is finite.

To see the superadditivity, fix any integers k and l. Assume that the message w and the corresponding channel input u^k generated by an encoding function $u_t(w, y^{t-1})$, $t = 0, 1, \cdots, k$, attains C_k, where y^k is the channel output induced by u^k. Moreover, assume that the message $\breve{w}$ and the corresponding channel input $\breve{u}^l$ generated by an encoding function $\breve{u}_t(\breve{w}, \breve{y}^{t-1})$, $t = 0, 1, \cdots, l$, attains C_l, where $\breve{y}^l$ is the channel output induced by $\breve{u}^l$. That is,

$$\begin{aligned} I(w; y^k) &\overset{(a)}{=} I(u^k \to y^k) &&= (k+1)C_k \\ I(\breve{w}; \breve{y}^l) &= I(\breve{u}^l \to \breve{y}^l) &&= (l+1)C_l, \end{aligned} \tag{B.16}$$

where (a) follows from [116].

Now we consider the horizon $(k+l+1)$, that is, time index i runs from 0 to $(k+l+1)$. In the times when $i = 0, 1, \cdots, k$, we use w and $u_t(w, y^{t-1})$, which generates channel output y^k; in the times when $i = k+1, \cdots, k+l+1$, we use $\breve{w}$ and $\breve{u}_t(\breve{w}, \breve{y}^{t-1})$, which generates channel output $\breve{y}^l$ (by stationarity of the colored Gaussian noise). Namely the input and output are

$$\begin{aligned} U^{k+l+1} &:= [u^{k\prime}, \breve{u}^{l\prime}]' \\ Y^{k+l+1} &:= [y^{k\prime}, \breve{y}^{l\prime}]'. \end{aligned} \tag{B.17}$$

Note that since u^k and $\breve{u}^l$ satisfy the power constraints, U^{k+l+1} also satisfies the power constraint.

Then we compute the mutual information between the message $W := [w, \breve{w}]$ and the stacked output Y^{k+l+1} as follows, assuming that w and $\breve{w}$ are independent.

$$\begin{aligned} I(W; Y^{k+l+1}) &= h(W) - h(W|Y^{k+l+1}) \\ &= h(w) + h(\breve{w}) - h(w|y^k, \breve{y}^l) - h(\breve{w}|w, y^k, \breve{y}^l) \\ &\geq h(w) + h(\breve{w}) - h(w|y^k) - h(\breve{w}|\breve{y}^l) \\ &= I(w; y^k) + I(\breve{w}; \breve{y}^l) \\ &= (k+1)C_k + (l+1)C_l. \end{aligned} \tag{B.18}$$

This further implies that $\{(T+1)C_T\}$ is superadditive. Then $\{C_T\}$ converges, and we denote the limit by C.

Now we show that $C_\infty = C$. Note first that $C_\infty \geq C$ since C is achievable and since C_∞ is the maximum achievable rate. To be more precise, recall that for any $\epsilon > 0$, there exists a sequence of $(2^{T(C_T - \epsilon/2)}, T)$ codes with vanishing PE_T as T tends to infinity [17]. For the same

ϵ, we can find T_0 such that for every $T > T_0$, it holds that

$$C - C_T = \frac{\epsilon}{2} - \delta_T \tag{B.19}$$

for some $0 \leq \delta_T \leq \epsilon/2$. Therefore, there exists a sequence of $(2^{T(C+\delta_T-\epsilon)}, T)$ codes, $T > T_0$, with vanishing PE_T as T tends to infinity. So C is achievable. On the other hand, it has been shown in [17] that, any sequence of $(2^{TR_T}, T)$ codes with vanishing PE_T must have

$$R_T < C_T + \epsilon_T \tag{B.20}$$

with $\epsilon_T > 0$ decaying to zero. Since $C \geq C_T$, we have

$$R_T < C + \epsilon_T. \tag{B.21}$$

Letting T go to infinity, we see that any achievable asymptotic rate R must hold $R < C$. However, C_∞ is achievable [116], so $C_\infty \leq C$. Thus we complete the proof.

B.4 Proof of Proposition 8: $K_\mathcal{E} = 0$

In this section, we prove that $K_\mathcal{E}$ has to be 0 to ensure the optimality in (4.85).

We first derive some properties of the communication system using the stationary GM inputs and the steady-state Kalman filtering. The system dynamics is given by

$$\begin{cases} u_t &= d'\tilde{s}_{s,t} + \mathcal{E}_t \\ s_{t+1} &= Fs_t + Gu_t \\ y_t &= Hs_t + N_t + u_t \\ \tilde{s}_{s,t+1} &= s_t - \hat{s}_{s,t} \\ \hat{s}_{s,t+1} &= F\hat{s}_{s,t} + L_s e_t \\ e_t &= y_t - H\hat{s}_{s,t} = (H + d')\tilde{s}_{s,t} + \mathcal{E}_t + N_t \\ \tilde{s}_{s,t+1} &= F\tilde{s}_{s,t} + Gu_t - L_s e_t, \end{cases} \tag{B.22}$$

where $\hat{s}_{s,0} = 0$ and $\tilde{s}_{s,0} = 0$. As before, the Kalman filter innovations $\{e_t\}$ will play an important role. The innovations process is white with variance asymptotically equal to

$$K_e = 1 + K_{\mathcal{E}} + (H + d')\Sigma_s(H + d')', \tag{B.23}$$

where $\Sigma_s := \mathbf{E}\tilde{s}_s\tilde{s}_s'$. Following the same derivation for Proposition 7, we know that the asymptotic information rate is given by

$$I(\mathcal{E}; y) = \frac{1}{2}\log K_e, \tag{B.24}$$

which is consistent with the result in [136].

We now invoke the equivalence between the colored Gaussian channel and the ISI channel $\mathcal{F}$, that is, instead of generating y by (B.22), we generate y by

$$\begin{cases} \tilde{y}_t &= u_t + Z_t \\ s_{c,t+1} &= Fs_{c,t} + G\tilde{y}_t \\ y_t &= Hs_{c,t} + \tilde{y}_t, \end{cases} \tag{B.25}$$

where $s_{c,0} = 0$. Since $Z^T = \mathcal{Z}_T N^T$, the mapping from (u, N) to y here is equivalent to that in (B.22). Therefore, (B.22) becomes

$$\begin{cases} u_t &= d'\tilde{s}_{s,t} + \mathcal{E}_t \\ \tilde{y}_t &= u_t + Z_t \\ s_{c,t+1} &= Fs_{c,t} + G\tilde{y}_t \\ y_t &= Hs_{c,t} + \tilde{y}_t \\ \hat{s}_{s,t+1} &= F\hat{s}_{s,t} + L_s e_t \\ e_t &= y_t - H\hat{s}_{s,t} = (H + d')\tilde{s}_{s,t} + \mathcal{E}_t + N_t \\ \tilde{s}_{s,t+1} &= F\tilde{s}_{s,t} + Gu_t - L_s e_t, \end{cases} \tag{B.26}$$

where $\hat{s}_{s,0} = 0$; see Fig. B.1 for the block diagram.

Our analysis of this system is facilitated by considering transfer functions. Note that

$$\begin{aligned} \mathcal{T}_{\mathcal{E}u} &= \mathbb{S} \\ \mathcal{T}_{Nu} &= \mathbb{T}\mathcal{Z}, \end{aligned} \tag{B.27}$$

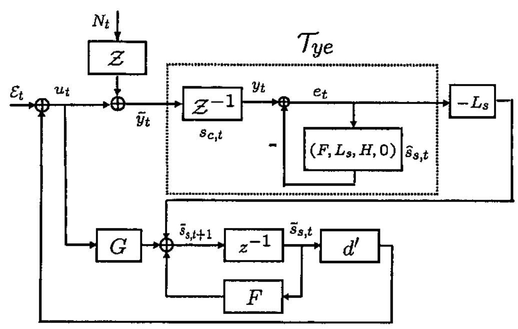


Figure B.1 Block diagram for the communication system using the GM inputs and Kalman filtering, where $s_{c,t}$ is the state for $\mathcal{Z}^{-1}$ with $s_{c,0} = 0$, and $\hat{s}_{s,t}$ is the state for system $(F, L_s, H, 0)$ with $\hat{s}_{s,0} = 0$.

where $\mathbb{S}$ is the sensitivity, and $\mathbb{T} := \mathbb{S} - 1$ is the complimentary sensitivity. (The sensitivity $\mathbb{S}$ here should not be confused with the sensitivity in Section 4.5.1.) Then we have

$$\begin{aligned} u &= \mathbb{S}\mathcal{E} + \mathbb{T}\mathcal{Z}N \\ \tilde{y} &= \mathbb{S}(\mathcal{E} + \mathcal{Z}N). \end{aligned} \tag{B.28}$$

Now assume that d and $K_{\mathcal{E}}$ form the *optimal* solution to (4.85), where $K_{\mathcal{E}} \neq 0$, for contradiction purpose. We can then compute the corresponding optimal Σ_s, L_s, $\mathbb{S}$, $\mathbb{T}$, etc. Fix the optimal L_s, $\mathbb{S}$, and $\mathbb{T}$. We will show that this leads to: 1) The whiteness of $\{\tilde{y}_t\}$; 2) $L_s = G$; 3) $K_{\mathcal{E}} = 0$ and hence contradiction.

1) For fixed optimal values of L_s, $\mathbb{S}$, and $\mathbb{T}$, suppose that we can have the freedom of choosing the power spectrum of $\mathcal{E}$ in (B.26). Since we have assumed the optimality of a white process $\{\mathcal{E}_t\}$, it must hold that any correlated process $\{\mathcal{E}_{c,t}\}$ does not lead to a larger mutual information than $\{\mathcal{E}_t\}$ does. Precisely, assume a stationary correlated process $\{\mathcal{E}_{c,t}\}$ replaces the white process $\{\mathcal{E}_t\}$ in (B.26). Then $\{\mathcal{E}_t\}$ yields the maximum achievable rate over all possible $\{\mathcal{E}_{c,t}\}$, i.e., it solves

$$\max_{\substack{L_s, \mathbb{S}, \mathbb{T} \text{ fixed}, \mathcal{S}_{\mathcal{E}_c}(e^{j2\pi\theta}) \\ s.t.\ \mathrm{E}u^2 \leq \mathcal{P}}} I(\mathcal{E}_c; \tilde{y}). \tag{B.29}$$

Since

$$I(\mathcal{E}_c; \tilde{y}) = h(\tilde{y}) - h(\tilde{y}|\mathcal{E}_c) = h(\tilde{y}) - h(\mathbb{S}\mathcal{Z}N) \tag{B.30}$$

and $h(\mathbb{S}\mathcal{Z}N)$ is fixed for fixed $\mathbb{S}$, the above optimization is equivalent to

$$\begin{aligned} &\max_{\mathcal{S}_{\mathcal{E}_c}(e^{j2\pi\theta})} \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \log \mathcal{S}_{\tilde{y}}(e^{j2\pi\theta}) d\theta. \\ &\scriptstyle s.t.\ \mathbf{E}u^2 = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{S}_{\mathbb{S}}(e^{j2\pi\theta})\mathcal{S}_{\mathcal{E}_c}(e^{j2\pi\theta}) + \mathcal{S}_{\mathbb{T}}(e^{j2\pi\theta})\mathcal{S}_{\mathcal{Z}}(e^{j2\pi\theta}) d\theta \le \mathcal{P} \end{aligned} \tag{B.31}$$

However, this optimization problem is equivalent to solving, for some $\mathcal{P}_1 \geq 0$,

$$\begin{aligned} &\max_{\mathcal{S}_{\mathcal{E}_c}(e^{j2\pi\theta})} \frac{1}{2}\int_{-\frac{1}{2}}^{\frac{1}{2}} \log\left(\mathcal{S}_{\mathbb{S}}(e^{j2\pi\theta})\mathcal{S}_{\mathcal{E}_c}(e^{j2\pi\theta}) + \mathcal{S}_{\mathbb{S}}(e^{j2\pi\theta})\mathcal{S}_{\mathcal{Z}}(e^{j2\pi\theta})\right) d\theta, \\ &\scriptstyle s.t.\ \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{S}_{\mathbb{S}}(e^{j2\pi\theta})\mathcal{S}_{\mathcal{E}_c}(e^{j2\pi\theta}) d\theta \le \mathcal{P}_1 \end{aligned} \tag{B.32}$$

which we identify as a new forward communication problem, see Fig. B.2. In this problem, we want to tune the power spectrum of $\mathbb{S}\mathcal{E}_c$, the effective channel input, to get the maximum rate. The optimal solution is given by waterfilling, namely, the power spectrum $\mathcal{S}_{\mathbb{S}}(e^{j2\pi\theta})\mathcal{S}_{\mathcal{E}_c}(e^{j2\pi\theta})$ needs to waterfill the power spectrum $\mathcal{S}_{\mathbb{S}}(e^{j2\pi\theta})\mathcal{S}_{\mathcal{Z}}(e^{j2\pi\theta})$. By optimality of $\{\mathcal{E}_t\}$, $K_{\mathcal{E}}\mathcal{S}_{\mathbb{S}}(e^{j2\pi\theta})$ is the waterfilling solution.

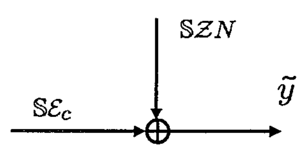


Figure B.2 An equivalent forward communication channel. Here $\mathbb{S}\mathcal{E}_c$ is the effective input, $\mathbb{S}\mathcal{Z}N$ is the effective channel noise, and $\tilde{y}$ is the output.

Since $\mathcal{S}_{\mathbb{S}}(e^{j2\pi\theta}) = 0$ for some θ if and only if $\mathbb{S}(z)$ has a zero for that θ on the unit circle, and since $\mathbb{S}(z)$ is a finite dimension transfer function with a finite number of zeros, the power spectrum $\mathcal{S}_{\mathbb{S}}(e^{j2\pi\theta})$ cannot have zero amplitude at any interval. This follows that the support of the channel input spectrum $K_{\mathcal{E}}\mathcal{S}_{\mathbb{S}}(e^{j2\pi\theta})$ is $[-1/2, 1/2]$.

In waterfilling, if the support of input spectrum is $[-1/2, 1/2]$, then the output spectrum must be flat. This is easily proven by contradiction. Thus, $\{\tilde{y}_t\}$ is a white process. Let us assume that its variance is σ^2.

2) Note that both $\tilde{y}$ and e have white spectrum, which imposes condition on the choice of L_s. The transfer function $\mathcal{T}_{ye}$ is illustrated in Fig. B.3, where we can see that its structure is a Kalman filter structure. To make e white, it is necessary to choose L_s to be the Kalman filter gain (cf. [57]), given by

$$L_s := \frac{F\Sigma_c H' + \sigma^2 G}{H\Sigma_c H' + \sigma^2}, \tag{B.33}$$

where Σ_c is the estimation error covariance matrix and is a nonnegative solution to the DARE

$$\Sigma_c = F\Sigma_c F' + \sigma^2 GG' - \frac{(F\Sigma_c H' + \sigma^2 G)(F\Sigma_c H' + \sigma^2 G)'}{H\Sigma_c H' + \sigma^2}. \tag{B.34}$$

Clearly, $\Sigma_c = 0$ is a solution to the DARE. By [57], it is also the unique nonnegative solution. Hence, we need to choose $L_s := G$.

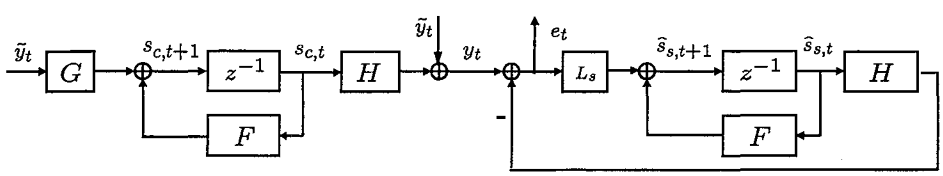


Figure B.3 The state-space representation of the transfer function $\mathcal{T}_{ye}$.

3) The fact that $L_s = G$ leads to reduction of system (B.26) or equivalently (B.22). We have

$$\begin{aligned} \tilde{s}_{t+1} &= (F - GH)\tilde{s}_t - GN_t \\ \Sigma_s &= (F - GH)\Sigma_s(F - GH)' + GG'. \end{aligned} \tag{B.35}$$

In the case that $(F - GH)$ is unstable, the closed-loop of (B.26) is unstable and cannot transmit information. In the case that $(F - GH)$ is stable, the steady-state of Σ_s depends only on (F, G, H) and is *independent* of the choice of d and $K_\mathcal{E}$, and thus (4.85) becomes

$$C_\infty = \max_{\substack{\Sigma_s \text{ fixed}, d\in\mathbb{R}^m, K_\mathcal{E}\in\mathbb{R} \\ s.t.\ \mathcal{P} = d'\Sigma_s d + K_\mathcal{E}}} \frac{1}{2}\log(1 + K_\mathcal{E} + (H + d')\Sigma_s(H + d')'). \tag{B.36}$$

This is equivalent to

$$\max_{\substack{d\in\mathbb{R}^m, K_\mathcal{E}\in\mathbb{R} \\ s.t.\ d'\Sigma_s d\leq\mathcal{P}-K_\mathcal{E}}} H\Sigma_s d, \tag{B.37}$$

which requires $K_\mathcal{E} = 0$.

B.5 Proof of Proposition 9: Finite dimensionality of the optimal scheme

i) To show that $C_{\infty,n}$ is non-decreasing as n increases, note that, an encoder (A, C) of dimension $(n+1)$ can be arbitrarily approximated by a sequence of encoders $\{(A_i, C_i)\}$ of dimension $(n+2)$ in the form of

$$\left(\left[\begin{array}{c|c} A & 0 \\ \hline 0 & 1 \end{array} \right], \left[\begin{array}{c|c} C & \frac{1}{i} \end{array} \right] \right), \tag{B.38}$$

and therefore the supremum in (4.87) with encoder dimension $(n+2)$ is no smaller than the supreme with encoder dimension $(n+1)$. So $C_{\infty,n}$ is increasing in n.

ii) By proposition 7 and the definition for $C_{\infty,m-1}(\mathcal{P})$, the optimization problem for solving $C_{\infty,m-1}(\mathcal{P})$ is given by

$$C_{\infty,m-1}(\mathcal{P}) = \sup_{\substack{A\in\mathbb{R}^{m\times m}, C \\ s.t. \Sigma=\mathbb{A}\Sigma\mathbb{A}'-\mathbb{A}\Sigma\mathbb{C}'(\mathbb{C}\Sigma\mathbb{C}'+1)^{-1}\mathbb{C}\Sigma\mathbb{A}' \\ \mathbb{D}\Sigma\mathbb{D}'=\mathcal{P}}} \frac{1}{2}\log(\mathbb{C}\Sigma\mathbb{C}'+1) \tag{B.39}$$

To compare it with $C_\infty(\mathcal{P})$, we rewrite (4.85) and (4.86) in another form, incorporating $K_\mathcal{E} = 0$. Define

$$\begin{aligned} \bar{\mathbb{A}} &:= \left[\begin{array}{c|c} F+Gd' & 0 \\ \hline Gd' & F \end{array} \right] \\ \bar{\mathbb{C}} &:= [d \quad H] \\ \bar{\mathbb{D}} &:= [d \quad 0] \\ \bar{\Sigma} &:= \left[\begin{array}{c|c} \Sigma_s & \Sigma_s \\ \hline \Sigma_s & \Sigma_s \end{array} \right]. \end{aligned} \tag{B.40}$$

It is then straightforward to verify that

$$\begin{aligned}\frac{1}{2}\log(1+(H+d')\Sigma_s(H+d')') &= \frac{1}{2}\log(1+\bar{\mathbb{C}}\bar{\Sigma}\bar{\mathbb{C}}') \\ d'\Sigma_s d &= \bar{\mathbb{D}}\bar{\Sigma}\bar{\mathbb{D}}' \\ \bar{\mathbb{A}}\bar{\Sigma}\bar{\mathbb{A}}' - \bar{\mathbb{A}}\bar{\Sigma}\bar{\mathbb{C}}'(\bar{\mathbb{C}}\bar{\Sigma}\bar{\mathbb{C}}'+1)^{-1}\bar{\mathbb{C}}\bar{\Sigma}\bar{\mathbb{A}}' &= \bar{\Sigma},\end{aligned} \tag{B.41}$$

which yields that

$$C_\infty(\mathcal{P}) = \sup_{\substack{d\in\mathbb{R}^m \\ s.t.\bar{\Sigma}=\bar{\mathbb{A}}\bar{\Sigma}\bar{\mathbb{A}}'-\bar{\mathbb{A}}\bar{\Sigma}\bar{\mathbb{C}}'(\bar{\mathbb{C}}\bar{\Sigma}\bar{\mathbb{C}}'+1)^{-1}\bar{\mathbb{C}}\bar{\Sigma}\bar{\mathbb{A}}' \\ \bar{\mathbb{D}}\bar{\Sigma}\bar{\mathbb{D}}'=\mathcal{P}}} \frac{1}{2}\log(1+\bar{\mathbb{C}}\bar{\Sigma}\bar{\mathbb{C}}') \tag{B.42}$$

Comparing (B.42) with (B.39), we conclude that $C_{\infty,m-1}(\mathcal{P}) \geq C_\infty(\mathcal{P})$. However, since for each (A, C), the channel input sequence is stationary by the steady-state characterization of the general coding structure, it holds that $C_{\infty,m-1}(\mathcal{P}) \leq C_\infty(\mathcal{P})$. Therefore, we have

$$C_{\infty,m-1}(\mathcal{P}) = C_\infty(\mathcal{P}). \tag{B.43}$$

Then ii) follows from i) immediately.

B.6 Proof of Proposition 10: Achieving C_∞ in the information sense

By Proposition 9, the optimization problem for solving $P_\infty(\mathcal{R})$ in (4.8) (which is equivalent to solving $C_\infty(\mathcal{P})$) can be reformulated as

$$[A^{opt}, C^{opt}, \Sigma^{opt}] := \arg \inf_{\substack{A\in\mathbb{R}^{m\times m}, C \\ s.t.\Sigma=\mathbb{A}\Sigma\mathbb{A}'-\mathbb{A}\Sigma\mathbb{C}'(\mathbb{C}\Sigma\mathbb{C}'+1)^{-1}\mathbb{C}\Sigma\mathbb{A}' \\ \log DI(A)=\mathcal{R}}} \mathbb{D}\Sigma\mathbb{D}', \tag{B.44}$$

for any desired rate $\mathcal{R}$. Without loss of generality, we may assume that (A, C) is in the observable canonical form, i.e.

$$A := \left[\begin{array}{c|c} 0_{n\times 1} & I_n \\ \hline a_n & a_{n-1}\cdots a_1 \end{array}\right]$$
$$C := \left[\begin{array}{cc} 1 & 0_{1\times n} \end{array}\right]. \tag{B.45}$$

Observe that $\det A = a_n$. Thus, $DI(A) = |\det A| = |a_n|$ if A does not contain stable eigenvalues, and $DI(A) > |\det A| = |a_n|$ otherwise.

As a consequence, if we search over A with a_n fixed to be $2^{\mathcal{R}}$ or $-2^{\mathcal{R}}$, we actually enforce $DI(A) \geq 2^{R}$. However, the optimal solution must satisfy $DI(A^{opt}) = 2^{\mathcal{R}}$, since otherwise the system has a rate equal to $R_{\infty,m-1} = \log DI(A^{opt}) > \mathcal{R}$, which would require more power than the case that $R_{\infty,m-1} = \mathcal{R}$; notice that (B.44) is a power minimization problem. To summarize, we can remove the constraint $\log DI(A) = \mathcal{R}$ by letting $a_n = \pm 2^{\mathcal{R}}$ in (B.45), and the optimal solution A does not contain stable eigenvalues. Furthermore, note that unit-circle eigenvalues do not generate any rate or power and hence can be removed. Thus, if A^{opt} has (n^*+1) unstable eigenvalues, we can solve the optimization problem with A having size (n^*+1) and the obtained optimal solution still achieves C_∞.

B.7 Proof of Proposition 11: Optimality in the analog transmission

The end-to-end distortion is given by

$$\begin{aligned}
\mathrm{MSE}(\hat{W}_t) &= \mathbf{E}(W - \hat{W}_t)(W - \hat{W}_t)' \\
&= \mathbf{E}(x_0 - \hat{x}_{0,t})(x_0 - \hat{x}_{0,t})' \\
&= \mathbf{E}(A^{-t-1}x_{t+1} - A^{-t-1}\hat{x}_{t+1})(A^{-t-1}x_{t+1} - A^{-t-1}\hat{x}_{t+1})' \\
&= \mathbf{E}A^{-t-1}\tilde{x}_{t+1}\tilde{x}'_{t+1}A'^{-t-1} \\
&= A^{-t-1}\Sigma_{x,t+1}A'^{-t-1},
\end{aligned} \tag{B.46}$$

where

$$\Sigma_{x,t+1} := [I, 0]\Sigma_{t+1}[I, 0]' \tag{B.47}$$

and the expectation is w.r.t. the randomness in W and $\hat{W}_t$. By rate-distortion theory, the above distortion needs an asymptotic rate R satisfying

$$\begin{aligned} R &\geq \lim_{t\to\infty} \frac{1}{2(t+1)} \log \frac{1}{\det MSE(\hat{W}_t)} \\ &= \lim_{t\to\infty} \frac{1}{2(t+1)} \log \frac{\det A^{2t+2}}{\det \Sigma_{x,t+1}} \\ &= \log|\det A|. \end{aligned} \tag{B.48}$$

From Proposition 10, we know that $\log|\det A^*|$ equals C_∞ and the average channel input power equals $\mathcal{P}$. Because C_∞ is the supremum of asymptotic rate, it follows that the equality in (B.48) is achieved. Then we see that the proposition holds.

B.8 Proof of Proposition 12: Optimality in digital transmission

It is sufficient to show that $R_{\infty,n}(A,C)$ is achievable for any fixed (A,C). To show this, for the fixed (A,C), construct the scheme in Fig. 4.2 and use $\mathcal{G}_T^*$, the Kalman-filter based optimal receiver. The closed-loop (4.58) is stabilized and will converge to its steady-state for large enough T.

We can then directly verify that Theorems 4.3 and 4.6 in [28] are applicable to the (steady-state) LTI system. These theorems assert that, if the closed-loop system is stabilized, then we can construct a sequence of codes to reliably (in the sense of vanishing probability of error) transmit the initial conditions associated with the open-loop unstable eigenvalues of A (denoted $a_0, \cdots, a_k$, if any), at a rate

$$R := (1-\epsilon)R_{\infty,n}(A,C) \tag{B.49}$$

for any $\epsilon > 0$, and in the meantime, $P_{\infty,n}(A,C) \leq \mathcal{P}$ holds. Therefore, we conclude that, for any (A,C), the portion of W that is associated with the unstable eigenvalues of A is transmitted reliably from the transmitter to the receiver at rate arbitrarily close to $R_{\infty,n}(A,C)$. Moreover, we notice that we can achieve $C_{\infty,n}$ by a sequence of purely unstable (A,C) (i.e. $k=n$), in which the initial condition W is the message being transmitted. This follows that W is transmitted at the capacity rate.

In addition, [28] showed that for any choice of x_0, it holds that

$$PE_T = 1 - \prod_{i=0}^{n}\left(1 - 2Q\left(\frac{\sigma_{T,i}^{-\epsilon}}{2}\right)\right), \tag{B.50}$$

where $\sigma_{T,i}$ is the square root of the ith eigenvalue of $\mathrm{MSE}(\hat{x}_{0,T})$, and

$$\begin{aligned} \mathrm{MSE}(\hat{x}_{0,T}) &= \mathbf{E}(x_0 - \hat{x}_{0,T})(x_0 - \hat{x}_{0,T})' \\ &= A^{-T-1}\Sigma_{x,T+1}A'^{-T-1}. \end{aligned} \tag{B.51}$$

Note that the expectation is w.r.t. the randomness in $\hat{x}_{0,T}$ only, different from (B.46), and that asymptotically Σ_{t+1} and hence $\Sigma_{x,T+1}$ are independent on the choice of x_0.

It then holds for each i,

$$\begin{aligned} (\sigma_{T,i})^2 &\leq \lambda_{\max}(\mathrm{MSE}(\hat{x}_{0,T})) \\ &= \lambda_{\max}(A^{-T-1}\Sigma_{x,T+1}A'^{-T-1}) \\ &\overset{(a)}{=} \lambda_{\max}(A'^{-T-1}A^{-T-1}\Sigma_{x,T+1}) \\ &\overset{(b)}{\leq} \bar{\sigma}(A'^{-T-1}A^{-T-1}\Sigma_{x,T+1}) \\ &\overset{(c)}{\leq} \bar{\sigma}(A'^{-T-1}A^{-T-1})\bar{\sigma}(\Sigma_{x,T+1}) \\ &= (\bar{\sigma}(A'A))^{-T-1}\,\bar{\sigma}(\Sigma_{x,T+1}) \end{aligned} \tag{B.52}$$

where $\lambda_{\max}(M)$ denotes the maximum eigenvalue of M, $\bar{\sigma}(M)$ denotes the maximum singular value of M, (a) follows from $\lambda(AB) = \lambda(BA)$, (b) follows from $|\lambda(A)| \leq \bar{\sigma}(A)$, and (c) is because the maximum singular value is an induced norm. Since $\Sigma_{x,T+1}$ converges to steady-state value exponentially, the above implies that, for T large enough, each $\sigma_{T,i}$ decays to zero exponentially as T increases.

Now using the union bound and the Chernoff bound, we have

$$\begin{aligned} PE_T &\leq \sum_{i=0}^{n} 2Q\left(\frac{\sigma_{T,i}^{-\epsilon}}{2}\right) \\ &\leq \sum_{i=0}^{n} \frac{4}{\sqrt{2\pi}\sigma_{T,i}^{-\epsilon}} \exp\left(-\frac{\sigma_{T,i}^{-2\epsilon}}{8}\right), \end{aligned} \tag{B.53}$$

and hence PE_T decreases to zero doubly exponentially since $\epsilon > 0$ and $\sigma_{T,i}$ decays exponentially. Thus we prove the proposition.

B.9 Numerical examples for optimal power computation

We list some of the numerical results of the optimal power computation for finite-impulse response (FIR) and infinite-impulse response (IIR) channels in Table B.1. Note that, Row 2 and Row 4 are first-order and second-order minimum-phase channels, and we observe that one unstable pole is enough for us to achieve the capacity. Row 3, Row 5, and Row 6 are a first-order non-minimum-phase channel, a second-order non-minimum-phase channel, and a third-order minimum-phase channel, respectively. Capacity of higher dimensional channels can also be computed with rather low complexity. For example, for a sixth-order channel with

$$\mathcal{Z}^{-1} = \frac{1+0.5z^{-1}-0.4z^{-2}+0.3z^{-3}+0.8z^{-4}+0.1z^{-5}-0.2z^{-6}}{1-0.8z^{-1}+0.6z^{-2}-0.4z^{-3}-0.2z^{-4}-0.1z^{-5}+0.3z^{-6}}, \tag{B.54}$$

we can compute that $P_{\infty,n} = 0.6003, 0.3830, 0.3781, 0.3781, 0.3781$ for $n = 2, 3, 4, 5, 6$, respectively, to achieve transmission rate $R = 1$ bit/channel use. Note that $n = 6$ corresponds to the capacity power. It takes less than 30 seconds to complete the computation for $n = 6$ case.

Table B.1 Capacity power for channels with transmission rate $R = 1$ bit/channel use

channel $\mathcal{F}$	$n=1$	$n=2$	$n=3$	capacity	water filling
$1-\frac{1}{2z}$	1.92	1.92	1.92	1.92	2.667
$1-\frac{2}{z}$	0.75	0.667	0.667	0.667	0.667
$1+\frac{1}{2z}+\frac{1}{2z^2}$	1.586	1.586	1.586	1.586	2.51
$1+\frac{5}{2z}-\frac{3}{2z^2}$	0.853	0.276	0.276	0.276	0.33
$\frac{(1+0.5z^{-1}-0.4z^{-2})}{(1+0.6z^{-2}-0.4z^{-3})}$	2.745	0.745	0.745	0.743	1.4689

APPENDIX C. PROOFS OF RESULTS IN CHAPTER 5

C.1 Proof of $C = \log \tilde{a}$

Assume without loss of generality that $S_t = s[j]$ and $S_{t+1} = s[i]$, both drawn from the stationary distribution of $\{S_t\}$. Then it holds that

$$\begin{aligned}
C &= \frac{1}{2}\mathbf{E}_{S_t,S_{t+1}} \log\left(1 + (S_{t+1})^2\gamma(S_t)\right) \\
&= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\pi[i]q_{ij}\log\left(1 + (S_{t+1})^2\gamma(S_t)\right) \\
&= \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\pi[i]q_{ij}\log\left(1 + s[i]^2\gamma(s[j])\right) \\
&= \sum_{i=1}^{m}\left(\sum_{j=1}^{m}\pi[i]q_{ij}\log a(s[i], s[j])\right) \\
&= \sum_{i=1}^{m}\log \bar{a}(s[i]) \\
&= \log \tilde{a}
\end{aligned}$$

C.2 Proof of Lemma 4: Stability of the closed-loop system

i) The dynamics of $x_t^{(j)}$ without external input is $x_{t+1}^{(j)} = a(S_{t-1}, S_t)^{-1}x_t^{(j)}$ if $S_{t-1} = s[j]$, or $x_{t+1}^{(j)} = x_t^{(j)}$ otherwise. So we have

$$x_{t+1}^{(j)} = \phi_t^{(j)}x_0^{(j)}$$

and hence Φ_t is the state transition matrix for (5.8).

By (5.4), on Ω_{TYP}, $n(j,t)$ grows approximately as $(t+1)\pi[j]$ for t large enough, then as t goes to infinity, $n(j,t)$ goes to infinity. Since $0 < a(S_{t-1}, S_t)^{-1} < 1$, $\phi_t^{(j)}$ goes to zero on Ω_{TYP} and $\phi_t^{(j)}$ is bounded between 0 and 1 on Ω.

ii) Conditioned on S_0^t and $\boldsymbol{x}_0$, we obtain that

$$\mathbf{E}x_{t+1}^{(j)} = \phi_t^{(j)}x_0^{(j)} \tag{C.1}$$

and hence $\mathbf{E}\boldsymbol{x}_t^{(j)}$ converges to zero on Ω_{TYP}. The boundedness of $\mathbf{E}\boldsymbol{x}_t^{(j)}$ on Ω for each t follows from i).

iii) We first show the independence of $x_t^{(j)}$ and $x_t^{(l)}$ when $j \neq l$. To show this, notice that, conditioned on $\boldsymbol{x}_0$ and $\{S_t\}$, both $x_t^{(j)}$ and $x_t^{(l)}$ are Gaussian. Then it is only needed to show they are uncorrelated, namely

$$\mathbf{E}x_t^{(j)}x_t^{(l)} = \mathbf{E}x_t^{(j)}\mathbf{E}x_t^{(l)}. \tag{C.2}$$

If $t = 0$, obviously (C.2) holds. Suppose that (C.2) holds for some t, then for $t+1$,

$$x_{t+1}^{(i)} = \begin{cases} a(S_{t-1}, S_t)^{-1}x_t^{(i)} - b(S_{t-1}, S_t)N_t & \text{if } S_{t-1} = s[i] \\ x_t^{(i)} & \text{otherwise.} \end{cases} \tag{C.3}$$

So

$$\begin{aligned} \mathbf{E}x_{t+1}^{(j)}x_{t+1}^{(l)} &= \begin{cases} \mathbf{E}x_t^{(j)}x_t^{(l)} & \text{if } S_{t-1} \neq s[j] \text{ and } S_{t-1} \neq s[l] \\ \mathbf{E}[a(S_{t-1}, S_t)^{-1}x_t^{(j)} - b(S_{t-1}, S_t)N_t]x_t^{(l)} & \text{if } S_{t-1} = s[j] \\ \mathbf{E}x_t^{(j)}[a(S_{t-1}, S_t)^{-1}x_t^{(l)} - b(S_{t-1}, S_t)N_t] & \text{if } S_{t-1} = s[l] \end{cases} \\ &= a(S_{t-1}, S_t)^{-1}\mathbf{E}x_t^{(j)}x_t^{(l)} \\ &= \mathbf{E}x_{t+1}^{(j)}\mathbf{E}x_{t+1}^{(l)}. \end{aligned} \tag{C.4}$$

Thus, (C.2) holds for any $t \in \mathbb{N}$.

iii) By the independence of $x_t^{(j)}$ and $x_t^{(l)}$ when $j \neq l$, Σ_t is a diagonal matrix. If $S_{t-1} = s[j]$ then

$$\begin{aligned} \mathbf{E}(x_{t+1}^{(j)})^2 &= a(S_{t-1}, S_t)^{-2}\mathbf{E}(x_t^{(j)})^2 + b(S_{t-1}, S_t)^2 \\ &\leq a^{-2}\mathbf{E}(x_t^{(j)})^2 + b^2, \end{aligned}$$

where $a := \min_{j,l} a(s[j], s[l])$ and $b := \max_{j,l} |b(s[j], s[l])|$; or if $S_{t-1} \neq s[j]$ then

$$\mathbf{E}(x_{t+1}^{(j)})^2 = \mathbf{E}(x_t^{(j)})^2.$$

Since $0 < a < 1$, for any t,

$$\mathbf{E}(x_t^{(j)})^2 \leq a^{-2T}(x_0^{(j)})^2 + \frac{1 - a^{-2T}}{1 - a^{-2}}b^2 \tag{C.5}$$

for some T (dependent on t and $\{S_t\}$). Since $\mathbf{E}(x_t^{(j)})^2 \geq \underline{b}^2$ where $\underline{b} := \min_{j,l} |b(s[j], [l])| > 0$, $\mathbf{E}(x_t^{(j)})^2$ is bounded both from above and from below by some positive constants for all t; note that the constants can be chosen independently on t, $\boldsymbol{x}_0$, and $\{S_t\}$. Notice that $\Sigma_t^{(j)} := \mathbf{E}(x_t^{(j)})^2 - (\mathbf{E}x_t^{(j)})^2$ is strictly positive since the randomness in the noise enters the system. Then, because $|\mathbf{E}x_t^{(j)}|$ decreases to zero monotonically, $\Sigma_t^{(j)}$ is uniformly bounded from above and from below by positive constants for any $\boldsymbol{x}_0$,

$\{S_t\}$, and t.

C.3 Proof of Proposition 14: Doubly exponential decay of error probability

Fix any $\{S_t\} \in \Omega_{TYP}$. Recall that

$$PE^{(j)}_{T|S} = Q_{T,1} + Q_{T,2}, \tag{C.6}$$

where $Q_{T,1}$ satisfies

$$Q_{T,1} \leq Q\left(\frac{\xi^{(j)}_T}{2\sqrt{\Sigma^{(j)}_{T+1}}}\alpha^{n(j,T)+(T+1)(\frac{1}{\epsilon}-1)(\pi[j]-\frac{n(j,T)}{T+1})} + \frac{x_0^{(j)}\phi_T^{(j)}}{\sqrt{\Sigma^{(j)}_{T+1}}}\right), \tag{C.7}$$

and $Q_{T,2}$ satisfies a similar inequality. Let

$$\delta_T := (T+1)\left(\frac{1}{\epsilon}-1\right)\left(\pi[j]-\frac{n(j,T)}{T+1}\right) + \log_\alpha\left\{\frac{\xi^{(j)}_T + 2\alpha^{-n(j,T)[1+(\frac{1}{\epsilon}-1)(\frac{(T+1)\pi[j]}{n(j,T)}-1)]}x_0^{(j)}\phi_T^{(j)}}{2\sqrt{\Sigma^{(j)}_{T+1}}}\right\}. \tag{C.8}$$

Then

$$Q_{T,1} = Q\left(\alpha^{n(j,T)+\delta_T}\right), \tag{C.9}$$

Since $n(j,T)/(T+1)$ converges to $\pi[j]$ when T goes to infinity, we have that $\delta_T/(T+1)$ vanishes when T goes to infinity. That is,

$$Q_{T,1} \leq Q\left(\alpha^{n(j,T)+o(T+1)}\right). \tag{C.10}$$

The Chernoff bound of the Q-function (cf. [59]) says that

$$Q(T) \leq \frac{1}{\sqrt{2\pi}T}\exp(-\frac{1}{2}T^2) = \exp\left[-\frac{1}{2}\left(T^2 + \log(2\pi T^2)\right)\right].$$

Then

$$\begin{aligned} Q_{T,1} &\leq \exp\left\{-\frac{1}{2}\left[\alpha^{2n(j,T)+o(T)} + \log(2\pi\alpha^{2n(j,T)+o(T+1)})\right]\right\} \\ &= \exp\left\{-\frac{1}{2}(\alpha^2)^{n(j,T)+\frac{1}{2}\log_\alpha\left[\alpha^{o(T+1)}-\alpha^{-2n(j,T)}\log(2\pi\alpha^{2n(j,T)+o(T+1)})\right]}\right\} \\ &= \exp\left(-\frac{1}{2}(\alpha^2)^{\pi[j](T+1)+o(T+1)}\right). \end{aligned}$$

Similarly,

$$Q_{T,2} := Q\left(\frac{0.5\xi_T^{(j)}}{\bar{a}[j]^{(T+1)(1-\epsilon)}\phi_T^{(j)}\sqrt{\Sigma_{T+1}^{(j)}}} - \frac{x_0^{(j)}\phi_T^{(j)}}{\sqrt{\Sigma_{T+1}^{(j)}}}\right) \leq \exp\left(-\frac{1}{2}(\alpha^2)^{\pi[j](T+1)+o(T+1)}\right),$$

and hence,

$$PE_{T|S}^{(j)} \leq 2\exp\left(-\frac{1}{2}(\alpha^{2\pi[j]})^{(T+1)+o(T+1)}\right). \tag{C.11}$$

To get the (asymptotic) decay exponent of $PE_{T|S}^{(j)}$, noticing that for T large enough, the Chernoff bound becomes tight, we can derive (following the steps similar to above) that

$$\begin{aligned}
\lim_{T\to\infty}\frac{1}{T+1}\log(\log(\frac{1}{Q_{1,T}})) &= \lim_{T\to\infty} -2\sum_{l=1}^{m} d(j,l,T)\log(a(s[j],s[l])) \\
&= 2\epsilon\sum_{l=1}^{m}\pi[j]p_{jl}\log(a(s[j],s[l])) \\
&= 2\epsilon\log\bar{a}[j].
\end{aligned}$$

It can be also shown that

$$\lim_{T\to\infty}\frac{1}{T+1}\log(\log(\frac{1}{Q_{2,T}})) = 2\epsilon\log\bar{a}[j].$$

It is easily seen that

$$\lim_{T\to\infty}\frac{1}{T}\log(\log(\frac{1}{\exp(-a^T)+\exp(-b^T)})) = \min(\log a, \log b) \tag{C.12}$$

for $a, b > 0$. (That is, the "average" decay exponent of $(\exp(-a^T)+\exp(-b^T))$ is the smaller decay exponent for $\exp(-a^T)$ and $\exp(-b^T)$.) Then (5.61) follows from (C.12). Finally, notice that asymptotically $PE_{T|S} = \sum_{j=1}^{m} PE_{T|S}^{(j)}$, then (5.62) follows from (C.12).

C.4 Proof of Proposition 15: Power computation

We study the recursion for $\mathbf{E}(x_t^{(j)})^2$. Assume $S_{t-2} = s[j]$ and $S_{t-1} = s[l]$, by (5.15),

$$\begin{aligned}
\mathbf{E}(x_t^{(j)})^2 &= a(s[j],s[l])^{-2}\mathbf{E}(x_{t-1}^{(j)})^2 + b(s[j],s[l])^2 \\
&\overset{(a)}{=} a(s[j],s[l])^{-2}\mathbf{E}(x_{t-1}^{(j)})^2 + a(s[j],s[l])^{-2}\gamma(s[j])^2 s[l]^2 \\
&= a(s[j],s[l])^{-2}(\mathbf{E}(x_{t-1}^{(j)})^2 + \gamma(s[j])^2 s[l]^2),
\end{aligned}$$

where (a) follows from (5.12). Subtracting both sides by $\gamma(s[j])$, we obtain

$$\begin{aligned}\mathbf{E}(x_t^{(j)})^2 - \gamma(s[j]) &= a(s[j],s[l])^{-2}\left(\mathbf{E}(x_{t-1}^{(j)})^2 + \gamma(s[j])^2 s[l]^2 - a(s[j],s[l])^2\gamma(s[j])\right) \\ &= a(s[j],s[l])^{-2}\left(\mathbf{E}(x_{t-1}^{(j)})^2 - \gamma(s[j])\right),\end{aligned}$$

and thus

$$\mathbf{E}(x_t^{(j)})^2 - \gamma(s[j]) = (\phi_{t-1}^{(j)})^2\left(\mathbf{E}(x_0^{(j)})^2 - \gamma(s[j])\right). \tag{C.13}$$

It follows from Lemma 4 that $\mathbf{E}(x_t^{(j)})^2 - \gamma(s[j])$ tends to zero with probability one, and hence $\mathbf{E}(x_t^{(j)})^2$ tends to $\gamma(s[j])$ with probability one. By the Cesaro mean, the time average of $\mathbf{E}(x_t^{(j)})^2$, denoted $\mathbf{E}(x^{(j)})^2$, tends to $\gamma(s[j])$ with probability one.

Since the channel input u is the multiplexing of $x_t^{(j)}$, the power of u is the power of $x_t^{(j)}$ averaged over j, therefore with probability one

$$\mathbf{E}u^2 = \sum_{i=1}^{m}\pi[i]\mathbf{E}(x^{(i)})^2. \tag{C.14}$$

So, (5.63) holds with probability one. In other words, the average channel input power is $\gamma(s[j])$ for any sequence $\{S_t\} \in \Omega_{TYP}$. On Ω_{TYP}^c, the power of $x_t^{(j)}$ is uniformly bounded by Lemma 4, and so is the power of u, which does not contribute to the average channel input power. Thus, the average power of u is as shown in (5.63), with the average taken over all $\{S_t\}$, $\boldsymbol{x}_0$, noise, and time.

C.5 Proofs for the modified cases

C.5.1 When A1) is not assumed

In this case, there exists some states that are assigned with zero power. Suppose $s[J]$ is the only such state; for cases with more than one such states, the same idea applies. We now have $a(s[i],[J]) = 1$ and $\bar{a}[J] = 1$. Then the Jth subsystem needs some modifications, as listed below. For the encoding, let $x_0^{(J)} = 0$ and make it known to the receiver. That is, the Jth subsystem is not used to transmit any message. It leads to zero signalling rate, zero transmission power, and zero probability of error associated with this subsystem. We then see that in this situation, our scheme behaves exactly as the capacity solution formula suggests. Therefore, it can be easily verified that the capacity is achieved: Simply note that 1) In Lemma 4, (5.34) holds for any $j \neq J$; 2) $C = \sum_{i=1}^{m}\log\bar{a}(s[i]) = \sum_{i\neq J}\log\bar{a}(s[i])$; and 3) $\mathbf{E}u^2 = \sum_{i=1}^{m}\pi[i]\mathbf{E}(x^{(i)})^2 = \sum_{i\neq J}^{m}\pi[i]\mathbf{E}(x^{(i)})^2$.

C.5.2 When A2) is not assumed

In this case there is some i with $s[i] = 0$. Then $a(s[j], [i]) = 1$, but $\bar{a}[j] > 1$ still holds for each j with $\gamma(s[j]) \neq 0$. To see this, assume otherwise for J, $\bar{a}[J] = 1$ but $\gamma(s[J]) > 0$. By (5.22) and (5.11), this would yield that, for each $l = 1, \cdots, m$, it must hold either i) $s[l] = 0$ or ii) $\pi[J]p_{Jl} = 0$. However, ii) is equivalent to $p_{Jl} = 0$ since $\pi[J] > 0$. Hence, for each l such that $s[l] \neq 0$, p_{Jl} has to be 0; i.e., the probability of $s[J]$ jumping to any nonzero state is *zero*. In other words, $s[J]$ must jump to $s[J]$ with probability one, which according to [125] would imply that $\gamma(s[J]) = 0$, a contradiction. So $\bar{a}[j] > 1$ holds as long as $\gamma(s[j]) \neq 0$.

Then we see that Lemma 4 still holds, but another ingredient is needed in its proof. In showing $\phi_t^{(j)}$ converges to zero with probability one, note that $a(S_{t-1}, S_t)$ is either less than 1 (i.e. "contractive") or equal to 1, and if for some $\{S_t\}$, $\phi_t^{(j)}$ does not converge to zero, each such sequence $\{S_t\}$ must have finite many "contractions" and thus form a zero-probability set. Then Lemma 1 as well as the main results hold.

C.6 Proof for the case of AWGN i.i.d. fading with infinite state

Denote the density of the random variable $A := \sqrt{\mathcal{P}S^2 + 1}$ as p_A, where $S \sim p_S$. Pick any $\epsilon > 0$ small enough. Uniformly partition the unit interval $[-\frac{1}{2}, \frac{1}{2}]$ into $\lfloor M_T \rfloor$ sub-intervals, where

$$M_T := \exp[(T+1)(1-\epsilon)\mathbf{E}_{A\sim p_A} \log A]. \tag{C.15}$$

Then the asymptotic signalling rate is

$$\begin{aligned} R &= \lim_{T\to\infty} \left(\frac{\log M_T}{T+1} - \frac{\log \xi_T}{T+1} \right) \\ &= \lim_{T\to\infty} \frac{(T+1)(1-\epsilon)\mathbf{E}_{A\sim p_A} \log A}{T+1} \\ &= (1-\epsilon)\mathbf{E}_{A\sim p_A} \log A \\ &= (1-\epsilon)\mathbf{E}_{S\sim p_S} \frac{1}{2} \log(1 + S^2\mathcal{P}) \\ &= (1-\epsilon)C_{iid,inf}, \end{aligned} \tag{C.16}$$

where $\xi_T := M_T / \lfloor M_T \rfloor$. Following the idea in the finite state-space case, to show the reliable communication with rate R, it suffices to show that the associated control system is stabilized in the sense of bounded and vanishing first moment $\mathbf{E}x_T$ and bounded second moment $\mathbf{E}(x_T - \mathbf{E}x_T)^2$, for every typical sequence. However, the stability of the control system can be easily proven; here for brevity we only show that $\mathbf{E}\boldsymbol{x}_T$ is bounded and vanishing. To this aim, note that

$$x_T = A(S_{T-1})^{-1} x_{T-1} - b(S_{T-1}) N_{T-1}. \tag{C.17}$$

The first moment evolves according to

$$\mathbf{E}x_T = A(S_{T-1})^{-1}\mathbf{E}x_{T-1} \tag{C.18}$$

when conditioned on the channel state sequence and initial condition, and hence

$$\mathbf{E}x_T = \prod_{j=0}^{T} A(S_{j-1})^{-1}x_0. \tag{C.19}$$

It is then clear that $\mathbf{E}x_T$ is bounded for any channel state sequence and vanishes for any typical channel state sequences. Finally, the power computation can be done as before. Thus we prove the optimality of the proposed scheme.

C.7 Multi-step delay case

We now briefly present the capacity-achieving feedback communication scheme for the general multi-step delay case, that is, there are $d \geq 1$ steps of delay between the receiver sending a signal and the transmitter receiving it. We also outline the main idea for the proof.

C.7.1 Communication setup

The communication setup to be used for information transmission analysis is shown in Fig. C.1. Let us assume that

$$\boldsymbol{z}_0 := [\boldsymbol{x}_0', \boldsymbol{x}_1', \cdots, \boldsymbol{x}_{d-1}']' \in \mathbb{R}^{dm} \tag{C.20}$$

is the message to be encoded at the transmitter and to be recovered at the receiver, where

$$\boldsymbol{x}_n := [x_n^{(1)}, x_n^{(2)}, \cdots, x_n^{(m)}]' \in \mathbb{R}^m \tag{C.21}$$

for $n = 0, 1, \cdots, d-1$. We store $\boldsymbol{z}_0$ in the $z^{-d}I_m$ block of the transmitter according to Fig. C.2, that is, $\boldsymbol{z}_0$ is the initial state of the $z^{-d}I_m$ block of the transmitter. The $z^{-d}I_m$ block at the receiver is initialized by zeros.

The signal $-\hat{\boldsymbol{x}}_{n,t+d}$ is the estimate of $\boldsymbol{x}_n$ at time t, for $n = 0, 1, \cdots, d-1$. Note that the channel state S_{t-d} is used at the transmitter side to generate the channel input u_t, and used at the receiver side at time $(t-d)$, which is consistent with the assumption of d-step-delayed DTRCSI. The operation of the communication setup starting from $t = 0$ is similar to the $d = 1$ case (details skipped). We use $A(S_j) := I_m$, $c'(S_j) := [1, 0, \cdots, 0]'$, and $v_j := 0$, if $j < 0$. We may also draw the extended communication setup where the feedback signal has bounded power, and it can be verified that the extended communication setup is an extension of the SK scheme. For this proposed scheme, the notions

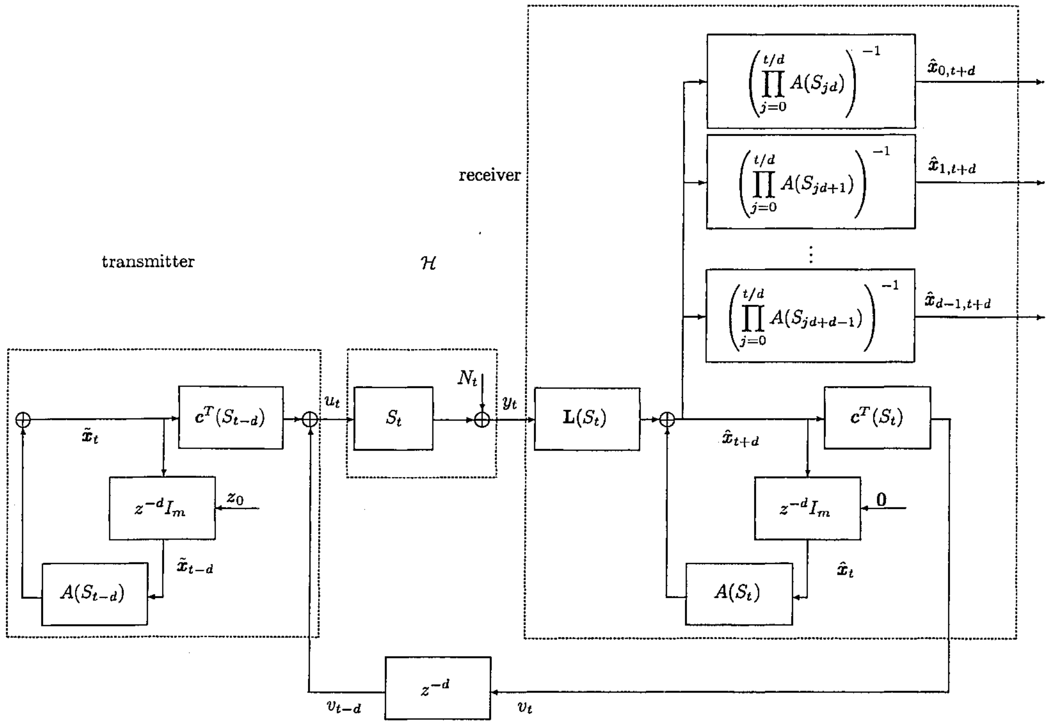


Figure C.1 The communication setup for multi-step delay case.

of the transmission rate, error probability, etc., can be generalized from the $d = 1$ case easily and are skipped here.

C.7.2 Control setup

Letting

$$x_t := \hat{x}_t + \tilde{x}_t, \tag{C.22}$$

we obtain the control setup, see Fig. C.3. The control setup is useful for stability analysis and power computation as before, and it operates in a fashion that there are d subsystems with no interaction between them, which allows us to achieve the capacity by applying the techniques of the $d = 1$ case.

C.7.3 Choice of parameters

Let $\gamma(\cdot)$ be the capacity-achieving power allocation that maps the channel state S_t to the channel input power $\gamma(S_t)$ and such that the average channel input power constraint holds.

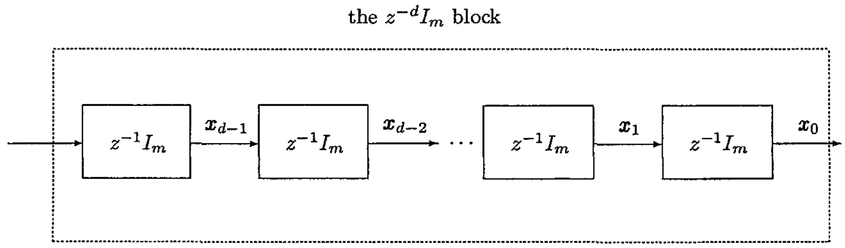


Figure C.2 Initialization of the transmitter memory.

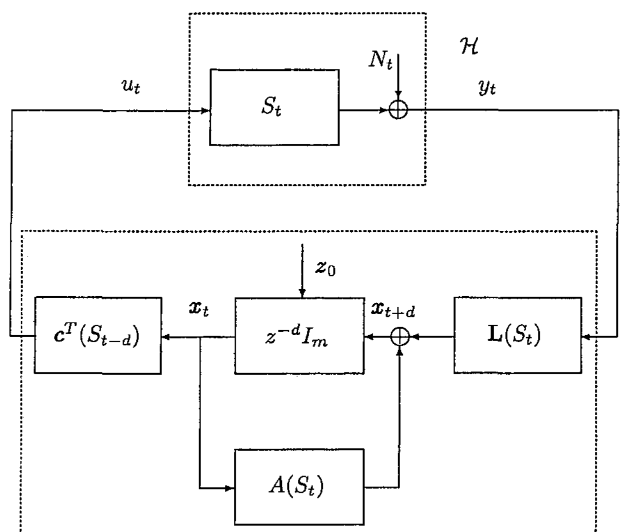


Figure C.3 The control setup for the multi-step delay case.

Supposing at time $(t-d)$ the channel state $S_{t-d} = s[i]$ for some i, we define

$$\begin{aligned} A(S_t) &:= A(S_{t-d}, S_t) && := \operatorname{diag}(\boldsymbol{a}(S_t)) && \in \mathbb{R}^{m\times m} \\ \boldsymbol{L}(S_t) &:= \boldsymbol{L}(S_{t-d}, S_t) && := [0, \cdots, 0, b(S_t), 0, \cdots, 0]' && \in \mathbb{R}^m \\ & \quad\ \boldsymbol{c}(S_{t-d}) && := [0, \cdots, 0, c(S_{t-d}), 0, \cdots, 0]' && \in \mathbb{R}^m, \end{aligned} \tag{C.23}$$

where $\boldsymbol{a}(S_t) := \boldsymbol{a}(S_{t-d}, S_t) \in \mathbb{R}^m$ is such that its i-th element is

$$a(S_t) := a(S_{t-d}, S_t) := \sqrt{\gamma(S_{t-d})(S_t)^2 + 1} \tag{C.24}$$

and all other elements are 1; $b(S_t)$, the i-th element of $\boldsymbol{L}(S_t)$, is

$$b(S_t) := b(S_{t-d}, S_t) := -\frac{\gamma(S_{t-d})S_t}{\sqrt{\gamma(S_{t-d})(S_t)^2+1}} = \frac{1}{a(S_{t-d}, S_t)} - a(S_{t-d}, S_t); \tag{C.25}$$

and $c(S_{t-d})$, the i-th element of $\boldsymbol{c}(S_{t-d})$, is

$$c(S_{t-d}) := 1. \tag{C.26}$$

Whenever S_t, $t < 0$, is encountered, it is treated as $s[1]$.

C.7.4 Achieving the capacity

We sketch the main idea of the proof for the d-step delay case. The proposed scheme can be viewed as a multiplexing scheme that multiplexes d feedback communication systems, each has only one feedback delay. In more detail, it holds that, both the communication setup and the control setup can be equivalently treated as d separate subsystems, where each subsystem is activated only once in every d steps and remains its previous states in the other $(d-1)$ steps, and at each time, only one subsystem is activated. Therefore, we can view each subsystem has only "one cumulative delay" of length d, during which this subsystem updates only once; note this is different from d delays during which the subsystem may update d times. Since the n-th subsystem outputs an updated estimate $\hat{\boldsymbol{x}}_{n,t}$ every d steps, the achievable rate for such a subsystem is $\log(|\tilde{a}|)/d$, which follows that the total achievable rate is still $\log(|\tilde{a}|)$. For the power computation, we notice that each subsystem uses power $\mathcal{P}$ but they occupy disjoint time slots, and hence the total average power is $\mathcal{P}$.

APPENDIX D. PROOFS OF RESULTS IN CHAPTER 6

D.1 Proof of Theorem 4: Optimality of the coding scheme for AWGN channels

In this section, we prove the optimality of the proposed coding scheme for the AWGN WDP-channel. To establish lossless interference cancelation, we need to show that, both the decoded message $\hat{W}_T$ and the (asymptotic) average channel input power are not affected by the interference sequence. It is therefore sufficient to prove that both the decoder state $\hat{x}_{0,T}$ and the channel input u_t are decoupled (or asymptotically decoupled) from the interference sequence. Then the theorem would follow from standard results of feedback communication over an AWGN channel *without* interference.

We derive the expressions for u_t and $\hat{x}_{0,t}$. We can express the encoder state in terms of the initial condition, interference, and channel outputs as

$$x_t = a^t(W + W_M) + a^t \sum_{j=0}^{t-1} a^{-j-1} L(\xi_j - y_j). \tag{D.1}$$

Hence

$$\sum_{j=0}^{t-1} a^{-j-1} L y_j = W + W_M - a^{-t} x_t + \sum_{j=0}^{t-1} a^{-j-1} L \xi_j. \tag{D.2}$$

On the other hand, noticing that $y_t = u_t + \xi_t + N_t$, the encoder state is also

$$\begin{aligned} x_t = {} & a x_{t-1} + L(\xi_{t-1} - u_{t-1} - \xi_{t-1} - N_{t-1}) \\ = {} & (a - Lc) x_{t-1} - L N_{t-1} \\ = {} & a^{-1} x_{t-1} - L N_{t-1} \\ = {} & a^{-t}(W + W_M) - \sum_{j=0}^{t-1} a^{-t+1+j} L N_j. \end{aligned} \tag{D.3}$$

Then the decoder state is

$$
\begin{aligned}
\hat{x}_{0,t-1} &= \sum_{j=0}^{t-1} a^{-j-1} L y_j \\
&= W + W_M - a^{-t} x_t + \sum_{j=0}^{t-1} a^{-j-1} L \xi_j \\
&= (1 - a^{-2t})(W + W_M) + a^{-2t} \sum_{j=0}^{t-1} a^{j+1} L N_j + \sum_{j=0}^{t-1} a^{-j-1} L \xi_j .
\end{aligned}
\tag{D.4}
$$

Therefore, by (6.7), we have

$$
\hat{x}_{0,T} = (1 - a^{-2T-2}) W + a^{-2T-2} \sum_{j=0}^{T} a^{j+1} L N_j , \tag{D.5}
$$

and thus the interference does not affect the decoder state at time T and the decoded message $\hat{W}_T$. Now note that

$$
u_t = c x_t = a^{-t}(W + W_M) - \sum_{j=0}^{t-1} a^{-t+1+j} L N_j . \tag{D.6}
$$

That is, the presence of interference incurs an extra power overhead of $a^{-2t} W_M^2$, which vanishes exponentially and is negligible, since the power due to the term $\sum_{j=0}^{t-1} a^{-t+1+j} L N_j$ approaches a nonzero constant and the coding length $(T+1)$ is sufficiently large. In summary, the interference does not affect either the decoded message or the asymptotic channel input power. This indeed leads to lossless interference canceling.

D.2 Proof of Theorem 5: Optimality of the coding scheme for Gaussian channels

In this section, we prove the optimality of the proposed coding scheme for general Gaussian WDP-channel. Similar to the AWGN case, the interference ξ does not affect x_t and hence the channel input u_t asymptotically. In addition, the term in $\hat{x}_{0,T}$ associated with ξ is known to the encoder before the transmission, and therefore the encoder can offset x_0 to completely cancel the term associated with ξ. To facilitate our analysis, we rewrite the communication system as a control system and then establish the above claims.

We first define for the coding scheme shown in Fig. 6.5 that

$$
\begin{aligned}
\tilde{s}_t &:= s_t - \bar{s}_t \\
\mathbb{X}_t &:= \begin{bmatrix} x_t \\ \tilde{s}_t \end{bmatrix} .
\end{aligned}
\tag{D.7}
$$

Then the dynamics of the encoder becomes

$$\begin{cases} \mathbb{X}_t &= \mathbb{A}_{cl}\mathbb{X}_{t-1} - LN_{t-1} = \mathbb{A}\mathbb{X}_{t-1} - Le_{t-1} \\ e_{t-1} &= \mathbb{C}\mathbb{X}_{t-1} + N_{t-1} \\ u_t &= \mathbb{D}\mathbb{X}_t, \end{cases} \tag{D.8}$$

where $\mathbb{A}_{cl} := \mathbb{A} - L\mathbb{C}$. Note that (D.8) is a control system (as indicated in Fig. 6.5), and is affected by the interference only through its initial condition $\mathbb{X}_0$. We now have

$$\mathbb{X}_t = \mathbb{A}_{cl}^t \mathbb{X}_0 - \sum_{j=0}^{t-1} \mathbb{A}_{cl}^{t-j-1} LN_j. \tag{D.9}$$

By $u_t = \mathbb{C}\mathbb{X}_t$, any L stabilizing the control system ensures that, the channel input power is asymptotically equal to $\sum_{j=0}^{t-1} \mathbb{A}_{cl}^{t-j-1} LL' \mathbb{A}_{cl}^{t-j-1\prime}$, independent of the interference. We can also verify that the term in $\hat{x}_{0,T}$ generated by x_0 is asymptotically x_0.

In addition, the interference would generate an extra term in $\hat{x}_{0,T}$ given by

$$\sum_{j=0}^{T} A^{-j-1} L_1 \xi_j - \sum_{j=0}^{T} \sum_{i=0}^{j-1} A^{-j-1} L_1 H (F - L_2 H)^{t-1-j} L_2 \xi_j, \tag{D.10}$$

if no compensation to the initial condition x_0 is used. Then we offset x_0 by this amount and therefore we achieve (asymptotic) lossless cancelation of the interference. The proof for the coding scheme achieving $C_\infty(\mathcal{P})$ follows from the reasoning in Chapter 4.

BIBLIOGRAPHY

[1] V. Borkar and S. Mitter. LQG control with communication constraints. In A. Paulraj, V. Roychowdhury, and C. Schaper, editors, *Communications, Computation, Control, and Signal Processing: A Tribute to Thomas Kailath*, pages 365–373. Kluwer Academic Publishers, Massachusetts, 1997.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization.* 2001. *http://www.stanford.edu/class/ee364/.*

[3] R. W. Brockett. Minimum attention control. *Proc. 36th IEEE Conf. Decision and Control*, 3:2628–2632, 1997.

[4] R. W. Brockett and D. Liberzon. Quantized feedback stabilization of linear systems. *IEEE Trans. Automat. Contr.*, 45(7):1279–1289, 2000.

[5] S.A. Butman. A general formulation of linear feedback communication systems with solutions. *IEEE Trans. Inform. Theory*, IT-15:392–400, 1969.

[6] S.A. Butman. Linear feedback rate bounds for regressive channels. *IEEE Trans. Inform. Theory*, IT-22:363–366, 1976.

[7] G. Caire and S. Shamai (Shitz). On the capacity of some channels with channel state information. *IEEE Trans. Inform. Theory*, 45(6):2007–2019, Sept. 1999.

[8] G. Caire and S. Shamai (Shitz). On the achievable throughput of a multi-antenna Gaussian broadcast channel. *IEEE Trans. Inform. Theory*, 49(7):1649–1706, July 2003.

[9] C. T. Chen. *Linear Systems Theory and Design.* Oxford University Press, New York, 3rd ed. edition, 1999.

[10] M. Chow and Y. Tipsuwan. Network-based control systems: A tutorial. *Proc. 27th Annual Conf. of IEEE Industrial Eletronics Society*, 3:1593–1602, Nov. 2001.

[11] A. Cohen and A. Lapidoth. The Gaussian watermarking game. *IEEE Trans. Inform. Theory*, 48(6):1639–1667, June 2002.

[12] A. Cohen and A. Lapidoth. Generalized writing on dirty paper. *Proc. IEEE International Symposium on Information Theory (ISIT)*, page 227, 2002.

[13] A. Cohen and R. Zamir. Writing on dirty paper in the presence of difference set noise. *Proc. 41st Annual Allerton Conference on Communication, Control and Computing*, Oct. 2003.

[14] M. H. M. Costa. Writing on dirty paper. *IEEE Trans. Inform. Theory*, 29(5):439–441, May 1983.

[15] O. L. V. Costa and M. D. Fragoso. Stability results for discrete-time systems with Markovian jumping parameters. *J. Math. Anal. Applicat.*, 179:154–178, 1993.

[16] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* John Wiley and Sons, New York, 1991.

[17] T.M. Cover and S. Pombra. Gaussian feedback capacity. *IEEE Trans. Inform. Theory*, IT-35:37–43, 1989.

[18] M. A. Dahleh and I. J. Diaz-bobillo. *Control of Uncertain Systems: A Linear Programming Approach.* Prentice Hall, 1995.

[19] D. Delchamps. Stabilizing a linear system with quantized state feedback. *IEEE Trans. Automat. Contr.*, 35(8):916–924, Aug. 1990.

[20] A. Dembo. On gaussian feedback capacity. *IEEE Trans. Inform. Theory*, 35:1072–1076, Sept. 1989.

[21] R. L. Dobrushin. Information transmission in a channel with feedback. *Theory of Probability and Its Applications*, 34:367–383, 1958.

[22] N.G. Dokuchaev and A.V. Savkin. Recursive state estimation via limited capacity communication channels. *Proc. 38th IEEE Conf. Decision and Control*, 5:4929 –4932, 1999.

[23] M. Egerstedt and X. Hu. A hybrid control approach to action coordination for mobile robots. *Automatica*, 38:125–130, Jan. 2002.

[24] N. Elia. Design of hybrid systems with guaranteed performance. *Proc. 39th IEEE Conf. Decision and Control*, 1:993 –998, 2000.

[25] N. Elia. Coarsest quantizer density for quadratic stabilization of two-input linear systems. *Proceedings of the 10th Mediterranean Conference on Control and Automation*, July 2002.

[26] N. Elia. Stabilization of systems over analog memoryless channels. *Proc. 40th Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2002.

[27] N. Elia. Feedback stabilization in the presence of fading channels. *Proc. of the 2003 American Control Conference*, pages 4438–4443, June 2003.

[28] N. Elia. When Bode meets Shannon: Control-oriented feedback communication schemes. *IEEE Trans. Automat. Contr.*, 49(9):1477–1488, Sept. 2004.

[29] N. Elia. The information cost of disturbance rejection. *Proc. 2005 Mediterranean Conference on Control and Automation*, June 2005.

[30] N. Elia. Linear remote stabilization over packet drop networks with ACK losses. *Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2005.

[31] N. Elia and J. Eisenbeis. Limitations of linear remote control over packet drop networks. *Proc. 43rd IEEE Conf. Decision and Control*, pages 5152 – 5157, Dec. 2004.

[32] N. Elia and S.K. Mitter. Stabilization of linear systems with limited information. *IEEE Trans. Automat. Contr.*, 46(9):1384–1400, Sept. 2001.

[33] P. Elias. Channel capacity without coding. *in Lectures on Communication System Theory, E. Baghdady Ed.*, 1961.

[34] A. Ephremides and B. Hajek. Information theory and communication networks: An unconsummated union. *IEEE Trans. Inform. Theory*, 44(6):2416–2434, Oct. 1987.

[35] W. Equitz and T. Cover. Successive refinement of information. *IEEE Trans. Inform. Theory*, 37:269–274, Mar. 1991.

[36] U. Erez, S. Shamai (Shitz), and R. Zamir. Capacity and lattice-strategies for cancelling known interference. *Proc. International Symposium on Information Theory and Applications (ISITA)*, Nov. 2000.

[37] U. Erez, S. Shamai (Shitz), and R. Zamir. Capacity and lattice-strategies for cancelling known interference. *Submitted to IEEE Trans. Inform. Theory*, 2002.

[38] U. Erez and S. ten Brink. A close-to-capacity dirty paper coding scheme. *IEEE Trans. Inform. Theory*, 51(10):3417–3432, Oct. 2005.

[39] F. Fagnani and S. Zampieri. Stability analysis and synthesis for scalar linear systems with a quantized feedback. *Proc. 40th IEEE Conf. Decision and Control*, 3:2204 –2209, Dec. 2001.

[40] J. A. Fax and R. M. Murray. Information flow and cooperative control of vehicle formations. *IEEE Trans. Automat. Contr.*, 49(9):1465 – 1476, Sept 2004.

[41] G. D. Forney. On the role of MMSE estimation in approaching the information theoretic limits of linear Gaussian channels: Shannon meets Wiener. *Proc. 41st Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2003.

[42] E. Frazzoli. *Robust hybrid control for autonomous vehicle motion planning.* PhD thesis, MIT, June 2001.

[43] R. G. Gallager. *Information Theory and Reliable Communication.* John Wiley and Sons, 1968.

[44] K. Gallivan, X. Rao, and P. Van Dooren. Singular Riccati equations stabilizing large-scale systems. *Lin. Alg. Appl.*, 2005. in press.

[45] M. Gastpar. On remote sources and channels with feedback. *Proc. 38th Annual Conference on Information Sciences and Systems (CISS)*, 2004.

[46] 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, May 1996.

[47] A. J. Goldsmith and P. P. Varaiya. Capacity of fading channels with channel side information. *IEEE Trans. Inform. Theory*, 43(6):1986–1992, Nov. 1997.

[48] S. Graham, G. Baliga, and P. R. Kumar. Issues in the convergence of control with communication and computing: Proliferation, architecture, design, services, and middleware. *Proc. 42nd IEEE Conf. Decision and Control*, pages 1466–1471, Dec. 2004.

[49] S. Graham and P. R. Kumar. The convergence of control, communication, and computation. *Proceedings of PWC 2003: Personal Wireless Communication, Lecture Notes in Computer Science, Volume 2775 / 2003, Springer-Verlag, Heidelberg, 2003.*, 2775:458–475, 2003.

[50] R. M. Gray. *Entropy and Information Theory.* Springer-Verlag, 1991.

[51] D. Guo, S. Shamai, and S. Verdu. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Trans. Inform. Theory*, 51(4):1261–1282, Apr. 2005.

[52] M. Horstein. Sequential transmission using noiseless feedback. *IEEE Trans. Inform. Theory*, IT-9:136–143, 1963.

[53] S. Ihara. Capacity of discrete time Gaussian channel with and without feedback - I. *Mem. Fac. Sci. Kochi Univ. (Math.)*, 9:21–36, 1988.

[54] S. Ihara. Capacity of discrete time Gaussian channel with and without feedback - II. *Japan Journal of Applied Mathematics*, 6(2):245–258, 1989.

[55] H. Ishii and B. A. Francis. Stabilizing a linear system by switching control with dwell time. *Proc. of the 2001 American Control Conference*, 3:1876 –1881, 2001.

[56] A. Jadbabaie, J. Lin, and A. S. Morse. Coordination of groups of mobile autonomous agents using nearest neighbor rules. *IEEE Trans. Automat. Contr.*, 48(6):988–1001, June 2003.

[57] T. Kailath, A. Sayed, and B. Hassibi. *Linear Estimation.* Prentice Hall, 2000, Englewood Cliffs, NJ, 2000.

[58] S. M. Kay. *Fundamentals of Statistical Signal Processing I: Estimation Theory.* Prentice-Hall PTR, Englewood Cliffs, N.J., 1998.

[59] S. M. Kay. *Fundamentals of Statistical Signal Processing II: Detection Theory.* Prentice-Hall PTR, Englewood Cliffs, N.J., 1998.

[60] Young-Han Kim. The feedback capacity of the first-order moving average Gaussian channel. *http://arxiv.org/abs/cs/0411036*, Nov. 2004.

[61] Young-Han Kim. On the feedback capacity of stationary Gaussian channels. *Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2005.

[62] T. E. Klein. *Capacity of Gaussian Noise Channels with Side Information and Feedback.* PhD thesis, MIT, Cambridge, MA, Feb. 2001.

[63] G. Kramer. *Directed Information for Channels with Feedback.* PhD thesis, Swiss Federal Institute of Technology (ETH), Zurich, Switzerland, 1998.

[64] G. Kramer. Feedback strategies for white Gaussian interference networks. *IEEE Trans. Inform. Theory*, 48(6):1423–1438, June 2002.

[65] G. Kramer. Capacity results for the discrete memoryless network. *IEEE Trans. Inform. Theory*, 49(1):4–21, Jan. 2003.

[66] F. R. Kschischang, B. J. Frey, and H. A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Trans. Inform. Theory*, 47(2):498 - 519, Feb. 2001.

[67] H. Kwakernaak and R. Sivan. *Linear Optimal Control Systems.* John Wiley & Sons, New York, 1972.

[68] S. K. Leung-Yan-Cheong L.H. Ozarow. An achievable region and outer bound for the gaussian broadcast channel with feedback. *IEEE Trans. Inform. Theory*, 30:667 – 671, 1984.

[69] D. Liberzon. Stabilization by quantized state or output feedback: a hybrid control approach. in Proceedings of the 15th IFAC World Congress, to appear, *http://black.csl.uiuc.edu/ liberzon/topics.html*, July 2002.

[70] D. Liberzon and R. W. Brockett. Nonlinear feedback systems perturbed by noise: steady-state probability distributions and optimal control. *IEEE Trans. Automat. Contr.*, 45(6):1116–1130, 2000.

[71] J. Liu. Quantized stabilization of nonlinear affine systems. Master's thesis, Iowa State University, Sept. 2002.

[72] J. Liu and N. Elia. Quantized feedback stabilization of non-linear affine systems. *International Journal of Control*, 77(3):239–249, Feb. 2004.

[73] J. Liu and N. Elia. Achieving the stationary feedback capacity for Gaussian channels with memory. *Submitted to IEEE Trans. Inform. Theory*, Oct. 2005.

[74] J. Liu, N. Elia, and S. Tatikonda. Capacity-achieving feedback communication scheme for Markov channels with channel state information. *Proc. 2004 IEEE International Symposium on Information Theory (ISIT)*, page 71, July 2004.

[75] J. Liu, V. Yadav, H. Sehgal, J. Olson, H. Liu, and N. Elia. Phase transitions on fixed connected graphs and random graphs in the presence of noise. *Proc. 43rd IEEE Conf. Decision and Control*, Dec. 2005.

[76] X. Liu and A. Goldsmith. Cross-layer design of distributed control over wireless networks. *Proc. 43rd Annual Allerton Conference on Communication, Control and Computing*, Sept. 2005.

[77] J. Lunze. Diagnosis of quantized systems based on a timed discrete-event model. *IEEE Trans. Syst., Man and Cyber., Part A,*, 30(3):322 –335, 2000.

[78] N. C. Martins and M. A. Dahleh. Feedback control in the presence of noisy channels: Bode-like fundamental limitations of performance. *Submitted to IEEE Trans. Automat. Contr.*, 2004.

[79] J. Massey. Causality, feedback, and directed information. *Proc. IEEE International Symposium on Information Theory and Applications (ISITA)*, pages 303–305, 1990.

[80] J. Massey and M. Sain. Codes, automata, and continuous systems: Explicit interconnections. *IEEE Trans. Automat. Contr.*, 12(6):644–650, Dec. 1967.

[81] S. K. Mitter. Dynamical view of information theory. *Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2005.

[82] S. K. Mitter and N. Newton. An information theoretic view of estimation. *LIDS-P-2514*, Nov. 2001.

[83] S. K. Mitter and N. Newton. Information and entropy flow in the Kalman-Bucy filter. *J. of Stat. Phys.*, 118:145–176, Jan. 2005.

[84] L. Moreau. Stability of multiagent systems with time-dependent communication links. *IEEE Trans. Automat. Contr.*, 50(2):169–182, Feb. 2005.

[85] P. Moulin and J. O'Sullivan. Information-theoretic analysis of information hiding. *Proc. IEEE International Symposium on Information Theory (ISIT)*, June 2000.

[86] G.N. Nair and R.J. Evans. State estimation via a capacity-limited communication channel. *Proc. 36th IEEE Conf. Decision and Control*, 1:866 –871, 1997.

[87] G.N. Nair and R.J. Evans. Communication-limited stabilization of linear systems. *Proc. 39th IEEE Conf. Decision and Control*, 1:1005 –1010, 2000.

[88] G.N. Nair, R.J. Evans, I. M. Y. Mareels, and W. Moran. Topological feedback entropy and nonlinear stabilization. *IEEE Trans. Automat. Contr.*, 49:1585 – 1597, 9 2004.

[89] R. Negi and J. M. Cioffi. Delay-constrained capacity with causal feedback. *IEEE Trans. Inform. Theory*, 48(9):2478–2494, Sept. 2002.

[90] R. Olfati-Saber, E. Franco, E. Frazzoli, and J. S. Shamma. Belief consensus and distributed hypothesis testing in sensor networks. *Workshop on Network Embedded Sensing and Control*, Oct. 2005.

[91] R. Olfati-Saber and R. M. Murray. Consensus protocols for network of dynamic agents. *IEEE Trans. Automat. Contr.*, 49(9):1520–1533, Sept. 2004.

[92] J. K. Omura. Optimum linear transmission of analog data for channels with feedback. *IEEE Trans. Inform. Theory*, 14(1):38–43, Jan. 1968.

[93] A. V. Oppenheim, A. S. Willsky, and S. H. Nawab. *Signals and Systems.* Prentice Hall, New Jersey, 2nd edition, 1996.

[94] E. Ordentlich. A class of optimal coding schemes for moving average additive Gaussian noise channels with feedback. *Proc. IEEE International Symposium on Information Theory (ISIT)*, page 467, June 1994.

[95] L. H. Ozarow. The capacity of the white Gaussian multiple access channel with feedback. *IEEE Trans. Inform. Theory*, 30:623–629, 1984.

[96] L.H. Ozarow. Random coding for additive Gaussian channels with feedback. *IEEE Trans. Inform. Theory*, 36(1):17–22, Jan. 1988.

[97] L.H. Ozarow. Upper bounds on the capacity of Gaussian channels with feedback. *IEEE Trans. Inform. Theory*, 36(1):156–161, Jan. 1990.

[98] A. Papoulis and S. U. Pillai. *Probability, Random Variables and Stochastic Processes.* McGraw-Hill, Boston, MA, 4th edition, 2002.

[99] J.W. Polderman and J.C. Willems. *Introduction to Mathematical Systems Theory.* Springer, 1997.

[100] G. Polya and G. Szego. *Problems and Theorems in Analysis, I.* Springer-Verlag, Englewood Cliffs, N.J., 1972.

[101] E. Ratzer. Low-density parity-check codes on Markov channels. *Presented at the Second IMA Conference on Mathematics in Communications*, 2002. http://www.inference.phy.cam.ac.uk/ear23/papers.

[102] A. Sahai. *Anytime Information Theory.* PhD thesis, MIT, Cambridge, MA, 2001.

[103] A. Sahai, S. Avestimehr, and P. Minero. Anytime communication over the Gilbert-Eliot channel with noiseless feedback. *Proc. IEEE International Symposium on Information Theory (ISIT)*, 2005.

[104] A. Sahai and S. Mitter. The necessity and sufficiency of anytime capacity for control over a noisy communication link: Parts I and II. *Submitted to IEEE Trans. Inform. Theory*, May 2005.

[105] J. P. Schalkwijk. Center of gravity information feedback. *IEEE Trans. Inform. Theory*, IT-14:324–331, 1968.

[106] J.P.M. Schalkwijk. A coding scheme for additive noise channels with feedback? Part II:Bandlimited signals. *IEEE Trans. Inform. Theory*, IT-12(2):183–189, Apr. 1966.

[107] J.P.M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback? Part I: No bandwidth constraint. *IEEE Trans. Inform. Theory*, IT-12(2):172–182, Apr. 1966.

[108] A. Shahar-Doron and M. Feder. On a capacity achieving scheme for the colored Gaussian channel with feedback. *Proc. 2004 IEEE International Symposium on Information Theory (ISIT)*, page 74, July 2004.

[109] C. E. Shannon. The zero-error capacity of a noisy channel. *IRE Trans. Inform. Theory*, IT-2:8–19, 1956.

[110] C. E. Shannon. The wonderful world of feedback. *First Shannon Lecture in IEEE Int. Symp. Information Theory*, 1973.

[111] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, M. I. Jordan, and S. S. Sastry. Kalman filtering with intermittent observations. *IEEE Trans. Automat. Contr.*, 49(9):1453 – 1464, Sept 2004.

[112] B. Sinopoli, L. Schenato, M. Franceschetti, K. Poolla, and S. S. Sastry. Estimation and control over lossy networks. *Proc. 43rd Annual Allerton Conference on Communication, Control and Computing*, Sept. 2005.

[113] S. Tatikonda. *Control Under Communication Constraints.* PhD thesis, MIT, Cambridge, MA, Aug. 2000.

[114] S. Tatikonda. When does the sum-product algorithm work? *Proc. 41st Annual Allerton Conference on Communication, Control, and Computing*, Oct. 2003.

[115] S. Tatikonda. Some robustness properties of the sum-product algorithm. *Proc. 43rd Annual Allerton Conference on Communication, Control, and Computing*, Sept. 2005.

[116] S. Tatikonda and S. Mitter. The capacity of channels with feedback Part I: the General Case. *Submitted to IEEE Trans. Inform. Theory*, Oct. 2001.

[117] S. Tatikonda and S. Mitter. Control over noisy channels. *IEEE Trans. Automat. Contr.*, 49(7):1196 – 1201, July 2004.

[118] S. Tatikonda and S. Mitter. Control under communication constraints. *IEEE Trans. Automat. Contr.*, 49(7):1056 – 1068, July 2004.

[119] S. Tatikonda, A. Sahai, and S. Mitter. Stochastic linear control over a communication channel. *IEEE Trans. Automat. Contr.*, 49(9):1549 – 1561, Sept. 2004.

[120] I. E. Telatar. Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications*, 10(6):585–595, Nov. 1999.

[121] J. Thomas. Feedback can at most double Gaussian mutiple access channel capacity. *IEEE Trans. Inform. Theory*, 33(5):711–716, Sept. 1987.

[122] J. C. Tiernan and J. P. Schalkwijk. An upper bound on the capacity of the bandlimited Gaussian autoregressive channel with noiseless feedback. *IEEE Trans. Inform. Theory*, IT-20:311–316, 1974.

[123] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I.* John Wiley and Sons, New York, 1968.

[124] L. Vandenberghe, S. Boyd, and S.-P. Wu. Determinant maximization with linear matrix inequality constraints. *SIAM Journal on Matrix Analysis and Applications*, 19(2):499–533, 1998.

[125] H. Viswanathan. Capacity of Markov channels with receiver CSI and delayed feedback. *IEEE Trans. Inform. Theory*, 45(2):761–771, March 1999.

[126] A. Wang and A. Chandrakasan. Energy-efficient DSPs for wireless sensor networks. *IEEE Signal Processing Magazine*, pages 68–78, July 2002.

[127] D. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:325–328, 1998.

[128] F. M. J. Willems. On Gaussian channels with side information at the transmitter. *Proc. of the Ninth Symposium on Information Theory*, 1988.

[129] F. M. J. Willems. Signalling for the Gaussian channel with side information at the transmitter. *Proc. IEEE International Symposium on Information Theory (ISIT)*, page 348, 2000.

[130] D. R. Wolf. *Information and Correlation in Statistical Mechanical Systems.* PhD thesis, U. of Texas, Austin, Austin, Texas, 1996.

[131] J. Wolfowitz. Signalling over a Gaussian channel with feedback and autoregressive noise. *Journal of Applied Probability*, 12(4), 1975.

[132] W. Wong and R. W. Brockett. Systems with finite communication bandwidth constraints I: State estimation problems. *IEEE Trans. Automat. Contr.*, 42(9):1294–1299, 1997.

[133] W. Wong and R.W. Brockett. Systems with finite communication bandwidth constraints II: Stabilization with limited information feedback. *IEEE Trans. Automat. Contr.*, 44(5):1049–1053, 1999.

[134] A. Wyner. On the schalkwijk-kailath coding scheme with a peak energy constraint. *IEEE Trans. Inform. Theory*, 14(1):129 – 134, Jan. 1968.

[135] K. Yanagi. Necessary and sufficient condition for capacity of the discrete-time Gaussian channel to be increased by feedback. *IEEE Trans. Inform. Theory*, 38:1788–1791, Nov. 1992.

[136] S. Yang, A. Kavcic, and S. Tatikonda. Feedback capacity of power constrained Gaussian channels with memory. *Submitted to IEEE Trans. Inform. Theory*, Nov. 2003.

[137] S. Yang, A. Kavcic, and S. Tatikonda. Feedback capacity of finite-state machine channels. *Submitted to IEEE Trans. Inform. Theory*, 51(3):799–810, Mar. 2005.

[138] W. Yu, A. Sutivong, D. Julian, T. Cover, and M. Chiang. Writing on colored paper. *Proc. IEEE International Symposium on Information Theory (ISIT)*, page 302, 2002.

[139] R. Zamir and A. Cohen. The rate loss in writing on dirty paper. *DIMACS (Center for Discrete Mathematics and Theoretical Computer Science) Workshop on Network Information Theory*, Mar. 2003.