On the Capacity of Relay Networks with Finite Memory Relays

by

Natanael Peranginangin

B.S., Electrical Engineering
Institut Teknologi Bandung, 1995

S.M., Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 2000

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Science in Electrical Engineering

at the
Massachusetts Institute of Technology
August 2004

Signature of Author _____
Department of Electrical Engineering and Computer Science, MIT

Certified by _____
Muriel Médard
Thesis Supervisor
Esther and Harold E. Edgerton Associate Professor
Department of Electrical Engineering and Computer Science, MIT

Accepted by _____
Arthur C. Smith
Chairman, Department Committee on Graduate Studies
Professor
Department of Electrical Engineering and Computer Science, MIT

On the Capacity of Relay Networks with Finite Memory Relays

by

Natanael Peranginangin

Submitted to the Department of Electrical Engineering and Computer Science
on August 24, 2004, in partial fulfillment of the
requirements for the degree of
Doctor of Science in Electrical Engineering

Abstract

In this thesis, we examine the effect of relay memory on the capacity of two types of relay channels.

In the first part of the thesis, we present a parallel relay channel model. Under this particular model, intermediate processing at the relays is distributed and cooperative. We derive the capacity by making use of the direct relation between capacity and estimation theory. We show that the capacity of the channel under distributed relay processing by a Kalman filter and that of the channel under optimal relay processing are equal. Using a one dimensional (1D) Kalman filter, processing at individual relays requires infinite memory, assuming that the channel is subject to inter-symbol interference (ISI). For a channel with ISI, we show that a two dimensional (2D) Kalman filter allows the memory for processing at individual relays to be finite.

In the second part of the thesis, we present a serial relay channel model. Under this particular model, one section of the channel is coupled with the next by a memoryless relay. Assuming the channel is subject to energy constraints and finite end-to-end noise power, we show that the capacity tends to infinity asymptotically in the number of relay stages. Given a finite number of relay stages, finding maximum mutual information subject to energy constraints alone is difficult. Thus, in addition to energy constraints, we propose entropy constraints. We give an explicit upper bound to capacity, assuming the channel is subject to the proposed set of constraints on the channel input as well as the relay outputs. We illustrate the use of our upper bound numerically and contrast it versus several lower bounds. Next, we relax the memoryless restriction, thus allow coding and decoding at the relays. We show two trade-offs concerning the length of relay memory and the number of relay stages.

Thesis Supervisor: Muriel Médard

Title: Associate Professor of Electrical Engineering

Acknowledgements

My advisor, Professor Muriel Médard, for her constant care and guidance. In trying times, she appealed to my best hopes, never to my worst fears.

Professor Ralf Koetter at the University of Illinois at Urbana-Champaign, for his humorous wit, ease, and unparalleled depth of thought.

Professor Vincent Chan, for his constant support throughout my thesis work. His sense of leadership instilled in me a love for the Laboratory and MIT.

Professor Robert Gallager, for helping me develop the attitude, values, and technical foundation for conducting a research. Although I did not always learn my lessons quickly, Professor Gallager encouraged me to improve my thesis work, by thinking carefully about things that matter and writing them in simple and clear terms.

My friend, David Santoso, for his unwavering friendship.

My parents, for what I have learned through their examples.

My four-months old daughter, Claudia, for what she has brought into our lives.

My wife, Eunike, for taking me as I am and being as she is. She is to me what a summer is to the farmer's year.

Contents

Chapter 1

Introduction

1.1 Purpose of the thesis

The problem of point-to-point communication over relay networks has spawned much research and debate in the information theory community. The body of literature pursuing information theoretic study of relay networks is vast.

An information theoretic study would naturally point to the question of capacity. Finding capacity for point-to-point communication over a set of relays is usually either trivial or hard. For the model involving a series of AWGN channels shown in Figure 1.1, with a relay coupling one AWGN channel with the next, the question of capacity is trivial. Reference [6] showed that capacity of the overall AWGN channel is that of the individual channel with the highest noise power. As demonstrated in [14, Ch. 14.7] and [47], for models involving the traditional AWGN relay channel shown in Figure 1.2 and the parallel AWGN relay channel shown in Figure 1.3, respectively, establishing capacity results is difficult.

There are more recent capacity approaches considering more tractable models than [14] and [47] but not as trivial as that of [6]. Reference [46] considers point-to-point communication where sender and receiver communicate over a network

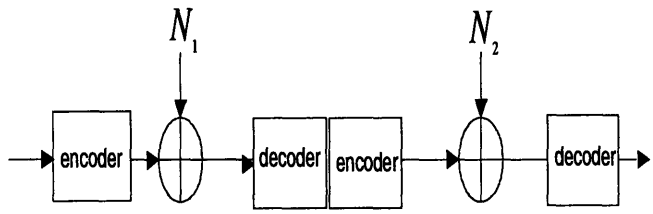

Figure 1.1: A serial AWGN relay channel

of fixed relay terminals. Reference [48] considers essentially the same communication model but allows relay nodes to be mobile. Reference [2] considers a high SNR model of point-to-point communication in which sender and receiver communicate over several layers of relay terminals. Finally, reference [5] considers a model where sender and receiver communicate over a single layer of relays but employ spatial multiplexing by means of multi-antennas at sender, receiver, and relay nodes.

Lately, there has been growing interest in using a joint source/channel approach for communication models involving relays with finite memory [9, 13, 15].

In this thesis we study a theme which has not received much attention in the context of capacity. The theme is best described by the following questions. Can we consider, from a capacity point of view, the effect of finite relay memory over the different types of relay channels? In particular, when faced with limited memory at

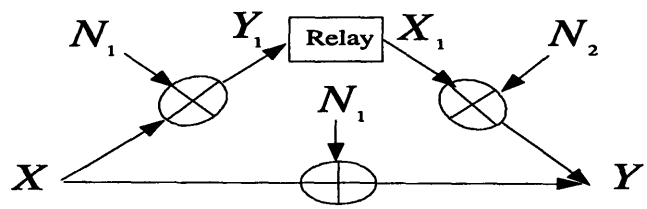

Figure 1.2: The traditional AWGN relay channel

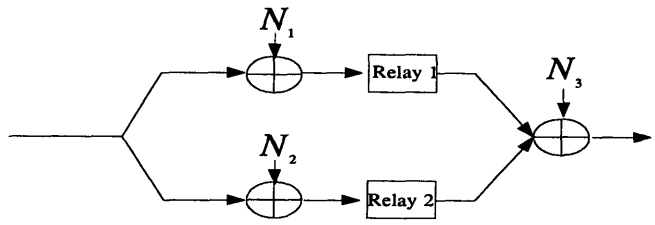

Figure 1.3: The parallel AWGN relay channel

relays, what is the maximum rate at which information can be reliably transmitted over the relays? Why we are concerned with such an issue, of all other issues surrounding point-to-point communication over relay networks, will be addressed in the following section.

1.2 Motivation for the thesis

As seen from information theory, a relay network is an idealized mathematical abstraction. Although such an abstraction may help—by bringing the essential mathematical structure with which issues are dealt with, it is not meant to enclose all the issues and complexity pertinent to the design of relay networks.

Many factors may be taken into account when defining the performance of a relay network such as implementation cost, communication protocol complexity, end-to-end delay, power efficiency. There are complicated trade-offs involved, and, speaking generally, it is difficult to arrive at concrete results if the problem is not greatly simplified .

A principle that works well is to isolate the issues which we would like to understand, using mathematical models representing the essential structure of the problem— of communication over a relay network—while eliminating unnecessary complexity. We will follow this principle, particularly by the following three key steps: making observations about two basic and fairly general relay commu-

nication models, sensing the issues we would like to understand from these observations, and investigating such issues through some simple mathematical models.

In this section, we describe two relay models for point-to-point communications, namely parallel and serial relay models. The two models are basic. Many relay networks are comprised of parallel and serial combination of relays. The central issue which will arise from our observations, as well as information theoretic models proposed in the next chapters, given our observations, is that of finite relay memory and its effect upon capacity.

1.2.1 A Parallel Relay Model

Figure 1.4 points to a model in which relays are deployed for communication between sender and far-receiver. We could view each relay as a sensor with some processing capability. Of many problems addressed in the sensor network literature, the problem of transporting data to a far-receiver via the n sensors has particularly received a lot of attention [10]. The following references provide some of the ideas concerning the problem of transporting data from sensors to a far receiver.

Reference [24] considers partitioning a sensor network into disjoint and equal-sized cells. The purpose is to eliminate redundant transmission by ensuring that only one sensor in each cell transmits. The environment in which sensors communicate between themselves is assumed as a fading environment.

Reference [22] considers flooding a sensor network with information from a source. One or a few sensors transmit data to a far distant receiver and the rest operate as repeaters. Sensors, chosen as repeaters, form several levels of relays extending through the entire network. Such an approach is intended to allow reliable transmission to far destinations that individual sensors cannot reach without rapid consumption of their own power resources. The sensor network is assumed to be impaired by fading.

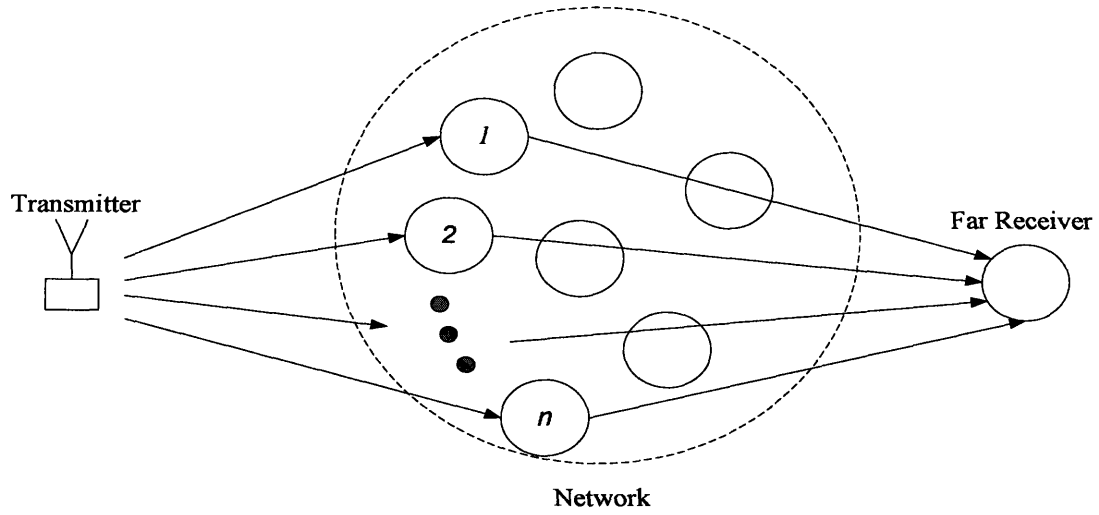

Figure 1.4: A parallel relay model

Reference [23] considers a model where individual sensors in the network pick up some correlated signals, encode these signals, and individually send the encoded signals over a multiple access channel to a far-receiver. The multiple access channel between sensors and the far-receiver is assumed to be a fading multiple access channel.

Observe that the communication approach shown in Figure 1.4 causes redundant streams to be sent to the far-receiver. The n intermediate receivers —interchangeably used to denote relay—individually transmit their output streams to the far-receiver, and thus the far-receiver receives n corrupted versions of the original stream from the sender. Next the far-receiver centrally processes the streams that it receives and reconstructs the original sender's information at its best. Hereafter, we denote such processing approach as the centralized processing approach.

Our view of reach-back communication considers the parallel relay approach shown in Figure 1.5. Assuming the n intermediate receivers have been indexed as intermediate receiver 1 to intermediate receiver n, the point-to-point communication approach shown in Figure 1.5 works as follows. Intermediate receiver 1

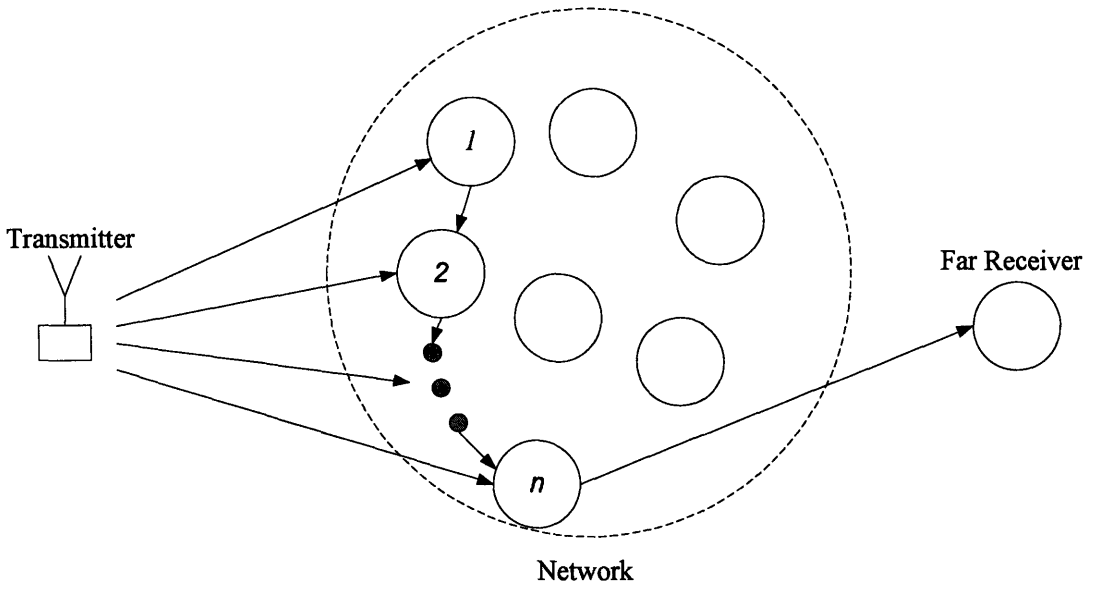

Figure 1.5: Processing which is distributed in space

processes the stream it receives from the sender and transmits an output stream to intermediate receiver 2. Intermediate receiver 2 receives streams from both sender and intermediate receiver 1, processes them, then forwards an output stream to intermediate receiver 3. Similarly, the processing progresses from one to the next intermediate receiver, until, in the final step, intermediate receiver n directs its output stream to the far-receiver.

By processing, as opposed to simply relaying, each intermediate receiver performs some operation on its received signals. Observe that the approach shown in Figure 1.5 points to processing which is distributed in *space*. Comparing Figure 1.4 with Figure 1.5, processing which is distributed in space ensures only one intermediate receiver—intermediate receiver n— communicates information to the far-receiver. We hereafter denote such a processing approach as the distributed processing approach.

The distributed processing approach brings about a number of benefits over the

centralized processing approach. First, the approach allows processing to be localized and distributed in the network of intermediate receivers, rather than centralized in the far-receiver. Second, the approach yields reduction in the redundancy of data being sent to the far-receiver, preserving communication resources such as bandwidth and power. Third, the processing approach can be tailored to work in a frequently changing topology of sensor networks, given that the far-receiver and the intermediate receivers form a spanning tree with the far-receiver being one of the end nodes.

In Chapter 2, we direct our attention to comparing the distributed processing approach to the centralized processing approach. Assuming a discrete-time channel model, we consider transmission capacity, rather than distortion, as the metric to contrast capacity under distributed versus centralized processing. We will investigate whether or not models for point-to-point communications with distributed processing lead to lower capacity than that when processing is centralized.

Why should we be concerned with the relation between the memory of relays and capacity of the communication over these relays? At this point, the aspect of memory and its utility for processing information at the relays is not apparent. In Chapter 2, we will adopt information theoretic models to represent the structure of point-to-point communication shown in Figure 1.4. For these models, the channels between individual intermediate receivers and the far-receiver are not subject to communication constraints; these channels are noiseless. The implication is two-fold. On one hand, such models imply that communication with centralized processing is optimal, since it allows the highest degree of freedom for processing information at the far-receiver. On the other hand, by elaborating on such models, we will find that communication with distributed processing between sender and far-receiver entails processing with memory at intermediate receivers.

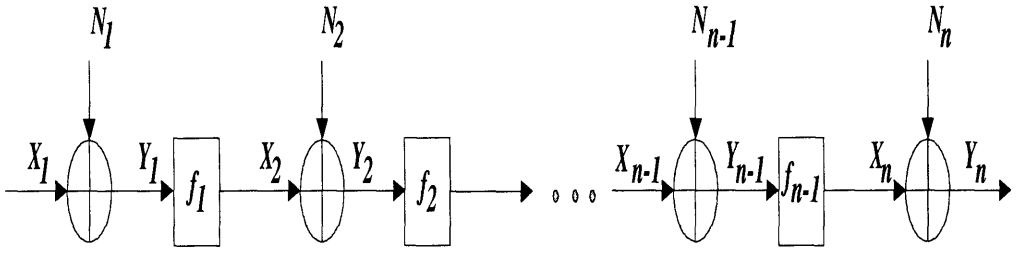

Figure 1.6: A cascade of n AWGN channels

1.2.2 A Serial Relay Channel

Consider the discrete-time channel shown in Figure 1.6. The channel is comprised of several sections, each with AWGN. Each section is coupled with the next by a *memoryless relay*. A *relay* is a function mapping the output of one section to the input of the next section. The mapping is assumed to be *memoryless* in the sense that the output at time n only depends on the input at time n.

The discrete time AWGN channel in Figure 1.6 arises in optical communication over a serial distribution network. In such a distribution network, a central problem is noise accumulation. The two major components of noise are receiver and amplifier noise. At sufficiently high signal-to-noise ratio (SNR), receiver noise, such as that from the dark current generated by an avalanche photodiode (APD), can be well modelled as AWGN, without the need to consider Poisson models that are more relevant in the low-photon regime count. Amplifier noise, such as amplified spontaneous emission (ASE) from Erbium-Doped Fiber Amplifiers (EDFAs) [7], is also well modelled as AWGN.

In order to counter noise, regeneration is performed at each relay stage. As used in the optical communication literature, the term regeneration denotes memoryless processing at individual relays, assuming that information is transmitted by means of on-off keying (OOK) or antipodal signalling; these two types of binary signalling are the most commonly used for optical transmissions. Such processing is aimed

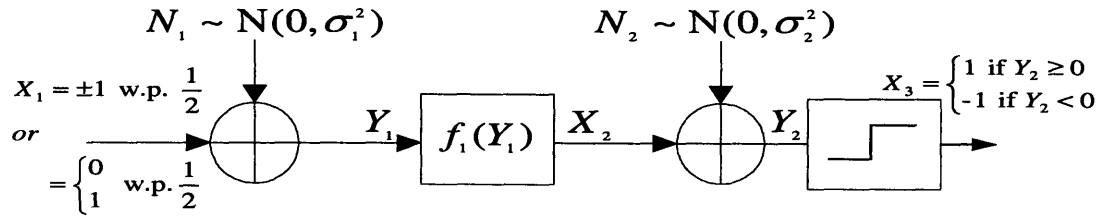

Figure 1.7: A cascade of two AWGN channels

at countering noise, and in effect lowering bit error rate (BER).

Concerning regeneration for optical communications, we observe the following. First, arbitrary processing, such as coding/recoding at source/destination and relays, would yield a better performance than that if processing at source/destination and relays is memoryless (symbol-wise). Second, if certain elements have already been placed in the fiber plant, what does this mean for its future performance? More sophisticated coding, decoding, and modulation than currently used will appear at the edge of fiber well before it appears in the middle. The two observations motivate us to consider a more general regeneration than currently addressed the optical communication literature. In other words, what if we allow arbitrary processing—more sophisticated coding, decoding, and modulation—at the edges of fiber, beyond that associated with OOK or antipodal signaling? The pursuit of higher rates for optical communication purposes will probably lead to the implementation of coding and recoding at the relays. On the road to realizing coding and recoding at the relays, implementing coding/decoding at the edges of fiber, while maintaining the relays to be memoryless, is in effect a natural milestone.

Before coming back to our discussion of regeneration with arbitrary processing at the source/destination, but memoryless at the relays, we present the following background on the problem of regeneration in optical communications. The problem of regeneration has been explored for years. It was originally investigated for

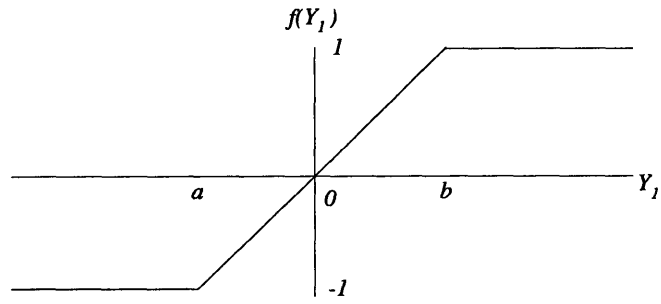

Figure 1.8: A hard-limiting function

coaxial systems [39], and is particularly relevant in optical transmission systems. Consider the channel model shown in Figure 1.7. For this model, given that the input is either OOK or antipodal and a hard-decision is placed at the output end of the channel, common relay functions studied in the optical communication literature are as follows:

- Linear amplifier

$$X_2 = f_1(Y_1) = \alpha Y_1 ,$$

  where α is some real valued constant,

- Hard-decision (the function in Figure 1.8, with $a = b$),

- A combination of hard-decision and linear amplifier (see Figure 1.8)

$$X_2 = f_1(Y_1) = \begin{cases} -1 & \text{for} Y_1 < a \\ \frac{2}{b-a} Y_1 - \frac{a+b}{b-a} & \text{for } a \leq Y_1 < b \\ 1 & \text{for } Y_1 \geq b \end{cases}$$

  where $a < b$ and $a, b \in \mathbb{R}$,

- An arbitrary nonlinear function.

There is no single nomenclature for regeneration, but a fairly common usage is that of R1 for linear amplification, R2 for linear amplification and hard-limiting, and R3

for linear amplification, hard-limiting, and retiming [49, 39, 50]. R2 systems are the common means of regeneration in optical systems [49, 32, 51, 52]. Many R3 systems have been demonstrated and proposed [53, 54, 55, 56]. Unlike R3, R1 and R2 systems assume that timing and jitter are perfect [57], and thus do not consider such issues.

The problem of transmitting signals through a channel subject to hard-limiter's noise [29, 38], of recovering original signals from noise contaminated signals [58, 30, 31, 59, 60, 44], and of transmitting various sinusoids [61] are classic problems in communication theory. More general non-linearities have been considered for noise [62], binary processes [42], phase modulated signals [36], noise contaminated signals [63] and Markov processes in general [64, 65, 66]. To our knowledge, none of these works establish the optimum memoryless relay function for regeneration of binary signals for the channel shown in Figure 1.7. Attempts at comparing the performances of different functions in regeneration have been limited [41]. The term optimal relay function denotes the optimum memoryless (symbol-wise) relay function that minimizes bit error rate (BER) for the channel shown in Figure 1.7 (assuming an antipodal or OOK input and a hard-decision at the output end of the channel).

There is a conjecture on the problem of regeneration under OOK signalling, popularly known as a 'folk theorem' in the optical communication community. Consider the discrete time channel shown in Figure 1.7 and assume an antipodal or OOK input and a hard-decision at the output end of the channel. Moreover, let SNR be defined as the ratio between input and noise powers at each section of the channel. For such a channel, the 'folk theorem' states that a hard-decision function is the best memoryless relay at low SNR and a linear amplification is the best at high SNR. In Chapter 3, we will present a serial relay model and use it to argue that such is not always the case.

More recently, [1] derived the optimum memoryless relay function $f_1(Y_1)$ for

the channel in Figure 1.7, assuming an antipodal or OOK input and a hard-decision at the output end of the channel, subject to energy constraints. Depending on the noise at the first and second section of the channel, the optimum function gradually varies from being a linear on one extreme, to being a hard-decision function on the other extreme. As such, the result has some likeness to that of the 'folk theorem'. It is different, however, for the gradual change of the shape of the optimum function is not dictated by SNR, but by the amount of noise power at each section of the channel. In Chapter 3, we will describe the optimum function derived in [1] and relate it to the main theorem of the chapter.

The issue of performing regeneration linearly or nonlinearly and that of whether to regenerate optically or electronically have spawned much research and debate. There are tradeoffs involved, most notably that between cost and BER. In this thesis, we isolate such issues and their inherent complexity. Rather, from a capacity point of view, we ask the following question: given memoryless processing at the relays but arbitrary processing at source/destination, what is the fundamental limit of communication? In effect, we are generalizing the current problem of regeneration with R1, R2, and R3 to that of regeneration with memoryless processing at the relays but arbitrary processing at both ends of the channel. This is the main thrust of Chapter 3. Our motivation stems from our view that the more general regeneration allows for:

- the increase in the latency of optical communications by coding/decoding at the source/destination ;
- economic feasibility at the current stage of development in optical communications, namely optical regeneration (without OEO) with memoryless relays is economically more feasible than
  - optical regeneration (without OEO) by coding/recoding at the relays;

- electronic regeneration (with OEO) using memoryless relays;
- electronic regeneration (with OEO) by coding/recoding at the relays.

1.3 Thesis Outline

In Chapter 2, we will study the problem of point-to-point and distributed communication over parallel relays. The channel model chosen to represent the structure of such a problem is the Single Input Multi Output (SIMO) with Additive White Gaussian Noise (AWGN).

In Section 2.2, we will propose a one dimensional (1D) Kalman filter as means of successively estimating the input sequence of the channel. One dimensional recursive estimation from one relay to the next by a 1D Kalman filter yields the same final estimate as the estimation in which the complete set of observed output is fused and processed in a centralized fashion. We will find the expression for the capacity of the centrally processed channel and show that it is the same as that of the channel processed by the 1D Kalman filter.

In Section 2.3, we will extend the model of Section 2.2 to a SIMO with AWGN and inter-symbol interference (ISI). ISI entails infinite memory for processing at each receiver if distributed processing is done by a 1D Kalman filter. In order to mitigate the problem of infinite processing memory, distributed processing is performed by a two dimensional (2D) Kalman filter. With a 2D Kalman filter, estimation proceeds from the first relay to the next and successively to the last one at each time step. Within the same time step, however, the last relay feeds back its estimate to the first, allowing the estimate of the next time step to be based on the present. The final estimate attained by the 2D Kalman filter with finite processing memory is the same as the estimate computed by 1D Kalman filter with infinite processing memory at the relays. Finally, we will find the expression for the ca-

pacity of the centrally processed channel and show that it is the same as that of the channel processed by a 2D Kalman filter.

In Chapter 3, we will study the problem of point-to-point communication over serial relays. We will adopt a model whereby the sender and far-receiver communicate over a series of AWGN channels resulting from the coupling of one AWGN channel with the next by a relay.

In Section 3.2, we will assume the relay functions to be memoryless. Under this assumption, and given that end-to-end noise power is finite and the serial relay channel is subject to energy constraints, we will show that capacity tends to infinity with the number of relay stages in the series. The particular model and assumptions considered in this section is aimed at demonstrating the following. A 'folk theorem' states that a hard-decision function is the best memoryless relay at low SNR and linear amplification function is the best at high SNR. Using the particular model and its assumptions, we will argue that such is not always the case.

In section 3.3, in addition to assuming the relay function to be memoryless, we will restrict the number of stages to be finite. We will derive an explicit upper bound to mutual information assuming that each section of the serial relay channel is subject to energy and entropy constraints. The energy and entropy constraints, together, are proposed to characterize a set of operating conditions for the serial relay channel. We will illustrate the use of our upper bound numerically. To illustrate the use of our upper bound, we will contrast such an upper bound versus several lower bounds.

In section 3.4, we will relax the memoryless restriction on the serial relay model of Section 3.3, hence allowing relays to perform operations with memory. We will show two tradeoffs. First, given that information is sent at a fixed rate R and given the requirement that end-to-end error probability is bounded above by some arbitrary quantity, we will show that the requisite length of memory N for decoding

purpose at each stage is monotonically decreasing in the number of relay stages n. Second, given that information is sent at an achievable rate R, we will show the rate at which mutual information per unit memory tends to R is higher as the number of relay stages n grows.

Chapter 2

On the Capacity of a Parallel Relay Channel

In this chapter, we consider discrete-time models motivated by the discussion in Section 1.2.1, and address the problems brought about by such models.

To compute the capacity of point-to-point communication with distributed and intermediate processing at the relays, we will make use of the direct relation between *transmission capacity* and *estimation theory*. As we make use of such a relation, it will be apparent that achieving reliable transmission between sender and far-receiver relates to memory at intermediate nodes.

We will also propose a particular algorithm for distributed processing at the relays. The simplicity of such an algorithm is another merit which will be more apparent as the chapter progresses. Under the models we adopt, we will investigate whether point-to-point communication involving such an algorithm is optimal, namely whether the resulting capacity is equal to that of point-to-point communication under centralized processing.

In general, the distributed processing problem falls in the class of problems involving the observation of events in space-time and making estimates or pre-

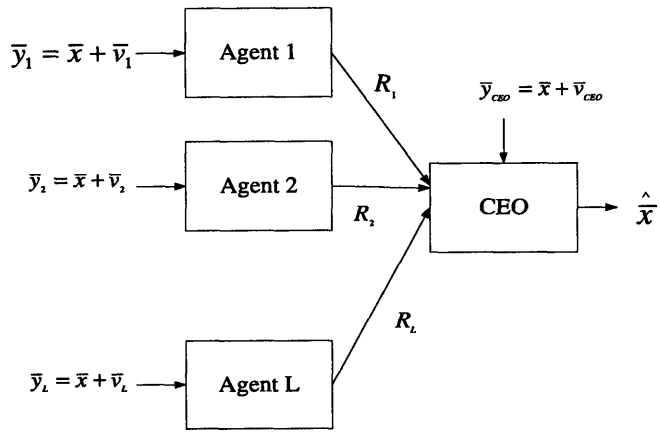

Figure 2.1: The CEO problem

dictions based on the phenomena observed. Often the receiver-estimator system is distributed in the sense that data are collected at several spatially separated sites and transmitted to a central estimator/decision maker over a communication channel with limited capacity.

A related model, in the context of source coding, is the *CEO* model shown in Figure 2.1. The CEO problem is a multi-terminal source coding problem. Its set-up is as follows: n agents receiving independently corrupted versions of an input process forward their corrupted observations to a central receiver—the CEO. There are however communication constraints associated with the communication between individual agents and the CEO: first, the total data rate at which the agents may communicate to the CEO is constrained by some finite quantity; second, the agents are not allowed to convene and collapse their corrupted observations. Under the communication constraints above, the CEO source coding problem is a hard problem. For the particular models used in this chapter, in contrast with the CEO model above, we assume that the links between relay nodes and far-receiver are noiseless, namely the links are not rate constrained.

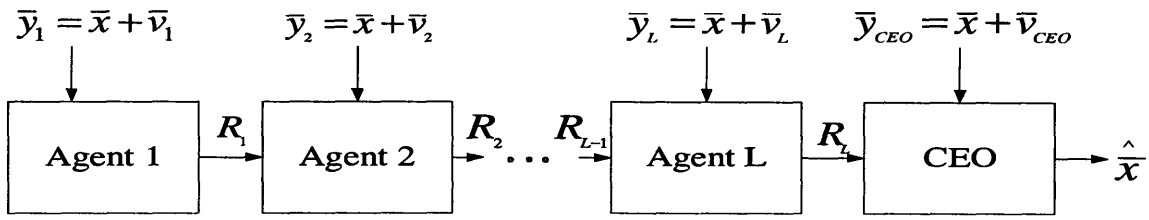

Figure 2.2: Successive encoding for the CEO problem

Reference [13] introduced and studied the CEO problem for discrete memoryless sources, contrasting distortion in the case when estimators convene for smoothing corrupted observations with that in the case when estimators do not convene. References [27] and [20] extends [13] to the special case of continuous Gaussian source and observation. More recently, [15] developed successive encoding strategies for the CEO problem, based on a generalization of Wyner-Ziv encoding. The successive encoding model for the CEO problem is shown in Figure 2.2. By successive encoding, estimators (or 'agents') are ordered and communicate—one to the next—over rate-constrained links, the final agent in the chain being termed as the CEO.

2.1 Channel Model: SIMO, AWGN

To address the issue of whether distributed processing yields the capacity equal to, or perhaps less than, that of centralized processing, we begin with the simplest model we can think of. As will become more obvious in the next section, the insights gained in this section will be of importance to the analysis of more involved models adopted in the next section

For the point-to-point communications shown in Figure 1.4, notably communication with centralized processing, and Figure 1.5, notably communication with

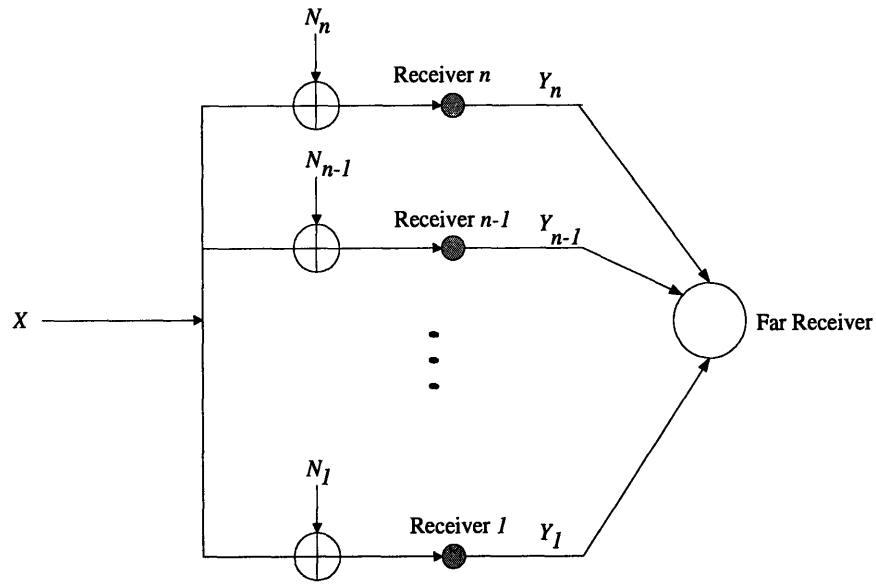

Figure 2.3: Channel under centralized processing

distributed processing, let us consider the information theoretic models shown in Figure 2.3 and Figure 2.4, respectively. Both channel models are discrete time AWGN channels with Single Input Multi Output (SIMO).

Such models, shown in Figure 2.3 and Figure 2.4, rely on the following two assumptions. First, the links between sender and the intermediate receivers are bandlimited and subject to AWGN. Second, the link between one intermediate receivers and the next, as well as the link between individual intermediate receiver and far-receiver, are noiseless.

The first assumption allows the channel models shown in Figure 2.3 and Figure 2.4 to be viewed as discrete time and continuous alphabet channels. The second assumption, in effect, allows the capacity problems associated with the communication models shown in Figure 2.3 and Figure 2.4 to be not as trivial as that of the serial AWGN relay channel shown in Figure 1.1 [6], but more tractable than that of the parallel AWGN relay channel shown in Figure 1.3 [47].

The sender transmits a symbol X with average energy constraint $E[X^2] \leq E$. There are n intermediate receivers, and the bandlimited channel between sender and intermediate receiver i, denoted as channel i, is corrupted by bandlimited AWGN N_i with average energy σ_i^2. Under the sampled model of the bandlimited channel, the relation between input symbol X and the output symbol at intermediate receiver i, Y_i, is given by

$$Y_i = X + N_i \qquad x = 1, \ldots, n \tag{2.1}$$

where

X a random variable with zero mean and variance E;

Y_i a random variable with zero mean and variance $(E + \sigma_i^2)$;

$N_i \sim \mathcal{N}(0, \sigma_i^2)$ for $i = 1, \ldots, n$ and N_i is independent of X and N_j, for $j \in \{1, 2, \ldots, n\} \setminus i$.

We consider two approaches for processing the Y_i's.

Option 1 : Centralized Processing

As Figure 2.3 shows, at each time step, intermediate receiver i, $i = 1, \ldots, n$, simultaneously relays its received signal, Y_i, $i = 1, \ldots, n$, to the far-receiver. Note the optimality of this approach, in account of the unlimited freedom that the far-receiver has, in processing the complete set of corrupted signals it receives.

Option 2 : Distributed Processing

As Figure 2.4 shows, this approach allows estimation to be done stage by stage at each intermediate receiver. In particular, intermediate receiver 1 receives Y_1, then produces $\hat{U}_1$, i.e., an estimate of X from Y_1, and transmits this estimate to intermediate receiver 2. Intermediate receiver 2 then produces $\hat{U}_2$, i.e., an estimate

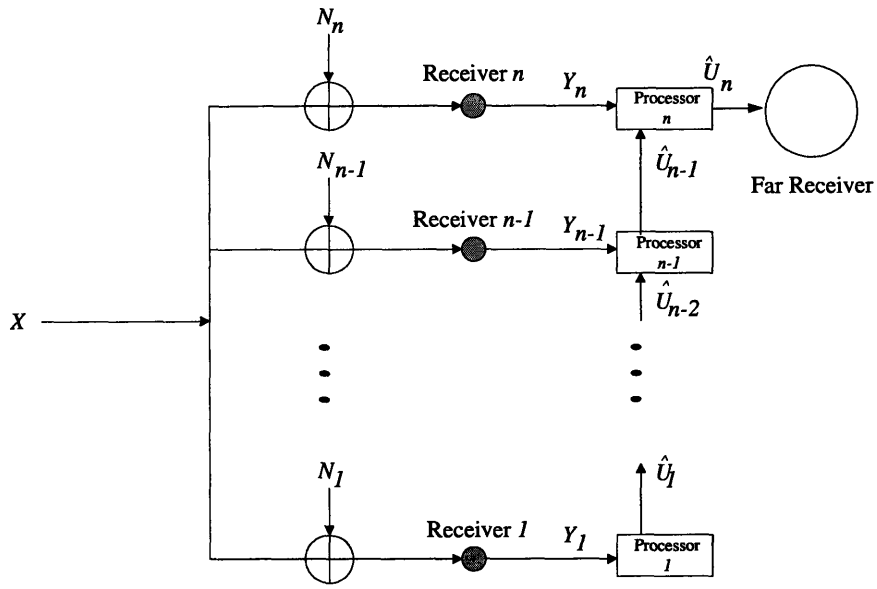

Figure 2.4: Channel under distributed processing.

of X from $\hat{U}_1$ and Y_2, and transmits the corresponding estimate to the subsequent intermediate receiver, i.e., intermediate receiver 3. In the same fashion processing carries on until the far-receiver receives an estimate, $\hat{U}_n$, i.e., an estimate of X from $\hat{U}_{n-1}$ and Y_{n-1}, from intermediate receiver n.

2.1.1 Distributed Processing by One Dimensional (1D) Kalman Filter

Since, at any time unit, it is possible to have a record of the measurements $\{Y_i\}_{i=1}^n$, distributed processing can be rephrased as the on-line estimation of the state variable of intermediate receiver n, i.e., $U_n = X$, from the measurements $\{Y_i\}_{i=1}^n$.

One such algorithm that yields an on-line unbiased linear least square error (LLSE) estimate of the state U_n is the 1D-Kalman filter [11], [28]. The estimate is

unbiased in the sense that

$$E[U_n - \hat{U}_n] = 0$$

where $E[\bullet]$ denotes expectation, and $\hat{U}_n = \hat{X}$ is the estimate of $U_n = X$ from $\{Y_i\}_{i=1}^n$, i.e., available measurements. The minimum error variance characteristic simply means that the quantity

$$E[(U_n - \hat{U}_n)^2]$$

can be minimized from the requirement that the estimate be the result of a linear operation on the available measurements [11], [28].

A particularly convenient form for the 1D-Kalman estimation algorithm can be developed in a recursive manner [11] [28], with a state-space model defined by state and output equations

$$U_{i+1} = U_i \tag{2.2}$$

$$Y_{i+1} = U_{i+1} + N_{i+1} \tag{2.3}$$

Note that the output equation (2.3) follows from (2.1) and (2.2).

Let us define

$\hat{U}[i|k]$ an LLSE estimate of U_i, based on observations from intermediate receiver 1 to intermediate receiver $k (k \leq i)$;

K_i receiver-varying Kalman gain (a scalar);

$\lambda_e[i|k] = E\left[(U_i - \hat{U}[i|k])^2\right]$, i.e., the error covariance, based on observations from intermediate receiver 1 to intermediate receiver $k (k \leq i)$;

σ_i^2 measurement noise variance at intermediate receiver i, i.e., $E[(N_i)^2]$;

The on-line 1D Kalman filtering algorithm is as follows.

1. Initialize the prediction and its associated error variance according to

$$\hat{U}[1|0] = 0 \tag{2.4}$$

$$\lambda_e[1|0] = E \tag{2.5}$$

and let $i = 1$.

2. If $i \leq n$ go to the next step, if $i > n$ then transmit $\hat{U}_n$ to the far-receiver and end process.

3. Port i computes the Kalman gain matrix

$$K_i = \lambda_e[i|i-1]\frac{1}{(\lambda_e[i|i-1] + \sigma_i^2)} \tag{2.6}$$

and generate the filtered estimate and its associated error covariance from the corresponding prediction quantities according to

$$\hat{U}[i|i] = \hat{U}[i|i-1] + K_i\left(Y_i - \hat{U}[i|i-1]\right) \tag{2.7}$$

$$\lambda_e[i|i] = \lambda_e[i|i-1] - K_i\lambda_e[i|i-1] \tag{2.8}$$

4. Port i generates the next prediction and its associated error covariance from the corresponding filtered quantities according to

$$\hat{U}[i+1|i] = \hat{U}[i|i] \tag{2.9}$$

$$\lambda_e[i+1|i] = \lambda_e[i|i] \tag{2.10}$$

5. Transmit $\hat{U}[i+1|i]$ and $\Lambda_e[i+1|i]$ to intermediate receiver $i+1$.

6. Increment i and go to step 2.

Moreover, let us define a new channel which is a twist from the *centralized* channel. Instead of having a channel with input X, and output $\bar{Y}$, i.e., the input-output

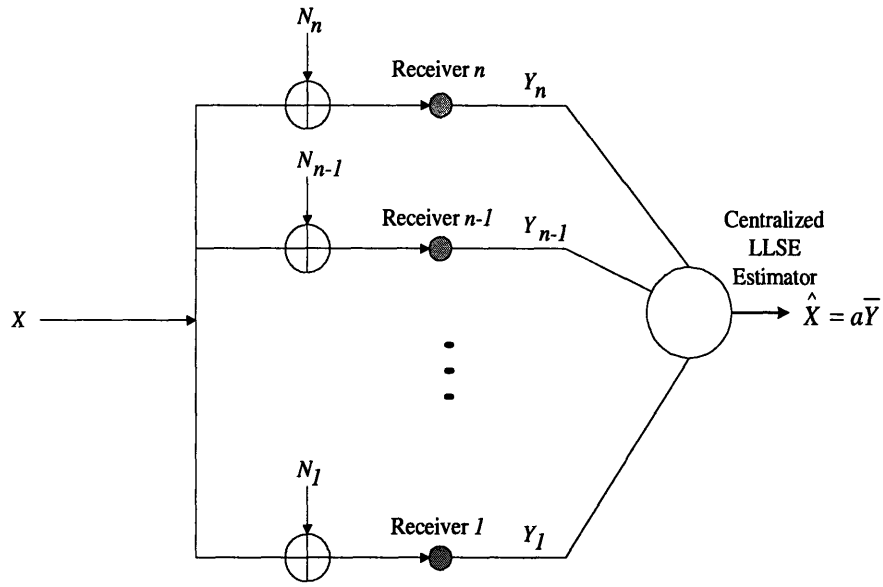

Figure 2.5: Reduced Channel

relation for the *centralized* channel, we reduce the channel to one which has input, X, and output, $\hat{X}$, where $\hat{X}$ is the Linear Least Square Error (LLSE) estimate of X, that is,

$$\hat{X} = \mathbf{a}\bar{Y}\,, \tag{2.11}$$

where $\mathbf{a}$ is a vector with the appropriate dimension such that $E\left[(X-\hat{X})^T(X-\hat{X})\right]$ is minimized. Let us name the channel with input and output relation in (2.11) as the *reduced* channel (see Figure 2.5).

The following theorem follows immediately.

Theorem 1 *Let $\hat{X}$ be the output of the* reduced *channel and $\hat{U}_n$ be an estimate attained by intermediate receiver n under distributed processing by 1D Kalman filter. Then*

$$\hat{X} = \hat{U}_n\,. \tag{2.12}$$

Theorem 1 asserts equality on the capacities of the following two channels: the reduced channel (see Figure 2.5) and the channel under distributed processing (see

Figure 2.4). However, the equality we desire is equality between the capacity of the channel under distributed processing and that of the channel under centralized processing (see Figure 2.3). What, therefore, remains to be shown is the equality that bridges the gap between Theorem 1 and the desired equality, i.e., the equality between the capacity of the channel under centralized processing and that of the reduced channel. We treat this problem in the following section.

2.1.2 Capacity of Channel under Distributed Vs. Centralized Processing

Consider the channel under *centralized* processing, i.e, a channel whose input X, noise $\bar{N} = \{N_i\}_{i=1}^n$, and output $\bar{Y} = \{Y_i\}_{i=1}^n$ are related by (2.1), with symbol energy constraint $E[X^2] \leq E$. Moreover, let X_G be a random variable such that $X_G \sim \mathcal{N}(0, E)$. Let $\mathbf{B} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T$, $\mathbf{N} = \begin{bmatrix} N_1 & N_2 & \cdots & N_3 \end{bmatrix}^T$ and

$$\bar{Y}_G = \mathbf{B}X_G + \mathbf{N}\,, \tag{2.13}$$

hence $\bar{Y}_G \sim \mathcal{N}(0, \Lambda_G)$, where Λ_G, a covariance matrix, is such that

$$|\Lambda_G| = E\begin{vmatrix} 1+\sigma_1^2 & 1 & \cdots & 1 \\ 1 & 1+\sigma_2^2 & \cdots & 1 \\ \vdots & 1 & \ddots & 1 \\ 1 & 1 & \ldots & 1+\sigma_n^2 \end{vmatrix}. \tag{2.14}$$

Note that

$$\begin{aligned} \mathbf{det}(\mathbf{cov}(\bar{Y})) &= \mathbf{det}(E[Y\bar{Y}^T]) \\ &= E[X^2]\begin{vmatrix} 1+\sigma_1^2 & 1 & \cdots & 1 \\ 1 & 1+\sigma_2^2 & \cdots & 1 \\ \vdots & 1 & \ddots & 1 \\ 1 & 1 & \ldots & 1+\sigma_n^2 \end{vmatrix} \\ &\leq |\Lambda_G| \end{aligned} \tag{2.15}$$

Finally let $\hat{X}_G$ be an LLSE estimate of X_G based on $\bar{Y}_G$ and $\lambda_e = E[(X_G - \hat{X}_G)^2]$, i.e., the estimation error variance.

Lemma 2

$$\lambda_e = \frac{1}{\frac{1}{E} + \sum_{k=1}^{n} \frac{1}{\sigma_k^2}}$$

Proof. From [11] and [28]

$$\lambda_e = \frac{1}{\left[\frac{1}{E} + \mathbf{B}^T E \left[\bar{N}\bar{N}^T\right]^{-1} \mathbf{B}\right]}$$

and with a few simple substitutions, the lemma follows. □

Using (2.13), (2.14), (2.15), and Lemma (2), we now find the maximum mutual information between the input X and output $\bar{Y}$ of the channel under centralized processing.

Lemma 3

$$\sup_{\bar{X}:E[X^2]\leq E} I(\bar{Y};X) = \frac{1}{2}\log\left(1 + \frac{E}{\frac{\Pi_{k=1}^{n}\sigma_k^2}{\sum_{i=1}^{n}\Pi_{k\in\{1,2,\ldots,n\}\setminus i}\sigma_k^2}}\right)$$

and the supremum is achieved by fixing $X = X_G \sim \mathcal{N}(0, E)$.

Proof.

$$\begin{aligned}
\sup_{\bar{X}:E[X^2]\leq E} I(\bar{Y};X) &\overset{a}{=} \sup_{\bar{X}:E[X^2]\leq E} [h(\bar{Y}) - h(\bar{Y}|X)] \\
&\overset{b}{\leq} h(\bar{Y}_G) - h(\bar{N}) \overset{c}{=} h(X_G) - h(X_G - \hat{X}_G|\bar{Y}_G) \\
&\overset{d}{=} \frac{1}{2}\log(2\pi e E) - \frac{1}{2}\log\left(2\pi e \frac{1}{\frac{1}{E} + \sum_{k=1}^{n}\frac{1}{\sigma_k^2}}\right) \\
&\overset{e}{=} \frac{1}{2}\log\left(1 + \frac{E}{\frac{\Pi_{k=1}^{n}\sigma_k^2}{\sum_{i=1}^{n}\Pi_{k\in\{1,2,\ldots,n\}\setminus i}\sigma_k^2}}\right)
\end{aligned}$$

(a) follows from the definition;

(b) combining (2.13), (2.14), and (2.15) , the entropy of $\bar{Y}$, given the constraint on the average energy of input symbol, is maximized by the entropy of $\bar{Y}_G$;

(c) follows from the definition and recall that (2.14) and (2.15) implies X_G is the random variable that satisfies (2.15) with equality;

(d) follows from Lemma 2;

(e) follows from simplifying the right hand side of (d).

From (2.13), the input that achieves the supremum on mutual information is $X_G \sim \mathcal{N}(0, E)$. □

Having found the maximum mutual information, the capacity of the channel under centralized processing is as follows.

Lemma 4 *Let C_c be the capacity of the channel under centralized processing, then*

$$C_c = \frac{1}{2}\log\left(1 + \frac{E}{\frac{\Pi_{k=1}^{n}\sigma_k^2}{\sum_{i=1}^{n}\Pi_{k\in\{1,2,\ldots,n\}\setminus i}\sigma_k^2}}\right)$$

Proof. Taking in account that the output of SIMO AWGN (*centralized* channel) is a vector instead of a scalar, and making the appropriate substitution, the proof is similar to the standard proof for Single Input Single Output (SISO) AWGN channel, [14, p.244–245]. □

We now establish an equality between the maximum mutual information of the channel under centralized processing, i.e., X and $\bar{Y}$, and that of the reduced channel, i.e., X and $\hat{X}$.

Lemma 5

$$\sup_{X:E[X^2]\leq E} I(X;\bar{Y}) = \sup_{X:E[X^2]\leq E} I(X;\hat{X})$$

Proof.

$$\begin{aligned}
\sup_{X:E[X^2]\leq E} I(X;\bar{Y}) &\stackrel{a}{=} I(X_G;\bar{Y_G}) \\
&\stackrel{b}{=} h(X_G) - h(\hat{X}_G|\bar{Y_G}) \\
&\stackrel{c}{=} h(X_G) - h(X_G - \hat{X}_G|\bar{Y_G}) \\
&\stackrel{d}{=} h(X_G) - h(X_G - \hat{X}_G) \\
&\stackrel{e}{=} h(X_G) - h(X_G - \hat{X}_G|\hat{X}_G) \\
&\stackrel{f}{=} I(X_G;\hat{X}_G) \\
&\stackrel{g}{=} \sup_{X:E[X^2]\leq E} I(X;\hat{X})
\end{aligned}$$

(a) follows from Lemma 3;

(b) follows from definition;

(c) follow since translation does not change differential entropy;

(d) follows from the fact that the error of an LLSE estimate is independent of the observed vector when the observation and the random variable to be estimated are jointly Gaussian;

(e) follows since translation does not change differential entropy;

(f) follows from definition;

(g) follows from data processing inequality, i.e., the fact that $\forall \bar{X} : E[X^2] \leq E$

$$\begin{aligned}
I(X;\hat{X}) &\leq I(X;\bar{Y}) \\
&\leq \sup_{X:E[X^2]\leq E} I(X;\bar{Y}).
\end{aligned}$$

□

In the following Lemma, we show an equality that is desirable, i.e., equality be-

tween the capacity of the channel under centralized processing and that of the reduced channel.

Lemma 6 *Let C_r be the capacity of the reduced channel and C_c be the capacity of the channel under centralized processing, then*

$$C_r = C_c = \frac{1}{2}\log\left(1 + \frac{E}{\frac{\Pi_{k=1}^{n}\sigma_k^2}{\sum_{i=1}^{n}\Pi_{k\in\{1,2,\ldots,n\}\setminus i}\sigma_k^2}}\right)$$

Proof. ($\Rightarrow$) Recall Lemma 3 and Lemma 5. Taking in account that the output of the reduced channel is $\hat{X}$, i.e., an LLSE estimate of $\{Y_i\}_{i=1}^{n}$, and making the apropriate changes, it is straight forward that the forward proof is similar to the forward proof of Lemma 4.

($\Leftarrow$) For the converse part, we modify the notations in (2.1) to account for the time sequence $\{j\}_{j=1}^{m}$ and henceforth follow the lines and notations of standard proof given in [14, pp. 246–247].

$$Y_{i_j} = X_j + N_{i_j},\ i = 1,\ldots,n\ j = 1,\ldots,m$$

where

Y_{i_j} the output symbol of intermediate receiver i at time j;

X_j the input symbol at time j;

N_{i_j} the additive noise on channel i at time j;

n the number of receiving intermediate receivers in the SIMO channel.

Let $\bar{Y}^m = \{\{Y_{i_j}\}_{i=1}^{n}\}_{j=1}^{m}$, $\bar{Y}_j = \{Y_{i_j}\}_{i=1}^{n}$, $\bar{X}^m = \{X_j\}_{j=1}^{m}$. Moreover let $\bar{N}^m = \{\{N_{i_j}\}_{i=1}^{n}\}_{j=1}^{m}$, $\bar{N}_j = \{N_{i_j}\}_{i=1}^{n}$, and $\hat{\bar{X}}^m$ be the LLSE estimate of $\bar{X}^m = \{\hat{X}_j\}_{j=1}^{m}$

when $\bar{Y}^m$ is observed. Now, consider any $(2^{mR}, m)$ code that satisfies the power constraint, i.e.,

$$\frac{1}{m}\sum_{j=1}^{m} X_j^2(w) \leq E$$

for the message index $w = 1, 2, \ldots, 2^{mR}$. Given that we can decode the index w from the LLSE vector estimate $\hat{\bar{X}}^m$ with low probability of error, we can apply Fano's inequality [14, p.38–40] to obtain

$$H(w|\hat{\bar{X}}^m) \leq 1 + mRP_e^{(m)} = m\epsilon_m, \tag{2.16}$$

where $\epsilon_m \to 0$ as $P_e^{(m)} \to 0$. Hence

$$\begin{aligned} mR &\overset{a}{=} H(w) \overset{b}{=} I(w; \hat{\bar{X}}^m) + H(w|\hat{\bar{X}}^m) \\ &\overset{c}{=} I(X^m; \hat{\bar{X}}^m) + m\epsilon_m \overset{d}{\leq} I(X^m; \bar{Y}^m) + m\epsilon_m) \\ &\overset{e}{=} h(\bar{Y}^m) - h(\bar{Y}^m|X^m) + m\epsilon_m \overset{f}{=} h(\bar{Y}^m) - h(\bar{N}^m) + m\epsilon_m \end{aligned}$$

$$\begin{aligned} &\overset{g}{\leq} \left(\sum_{j=1}^{m} h(\bar{Y}_j)\right) - h(\bar{N}^m) + m\epsilon_m \overset{h}{=} \left(\sum_{j=1}^{m} h(\bar{Y}_j)\right) - \left(\sum_{j=1}^{m} h(\bar{N}_j)\right) + m\epsilon_m \\ &\overset{i}{=} \left(\sum_{j=1}^{m} I(\bar{X}_j; \bar{Y}_j)\right) + m\epsilon_m \end{aligned} \tag{2.17}$$

(a) follows since w is a discrete random variable with uniform probability mass function;

(b) follows from definition;

(c) follows from data processing inequality and (2.16);

(d) follows from data processing inequality, i.e., $I(X; \hat{\bar{X}}^m) \leq I(X; \bar{Y}^m)$;

(e) follows from definition;

(f) follows since translation does not change differential entropy;

(g) follows from the chain rule for entropy and the fact that conditioning reduces entropy;

(h) follows from the chain rule for entropy;

(i) follows from definition;

Here $X_i = X_i(w)$, where w is drawn according to the uniform distribution on the set of message indices, $\{1,2,\ldots,2^{mR}\}$. Now, let E_j be the average power of i^{th} column of the codebook, i.e.,

$$E_j = \frac{1}{2^{mR}} \sum_{w} X_j^2$$

Continuing with the inequalities of the converse, we obtain

$$\begin{aligned} mR &\overset{a}{=} \sum_{j=1}^{m} \left[h(\bar{Y}_j) - h(\bar{N}_j)\right] + m\epsilon_m \\ &\overset{b}{=} \sum_{j=1}^{m} \left[\frac{1}{2}\log\left(1 + \frac{E_j}{\frac{\Pi_{k=1}^{n}\sigma_k^2}{\sum_{i=1}^{n}\Pi_{k\in\{1,2,\ldots,n\}\setminus i}\sigma_k^2}}\right)\right] + m\epsilon_m \end{aligned}$$

(a) follows from (2.17);

(b) follows from Lemma 3;

Since each of the codewords satisfies the power constraint, so does their average, and hence

$$\frac{1}{m}\sum_{j=1}^{m} E_j \leq E$$

Since $f(x) = \frac{1}{2}\ln(1+x)$ is a concave function of x, we can apply Jensen's inequality to obtain

$$\frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{2}\log\left(1+\frac{E_j}{\frac{\Pi_{k=1}^{n}\sigma_k^2}{\sum_{i=1}^{n}\Pi_{k\in\{1,2,\ldots,n\}\setminus i}\sigma_k^2}}\right)\right] \leq \frac{1}{2}\log\left[1+\frac{1}{m}\sum_{j=1}^{m}\frac{E_j}{\frac{\Pi_{k=1}^{n}\sigma_k^2}{\sum_{i=1}^{n}\Pi_{k\in\{1,2,\ldots,n\}\setminus i}\sigma_k^2}}\right]$$

$$\leq \frac{1}{2}\log\left[1+\frac{E}{\frac{\Pi_{k=1}^{n}\sigma_k^2}{\sum_{i=1}^{n}\Pi_{k\in\{1,2,\ldots,n\}\setminus i}\sigma_k^2}}\right]$$

Thus, $R \leq \frac{1}{2}\log\left[1+\frac{E}{\frac{\Pi_{k=1}^{n}\sigma_k^2}{\sum_{i=1}^{n}\Pi_{k\in\{1,2,\ldots,n\}\setminus i}\sigma_k^2}}\right] + m\epsilon_m$, $\epsilon_m \to 0$, and we have the required converse for the channel under distributed processing. □

Theorem 7 *Let C_d be the capacity of the channel under distributed processing, then*

$$C_d = \frac{1}{2}\log\left(1+\frac{E}{\frac{\Pi_{k=1}^{n}\sigma_k^2}{\sum_{i=1}^{n}\Pi_{k\in\{1,2,\ldots,n\}\setminus i}\sigma_k^2}}\right)$$

Proof. By Theorem 1 and Lemma 6, Theorem 7 follows immediately. □

2.2 Channel Model: AWGN, SIMO, ISI

Let us refer to Figure 1.4 and Figure 1.5. What if the channels between sender and the intermediate receivers are subject to many changes which cannot be controlled or predicted? Such is the case when obstacles appear and disappear in the environment between sender and intermediate receivers; the phenomenon leads to echoes (multipath).

The channel models which account for multipath effect in the point-to-point communication, shown in Figure 1.4 and Figure 1.5, are the Single Input Multi Output (SIMO) discrete time channels subject to AWGN and inter-symbol interference (ISI). Such models, shown in Figure 2.6 and 2.7, rely on the same assumptions

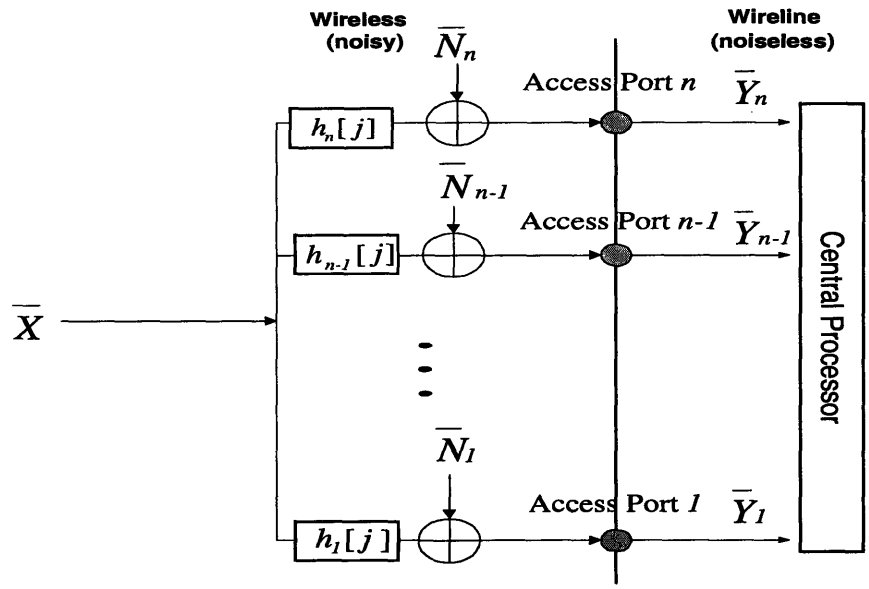

Figure 2.6: Channel under Centralized Processing

as those corresponding to the models in Section 2.1. One additional assumption, however, is that the channels are subject to ISI. ISI is perhaps inadequate for characterizing multipath since it only captures one ensuing aspect of multipath, namely non-ideal frequency responses of the links between sender and intermediate receivers. Representing multipath effect with ISI, nevertheless, allows for tractability and poses a structure that will yield insights into distributed processing of information at the intermediate receivers.

Let us look at the Single Input Multi Output (SIMO) channel under centralized processing, as shown in Figure 2.6, and distributed processing, as shown in Figure 2.7. A sender transmits a sequence of symbols, $\bar{X} = \{X_j\}_{j=1}^{\infty}$. The output process at each intermediate receiver is the input process which is corrupted by a bandlimited AWGN process. Under the sampled model of the bandlimited channel, the relation

between the input and output process at each intermediate receiver is

$$Y_{i_j} = \sum_{k=0}^{M_i} h_{i_k} X_{j-k} + N_{i_j} \quad i = 1, \ldots, n \quad j = 1, \ldots, \infty \tag{2.18}$$

where

Y_{i_j} be the output at intermediate receiver i and time j;

X_j be the zero mean input at time j satisfying the average symbol energy constraint;

$$E[X_j^2] \leq E, \; i = 1, \ldots, \infty; \tag{2.19}$$

$\{h_{i_k}\}_{k=0}^{M_i}$ be the finite impulse response (FIR) (with memory M_i) of the ISI corresponding to intermediate receiver i;

N_{i_j} be an additive white Gaussian noise with mean zero and average energy $E[N_{i_j}^2] = \sigma^2$ for all i, j.

Note that the same average noise energy for all i, j, i.e., $E[N_{i_j}^2] = \sigma^2$, is not a limiting assumption. In the case where $E[N_{i_j}^2]$ varies with i and j, it can always be normalized to σ^2 by some appropriate scaling of the FIR filter $\{h_{i_j}\}_{j=0}^{M_i}$ for all i, j.

In the same spirit as that of the previous section, we seek to find whether distributed processing shown in Figure 2.6 leads to the same capacity as that of centralized processing shown in Figure 2.7.

2.2.1 Distributed Processing by Two Dimensional (2D) Kalman Filter

As Figure 2.7 shows, a problem with the channel under distributed processing by 1D Kalman filter, when channel is subject to ISI, is as follows. For each intermediate receiver to compute distributed and bottom-up linear least square error (LLSE) estimates of the input process $\{X_j\}_{j=1}^{\infty}$, the intermediate receiver i $(i = 1, \ldots, n)$ is

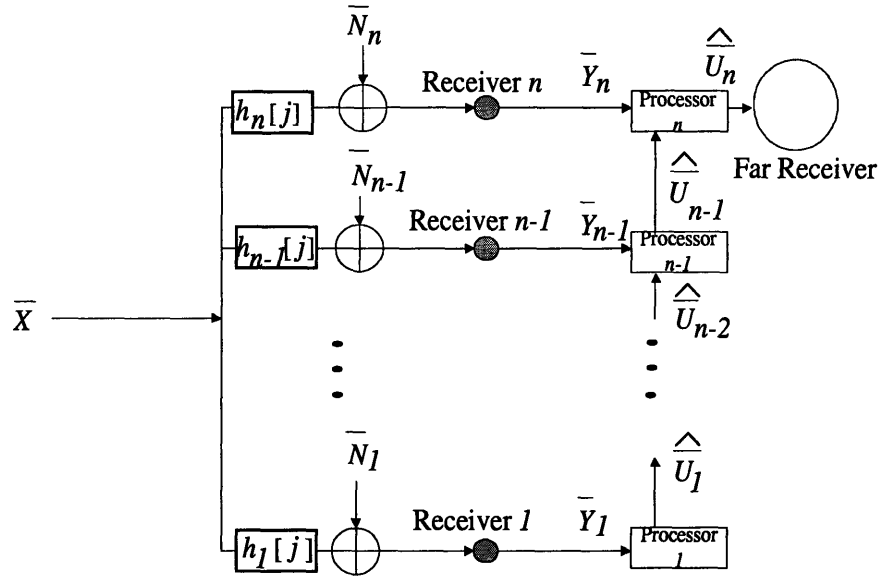

Figure 2.7: Channel under Distributed Processing by 1D Kalman Filter

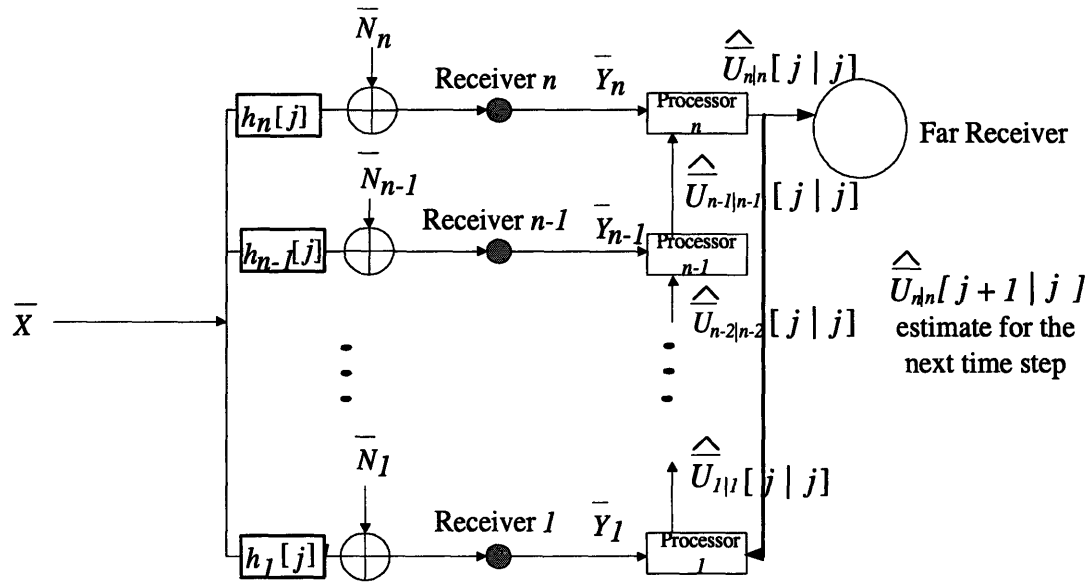

Figure 2.8: Channel under Distributed Processing by 2D Kalman Filter

forced to store an infinite sequence of the observed output process $\{Y_{i_j}\}_{j=1}^{\infty}$. With an infinite sequence of observed output process being stored at each intermediate receiver, one dimensional receiver-by-receiver Kalman filtering is then performed from the bottom intermediate receiver to the top one, thus attaining the final estimate, i.e., $\{\hat{X}_j\}_{j=1}^{\infty}$. The implications are twofold. First, infinite memory for holding an infinite sequence of the observed output process is required at individual intermediate receivers. Second, there is an infinite delay before $\{\hat{X}_j\}_{j=1}^{\infty}$, i.e., the estimate of $\{X_j\}_{j=1}^{\infty}$, can finally be sent to the far-receiver.

To allow for finite processing memory and delay, we consider the case when the channel under distributed processing performs two dimensional (2D) Kalman filtering (see Figure 2.8) in order to compute the LLSE estimate of the input process.

By a 2D Kalman filter, the channel under distributed processing performs two dimensional filtering operation in time and space. With some finite processing memory, equal to $\max_i M_i$, i.e., the maximum memory among all the FIR filters representing ISI, estimation is performed receiver-by-receiver from the bottom intermediate receiver to the top one, i.e., intermediate receiver 1 to intermediate receiver n, respectively. However, at each time step, intermediate receiver n feeds back its estimate to the first intermediate receiver, allowing the estimate of the next time step to be based on the estimate of the present time step. Note that the finite processing memory requirement for 2D Kalman filtering entails a time spread (processing delay), of the same amount as the processing memory, in the final step of estimating the input process at the far-receiver side.

We will show that 2D Kalman filter yields the same LLSE estimate of the input process as that which is resulted from the 1D Kalman filter when an infinite sequence of observed output process is in store at each intermediate receiver.

Let us return to the model shown in Figure 2.8. Such a model of distributed processing under 2D Kalman filter corresponds to the point-to-point and distributed communication shown in Figure 2.9. Processing at the intermediate nodes is not

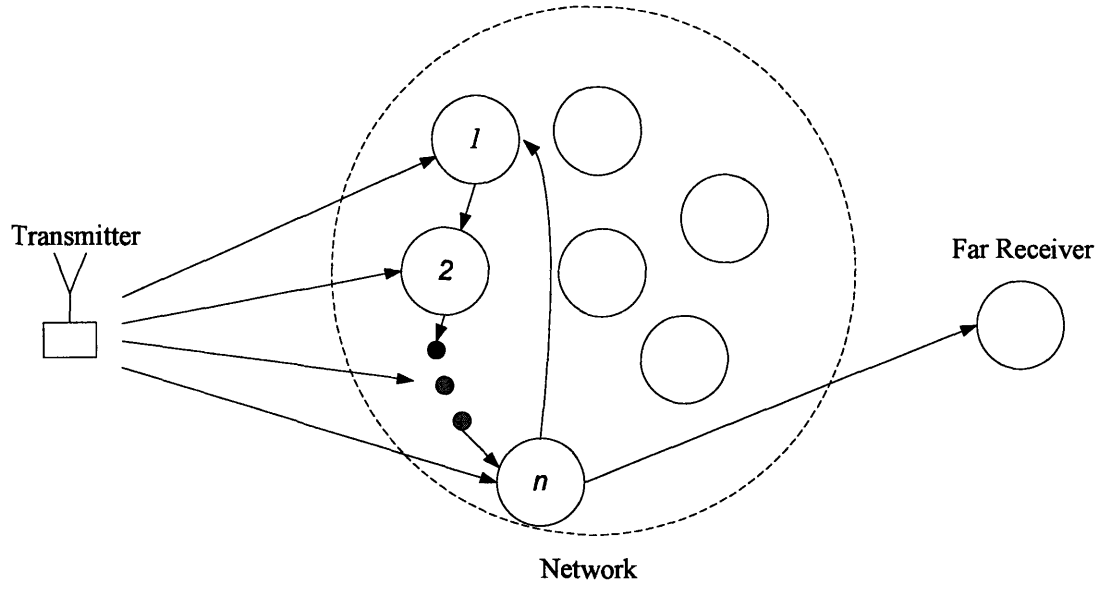

Figure 2.9: Distributed Processing in Time and Space

only distributed in *time*, but also in *space*.

Similar to distributed processing in space alone, distributed processing in time and space is localized within the intermediate nodes; the far-receiver is excluded from taking part in the intermediate processing. However, there is a difference. Under distributed processing in time and space, intermediate receiver n feeds back its output stream to intermediate receiver 1, at each unit of time, besides sending the same stream to the far-receiver. We will show that this feedback allows processing with finite memory at the intermediate receivers.

2.2.2 A Base Model for Estimation

As (2.18), i.e., the sampled model of a bandlimited channel, suggests, the base model chosen to represent the relation between input and observed output of the

channel corresponding to intermediate receiver i ($i = 1, \ldots, n$) is

$$Y_{ij} = \sum_{k=0}^{M} h_{ik} X_{j-k} + N_{ij} \tag{2.20}$$

where

Y_{ij} be the output at intermediate receiver i and time j;

X_j be the input at time j such that $\{X_j\}_{j=1}^{\infty}$ forms a white wide sense stationary (WSS) input process with zero mean and unit variance;

$\{h_{ik}\}_{k=0}^{M_i}$ be the finite impulse response (with memory M_i) of the ISI corresponding to intermediate receiver i;

M be $\max_i M_i$, i.e., the maximum memory among all FIR filters representing ISI, or may well be interpreted as the ensuing time spread (processing delay) in the estimation of input process;

N_{ij} be a bandlimited AWGN such that $\{N_{ij}\}_{j=1}^{\infty}$ a white Gaussian noise process with mean zero and average energy $E[N_{ij}{}^2] = \sigma^2$ for all i, j.

Moreover, we assume that $X_{-M} = \ldots = X_0 = 0$.

Denote $\bar{U}_{ij}$ to be the vector of states at time j for the channel associated to intermediate receiver i. Then (2.20) suggests that the dynamics of the input process can be represented by the state equation

$$\bar{U}_{i(j+1)} = \mathbf{F}\bar{U}_{ij} + \mathbf{G}X_{j+1} \tag{2.21}$$

$$\bar{U}_{(i+1)j} = \bar{U}_{ij} \tag{2.22}$$

where

$$\bar{U}_{i(j+1)} = \begin{bmatrix} X_{j+1} & X_j & \cdots & X_{j-M+1} \end{bmatrix}^T$$

$$\bar{U}_{ij} = \begin{bmatrix} X_j & X_{j-1} & \cdots & X_{j-M} \end{bmatrix}^T$$

$$\bar{U}_{(i+1)j} = \begin{bmatrix} X_j & X_{j-1} & \cdots & X_{j-M} \end{bmatrix}^T$$

and

$$\mathbf{F} = \begin{bmatrix} 0 & \cdots & \cdots & \cdots & 0 \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & & \\ 0 & \cdots & 0 & 1 & 0 \end{bmatrix}, \mathbf{G} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.23}$$

Relating (2.20) to (2.21) and (2.22), the channel output at time j can be described in terms of the state vector at time j

$$Y_{ij} = \mathbf{C_i}\bar{U}_{ij} + N_{ij} \tag{2.24}$$

where $\mathbf{C_i}$ is an $1 \times (M+1)$ matrix, i.e., $\mathbf{C_i} = \begin{bmatrix} h_{i0} & h_{i1} & \cdots & h_{iM} \end{bmatrix}$.

We now state the estimation problem: given the foregoing channel model and the statistics of the input signal and the measurement noise, we desire to obtain an on-line estimation procedure that yields an LLSE estimate of the symbol X_j at some delayed time $(j+M)$.

2D-Kalman Filtering Algorithm

Since, at any time j, it is possible to have a record of the measurements $\{Y_{i1}\}_{i=1}^{n}$, $\{Y_{i2}\}_{i=1}^{n}, \ldots, \{Y_{ij}\}_{i=1}^{n}$, the preceding estimation problem can be rephrased to be the on-line estimation of the $(M+1)$ components of the state vector of intermediate receiver n at time j, i.e., $\bar{U}_{nj} = \begin{bmatrix} X_j & X_{j-1} & \cdots & X_{j-M} \end{bmatrix}^T$, from the measurements $\{Y_{i1}\}_{i=1}^{n}, \{Y_{i2}\}_{i=1}^{n}, \ldots, \{Y_{ij}\}_{i=1}^{n}$.

One such algorithm that yields an on-line unbiased LLSE estimate, i.e., $\hat{\bar{U}}_{nj} = \begin{bmatrix} \hat{X}_j & \hat{X}_{j-1} & \cdots & \hat{X}_{j-M} \end{bmatrix}^T$, of the complete state vector $\bar{U}_{nj}$ from the available measurements at time j, i.e., $\{Y_{i1}\}_{i=1}^{n}$, $\{Y_{i2}\}_{i=1}^{n}$ is the 2D-Kalman filter [11], [28].

A particularly convenient form for the 2D-Kalman estimation algorithm can be developed in a recursive manner [11], [28]. More precisely, for the estimation model defined by (2.21), (2.22), and (2.24), let us define

$\hat{\bar{U}}_{i|t}[j|k]$ an LLSE estimate of $\bar{U}_{ij}$, based on observations from intermediate receiver 1 to intermediate receiver $t(t \leq i)$ which spans from time 1 to time $k(k \leq j)$, i.e., $\{\{Y_{p\,q}\}_{p=1}^{l}\}_{q=1}^{k}$;

$\bar{K}_{ij}$ time and receiver-varying Kalman gain (an $(M+1) \times 1$ vector);

$\mathbf{\Lambda_{e_{i|t}}}[j|k]$ error covariance matrix, based on observations from intermediate receiver 1 to intermediate receiver $t(t \leq i)$ which spans from time 1 to time $k(k \leq j)$, $\{\{Y_{p\,q}\}_{p=1}^{l}\}_{q=1}^{k}$, i.e.,
$E\left[(\bar{U}_{ij} - \hat{\bar{U}}_{i|t}[j|k])(\bar{U}_{ij} - \hat{\bar{U}}_{i|t}[j|k])^T\right]$;

σ^2 measurement noise covariance = $E[(N_{ij})^2]$ for all i,j;

$\mathbf{I}, \mathbf{0}$ the $(M+1) \times (M+1)$ identity matrix and the column vector with all its $(M+1)$ components being 0, respectively.

The algorithm is as follows.

1. Initialize the prediction and its associated error variance according to

$$\hat{\bar{U}}_{1|1}[1|0] = \mathbf{0} \tag{2.25}$$

$$\mathbf{\Lambda_{e_{1|1}}}[1|0] = \mathbf{I} \tag{2.26}$$

   and let $j = 1$.

2. Let $i = 1$

3. Port 1 computes the Kalman gain matrix

$$\bar{K}_{1_j} = \mathbf{\Lambda_{e_{1|1}}}[j|j-1]\mathbf{C_1}^T \left(\mathbf{C_i}\mathbf{\Lambda_{e_{1|1}}}[j|j-1]\mathbf{C_1}^T + \sigma^2\right)^{-1} \tag{2.27}$$

and generate the filtered estimate and its associated error covariance from the corresponding prediction quantities according to

$$\hat{U}_{1|1}[j|j] = \hat{U}_{1|1}[j|j-1] + \bar{K}_{1_j}\left(Y_{1_j} - \mathbf{C_1}\hat{U}_{1|1}[j|j-1]\right) \quad (2.28)$$

$$\mathbf{\Lambda}_{\mathbf{e}_{1|1}}[j|j] = \mathbf{\Lambda}_{\mathbf{e}_{1|1}}[j|j-1] - \bar{K}_{1_j}\mathbf{C_1}\mathbf{\Lambda}_{\mathbf{e}_{1|1}}[j|j-1] \quad (2.29)$$

4. While $i \leq n$ perform as follows. If $i > n$ go to step 5.

   (a) Port i generates the next prediction and its associated error covariance from the corresponding filtered quantities according to

   $$\hat{U}_{i+1|i}[j|j] = \hat{U}_{i|i}[j|j] \quad (2.30)$$

   $$\mathbf{\Lambda}_{\mathbf{e}_{\mathbf{i+1|i}}}[j|j] = \mathbf{\Lambda}_{\mathbf{e}_{\mathbf{i|i}}}[j|j] \quad (2.31)$$

   (b) Transmit $\hat{U}_{i+1|i}[j|j]$ and $\mathbf{\Lambda}_{\mathbf{e}_{\mathbf{i+1|i}}}[j|j]$ to intermediate receiver $i+1$.

   (c) Increment i.

   (d) Port i computes the Kalman gain matrix

   $$\bar{K}_{ij} = \mathbf{\Lambda}_{\mathbf{e}_{\mathbf{i|i-1}}}[j|j]\mathbf{C_i}^T\left(\mathbf{C_i}\mathbf{\Lambda}_{\mathbf{e}_{\mathbf{i|i-1}}}[j|j]\mathbf{C_i}^T + \sigma^2\right)^{-1} \quad (2.32)$$

   and generate the filtered estimate and its associated error covariance from the corresponding prediction quantities according to

   $$\hat{U}_{i|i}[j|j] = \hat{U}_{i|i-1}[j|j] + \bar{K}_{ij}\left(Y_{ij} - \mathbf{C_i}\hat{U}_{i|i-1}[j|j]\right) \quad (2.33)$$

   $$\mathbf{\Lambda}_{\mathbf{e}_{\mathbf{i|i}}}[j|j] = \mathbf{\Lambda}_{\mathbf{e}_{\mathbf{i|i-1}}}[j|j-1] - \bar{K}_{ij}\mathbf{C_i}\mathbf{\Lambda}_{\mathbf{e}_{\mathbf{i|i-1}}}[j|j-1] \quad (2.34)$$

   (e) Go to step 4.

5. Port n generates the next prediction and its associated error covariance from the corresponding filtered quantities according to

$$\hat{U}_{1|1}[j+1|j] = \mathbf{F}\hat{U}_{n|n}[j|j] \quad (2.35)$$

$$\mathbf{\Lambda}_{\mathbf{e}_{1|1}}[j+1|j] = \mathbf{F}\mathbf{\Lambda}_{\mathbf{e}_{\mathbf{i|i}}}[j|j]\mathbf{F}^T + \mathbf{G}\mathbf{G}^T \quad (2.36)$$

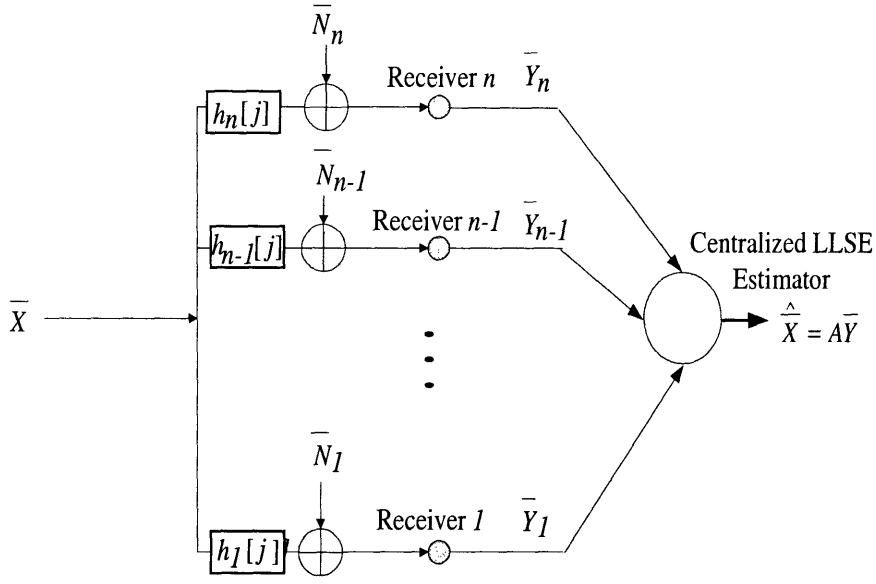

Figure 2.10: Reduced Channel

6. Transmit $\hat{U}_{1|1}[j+1|j]$ and $\mathbf{\Lambda}_{\mathbf{e}_{1|1}}[j+1|j]$ to intermediate receiver 1.

7. Increment j and go to step 2.

We shall refer to the channel whose input-output relationship is described by (2.18) as the channel under centralized processing. Now, let

$\bar{Y}$ be the column vector of output process $\{\{Y_{ij}\}_{j=1}^{\infty}\}_{i=1}^{n}$, i.e., infinite sequence of output symbols at intermediate receiver i;

$\bar{X}$ be the column vector of the zero mean input process $\{X_j\}_{j=1}^{\infty}$.

Moreover, let us define a new channel which is a twist from the channel under centralized processing. Instead of having a channel with input, $\bar{X}$, and output $\bar{Y}$, i.e., the input-output relationship for the channel under centralized processing, we reduce the channel to one which has input, $\bar{X}$, and output, $\hat{\bar{X}}$, where $\hat{\bar{X}}$ is the LLSE estimate of $\bar{X}$, that is,

$$\hat{\bar{X}} = \mathbf{A}\bar{Y}\,, \tag{2.37}$$

where **A** is a matrix with the appropriate dimension such that $E\left[(\bar{X}-\hat{\bar{X}})^T(\bar{X}-\hat{\bar{X}})\right]$ is minimized. Let us name the channel with input and output relation in (2.37) as the *reduced* channel (see Figure 2.10).

The following theorem follows immediately.

Theorem 8 *Let $\hat{\bar{X}}$ be the output of the reduced channel and $\hat{\bar{U}}_{nj}^{M+1}$ is the $(M+1)$-th component of the state vector estimate attained by intermediate receiver n at time j, i.e., $\bar{U}_{nj}$, under distributed processing by 2D Kalman filter. Then*

$$\hat{\bar{X}} = \{\hat{\bar{U}}_{nj}^{M+1}\}_{j=M+1}^{\infty} \tag{2.38}$$

Theorem 8 asserts equality on the capacities of the following two channels: the reduced channel (see Figure 2.10) and the channel under distributed processing (see Figure 2.8). However, the equality we desire is equality between the capacity of the channel under distributed processing and that of the channel under centralized processing (see Figure 2.6). What, therefore, remains to be shown is the equality that bridges the gap between Theorem 8 and the desired equality, i.e., the equality between the capacity of the channel under centralized processing and that of the reduced channel. We treat this problem in the following section.

2.2.3 Capacity of Channel under Distributed Vs. Centralized Processing

Before showing the desired equality, i.e., the equality between the capacity of the channel under centralized processing and that of the reduced channel, we must first find an expression for the capacity of the channel under centralized processing.

Theorem 9 *Let C_c be the capacity of the channel under centralized processing. Then*

$$C_c = (2\pi)^{-1}\int_0^{\pi}\log\left[\max\left(\Theta\frac{\Pi_{i=1}^{n}|H_i(\lambda)|^{-2}}{\sum_{k=1}^{n}\Pi_{i\in\{1,2,\ldots,n\}\setminus i}|H_i(\lambda)|^{-2}},1\right)\right]d\lambda \tag{2.39}$$

where $H_i(\lambda)$ is the channel transfer function given by

$$H_i(\lambda) = \sum_{j=0}^{M_i} H_{ij} \exp^{-lj\lambda}, \quad l = \sqrt{-1} \tag{2.40}$$

(periodic in λ with period 2π) and where the parameter θ is the solution of

$$\int_{0(H_i(\lambda)\neq 0\ \forall i,j)}^{\pi} \max\left(\Theta - \frac{\Pi_{i=1}^{n} |H_i(\lambda)|^{-2}}{\sum_{k=1}^{n} \Pi_{i\in\{1,2,\ldots,n\}\setminus i} |H_i(\lambda)|^{-2}}, 0\right) d\lambda = \frac{\pi E}{\sigma^2} \tag{2.41}$$

Moreover, the capacity-achieving q_N, the inputs $\{X_j\}_{j=\infty}^{\infty}$, are correlated Gaussian random variables with mean zero and covariances r_n, $-\infty \leq n \leq \infty$, given by

$$r_n = E[X_{k+n}X_k] = (\pi)^{-1} \int_0^{\pi} S_X(\lambda) \cos(n\lambda) d\lambda \tag{2.42}$$

where the input power spectral density satisfies

$$S_X(\lambda) = \begin{cases} \sigma^2(\theta - K(\lambda)^{-2}), & \theta K(\lambda)^2 > 1 \quad |\lambda| \leq \pi, \\ 0, & otherwise \end{cases} \tag{2.43}$$

with

$$K(\lambda) = \frac{\Pi_{i=1}^{n} |H_i(\lambda)|^{-2}}{\sum_{k=1}^{n} \Pi_{i\in\{1,2,\ldots,n\}\setminus i} |H_i(\lambda)|^{-2}}$$

In particular, capacity is achieved when all inputs X_j, $-\infty \leq j \leq \infty$, have the same average energy $E[X_j^2] = r_0 = E$.

Proof. The proof is an extension on the proof for the capacity of the single input and single output (SISO) discrete-time Gaussian channel with intersymbol interference. An approach that can be taken in proving the theorem is the eigenvalue decomposition approach [17, 26]. However, we take on an approach that utilizes decomposition technique on the dicrete Fourier transform (DFT) domain [18]. We will only show the necessary modification and the rest follows immediately from [18].

Consider the N-circular Gaussian Channel (NCGC) defined by [18]. Then discrete Fourier transform (DFT) on the NCGC channel which is derived from (2.18),

followed by an appropriate orthogonal decomposition (described in [18]) on the transform domain, yields

$$Y'_{ij} = X'_j + N'_{ij} \quad i = 1, \ldots, n \quad j = 0, \ldots, (K-1) \tag{2.44}$$

with block-energy constraint becomes, in the transform domain,

$$\sum_{j=0}^{K-1} E[X'_j] \leq K^2 E \tag{2.45}$$

and the N'_{ij} in (2.44) are statistically independent Gaussian random variables with mean zero and variance

$$\sigma_{ij}^2 = K\sigma^2 \left|\tilde{h}_{ij}\right|^{-2}, \quad i = 1, \ldots, n \quad j = 0, \ldots, (K-1) \tag{2.46}$$

From (2.44), there are n SIMO channels for every index j, for which the input and output relation is described by

$$Y'_{ij} = X'_j + N'_{ij} \quad i = 1, \ldots, n \tag{2.47}$$

Recalling Lemma 3 and Lemma 4, for every j, we have an equivalent channel, i.e.,

$$Y'_j = X'_j + N'_j \quad j = 1, \ldots, K \tag{2.48}$$

where the N'_j in (2.48) are statistically independent Gaussian random variables with mean zero and variance

$$\sigma_j^2 = K\sigma^2 \frac{\Pi_{i=1}^n \left|\tilde{h}_{ij}\right|^{-2}}{\sum_{k=1}^n \Pi_{i \in \{1,2,\ldots,n\} \setminus i} \left|\tilde{h}_{ij}\right|^{-2}} \tag{2.49}$$

Thus it follows from (2.48) that the equivalent transform domain channel model for the NCGC is a set of K parallel discrete memoryless additive Gaussian noise channels where the channel inputs $X'_j, j = 0, \ldots, (K-1)$, satisfy (3.23). This equivalence implies

$$\sup_{q_K} I\left(\{X_j\}_{j=0}^{K-1}; \{\{\tilde{Y}_{ij}\}_{i=1}^n\}_{j=0}^{K-1}\right) = \sup_{Q_K} I\left(\{X'_j\}_{j=0}^{K-1}; \{\{Y'_{ij}\}_{i=1}^n\}_{j=0}^{K-1}\right) \tag{2.50}$$

where Q_K is the class of probability densities for $\{X_j\}_{j=0}^{K-1}$ satisfying block-energy constraint (3.23). To write (2.50), we have made use of the fact that the average mutual information between two sequences is invariant to any succession of reversible transformations of one or both of the sequences [16, p.30]. Application of a theorem [16, Theorem 7.5.1] to this transform-domain channel provides the solution for C_c, i.e., the capacity of SIMO NCGC channel, in the form of the parametric expression

$$\tilde{I}_K(E) = (2K)^{-1} \sum_{j=0}^{K-1} \log \left[\max \left(\Theta \frac{\Pi_{i=1}^{n} \left|\tilde{h}_{ij}\right|^{-2}}{\sum_{k=1}^{n} \Pi_{i \in \{1,2,\ldots,n\} \setminus i} \left|\tilde{h}_{ij}\right|^{-2}}, 1 \right) \right] \tag{2.51}$$

where the parameter θ is the solution of

$$\sum_{j=0}^{K-1} \max \left(\Theta - \frac{\Pi_{i=1}^{n} \left|\tilde{h}_{ij}\right|^{-2}}{\sum_{k=1}^{n} \Pi_{i \in \{1,2,\ldots,n\} \setminus i} \left|\tilde{h}_{ij}\right|^{-2}}, 0 \right) = \frac{KE}{\sigma^2} \tag{2.52}$$

With (2.52), the modification, through which the capacity result in [18] extends directly to the centralized SIMO channel, is complete. □

We now establish an equality between the maximum mutual information of the channel under centralized processing, i.e., $\bar{X}$ and $\bar{Y}$, and that of the reduced channel, i.e., $\bar{X}$ and $\hat{\bar{X}}$.

Let $\bar{X}_G = \{X_j^G\}_{j=1}^{\infty}$ be the input sequence that achieves the resulted capacity in Theorem 9, $\bar{Y}_G = \{\{Y_{ij}^G\}_{j=1}^{\infty}\}_{i=1}^{n}$ is the output sequence which satisfies

$$Y_{ij}^G = \sum_{k=0}^{M} h_{ik} X_{j-k}^G + N_{ij},$$

and $\hat{\bar{X}}_G = \{\hat{X}_j^G\}_{j=1}^{\infty}$ be the estimate of $\bar{X}_G$ based on the observation $\bar{Y}_G$.

Lemma 10

$$\sup_{\bar{X}: E[X_i^2] \le E, \forall i} I(\bar{X}; \bar{Y}) = \sup_{\bar{X}: E[X_i^2] \le E, \forall i} I(\bar{X}; \hat{\bar{X}})$$

Proof.

$$\begin{aligned}
\sup_{\bar{X}:E[X_i^2]\leq E,\forall i} I(\bar{X};\bar{Y}) &\stackrel{a}{=} I(\bar{X}_G;\bar{Y}_G) \\
&\stackrel{b}{=} h(\bar{X}_G) - h(\bar{X}_G|\bar{Y}_G) \\
&\stackrel{c}{=} h(\bar{X}_G) - h(\bar{X}_G - \hat{\bar{X}}_G|\bar{Y}_G) \\
&\stackrel{d}{=} h(\bar{X}_G) - h(\bar{X}_G - \hat{\bar{X}}_G) \\
&\stackrel{e}{=} h(\bar{X}_G) - h(\bar{X}_G - \hat{\bar{X}}_G|\hat{\bar{X}}_G) \\
&\stackrel{f}{=} I(\bar{X}_G;\hat{\bar{X}}_G) \\
&\stackrel{g}{=} \sup_{\bar{X}:E[X_i^2]\leq E,\forall h} I(\bar{X};\hat{\bar{X}})
\end{aligned}$$

(a) follows from Lemma 9;

(b) follows from definition;

(c) follow since translation does not change differential entropy;

(d) follows from the fact that the error of an LLSE estimate is independent of the observed vector when the observation and the random vector to be estimated are jointly Gaussian;

(e) follows since translation does not change differential entropy;

(f) follows from definition;

(g) follows from data processing inequality, i.e., the fact that $\forall \bar{X} : E[X_i^2] \leq E, \forall i$,

$$\begin{aligned}
I(\bar{X};\hat{\bar{X}}) &\leq I(\bar{X};\bar{Y}) \\
&\leq \sup_{\bar{X}:E[X_i^2]\leq E,\forall i} I(\bar{X};\bar{Y}) .
\end{aligned}$$

□

Having found an equality between the maximum mutual information of the channel under centralized processing with that of the reduced channel, the following lemma follows.

Lemma 11 *Capacity of reduced channel, C_r, is equal to capacity of the channel under centralized processing, C_c.*

Proof. Recall Theorem 9 and Lemma 10. The fact that LLSE estimation in the time domain maps to LLSE estimation in the discrete Fourier transform (DFT) domain allows the coding theorem in Theorem 9 and Theorem 6 to apply, hence Lemma 11 follows immediately. □

Theorem 12 *Capacity of distributed channel, C_d, is equal to capacity of channel under centralized processing, C_c.*

Proof. By Theorem 8 and Lemma 11, Theorem 12 follows immediately. □

Note that, from (2.43), the capacity-achieving input process is a colored Gaussian process. On the other hand, distributed processing by 2D Kalman filter (Section 3.2.1) stipulates the input process to be white, i.e, a sequence of independent random variables. Thus, estimating the capacity-achieving input process by means of 2D Kalman filter requires a causal shaping filter. With a shaping filter, a unit energy white Gaussian process is shaped into a zero mean and colored Gaussian process with power spectral density as given in (2.43). In what follows, we assert the existence of such a filter.

Lemma 13 *For any colored wide sense stationary process (WSS) $\{V_j\}_{j=1}^{\infty}$, there exists a causal filter with memory H and impulse response $\{g_i\}_{i=0}^{H}$. Moreover,*

$$V_j = \sum_{k=0}^{H} g_k X_{j-k}$$

and $\{X_j\}_{j=1}^{\infty}$ is a white WSS process with $E\left[(X_j)^2\right] = 1$, $j = 1, \ldots, \infty$.

Proof. From [28] or [11], spectral factorization, e.q. Gram-Schmidt orthogonalization, will yield a causal filter which satisfies the statement of the theorem. □

Recall (2.20). The input sequence is the white WSS process, $\{X_j\}_{j=1}^{\infty}$. By Lemma 13, we can perform 2D Kalman filtering on the colored WSS process $\{V_j\}_{j=1}^{\infty}$ by replacing $\{\{h_{ij}\}_{i=1}^{n}\}_{j=0}^{M}$ with $\{\{\tilde{h}_{ij}\}_{i=1}^{n}\}_{j=0}^{M+H}$ where

$$\tilde{h}_{ij} = \sum_{k=0}^{\min\{M,H\}} g_k h_{i(j-k)}\,, \quad j = 0,\ldots,(M+H)$$

and replace M with $(M+H)$. In the case where $H = \infty$, a truncation strategy would be required to find the most sensibly finite amount of processing memory, $\tilde{H}$, to assign in the state vector of the state-space estimation model. The strategy is aimed at narrowing the error covariance gap between the two estimates, i.e., $\hat{\hat{X}}$ and $\{\bar{U}_{nj}^{M+\tilde{H}+1}\}_{j=M+\tilde{H}+1}^{\infty}$ (the sequence of the $(M+\tilde{H}+1)$-th component of the state vector estimate). One approach to truncating H and finding a causal filter with finite memory $\tilde{H}$, $\{g_i\}_{i=0}^{\tilde{H}}$, is a decomposition approach using the prolate spheroidal wave functions [12]. This problem is not elaborated in this chapter and is subject to further study.

In closing our discussion, we make the following remarks.

1. Distributed processing by 2D Kalman filter requires the rate of the feedback link to be n times the line rate, with n being the number of intermediate receivers.

2. For a rate below capacity, distributed processing by Kalman filter also works. Instead of consisting of Gaussian symbols attaining maximum mutual information, code book consists of Gaussian symbols with mutual information less than the maximum.

3 If the channel between an intermediate receiver and the far- receiver is subject to noise, the capacity result and the equality between capacity of the channels under centralized and distributed processing by Kalman filter do not hold. The capacity of such a relay model in difficult to find [47].

2.3 Conclusions

We have studied the problem of distributed processing for point-to-point communication over parallel relays. Such processing for relay communication purposes is relevant to a problem in the sensor networks literature, notably the problem of transporting data to a far-receiver via intermediate sensors. The model we chose to represent the structure of the problem is the single-input multi-output AWGN channel with and without inter-symbol interference (ISI). Our results point to the implementation of distributed processing via one dimensional (1D) and two dimensional (2D) Kalman filters. With the 1D Kalman filter, processing proceeds in *space*, namely from one to the next intermediate receiver. With the 2D Kalman filter, in addition to progressing in *space*, processing proceeds in *time*, namely from one to the next time step. We find the capacity of the distributively and optimally—that is, centrally—processed channels to be the same. In computing the capacity, we make use of the direct relation between capacity and estimation theory. Under 1D and 2D Kalman filters, processing is linear. However, the 2D Kalman filter mitigates a problem—of infinite memory for processing at each intermediate receiver—imposed by the 1D Kalman filter. Under the 2D Kalman filter, intermediate processing is done with finite memory.

When the capacity achieving input process (a wide sense stationary stochastic process) has infinite memory, a truncation strategy is needed. The truncation strategy will allocate some sensible finite amount of processing memory in the state vector of the state-space estimation model. By sensible, we mean the strategy will reasonably narrow the error covariance gap between the two estimates, i.e., the estimate when the state vector has infinite memory and that when the memory of the state vector is limited to some finite amount. One approach to truncating memory is a decomposition approach using the prolate spheroidal wave functions [12]. This problem is not elaborated in this chapter and is subject to further study.

An interesting extension to study is that of distributed processing on a channel that is not perfectly known at the intermediate receivers, namely a channel whereby the intermediate receivers perform channel estimation with some error. A cumulative error would build up in the case of such a channel. The results based on [19] may be extended to this case.

Chapter 3

On the Capacity of a Serial Relay Channel

3.1 Introduction

Consider the discrete-time channel shown in Figure 3.1. In this chapter, we address the problem of regeneration for such a channel, motivated by the discussion in Section 1.2.2. Note that the regeneration we are concerned with is that which involves arbitrary processing at source/ destination but memoryless processing at the relays. The natural approach with which we study the problem is the capacity approach.

3.2 Infinite Series of AWGN Channels with Memoryless Relay Functions

Assume that n relays are placed uniformly between identical AWGN channel sections, and assume that the AWGN noise power at each section is $\frac{\sigma^2}{n}$. With the assumption above, we obtain the serial relay channel shown in Figure 3.2.

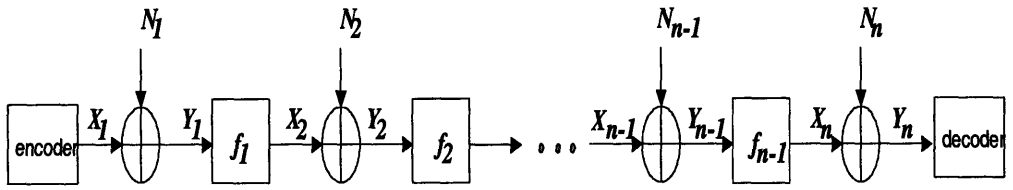

Figure 3.1: Regeneration with coding/decoding at the sender/receiver and memoryless processing at the relays.

A caveat for the channel model considered in this section is as follows. The model—and therefore the result obtained, given such a model—depends on the idealization that the end-to-end sum of noise power is constant and evenly divided by the number of relay stages n. Such an idealization is not compatible at all with the serial distribution network for optical communication purposes. In such a network, inserting an additional relay on the channel adds noise to the overall system. Adding more and more relays across the channel will cause the noise power to increase unboundedly.

Although the channel model is based on an idealization that is not compatible with the serial channel model for optical communication purposes, the ensuing capacity result, nevertheless, carries an insight. The insight is summarized as follows. Given that the serial relay channel is subject to memoryless relay functions, capacity tends to infinity in the number of relay stages. The capacity result depends on the choice of input distribution and relay functions used, and moreover achieved without coding/decoding at the source/destination. We will argue that such a capacity result does not support the view taken by the 'folk theorem'—discussed in Section 1.3.2.

Let us characterize the discrete-time memoryless channel shown in Figure 1.6. Let X_1 be a random variable with some density function. For all $i = 1, 2, \ldots, n$, let N_i be independent zero mean Gaussian random variable with variance $\frac{\sigma^2}{n}$, $N_i \sim \mathcal{N}(0, \frac{\sigma^2}{n})$. Moreover, for $i = 1, 2, \ldots, n-1$, let the relay function $f_i : \mathbb{R} \to \mathbb{R}$ be a

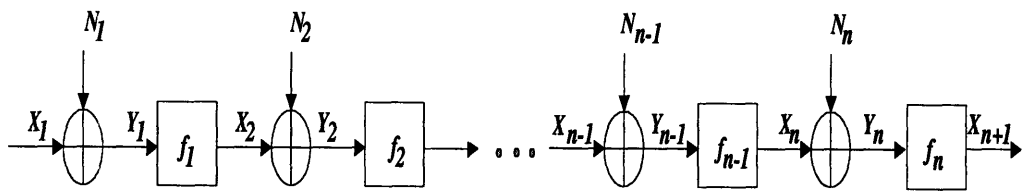

Figure 3.2: A series of n AWGN channels subject to energy constraints

memoryless mapping. By memoryless, the mapping at time n only depends on the relay's input at time n. Then,

$$Y_i = X_i + N_i \quad \forall\, i = 1,2,\ldots,n \tag{3.1}$$

$$X_j = f_{j-1}(Y_{j-1}) \quad \forall\, j = 2,\ldots,n+1 \tag{3.2}$$

To preclude signals of infinite energy traversing across the serial relay channel, we constrain the energy at the input of the channel as well as at the outputs of individual relays to be of some finite quantities. Namely, the channel is subject to energy constraints

$$\mathrm{E}[X_i^2] \leq 1 \quad \forall\, i = 1,2,\ldots,n+1 \tag{3.3}$$

For simplicity, we assume the mapping $f_{j-1} : \mathbb{R} \to \mathbb{R}$ preserves X_j, $\forall\, j = 2,\ldots,n$, to be zero-mean. This assumption is not a limiting assumption since we can always force X_j to be zero-mean by adding the appropriate constant.

Let C be the capacity of such a channel and the following theorem follows.

Theorem 14 *Asymptotically in the number of stages in the cascade,*

$$C \underset{n\to\infty}{\longrightarrow} \infty. \tag{3.4}$$

Moreover, when the number of stages, n, is infinite, the capacity in (3.4) is achieved without coding/decoding at the source/destination.

Proof. To see that (3.4) holds, consider the following scheme. Consider an M-ary signalling scheme, that is, X_1 is an M-ary (where M is assumed as an even and strictly positive integer) random variable. The probability density function (PDF) of X_1 is

$$P_{X_1}(x) = \frac{1}{M} \sum_{i=-(M/2)+1}^{M/2} \delta\left(x - \frac{2i-1}{M-1}\right). \tag{3.5}$$

The average energy of X_1 is

$$E\left[{X_1}^2\right] = \frac{M^2-1}{3(M-1)^2} \leq 1 \tag{3.6}$$

Define the relay function $f_i(Y_i), \forall\ i = 1,2,\ldots,n$, by

$$f_i(Y_i) = \begin{cases} \frac{2j-1}{M-1} & \text{if } \frac{2j-2}{M-1} < Y_i \leq \frac{2j}{M-1} \\ & \text{and } j = \frac{-M+4}{2}, \ldots, \frac{M-2}{2} \\ -1 & \text{if } Y_i \leq \frac{-M+2}{M-1} \\ 1 & \text{if } Y_i \geq \frac{M-2}{M-1} \end{cases}. \tag{3.7}$$

Note that

$$E\left[f_i(Y_i)^2\right] \leq 1, \tag{3.8}$$

hence, from (3.6) and (3.8), constraint (3.3), $\forall\ i = 1,2,\ldots,n$, is satisfied. Now,

$$\begin{aligned} Pr(\text{error}) &= Pr(X_{n+1} \neq X_1) && (3.9) \\ &= 1 - Pr(X_{n+1} = X_1) \\ &\overset{(a)}{\leq} 1 - \left[\Pi_{i=1}^{n} Pr(X_{i+1} = X_i)\right] \\ &= 1 - \left[Pr(X_2 = X_1)\right]^n \\ &= 1 - \left[1 - Pr(X_2 \neq X_1)\right]^n \\ &\overset{(b)}{=} 1 - \left[1 - \frac{2(M-1)}{M}\, Q\left(\frac{\sqrt{n}}{(M-1)\sigma}\right)\right]^n, && (3.10) \end{aligned}$$

where

(a) follows since $Pr(X_{n+1} = X_1) \geq \Pi_{i=1}^{n} Pr(X_{i+1} = X_i)$,

(b) follows since $Pr(\text{error}|\text{outer}) = \mathcal{Q}\left(\frac{\sqrt{n}}{(M-1)\sigma}\right)$ for the two outer sample signals of X_1, and $Pr(\text{error}|\text{inner}) = 2\mathcal{Q}\left(\frac{\sqrt{n}}{(M-1)\sigma}\right)$ for the $M-2$ inner sample signals,

where the $\mathcal{Q}$ function is defined by

$$\mathcal{Q}(x) = \frac{1}{2\pi}\int_x^\infty e^{-\xi^2/2}d\xi.$$

Asymptotically (in the number of channels in the cascade),

$$Pr(\text{error}) \underset{n\to\infty}{\longrightarrow} 0, \tag{3.11}$$

since, on the right hand side of (3.10), $\mathcal{Q}\left(\frac{\sqrt{n}}{(M-1)\sigma}\right) \underset{n\to\infty}{\longrightarrow} 0$.

Equation (3.11) suggests that the cascade of channel becomes noiseless as the number of relay functions tends to infinity. Information can be sent reliably across the cascade, without coding, at the rate of $\log M$ (in bits/channel use). Since such a rate holds for any M, i.e., the number of sample signals in the PDF of X_1, the asymptotic (in the number of channels in the cascade) capacity in (3.4) follows. □

Theorem 14 implies the following. As demonstrated in the proof, such an asymptotic capacity result depends on the choice of relay functions used, and is achieved without coding/decoding at the source/destination. For the asymptotic model considered, note that the SNR per section tends to infinity with the number of relay stages n. From the discussion in Section 1.3.2, recall that the 'folk theorem' points to the use of linear relay functions when SNR is high. Yet, fixing the relays to be unit amplification functions and $X_1 \sim \mathcal{N}(0,1)$, thus satisfying the set of energy constraints, only yields a finite capacity, i.e., $\frac{1}{2}\ln\left(1+\frac{1}{\sigma^2}\right)$ [nats/channel use].

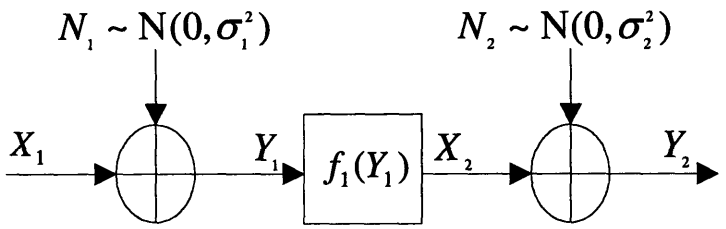

Figure 3.3: A series of 2 AWGN channels

3.3 Finite series of AWGN Channels with Memoryless Relay Functions

In a serial distribution network for optical communication purposes, the two major components of noise are the receiver and relay noise. This means that inserting an additional relay on the serial network adds noise to the overall system. In this section, we consider a discrete-time channel model which is more compatible—than the model in the previous section—for such a serial network. The model is shown in Figure 3.1. For this model, we assume that the number of relay stages n is finite, processing at the source/destination is arbitrary but processing at the relays is memoryless.

If we consider the types of memoryless relay functions that include the optimum function which was derived in [1] (see (3.25)) and, moreover, the hard-decision and linear amplification functions, what does this say about the performance limits? The question motivates us to consider the maximum mutual information of the channel shown in Figure 3.1, assuming an arbitrary power-limited input distribution and a family of relay functions that includes those described previously.

Finding the maximum mutual information of the channel in Figure 3.1, subject to energy constraints alone, is problematic. We illustrate the point with the following example. Consider the serial relay channel shown in Figure 3.3. Such a channel

is obtained from the channel shown in Figure 3.2 by setting $n = 2$. To preclude signals of infinite energy traversing the channel, let us constrain the energy of X_1 as well as the energy of X_2 such that $\mathrm{E}[X_1^2] \leq 1$ and $\mathrm{E}[X_2^2] \leq 1$. These are the same constraints imposed on the serial channel of the previous section. Given such a channel, what is its maximum mutual information?

A natural approach for finding the maximum mutual information begins with fixing the non-linear and memoryless relay function f_1 to be some initially chosen function. Given that f_1 is fixed, one would use one of the following algorithms [14, Ch. 8.3]: constrained maximization using calculus and the Kuhn-Tucker conditions; the Frank-Wolfe gradient search algorithm; or an iterative algorithm developed by Arimoto and Blahut, to search for the input distribution $p_{X_1^*}$ that maximizes mutual information. If the search converges to the maximizing $p_{X_1^*}$, then another search begins, namely we fix the input distribution to be $p_{X_1^*}$ and, by using some search algorithm, search for the non-linear and memoryless relay function f_1^* that maximizes mutual information. If the search converges to the maximizing f_1^* then fix it as the relay function and search for $p_{X_1^*}$ again. So the search continues until it converges, assuming that it does converge, and produce the pair of $p_{X_1^*}$ and f_1^* which maximizes mutual information simultaneously, that is, given f_1^*, there is no other distribution that attains higher mutual information than that of $p_{X_1^*}$, and vice versa.

The problem with the approach above is as follows. For the channel in Figure 3.1, the problem of finding an input distribution $p_{X_1^*}$ which maximizes mutual information, given that the relay function f_1 is fixed and the channel is subject to the energy constraints at the input as well as the relay, is a convex optimization problem. A number of algorithms have been developed for solving such a problem, making use of either the saddle point or the alternate maximization approach. A globally optimal solution exists since mutual information is a concave function of the input distribution, given that the relay function—thus the channel

transition probability $p_{Y_2|X_1}$—is fixed. It is not the case, however, for the following search. Given that the input distribution p_{X_1} is fixed and the channel is subject to some energy constraints at the input as well as the relay, the problem of finding a relay function f_1^* which maximizes mutual information may not be a convex optimization problem. Fixing the input distribution p_{X_1} and subject to some energy constraints at the input as well as the relay, mutual information is a convex function of the transition probability $p_{Y_2|X_1}$. This, however, is not enough to guarantee that a globally optimal solution exists. This is due to the fact that the mathematical relation between the relay function f_1 and the transition probability $p_{Y_2|X_1}$ is not a straight-forward one. As a consequence, we do not know whether the overall search—for the pair of $p_{X_1^*}$ and f_1^* which maximizes mutual information simultaneously—converges to a globally optimal solution. Even if the overall search—namely the search, which alternate from finding the maximizing input distribution given that the relay is fixed, and moreover finding the maximizing relay given that the input distribution is fixed—is convergent, this search leads only to a local maximum, instead of a global maximum.

The fact that we do not know whether the standard approach for finding maximum mutual information converges to a globally optimal solution, given that the channel in Figure 3.1 is only constrained in energy, motivates us to consider a different set of operating conditions: a set of operating conditions which accounts for the interplay amongst noise and input energy of individual sections of the channel, as well as a certain characteristic of the individual relay functions. To such an extent, we propose a set of constraints involving energy and differential entropy. We will show that the set of constraints allows us to derive an upper bound to mutual information in terms of the noise at each section of the channel, the energy at the first and the last section of the channel, and finally the relation between the differential entropy of the output and input of the individual relay functions.

We characterize the discrete memoryless channel shown in Figure 3.1 as fol-

lows. Let X_1 be a random variable with some density function. For $i = 1,2,\ldots,n$, let N_i be an independent zero mean Gaussian random variable with variance σ_i^2, $N_i \sim \mathcal{N}(0,\sigma_i^2)$. Moreover, for $i = 1,2,\ldots,n-1$, let $f_i : \mathbb{R} \to \mathbb{R}$ be a *memoryless* mapping. By memoryless, the mapping at time n only depends on the relay's input at time n. Then,

$$Y_i = X_i + N_i \quad \forall\, i = 1,2,\ldots,n \tag{3.12}$$

$$X_j = f_{j-1}(Y_{j-1}) \quad \forall\, j = 2,\ldots,n \tag{3.13}$$

To preclude signals of infinite energy traversing across the first and the last section of the serial channel, the channel is subject to energy constraints. Given the constants $\eta \in \mathbb{R}^+$, and $\xi \in \mathbb{R}^+$, the cascade of AWGN channels is subject to energy constraints, namely

$$E[X_1^2] \leq \xi \tag{3.14}$$

$$E[X_n^2] \leq \eta \tag{3.15}$$

The set of energy constraints above, i.e., (3.14) and (3.15), allows for the case where noise-added signals with arbitrarily large energy traverse across the 2^{nd} up to the $(n-1)^{th}$ section of the channel. As such, a theorem which we will derive, i.e., Theorem 17, yet demonstrates that the mutual information is bounded from above.

As discussed in the beginning of this section, to mitigate the problem of finding the upper bound to mutual information, we propose a set of constraints that does not only constrain the energy, but also the differential entropy of the output of each relay function in terms of that of its input. Namely, given the constants $\beta_i \in \mathbb{R}^+$, $\forall i = 1,2,\ldots,n-1$,

$$\begin{aligned} h(X_{i+1}) &\geq h(\beta_i Y_i) \\ &= \ln \beta_i + h(Y_i) \end{aligned} \tag{3.16}$$

In what follows, we proceed with the steps for the derivation of an upper bound

to the mutual information of the series of channel shown in Figure 3.1, subject to constraints (3.14), (3.15), and (3.16). Let us begin with the following theorem.

Theorem 15 (Conditional Entropy-Power Inequality) *Let X and Y be two random variables which are conditionally independent given a random variable W. Then*

$$e^{2h(X+Y|W)} \geq e^{2h(X|W)} + e^{2h(Y|W)}, \tag{3.17}$$

Proof. The proof is given in Appendix A. □

We shall now apply Theorem 15 to the cascaded channel, that is, a channel whose model is specified by (3.12) and (3.13), constraints (3.14), (3.15), and (3.16). Before proceeding with Lemma (16), we introduce a quantity which we call the conditional entropy-power. Suppose W and Z are two random variables, we define $\overline{W|Z}$ as the conditional entropy-power of W given Z, that is,

$$\overline{W|Z} \triangleq \frac{e^{2h(W|Z)}}{2\pi e}. \tag{3.18}$$

Lemma 16 *Given the the serial channel in Figure 3.1,*

$$h(Y_n|X_1) \geq \frac{1}{2}\ln\left(2\pi e\left(\sigma_n^2 + \sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=1}^{n-1}\beta_k^2\right)\right)\right). \tag{3.19}$$

Proof. We make use of induction. By induction, the base step, i.e., $i = 2$, is as follows.

$$\begin{aligned} h(Y_2|X_1) &\overset{(a)}{=} \frac{1}{2}\ln 2\pi e\,\overline{Y_2|X_1} \\ &\overset{(b)}{\geq} \frac{1}{2}\ln 2\pi e(\overline{X_2|X_1} + \overline{N_2|X_1}) \\ &\overset{(c)}{\geq} \frac{1}{2}\ln 2\pi e(\beta_1^2\sigma_1^2 + \sigma_2^2) \end{aligned}$$

where

(a) follows from the definition of conditional entropy-power in (3.18),

(b) holds by applying Theorem 15 for the random variables X_2, N_2, and $Y_2 = X_2 + N_2$ conditional on X_1,

(c) follows for the following two reasons. The first is the fact that $h(N_2|X_1) = h(N_2) = \frac{1}{2}\ln 2\pi e\sigma_2^2$, hence, by (3.18), $\overline{N_2|X_1} = \sigma_2^2$. The second is from *data processing inequality*, i.e., $I(X_1;Y_1) \geq I(X_1;X_2)$. Data processing inequality and the entropy constraint (3.16) imply

$$\begin{aligned} h(Y_1|X_1) - h(X_2|X_1) &\leq h(Y_1) - h(X_2) \\ &\leq -\ln\beta_1 \end{aligned}$$

hence

$$\begin{aligned} h(X_2|X_1) &\geq h(Y_1|X_1) + \ln\beta_1 \\ &= h(N_1) + \ln\beta_1 \\ &= \frac{1}{2}\ln 2\pi e(\beta_1^2\sigma_1^2). \end{aligned}$$

By definition of entropy power in (3.18), $\overline{X_2|X_1} \geq \beta_1^2\sigma_1^2$

Now, the induction step for $i = n$ is as follows.

$$\begin{aligned} h(Y_n|X_1) &\overset{(a)}{=} \frac{1}{2}\ln 2\pi e\,\overline{Y_n|X_1} \\ &\overset{(b)}{\geq} \frac{1}{2}\ln 2\pi e(\overline{X_n|X_1} + \overline{N_n|X_1}) \\ &\overset{(c)}{\geq} \frac{1}{2}\ln 2\pi e\left(\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=1}^{n-1}\beta_k^2\right) + \sigma_n^2\right) \end{aligned}$$

where

(a) follows from the definition of conditional entropy-power in (3.18),

(b) holds by applying Theorem 15 for the random variables X_n, N_n, and $Y_n = X_n + N_n$ conditional on X_1,

(c) follows for the following two reasons. The first is the fact that $h(N_n|X_1) = h(N_2) = \frac{1}{2}\ln 2\pi e\sigma_n^2$, hence, by (3.18), $\overline{N_n|X_1} = \sigma_n^2$. The second is from *data processing inequality*, i.e., $I(X_1;Y_{n-1}) \geq I(X_1;X_n)$, entropy constraint (3.16), and our induction hypothesis, i.e., $h(Y_{n-1}|X_1) \geq \frac{1}{2}\ln\left(2\pi e\left(\sigma_{n-1}^2 + \sum_{i=1}^{n-2}\sigma_i^2\left(\Pi_{k=1}^{n-2}\beta_k^2\right)\right)\right)$. Data processing inequality, the entropy constraint (3.16), and the induction hypothesis imply

$$\begin{aligned} h(Y_{n-1}|X_1) - h(X_n|X_1) &\leq h(Y_{n-1}) - h(X_n) \\ &\leq -\ln\beta_{n-1} \end{aligned}$$

hence

$$\begin{aligned} h(X_n|X_1) &\geq h(Y_{n-1}|X_1) + \ln\beta_{n-1} \\ &= \frac{1}{2}\ln\left(2\pi e\left(\sigma_{n-1}^2 + \sum_{i=1}^{n-2}\sigma_i^2\left(\Pi_{k=1}^{n-2}\beta_k^2\right)\right)\right) + \ln\beta_{n-1} \\ &= \frac{1}{2}\ln\left(2\pi e\left(\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=1}^{n-1}\beta_k^2\right)\right)\right) \end{aligned}$$

By definition of entropy power in (3.18), $\overline{X_n|X_1} \geq \sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=1}^{n-1}\beta_k^2\right)$,

completing the proof. □

Now, let $\mathcal{C}$ be a set of all pairs $(X_1, f_i : \mathbb{R} \to \mathbb{R}\ \forall i = 1,\ldots,n-1)$ which obey (3.14), (3.15) and (3.16), namely $\mathcal{C} = \{(X_1, f : \mathbb{R} \to \mathbb{R}\ \forall i = 1,\ldots,n-1)$: (3.14), (3.15) and (3.16) are satisfied$\}$.

Theorem 17 *(i) For* $\eta \leq \left(\Pi_{k=1}^{n-1}\beta_k^2\right)\xi + \sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)$ *and* $\eta \geq \sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)$,

$$\sup_{(X_1,\, f:\mathbb{R}\to\mathbb{R})\in\mathcal{C}} I(X_1;Y_n) = \frac{1}{2}\ln\left(1 + \frac{\eta - \sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)}{\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right) + \sigma_n^2}\right),$$

and the supremum is attained by fixing $X_1 \sim \mathcal{N}\left(0, \frac{\eta - \sum_{i=1}^{n-1} \sigma_i^2 \left(\Pi_{k=i}^{n-1} \beta_k^2\right)}{\Pi_{k=1}^{n-1} \beta_k^2}\right)$ *and* $f_{j-1}(Y_{j-1}) = \beta_{j-1}, \forall\, j = 2, \ldots, n$.

In particular for $\eta = \left(\Pi_{k=1}^{n-1} \beta_k^2\right) \xi + \sum_{i=1}^{n-1} \sigma_i^2 \left(\Pi_{k=i}^{n-1} \beta_k^2\right)$,

$$\sup_{(X_1,\, f:\mathbb{R}\to\mathbb{R}) \in \mathcal{C}} I(X_1; Y_n) = \frac{1}{2} \ln \left(1 + \frac{\xi}{\sigma_1^2 + \sum_{i=2}^{n} \frac{\sigma_i^2}{\left(\Pi_{k=1}^{i-1} \beta_k^2\right)}}\right),$$

and the supremum is attained by fixing $X_1 \sim \mathcal{N}(0, \xi)$ *and* $f_{j-1}(Y_{j-1}) = \beta_{j-1}$, $\forall\, j = 2, \ldots, n$.

(ii) For $\eta > \left(\Pi_{k=1}^{n-1} \beta_k^2\right) \xi + \sum_{i=1}^{n-1} \sigma_i^2 \left(\Pi_{k=i}^{n-1} \beta_k^2\right)$,

$$\sup_{(X_1,\, f:\mathbb{R}\to\mathbb{R}) \in \mathcal{C}} I(X_1; Y_n) < \frac{1}{2} \ln \left(1 + \frac{\eta - \sum_{i=1}^{n-1} \sigma_i^2 \left(\Pi_{k=i}^{n-1} \beta_k^2\right)}{\sum_{i=1}^{n-1} \sigma_i^2 \left(\Pi_{k=i}^{n-1} \beta_k^2\right) + \sigma_n^2}\right).$$

Proof. From the definition of mutual information

$$I(X_1; Y_n) = h(Y_n) - h(Y_n | X_1) \tag{3.20}$$

By constraint (3.15)

$$h(Y_n) \leq \frac{1}{2} \ln \left(2\pi e \left(\eta + \sigma_n^2\right)\right) \tag{3.21}$$

Combining (3.20), (3.21), and (3.19) of Lemma 16,

$$I(X_1; Y_n) \leq \frac{1}{2} \ln \left(\frac{\eta + \sigma_n^2}{\sum_{i=1}^{n-1} \sigma_i^2 \left(\Pi_{k=i}^{n-1} \beta_k^2\right) + \sigma_n^2}\right) \tag{3.22}$$

Consider the following two cases.

(i) Case 1: $\sum_{i=1}^{n-1} \sigma_i^2 \left(\Pi_{k=i}^{n-1} \beta_k^2\right) < \eta < \left(\Pi_{k=1}^{n-1} \beta_k^2\right) \xi + \sum_{i=1}^{n-1} \sigma_i^2 \left(\Pi_{k=i}^{n-1} \beta_k^2\right)$.

Before proceeding with the proof of the first case, we note that from 3.22, the

lower bound on η, that is, $\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)\leq\eta$, is needed to guarantee that the constraints (3.14), (3.15) and (3.16) are mutually compatible. Otherwise, i.e., when $\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)>\eta$, the constraints (3.14), (3.15) and (3.16) become mutually incompatible, leading to $I(X_1;Y_n)$ being bounded from above by a negative quantity.

Now, from (3.22),

$$\begin{aligned} I(X_1;Y_n) &\leq \frac{1}{2}\ln\left(\frac{\eta}{\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)+\sigma_n^2}\right) \\ &= \frac{1}{2}\ln\left(1+\frac{\eta-\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)}{\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)+\sigma_n^2}\right) \end{aligned}$$

The upper bound is satisfied by fixing $X_1\sim\mathcal{N}\left(0,\frac{\eta-\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)}{\Pi_{k=1}^{n-1}\beta_k^2}\right)$ and $f_{j-1}(Y_{j-1})=\beta_{j-1},\forall\, j=2,\ldots,n$.

In particular, when $\eta=\left(\Pi_{k=1}^{n-1}\beta_k^2\right)\xi+\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)$,
then from (3.22),

$$\begin{aligned} I(X_1;Y_n) &\leq \frac{1}{2}\ln\left(\frac{\eta+\sigma_n^2}{\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)+\sigma_n^2}\right) \\ &= \frac{1}{2}\ln\left(1+\frac{\left(\Pi_{k=1}^{n-1}\beta_k^2\right)\xi}{\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)+\sigma_n^2}\right) \\ &= \frac{1}{2}\ln\left(1+\frac{\left(\Pi_{k=1}^{n-1}\beta_k^2\right)\xi}{\sigma_1^2+\sum_{i=2}^{n}\frac{\sigma_i^2}{\left(\Pi_{k=1}^{i-1}\gamma_k^2\right)}}\right) \end{aligned}$$

The upper bound is satisfied by fixing $X_1\sim\mathcal{N}(0,\xi)$ and $f_{j-1}(Y_{j-1})=\beta_{j-1}$, $\forall\, j=2,\ldots,n$.

(ii) Case 2: $\eta>\left(\Pi_{k=1}^{n-1}\beta_k^2\right)\xi+\sum_{i=1}^{n-1}\sigma_i^2\left(\Pi_{k=i}^{n-1}\beta_k^2\right)$.

From (3.22),

$$\begin{aligned} I(X_1;Y_n) &\leq \frac{1}{2}\ln\left(\frac{\eta}{\sum_{i=1}^{n-1}\sigma_i^2\left(\prod_{k=i}^{n-1}\beta_k^2\right)+\sigma_n^2}\right) \\ &= \frac{1}{2}\ln\left(1+\frac{\eta-\sum_{i=1}^{n-1}\sigma_i^2\left(\prod_{k=i}^{n-1}\beta_k^2\right)}{\sum_{i=1}^{n-1}\sigma_i^2\left(\prod_{k=i}^{n-1}\beta_k^2\right)+\sigma_n^2}\right), \end{aligned}$$

completing the proof. □

For case (i) of Theorem 17, we have shown a tight upper bound on mutual information for the channel shown in Figure 3.1, given that the pair $(X_1, f_i\ \forall i = 1,\ldots,n-1)$ satisfies constraints (3.14), (3.15), and (3.16). For case (ii) of Theorem 17, we are however unable to show that the upper bound is tight.

3.3.1 Numerical Results

As opposed to case (i) of Theorem 17, case (ii) of of Theorem 17 applies to a region where the upper bound is not exactly the the maximum mutual information, subject to constraints (3.14), (3.15), and (3.16). In this region, we will use the upper bound– given by case (ii) of of Theorem 17– to investigate how several reasonable choices of pairs of input distribution and relay function perform. For this purpose, we will numerically compute the mutual information, subject to constraints (3.14), (3.15), and (3.16), attained by the following pairs of input distribution and relay function:

- an antipodal input and an optimum relay function derived in [1];
- an antipodal input and a relay function performing linear amplification;
- an antipodal input and a relay function performing minimum mean-square error (MMSE) estimator;
- a Gaussian input and a relay function performing linear amplification

Note that—by construction—these mutual information are lower bounds to maximum mutual information subject to constraints (3.14), (3.15), and (3.16). We will numerically contrast these lower bounds versus the upper bound given by case (ii) of Theorem 17.

The channel we consider is the channel with two sections—each with AWGN—shown in Figure 3.3 and subject to constraints (3.14), (3.15), and (3.16). As will be discussed later, the main result of this section is as follows. We will numerically demonstrate that lower bound attained by the pair of Gaussian input and linear relay function approaches the upper bound more closely than that attained by other choices of pair of input and relay function which we consider. The lower bound attained by a pair of Gaussian input and linear relay function is closer to the upper bound as the noise power of the first section decays and that of the second section grows. In the context of a two-stage serial distribution network for optical communications, the noise power of the second section is likely to outgrow that of the first section. This is due to the fact that the major noise components are the noise from the relay and that from the receiver. For optical communication purposes, this means that the choice of a code book consisting of alphabets, each with a distribution approaching that of Gaussian, at the source and a linear function at the relay outperforms other reasonable choices of code book and relay functions.

In the following, we describe in more details the numerical steps with which we demonstrate the result above.

Numerical Procedure

1. Consider the two-stage serial channel shown in Figure 3.3. Assign the AWGN noise power σ_1^2 and σ_2^2 to take on one of these values : $100, 200, 300, 400$. Next,

assign the energy constraints (3.14) and (3.15) by fixing

$$E[X_1^2] \leq 1 \tag{3.23}$$

$$E[X_2^2] = 1 + \sigma_1^2 \tag{3.24}$$

2. Let X_1 be an antipodal random variable, that is, X_1 takes on one of the equiprobable sample values $+1$ and -1. Moreover, let

$$f_1(Y_1) = \begin{cases} -g(-Y_1) & \text{for } Y_1 < 0 \\ g(Y_1) & \text{for } Y_1 \geq 0 \end{cases} \tag{3.25}$$

where

$$g(Y_1) = \sqrt{\sigma_2^2 \text{LambertW}\left(2\frac{\left(\tanh\left(\frac{Y_1}{\sigma_1^2}\right)\right)^2}{(\lambda 2\sigma_2^2\sqrt{\pi})^2}\right)},$$

and λ is chosen such that

$$\mathrm{E}\left[f_1(Y_1)^2\right] = 1 + \sigma_1^2. \tag{3.26}$$

Note that the function LambertW(x) is Lambert's W function[1] that is analytic at 0 and defined in terms of its inverse, that is, for $y = \text{LambertW}(x)$,

$$x = y\exp(y).$$

A plot of $f_1(Y_1)$ versus several values of the pair $\{\sigma_1^2, \sigma_2^2\}$ is depicted in Figure 3.5. Later in this section, we will describe what motivates us to choose the pair $\{X_1, f_1(Y_1)\}$ given above. Also note that the pair $\{X_1, f_1(Y_1)\}$ satisfies energy constraints (3.23) and (3.24) with equality.

3. Given the pair $\{X_1, f_1(Y_1)\}$ above, compute $h(X_2)$ and $h(Y_1)$. Next, compute β_1, that is, $\beta_1 = \exp(h(X_2) - h(Y_1))$. Now assign the resulted β_1 to the entropy constraint (3.16) for the channel shown in Figure 3.3 by fixing

$$h(X_2) \geq h(\beta_1 Y_1) \tag{3.27}$$

[1]Discovered by J.H. Lambert, a French-born German mathematician (1728-1777) who did the 1^{st} rigorous proof on the fact that π is an irrational number.

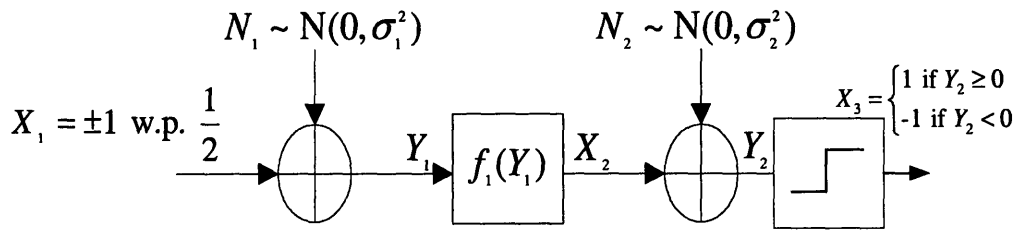

Figure 3.4: A series of 2 AWGN channels given antipodal input and a hard-decision function at the output end

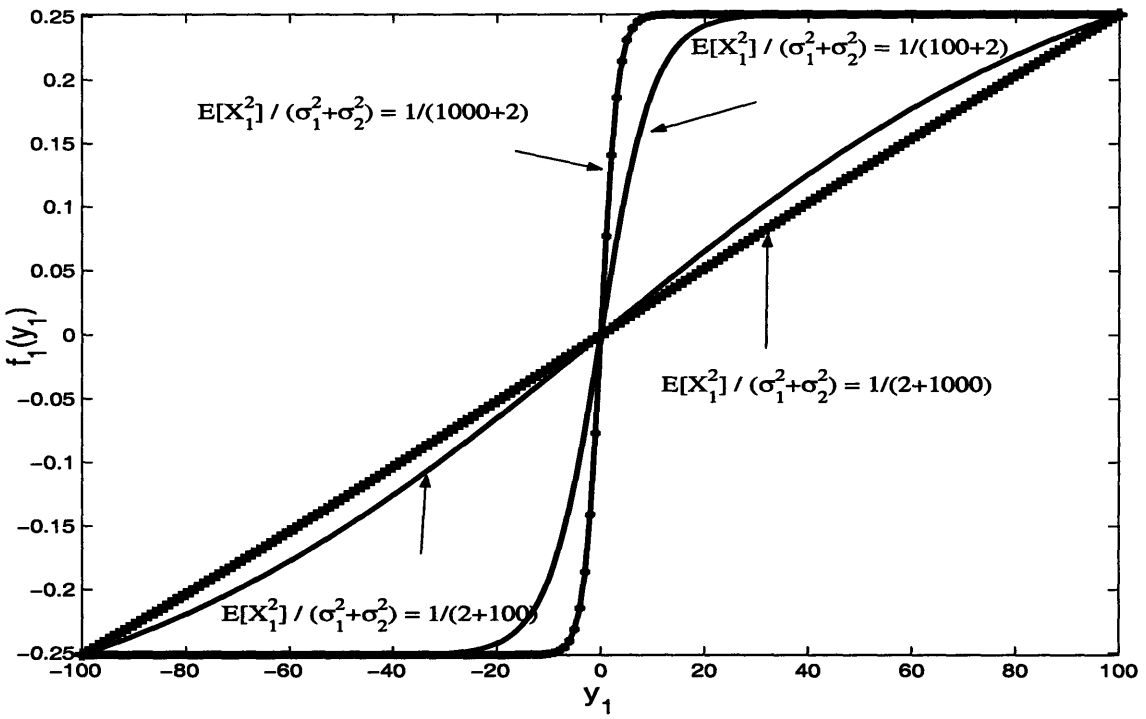

Figure 3.5: $f_1(Y_1)$ given by (3.25) as a function of the pair (σ_1^2, σ_2^2)

4. Let $I^{UB}(X_1; Y_2)$ be the upper bound to mutual information of the channel shown in Figure 3.3 subject to constraints (3.23), (3.24), and (3.27). As shown in Appendix B, the computed values of β_1 yield the case where $1 + \sigma_1^2 = \eta > \beta_1(1 + \sigma_1^2)$. Given such a case, we make use of case (ii) of Theorem 17 to

compute $I^{UB}(X_1; Y_2)$.

5. Compute four explicit lower bounds to maximum mutual information of the channel in Figure 3.3. These lower bounds are:

   (a) $I_{Lambert}(X_1; X_3)$, that is, the mutual information between X_1 and X_3 for the channel shown in Figure 3.4, given that the relay function $f_1(Y_1)$ is defined by (3.25) and X_1 takes on one of the equiprobable sample values $+1$ and -1. Later in this section, we will describe what motivates us to choose the pair $\{X_1, f_1(Y_1)\}$ given above.

   (b) $I_{Linear}(X_1; X_3)$, that is, the mutual information between X_1 and X_3 for the channel shown in Figure 3.4, given that the relay function is $f_1(Y_1) = Y_1$ and X_1 takes on one of the equiprobable sample values $+1$ and -1.

   (c) $I_{MMSE}(X_1; X_3)$, that is, the mutual information between X_1 and X_3 for the channel shown in Figure 3.4, given that X_1 takes on one of the equiprobable sample values $+1$ and -1 and the relay function $f_1(Y_1)$ is a minimum mean-square error (MMSE) estimator, that is,

   $$f_1(Y_1) = \alpha \tanh\left(\frac{Y_1}{\sigma_1^2}\right), \tag{3.28}$$

   where α is chosen such that

   $$\mathrm{E}[f_1(Y_1)^2] = 1 + \sigma_1^2. \tag{3.29}$$

   Later in this section, we will describe what motivates us to choose the pair $\{X_1, f_1(Y_1)\}$ given above.

   (d) $I_{Gaussian}(X_1; Y_2)$, that is, the mutual information between X_1 and Y_2 for the channel shown in Figure 3.3, given that the relay function is $f_1(Y_1) = Y_1$ and X_1 is a Gaussian random variable $\mathcal{N}(0, 1)$

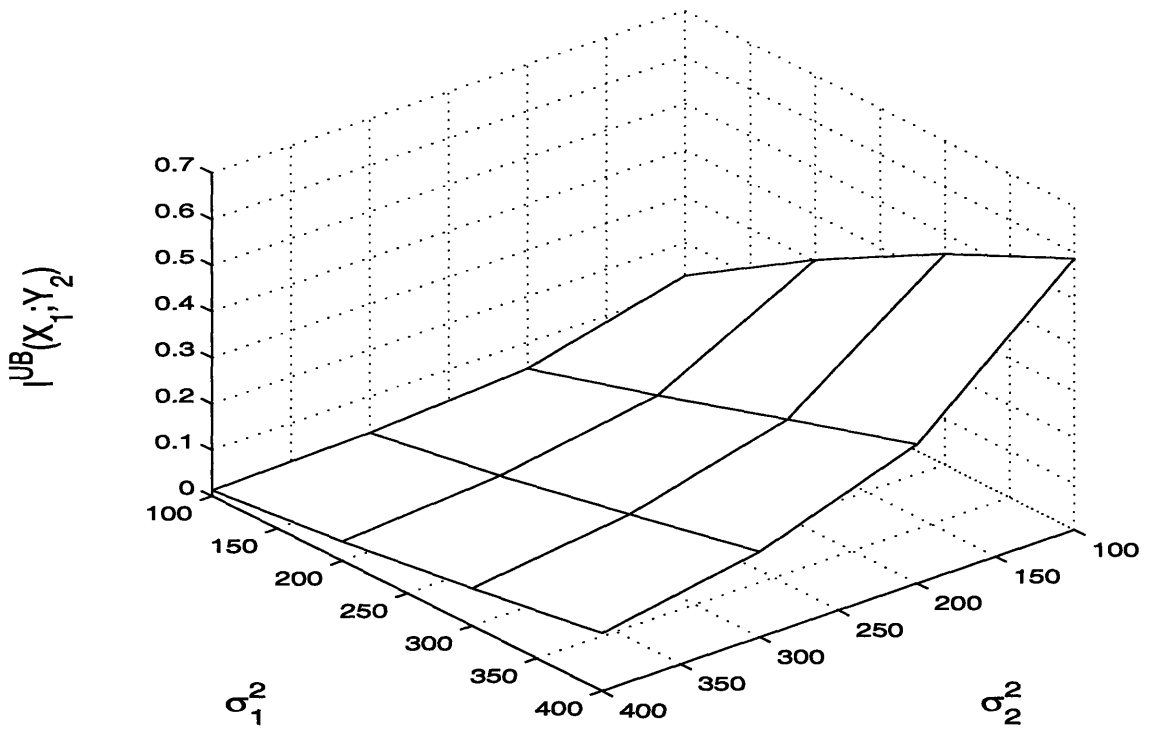

Figure 3.6: $I^{UB}(X_1;Y_2)$ as a function of the pair (σ_1^2,σ_2^2)

Note that the systems considered for computing the four lower bounds above are also subject to constraints (3.23), (3.24), and (3.27)—refer to Appendix B for the numerical results supporting the fact that the constraints are satisfied. In particular, energy constraints (3.23) and (3.24) are satisfied with equality.

6. Find the maximum of the four lower bounds above. Next, we contrast the upper bound $I^{UB}(X_1;Y_2)$ versus the maximum of the four lower bounds.

Upper bound to Mutual Information

In the following, we will compute $I^{UB}(X_1;Y_2)$, that is, the upper bound to mutual information of the channel shown in Figure 3.3 subject to constraints (3.23), (3.24), and (3.27).

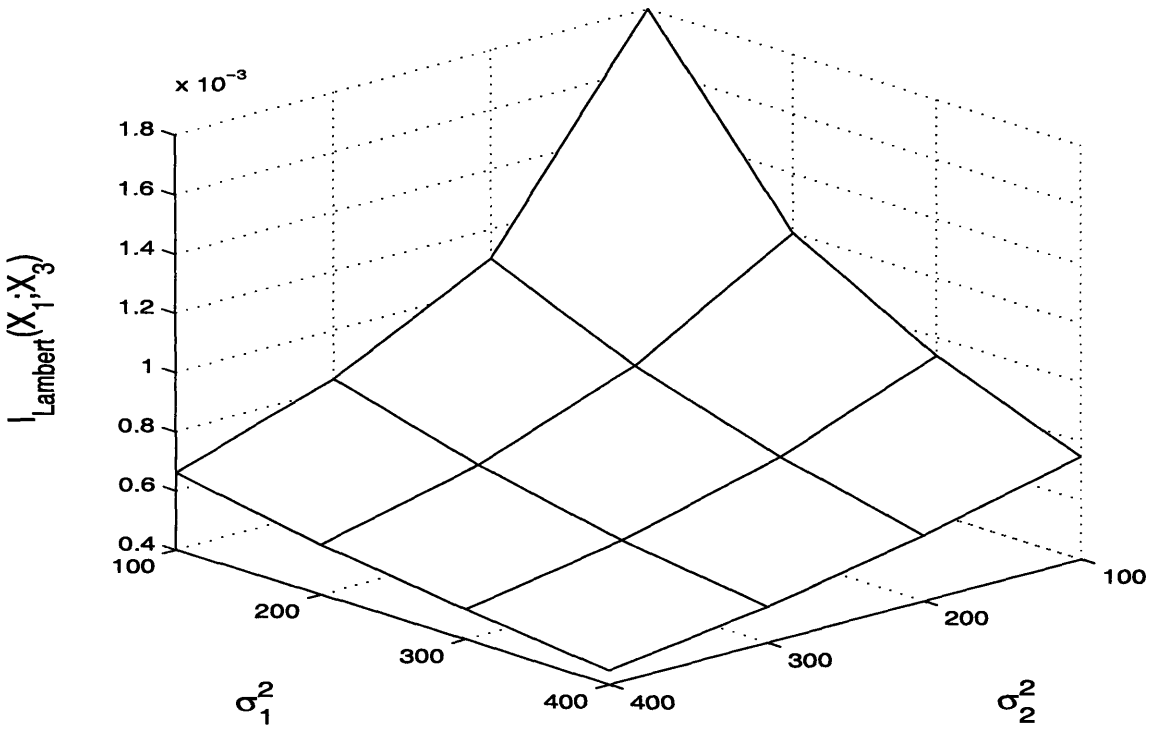

Figure 3.7: $I_{Lambert}(X_1; X_3)$ as a function of the pair (σ_1^2, σ_2^2)

To derive p_{X_2}, given that X_1 X_1 takes on one of the equiprobable sample values $+1$ and -1, and $f_1(Y_1)$ is given by (3.25), we make use of the following lemma.

Lemma 18 *Suppose that g is monotonic and that for some function h and all x in the range I of X we have*

$$y = g(x) \text{ if and only if } x = h(y).$$

Assume that h has first derivative $\frac{dh(y)}{dy}$. Then the PDF of Y in the region where $p_Y(y) > 0$ is given by

$$p_Y(y) = p_X(h(y)) \left| \frac{dh(y)}{dy} \right|$$

Proof. A proof for the lemma is given in [25]. □

First, for each value of σ_1^2, we numerically compute λ using (3.26). Next we apply Lemma 18 to derive p_{X_2} in terms of the function expressed in (3.25) and p_{Y_1},

with p_{Y_1} being a Gaussian PDF with mean zero and variance $1+\sigma_1^2$.

Having derived p_{X_2}, we can now compute $h(X_2)$ and $h(Y_1)$ numerically. At this point, computing β_1 is straight forward, that is, $\beta_1 = \exp(h(X_2) - h(Y_1))$. Let us assign β_1 onto the entropy constraint (3.16), that is, $h(X_2) \geq h(\beta_1 Y_1)$. As shown in Appendix B, the computed values of β_1 yield the case where $1+\sigma_1^2 = \eta > \beta_1(1+\sigma_1^2)$. Given such a case, we make use of case (ii) of Theorem 17 to compute $I^{UB}(X_1;Y_2)$ for the channel shown in Figure 3.3, subject to constraints (3.23), (3.24), and (3.27).

Figure 3.6 shows a 3D plot of $I^{UB}(X_1;Y_2)$ as a function of the pair (σ_1^2,σ_2^2). We observe that $I^{UB}(X_1;Y_2)$ grows with σ_1^2, given that σ_2^2 is fixed, and decays with σ_2^2, given that σ_1^2 is fixed. Appendix B presents the numerical figures leading to the plot in Figure 3.6.

Lower Bound to Maximum Mutual Information under Antipodal Input and the Function given by (3.25)

Consider the channel shown in Figure 3.4. Let X_1 be an antipodal random variable, that is, X_1 takes on one of the equiprobable sample values $+1$ and -1. Moreover, let $f_1(Y_1)$ be the function given by (3.25). Note that the pair $\{X_1, f_1(Y_1)\}$ satisfies constraints (3.23), (3.24), and (3.27), all with equality.

What motivates us to consider the pair $\{X_1, f_1(Y_1)\}$ above is as follows. For the channel shown in Figure 3.4 (given an antipodal input and a hard-decision function at the output end), subject to energy constraints (3.23) and (3.24) and a functional constraint that $f_1(Y_1)$ is sign-preserving and symmetric, reference [1] derived the function $f_1(Y_1)$ given by (3.25) and demonstrated that such a function minimizes error probability, that is, $\Pr(X_1 \neq X_3)$.

Since relay function $f_1(Y_1)$ is sign-preserving and symmetric, moreover X_1 and

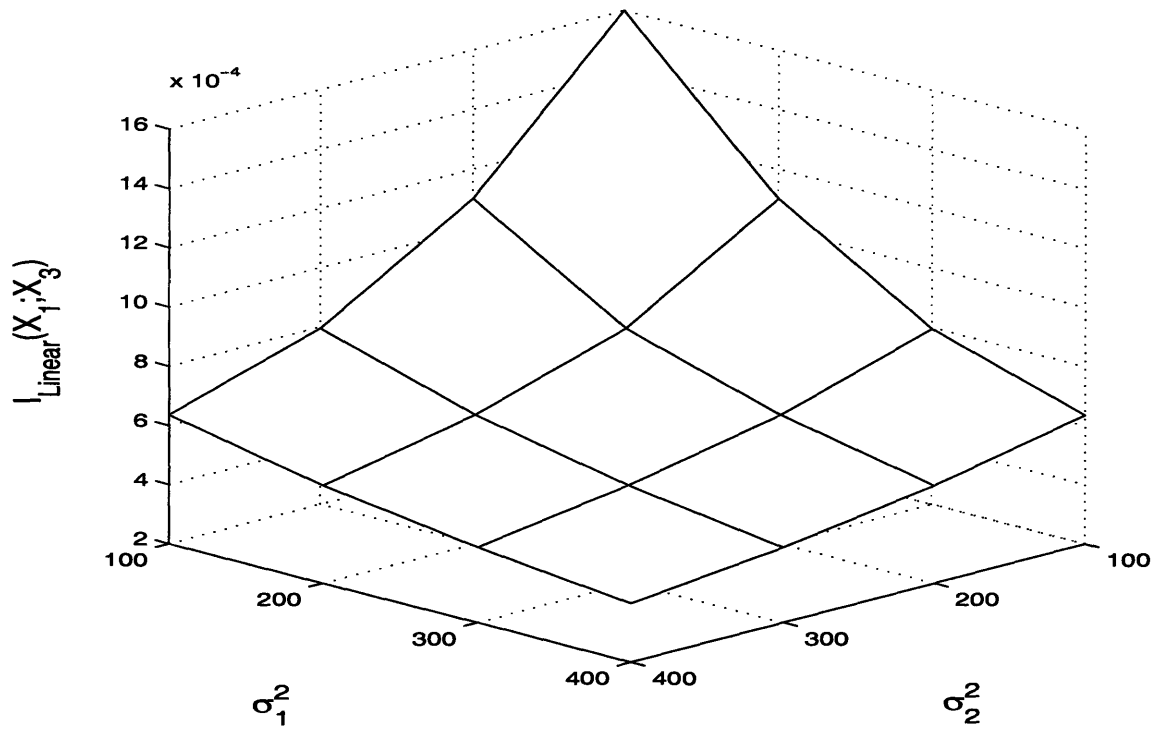

Figure 3.8: $I_{Linear}(X_1; X_3)$ as a function of the pair (σ_1^2, σ_2^2)

X_3 are discrete random variables, it is not difficult to show that

$$\begin{aligned} \text{Pr(error)} &\triangleq \Pr(X_1 \neq X_3) \\ &= \Pr(X_1 \neq X_3 | X_1 = 1) = \Pr(X_1 \neq X_3 | X_1 = -1) \end{aligned}$$

Similarity in the conditional error probabilities, i.e., conditional on $X_1 = 1$ and $X_1 = -1$, implies that the channel in Figure 3.4 is a binary symmetric channel (BSC) and

$$I_{Lambert}(X_1; X_3) = \ln(2)\,(1 - H(\Pr(\text{error}))) \quad \text{[Nats/Channel Use]}, \tag{3.30}$$

where $H(\Pr(\text{error})$ denotes discrete entropy of the error probability and $I_{Lambert}(X_1; X_3)$ is the mutual information between X_1 and X_3 when relay function is the function given by (3.25). From (3.30) and the fact that the function $f_1(Y_1)$ given by (3.25)

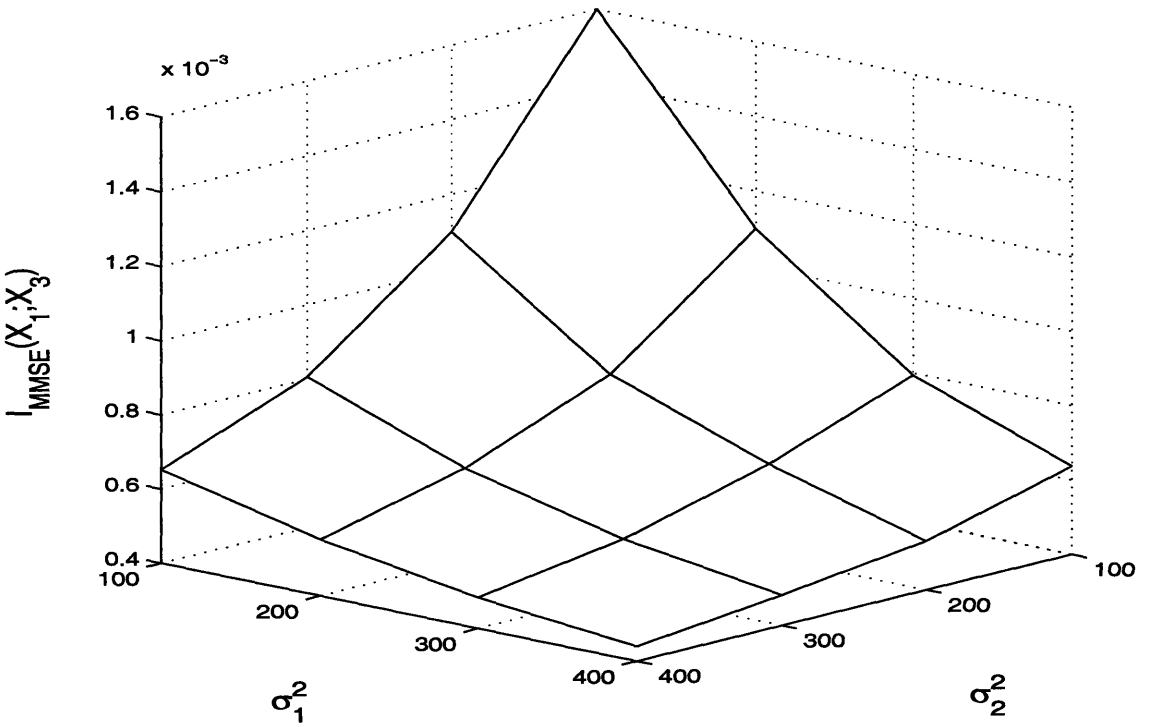

Figure 3.9: $I_{MMSE}(X_1; X_3)$ as a function of the pair (σ_1^2, σ_2^2)

minimizes error probability [1], it follows that $I_{Lambert}(X_1; X_3)$ maximizes mutual information of the channel shown in Figure 3.4, subject to energy constraints (3.23) and (3.24) and a functional constraint that $f_1(Y_1)$ is sign-preserving and symmetric.

We compute the mutual information $I_{Lambert}(X_1; X_3)$ of the channel shown in Figure 3.4 as follows.

The expression for Pr(error) is derived in [1], that is,

$$\begin{aligned} \text{Pr(error)} &= \int_{-\infty}^{0} dx \mathcal{N}\left(x; 1, \sigma_1^2\right) \\ &\quad + \int_{-\infty}^{0} dy_1 \left(\mathcal{N}\left(x; -1, \sigma_1^2\right) - \mathcal{N}\left(y_1; 1, \sigma_1^2\right)\right) \int_{-f_1(x)}^{-\infty} dy \mathcal{N}\left(y; 0, \sigma_2^2\right), \end{aligned} \tag{3.31}$$

where $\mathcal{N}(x; m, \mu) \triangleq \frac{1}{\sqrt{2\pi\mu}} \exp\left(-\frac{(x-m)^2}{2\mu}\right)$ and $f_1(x)$ is the function given by (3.25).

We numerically compute the error probability by making use of (3.31). Having

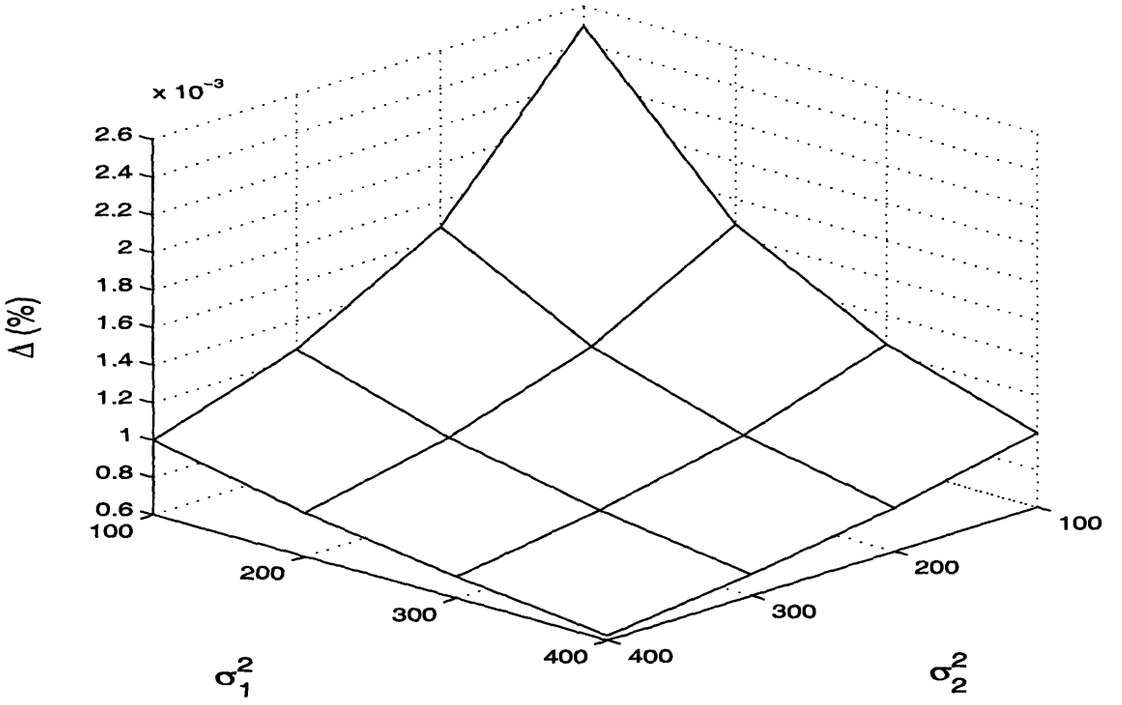

Figure 3.10: $I_{Gaussian}(X_1; Y_2)$ as a function of the pair (σ_1^2, σ_2^2)

computed the error probability, we can readily compute $I_{Lambert}(X_1; X_3)$ by making use of (3.30).

Figure 3.7 shows a 3D plot of $I_{Lambert}(X_1; X_3)$ as a function of the pair (σ_1^2, σ_2^2). We observe that $I_{Lambert}(X_1; X_3)$ decays with σ_1^2 and σ_2^2. In particular, $I_{Lambert}(X_1; X_3)$ decays with σ_1^2, given that σ_2^2 is fixed, at a faster rate than that of σ_2^2, given that σ_1^2 is fixed. Appendix B presents the numerical figures leading to the plot in Figure 3.7.

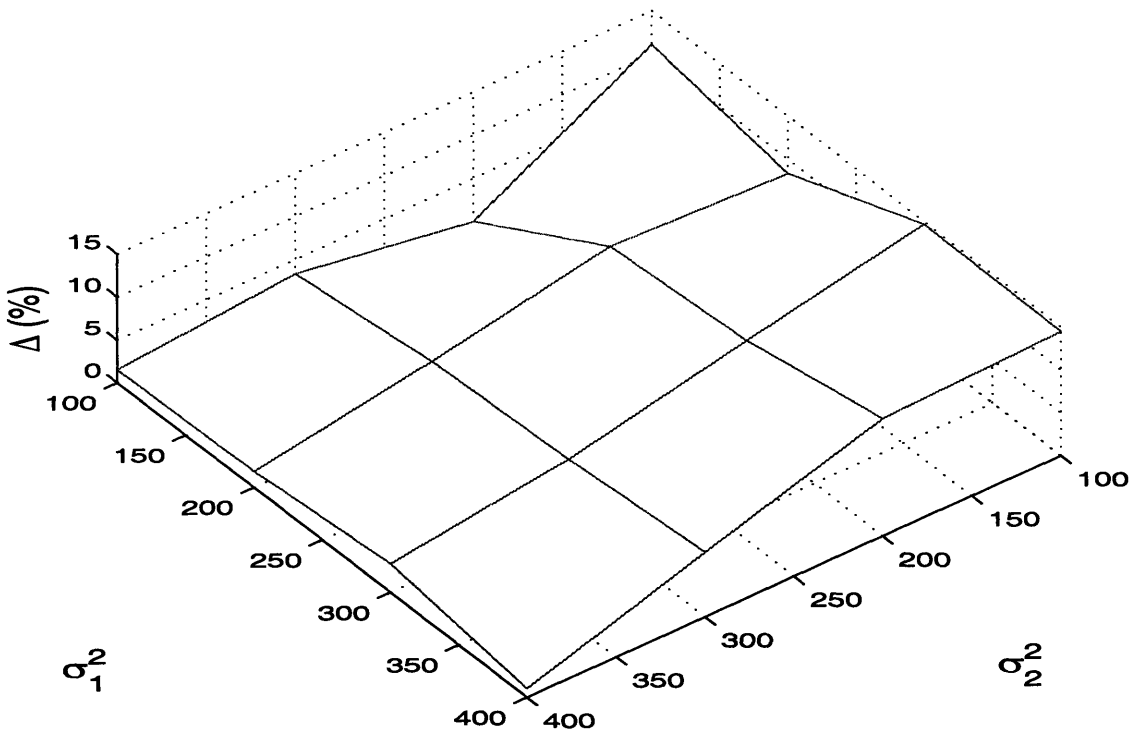

Figure 3.11: $\frac{I_{Lambert}(X_1;X_3)-I_{Linear}(X_1;X_3)}{I_{Lambert}(X_1;X_3)}$ as a function of the pair (σ_1^2,σ_2^2)

Lower Bound to Maximum Mutual Information under Antipodal Input and Linear Amplifier

Consider the channel shown in Figure 3.4. Let X_1 be antipodal, i.e., X_1 takes on equiprobable sample values of $+1$ and -1, and $f_1(Y_1) = Y_1$. Note that the pair $\{X_1, f_1(Y_1)\}$ satisfies energy constraints (3.23), (3.24), and (3.27).

Since relay function $f_1(Y_1)$ is sign-preserving and symmetric, moreover X_1 and X_3 are discrete random variables, the computation of $I_{Linear}(X_1;X_3)$, that is the mutual information of the channel shown in Figure 3.4, given antipodal input and the linear amplifier, follows the same lines as given the pair of antipodal input and the function given by (3.25).

Figure 3.8 shows a 3D plot of $I_{Linear}(X_1;X_3)$ as a function of the pair (σ_1^2,σ_2^2).

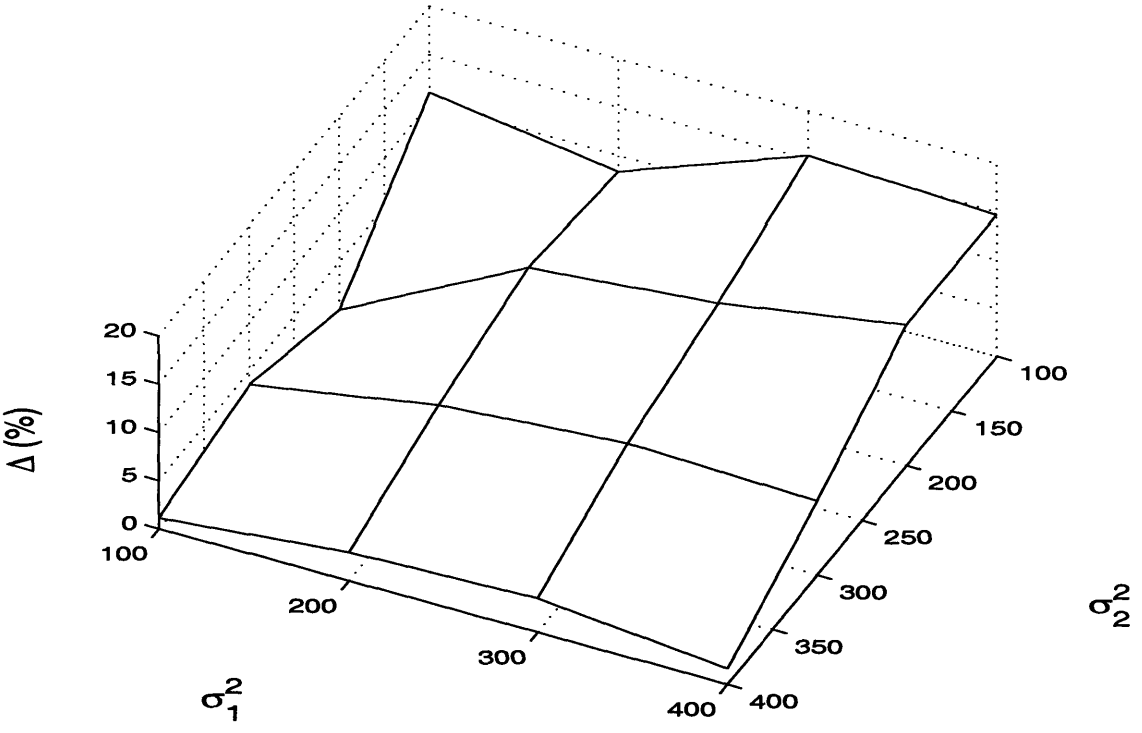

Figure 3.12: $\frac{I_{Lambert}(X_1;X_3)-I_{MMSE}(X_1;X_3)}{I_{Lambert}(X_1;X_3)}$ as a function of the pair (σ_1^2,σ_2^2)

We observe that $I_{Linear}(X_1; X_3)$ decays with σ_1^2 and σ_2^2. Appendix B presents the numerical figures leading to the plot in Figure 3.8.

Lower Bound to Maximu Mutual Information under Antipodal Input and MMSE Estimator

Consider the channel shown in Figure 3.4. Let X_1 be antipodal, i.e., X_1 takes on equiprobable sample values of $+1$ and -1, and and $f_1(Y_1)$ be a minimum mean square error (MMSE) estimator, that is,

$$f_1(Y_1) = \alpha \tanh\left(\frac{Y_1}{\sigma_1^2}\right), \tag{3.32}$$

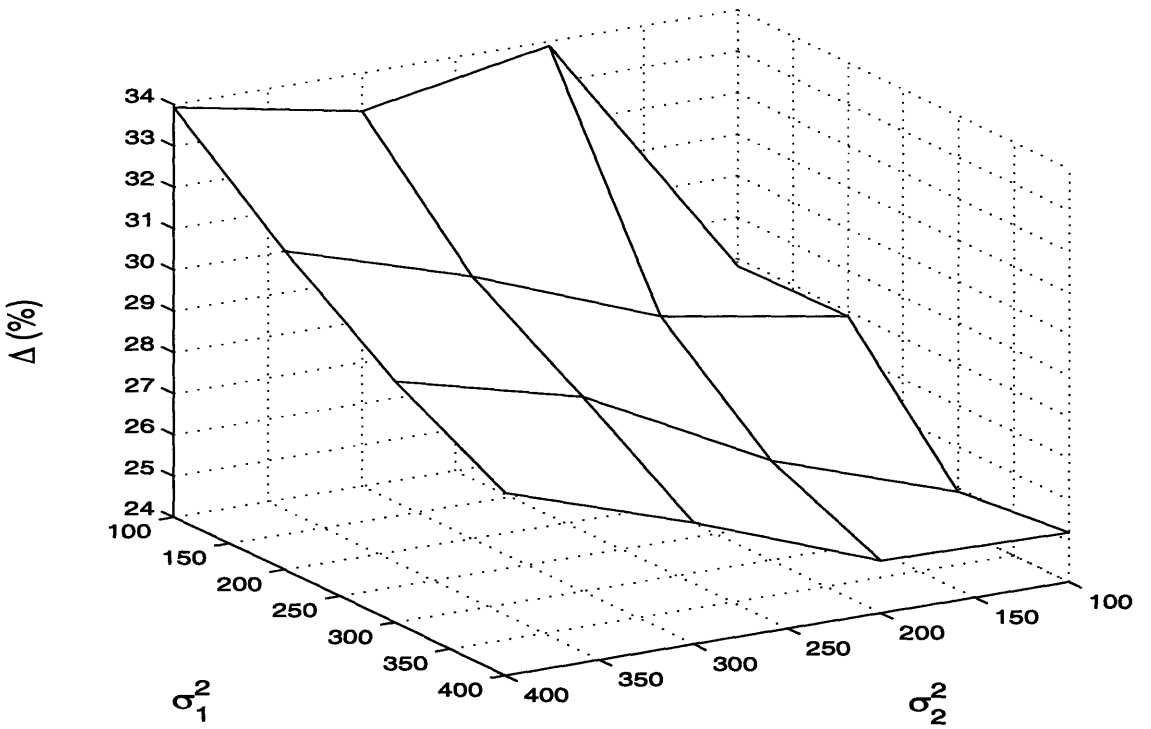

Figure 3.13: $\frac{I_{Gaussian}(X_1;Y_2)-I_{Lambert}(X_1;X_3)}{I_{Gaussian}(X_1;Y_2)}$ as a function of the pair (σ_1^2, σ_2^2)

where α is a constant chosen such that

$$\mathrm{E}[f_1(Y_1)^2] = 1 + \sigma_1^2. \tag{3.33}$$

Note that the pair $\{X_1, f_1(Y_1)\}$ satisfies energy constraints (3.23), (3.24), and (3.27).

What motivates us to consider $f_1(Y_1)$ given by (3.32) is as follows. When concerned with the relay function $f_1(Y_1)$ that minimizes mean square error distortion, i.e., $\mathrm{E}\left[(X_3 - X_1)^2\right]$, for the channel shown in Figure 3.4 (given antipodal input and a hard-decision at the output end), subject to energy constraints (3.23) and (3.24), reference [1] shows that the optimum relay function is the function given by (3.32).

Since relay function $f_1(Y_1)$ is sign-preserving and symmetric, moreover X_1 and X_3 are discrete random variables, the computation of $I_{MMSE}(X_1; X_3)$, that is the

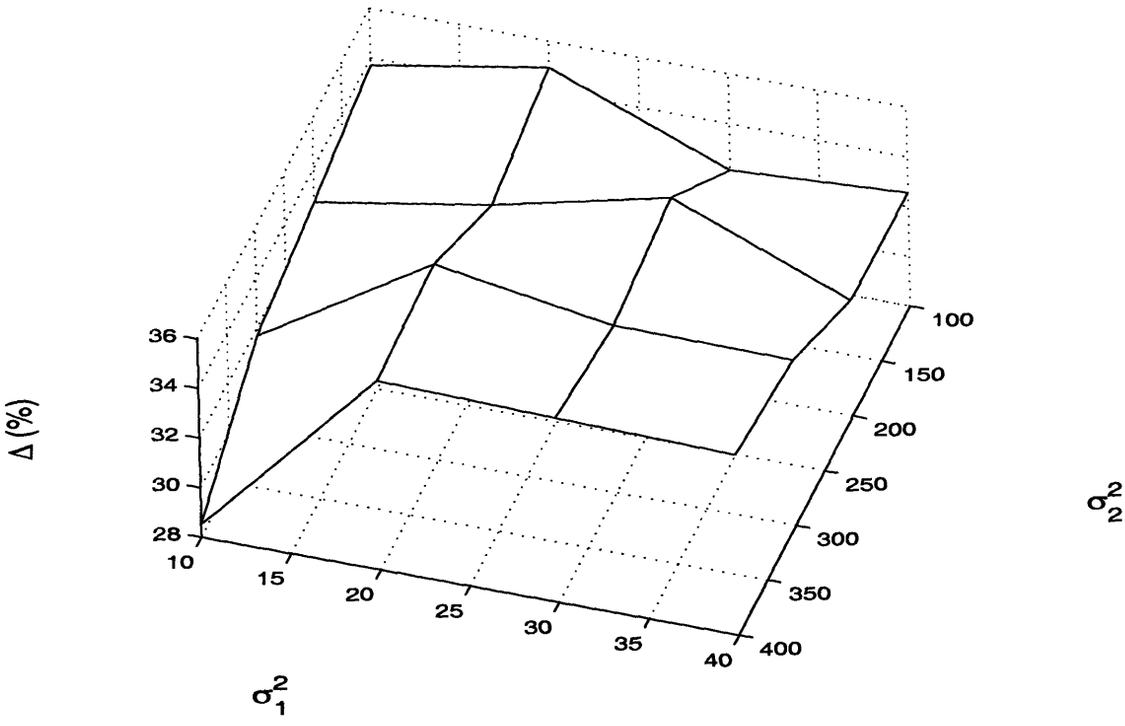

Figure 3.14: $\frac{I_{Gaussian}(X_1;Y_2)-I_{Lambert}(X_1;X_3)}{I_{Gaussian}(X_1;Y_2)}$ as a function of the pair (σ_1^2, σ_2^2)

mutual information of the channel shown in Figure 3.4, given antipodal input and MMSE estimator, follows the same lines as given the pair of antipodal input and the function given by 3.25.

Figure 3.9 shows a 3D plot of $I_{MMSE}(X_1; X_3)$ as a function of the pair (σ_1^2, σ_2^2). We observe that $I_{MMSE}(X_1; X_3)$ decays with σ_1^2 and σ_2^2. Appendix B presents the numerical figures leading to the plot in Figure 3.9.

Lower Bound to Maximum Mutual Information under Gaussian Input and Linear Amplifier

Consider the channel shown in Figure 3.3. Let X_1 be a Gaussian random variable $\mathcal{N}(0,1)$ and $f_1(Y_1) = Y_1$. Let $I_{Gaussian}(X_1; Y_2)$ denote the mutual information at-

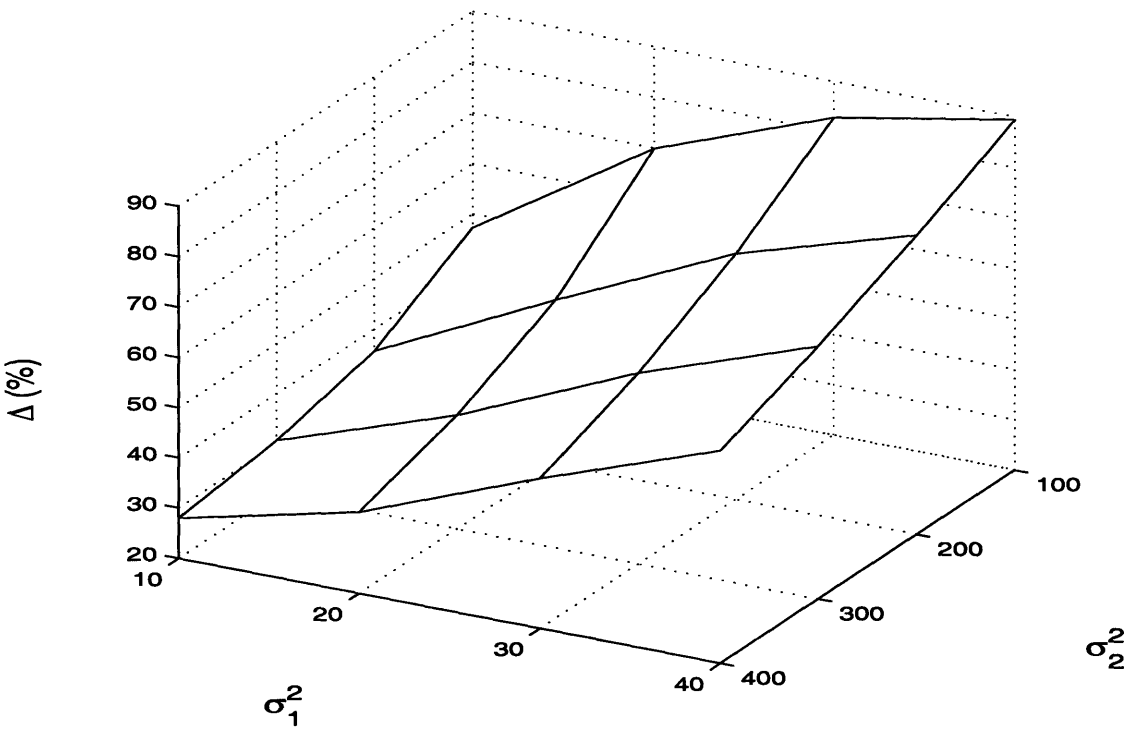

Figure 3.15: $\frac{I^{UB}(X_1;Y_2)-I_{Gaussian}(X_1;Y_2)}{I^{UB}(X_1;Y_2)}$ with $\sigma_1^2 = 10, 20, 30, 40$ and $\sigma_2^2 = 100, 200, 300, 400$

tained under such a pair. For such a channel,

$$I(X_1;Y_2) = \frac{1}{2}\ln\left(1+\frac{1}{\sigma_1^2+\sigma_2^2}\right)$$

Note that the pair $\{X_1, f_1(Y_1)\}$ satisfies energy constraints (3.23), (3.24), and (3.27).

Figure 3.10 shows a 3D plot of $I_{Gaussian}(X_1; X_3)$ as a function of the pair (σ_1^2, σ_2^2). We observe that $I_{Linear}(X_1; X_3)$ decays with σ_1^2 and σ_2^2.

Discussion

Let us turn to Figure 3.11. Figure 3.11 shows a 3D plot of $\frac{I_{Lambert}(X_1;X_3)-I_{Linear}(X_1;X_3)}{I_{Lambert}(X_1;X_3)}$ as a function of the pair (σ_1^2,σ_2^2).

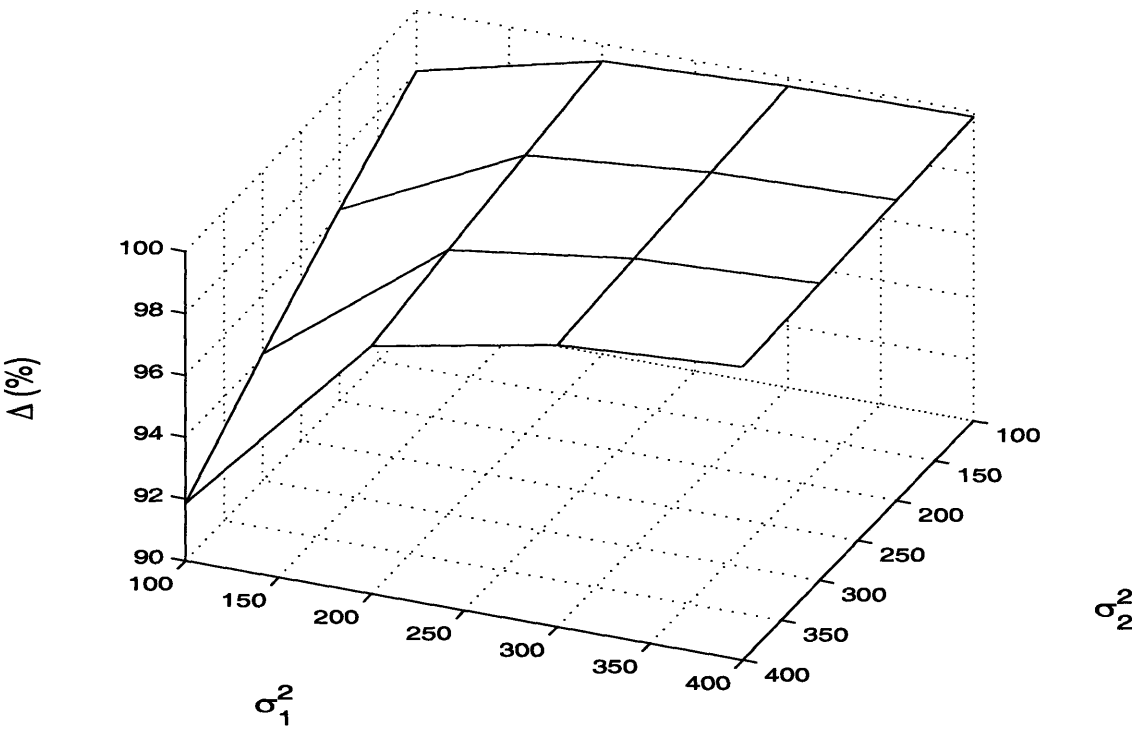

Figure 3.16: $\frac{I^{UB}(X_1;Y_2)-I_{Gaussian}(X_1;Y_2)}{I^{UB}(X_1;Y_2)}$ as a function of the pair (σ_1^2, σ_2^2)

As discussed earlier in this section, $I_{Lambert}(X_1; X_3)$ maximizes mutual information of the channel shown in Figure 3.4 (given antipodal input and a hard-decision at the output end), subject to energy constraints (3.23) and (3.24) and a functional constraint that $f_1(Y_1)$ is sign-preserving and symmetric. This is illustrated by Figure 3.11. The quantity $(I_{Lambert}(X_1; X_3) - I_{Linear}(X_1; X_3))$ is always nonnegative.

When concerned with the relay function $f_1(Y_1)$ that minimizes mean square error distortion, i.e., $\mathrm{E}\left[(X_3 - X_1)^2\right]$ for the channel shown in Figure 3.4 (given antipodal input and a hard-decision at the output end), subject to energy constraints (3.23) and (3.24), reference [1] shows that the optimum relay function is the MMSE estimator given by (3.32). Figure 3.12 shows a 3D plot of $\frac{I_{Lambert}(X_1;X_3)-I_{MMSE}(X_1;X_3)}{I_{Lambert}(X_1;X_3)}$ as a function of the pair (σ_1^2,σ_2^2). The plot demonstrates that $I_{MMSE}(X_1; X_3)$ is

below—although closely approximates—$I_{Lambert}(X_1; X_3)$. For the channel shown in Figure 3.4, subject to energy constraints (3.23) and (3.24), this implies that minimizing minimum mean square error does not imply maximizing capacity.

Figure 3.13 shows 3D plots of $\frac{I_{Gaussian}(X_1;Y_2)-I_{Lambert}(X_1;X_3)}{I_{Gaussian}(X_1;X_3)}$ as a function of the pair (σ_1^2, σ_2^2). The plot demonstrates that $I_{Gaussian}(X_1; Y_2)$, a lower bound to

$$\sup_{(X_1,\, f_1(Y_1)\text{: (3.23),(3.24),(3.27)})} I(X_1; Y_n)$$

for the channel shown in Figure 3.3, is greater than $I_{Lambert}(X_1; X_3)$ for the pairs of (σ_1^2,σ_2^2) considered. It follows that $I_{Gaussian}(X_1; Y_2)$ is the maximum of the four lower bounds that we consider. We shall therefore compare $I^{UB}(X_1; Y_2)$ with $I_{Gaussian}(X_1; Y_2)$.

Figure 3.15 shows a 3D plot of $\frac{I^{UB}(X_1;Y_2)-I_{Gaussian}(X_1;Y_2)}{I^{UB}(X_1;Y_2)}$ with $\sigma_1^2 = 10, 20, 30, 40$ and $\sigma_2^2 = 100, 200, 300, 400$. Figure 3.16 shows a 3D plot of $\frac{I^{UB}(X_1;Y_2)-I_{Gaussian}(X_1;Y_2)}{I^{UB}(X_1;Y_2)}$ with $\sigma_1^2 = 100, 200, 300, 400$ and $\sigma_2^2 = 100, 200, 300, 400$.

Figure 3.15 and 3.16 demonstrate that $I^{UB}(X_1; Y_2) - I_{Gaussian}(X_1; X_3)$ tends to zero as σ_2^2 grows. Particularly, such a difference decays, thus the upper bound is tighter, with σ_2^2, given that σ_1^2 is fixed, and grows, thus the upper bound is looser, with σ_1^2, given that σ_2^2 is fixed.

The insight above brings about the main result of this section. In the context of a two-stage serial distribution network for optical communications, it is reasonable to assume that such a distribution network operates under a noise region in which the noise power of the second section outgrows that of the first section. This is due to the fact that the dominant noise components, such as amplified spontaneous emission (ASE) from Erbium-Doped Fiber Amplifiers (EDFA) and the dark current from avalanche photodiode (APD), are generated at the relay and the receiver [7]. For optical communication purposes, this means that the choice of a code book, consisting of alphabets whose distributions are close to being Gaussian, at the source and a linear function at the relay outperforms other reasonable

choices of code book and relay function.

3.4 Cascade of AWGN Channels by Relay Functions with Memory

Let us recall the result in Section 3.2. Asymptotically in the number of relay stages, capacity grows to infinity, while, on the other hand, the memory length of each relay function decays to one—the relay is memoryless. The result implies that infinite memory requirement for decoding purpose at the destination can be distributed amongst the relay functions.

In this section, we revisit the result of Section 3.2, from a view that relay function is with memory, instead of memoryless as assumed in Section 3.2. For the model considered in this section, however, we further assume that the number of relay stages is finite. The assumptions above imply that coding/decoding are performed at the source/destination as well as the relay. The problem we would like to investigate is described by the following question: what is the the relation of the number of relay stages to the length of memory for decoding purpose at the relays. Later in this section, we will derive a theorem concerning such a relation. We will demonstrate that, asymptotically in the number of relay stages, the theorem agrees with that of Section 3.2.

Let us turn to the cascade in Figure 3.17. Such is a cascade obtained by placing a decoding function f_n and an encoding function g_1 at the input and output ends, and by placing decoding function f_j, $\forall j = 1 \ldots n$, and encoding function g_i, $\forall i = 1 \ldots n-1$, at each stage of the cascade in Figure 1.6.

Referring to Figure 3.17, let N be the length of relay memory and R be the transmission rate. For some $N \in \mathbb{N}$ and $R \in \mathbb{R}^+$, W_i, $\forall i = 1, \ldots, n+1$, is a random variable taking on one of the discrete sample values in the set $\mathcal{W} = \{w_1, \ldots, w_{2^{NR}}\}$.

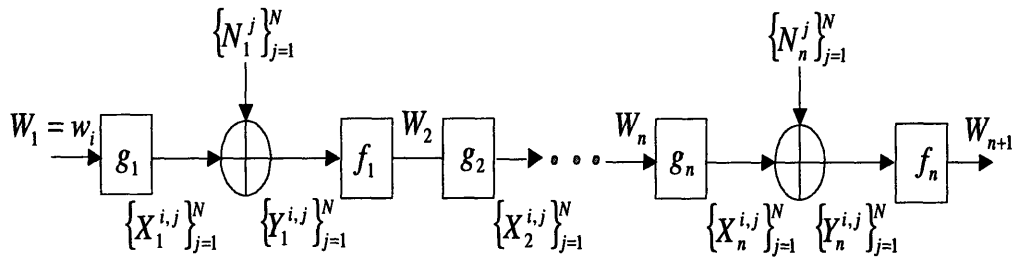

Figure 3.17: A cascade with encoding and decoding functions at each stage of the cascade in Figure 1.6.

Function $g_i : \mathcal{W}_i \rightarrow \mathbb{R}^N$, $\forall i = 1,\ldots,n$, maps each sample value $w_j \in \mathcal{W}_i$ to one of the random codewords $\left(X_i^{j,1},\ldots,X_i^{j,N}\right) \in \mathbb{R}^N$, $\forall j = 1,\ldots,2^{NR}$. Furthermore, we assume $\left\{X_i^{j,k}\right\}$, $\forall i = 1,\ldots,n$, $\forall j = 1,\ldots,2^{NR}$, and $\forall k = 1,\ldots,N$, is a collection of independent and identically distributed zero-mean Gaussian random variables with $E\left[\left(X_i^{j,k}\right)^2\right] \leq \xi$. For $i = 1,\ldots,n$, the function $f_i : \mathbb{R}^N \rightarrow \mathcal{W}$ maps $(Y_i^1,\ldots,Y_i^N) \in \mathbb{R}^N$ to one of the discrete sample values in $\mathcal{W}$ according to the maximum likelihood decision rule, namely choose $w_j \in \mathcal{W}$ if

$$Pr(Y_i^1,\ldots,Y_i^N|w_j) \geq Pr(Y_i^1,\ldots,Y_i^N|w_k)\ \forall k \neq j,\ 1 \leq k \leq 2^{NR}$$

Lastly, for $i = 1,2,\ldots,n$, let N_i be independent zero mean Gaussian random variable with equal variances, $N_i \sim \mathcal{N}(0,\frac{\sigma^2}{n})$.

The capacity of such a cascade is known [6], that is,

$$C = \min_{j=1\ldots n} I(X_i;Y_i) = \frac{1}{2}\ln\left(1+\frac{n\xi}{\sigma^2}\right) \tag{3.34}$$

Now, let n be the number of stages, N be the memory length of the decoding function f_i, $\forall i = 1,\ldots,n$, P_e be the end-to-end error probability, P_e^{UB} be the upper bound on end-to-end error probability, and R be the transmission rate. Then, by the following theorem, we claim a tradeoff between the number of stages in the cascade and the memory length of the decoding function at each stage.

Theorem 19 *Assuming finite end-to-end noise power σ^2 and given any rate $R \in \mathbb{R}^+$ for which $P_e^{UB} \in [0, \frac{1}{2}]$ is arbitrarily small,*

$$\exists \text{ a pair } (n_0, N_0) \text{ s.t. } P_e \leq P_e^{UB} \text{ for } \forall n \geq n_0 \text{ and } \forall N \geq N_0,$$

Moreover, N decreases monotonically as n grows .

Proof. An upper bound for error probability of the AWGN channel was derived in [16], assuming the receiver performs maximum likelihood decoding. We will extend such a result to the cascade shown in Figure 3.17.

Let the end-to-end error probability $P_e \triangleq P(W_1 \neq W_{n+1})$. It follows

$$\begin{aligned} P_e &\overset{(a)}{\leq} \sum_{j=1}^{n} P(W_j \neq W_{j+1}) \\ &\overset{(b)}{=} P_e^{UB}[n] \end{aligned} \tag{3.35}$$

with

$$\begin{aligned} P_e^{UB}[n] &= nC[n] \exp\left(-NE[n]\right) \\ E[n] &= (1-\beta[n]) + \frac{1}{2}\ln\left(\beta[n] + \frac{n\,SNR}{4} - R\right) \\ C[n] &= \pi N e^2 (1-\beta[n])^2 \\ \beta[n] &= \frac{1}{2}\left(1 - \frac{n\,SNR}{2} + \sqrt{1 + \frac{n^2 SNR^2}{4}}\right) \\ SNR &= \frac{\xi}{\sigma^2}. \end{aligned} \tag{3.36}$$

where

(a) follows from applying the union bound on error probability,

(b) follows from a direct application of Gallager's upper bound [16, Eq.(129),(138),(139)] on the probability of error for each stage of the cascade, i.e., the channel between W_j and W_{j+1}, $\forall j = 1 \ldots n$.

From (3.34), there exists n_0 such that $n_0 = \inf\{n : C \leq R\}$ for any given $R \in \mathbb{R}^+$. Next, fixing $n \geq n_0$, (3.36) implies that $\exists N_0$ such that P_e^{UB} monotonically decreases for some $N \geq N_0$, hence $\exists N_1 \geq N_0$ such that, for $N \geq N_1$, P_e^{UB} is less than or equal to the target upper bound on error probability, completing the first statement of the theorem.

From (3.36), it can be seen that $\exists$ a pair (n_2, N_2) such that, for $n \geq n_2$ and $N \geq N_2$, P_e^{UB} monotonically decreases with n and N. Hence, choosing $n \geq \max\{n_0, n_2\}$ and $N \geq \max\{N_1, N_2\}$, the target upper bound on error probability is satisfied and, moreover, P_e^{UB} monotonically decreases with n and N. As such, (3.36) implies that the target upper bound on error probability P_e^{UB} remains satisfied, if n is further increased while N is decreased by some integer value for which the inequality $P_e \leq P_e^{UB}$ holds still, completing the proof. □

Let us now look at what happens when the number of stages in the cascade shown in Figure 3.17 tends to infinity. Theorem 19 tell us that we can achieve rate R, with $0 \leq R < \infty$, at an arbitrarily small error probability, with individual decoding functions whose memory N tends to 1. This asymptotic view agrees with that of Section 3.2, that is, we can achieve infinite capacity without coding as $n \to \infty$.

In what follows we would like to gain more insight by applying Fano's inequality to the cascade show in Figure 3.17. Let us assume that each sample value $w_i \in \mathcal{W}, i = 1, \ldots, 2^{NR}$, is equiprobable. The following lemma follows.

Lemma 20 *For the channel shown in Figure 3.17,*

$$\log M - P_e \log(M-1) - H(P_e) \leq I(W_1, W_{n+1}) \text{ [bit]}$$

where P_e is the end-to-end error probability, i.e., the probability that $W_1 \neq W_{n+1}$, and M is the size of the signal set for W_1, that is, $|\mathcal{W}|$.

Proof. By applying Fano's inequality, we will show a lower bound on the mutual information between W_1 and W_{n+1}. An interesting fact to note is that the lower bound is expressed only in terms of the size of the signal set for W_1, i.e., M, and the probability of error, i.e., the probability that $W_1 \neq W_{n+1}$.

For the lower bound, let us define an error indicator variable E, that is, $E = 1$ if $W_1 \neq W_{n+1}$ and $E = 0$ else. Let the probability of error $P_e \triangleq P(E = 1)$. Now,

$$\begin{aligned} I(W_1; W_{n+1}) &\overset{(a)}{=} H(W_1) + H(W_1|W_{n+1}) \\ &\overset{(b)}{\geq} \log|\mathcal{W}| - P_e \log(|\mathcal{W}| - 1) - H(P_e) \\ &\overset{(c)}{=} \log M - P_e \log(M-1) - H(P_e) \end{aligned} \tag{3.37}$$

where

(a) follows from the definition of mutual information,

(b) follows since $H(W_1) = \log|\mathcal{W}|$ and, by Fano's inequality, $H(W_1|W_{n+1}) \leq P_e \log(|\mathcal{W}| - 1) + H(P_e)$, with

$$H(P_e) = P_e \log\left(\frac{1}{P_e}\right) + (1 - P_e) \log\left(\frac{1}{1 - P_e}\right),$$

(c) from the fact that $|\mathcal{W}| = M$, completing the proof. □

Let us turn to the cascade shown in Figure 3.17 again and consider two such cascades with different number of stages. Assuming $n_A > n_B$ and $n_A, n_B \in \mathbb{N}$, cascade A is a cascade with n_A stages and cascade B is a cascade with n_B stages. Next, suppose that we are sending information at the same achievable rate R, i.e., $R \leq \left[\frac{1}{2}\ln\left(1 + \frac{n_B \xi}{\sigma^2}\right)\right]$, on both cascade A and B. Finally, let P_e be the end-to-end error probability, i.e., the probability that $W_1 \neq W_{n+1}$. The following theorem follows.

Theorem 21 *Lower and upper bounds on* $\frac{I(W_1;W_{n_A+1})}{N}$ *and* $\frac{I(W_1;W_{n_B+1})}{N}$ *for cascade A and B, respectively, are :*

$$R(1-P_e^{UB}[n_A])-\frac{1}{N} \leq \frac{I(W_1;W_{n_A+1})}{N} \leq R; \tag{3.38}$$

$$R(1-P_e^{UB}[n_B])-\frac{1}{N} \leq \frac{I(W_1;W_{n_B+1})}{N} \leq R. \tag{3.39}$$

with N being the memory length of each decoding function f_i, $i = 1,\ldots,n$, *and* P_e^{UB} *being specified by (3.35) and (3.36).*

Both lower bounds expressed in (3.38) and (3.39) approach the upper bound R with N. In particular, the lower bound for cascade A approaches R faster, with N, than that for cascade B.

Proof. From Lemma 20,

$$\begin{aligned} I(W_1;W_{n+1}) &\overset{(a)}{\geq} \log(|\mathcal{W}|)-P_e\log(|\mathcal{W}|)-H(P_e) \\ &\overset{(b)}{\geq} NR(1-P_e)-1 \end{aligned} \tag{3.40}$$

where

(a) follows from Lemma 20,

(b) follows since $\log(|\mathcal{W}|) = NR$ and $H(P_e) \leq 1$.

Moreover,

$$\begin{aligned} I(W_1;W_{n+1}) &\overset{(a)}{=} H(W_1)-H(W_1|W_{n+1}) \\ &\overset{(b)}{\leq} H(W_1) \\ &\overset{(c)}{\leq} NR \end{aligned} \tag{3.41}$$

where

(a) follows from the definition of mutual information,

(b) follows since discrete entropy $H(W_1|W_{n+1}) \geq 0$,

(c) follows since $H(W_1) = \log(|\mathcal{W}|) = NR$.

From (3.40) and (3.41), it follows

$$NR(1-P_e) - 1 \leq I(W_1; W_{n+1}) \leq NR. \tag{3.42}$$

Dividing both sides of (3.42) with N,

$$R(1-P_e) - \frac{1}{N} \leq \frac{I(W_1; W_{n+1})}{N} \leq R, \tag{3.43}$$

Note that, as $N \to \infty$, thus $P_e \to 0$, the lower bound meets the upper bound and

$$\lim_{N\to\infty} \frac{I(W_1; W_{n+1})}{N} = R.$$

Applying (3.43) to cascade A, we get

$$R(1-P_e^{UB}[n_A]) - \frac{1}{N} \leq \frac{I(W_1; W_{n_A+1})}{N} \leq R. \tag{3.44}$$

since from (3.35), $P_e \leq P_e^{UB}[n_A]$, with $P_e^{UB}[n_A] = n_A C[n_A] \exp[-NE[R, n_A]]$. Similarly, applying (3.43) to cascade B, we get

$$R(1-P_e^{UB}[n_B]) - \frac{1}{N} \leq \frac{I(W_1; W_{n_B+1})}{N} \leq R. \tag{3.45}$$

since from (3.35), $P_e \leq P_e^{UB}$, with $P_e^{UB} = n_B C[n_B] \exp[-NE[R, n_B]]$

Now, let us contrast how the lower bounds approache R in the limit as the memory length of individual decoding functions f_i, $i = 1, \ldots, n$, $N \to \infty$ on both cascades. Taking derivatives of the lower bounds in (3.44) and (3.45) with respect to N,

$$\begin{aligned} \frac{d\left(R(1-P_e^{UB}[n_A]) - \frac{1}{N}\right)}{dN} &= n_A R K_A (NE[n_A] - 1) \exp[-NE[n_A]] + \frac{1}{N^2} \\ &= O(\exp[-NE[n_A]]), \end{aligned} \tag{3.46}$$

$$\begin{aligned}\frac{d\left(R(1-P_e^{UB}[n_B])-\frac{1}{N}\right)}{dN} &= RK_B(NE[n_B]-1)\exp\left[-NE[n_B]\right]+\frac{1}{N^2}\\ &= O(\exp\left[-NE[n_B]\right]), \qquad (3.47)\end{aligned}$$

for some constants K_A for cascade A and K_B for cascade B. From (3.36) and the assumption that $n_A > n_B$, it follows that the lower bound of (3.44) approaches R faster with N than that of (3.45). Such is true since

$$\frac{d\left(R(1-P_e^{UB}[n_A])-\frac{1}{N}\right)}{dN} < \frac{d\left(R(1-P_e^{UB}[n_B])-\frac{1}{N}\right)}{dN},$$

and the second derivatives for the lower bounds are negative $\forall N$, completing the proof. □

Although hardly a strong argument, in fact, almost an abuse of interpretation, the fact that the lower bound of (3.38) approaches R faster than that of (3.39), with respect the to memory length N of individual decoders f_i, $i = 1,\ldots,n$, gives us a 'feel' the rate at which $\frac{I(W_1;W_{n_A+1})}{N} \to R$ on cascade A is faster than the rate at which $\frac{I(W_1;W_{n_B+1})}{N} \to R$ on cascade B.

3.5 Conclusions

We have considered point-to-point communication over a series of AWGN channels. In such a series, one section of the channel is coupled with the next by a *relay function*. This communication set-up is particularly relevant in optical communications.

We have studied the maximum rate at which information can be reliably transmitted over a series of AWGN channels with memoryless relays. We have shown the following two results. First, assuming end-to-end noise power is finite and the serial relay channel is subject to energy constraints, we have shown that the capacity tends to infinity with the number of relay stages in the series. This result, though based on a set of assumptions that is not compatible at all with the serial

channel model for optical communication purposes, points to the following insight. A 'folk theorem' stated that a hard-decision function is the best memoryless relay at low SNR and linear amplification function is the best at high SNR. Using the particular model used, we have argued that such is not always the case. Second, given a finite number of relay stages, finding maximum mutual information subject to energy constraints alone is difficult. In addition to energy constraints, we have proposed entropy constraints at the output of individual relays. We have derived an explicit upper bound to mutual information, subject to these energy and entropy constraints. For a particular series of two AWGN channels, we have numerically demonstrated that lower bound attained by the pair of Gaussian input and linear relay function approaches the upper bound more closely than that attained by other choices of pair of input and relay function that we consider. In the context of a two-stage serial distribution network for optical communications, the noise power of the second section is likely to outgrow that of the first section. This is due to the fact that the major noise components are the noise from the relay—such as amplified stimulated emission (ASE) from erbium-doped fiber amplifier (EFDA)—and that from the receiver—such as the noise from the dark current generated by avalanche photodiode (APD). For optical communication purposes, this means that the choice of a code book, consisting of alphabets whose distributions are close to being Gaussian, at the source and a linear function at the relay outperforms other reasonable choices of code book and relay function.

Next, we have revisited the result from the model for infinite series of AWGN channels with memoryless relays, from a view that relay function is with memory, rather than memoryless—thus allowing coding/decoding at individual relays. We have shown two tradeoffs. First, given that information is sent at a fixed rate R and given the requirement that end-to-end error probability is bounded above by some arbitrary quantity, we have shown that the requisite length of memory N for decoding purpose at each stage is monotonically decreasing in the number of relay

stages n. Second, given that information is sent at an achievable rate R, we have shown the rate at which mutual information per unit memory tends to R is higher as the number of relay stages n grows.

Chapter 4

Conclusions and Direction on Further Research

Of all the issues surrounding point-to-point communication over relay networks, we have studied a theme which has not received much attention in the context of capacity. The theme is best described by the following questions. Can we consider, from a capacity point of view, the effect of finite relay memory over the different types of relay channels? In particular, when faced with limited memory at relays, what is the maximum rate at which information can be reliably transmitted over the relays?

In this thesis, we have proposed two capacity approaches on point-to-point communication over a set of relay nodes with finite processing memory. The essential difference between the first and the second approach is that point-to-point communication in the first utilizes of a set of parallel processing nodes, while in the second a set of serial processing nodes. In what follows, we state conclusions and direction for further study.

4.1 On the Capacity of a Parallel Relay Channel

We have studied the problem of distributed processing for point-to-point communication over parallel relays. Such a processing for relay communication purpose is relevant to a problem in the sensor networks literature, notably the problem of transporting data to a far-receiver via the intermediate sensors. The model we have chosen to represent the structure of such a problem is the single-input multi-output AWGN channel with and without inter-symbol interference (ISI). Our results point to the implementation of distributed processing via one dimensional (1D) and two dimensional (2D) Kalman filter. With 1D Kalman filter, processing proceeds in *space*, namely from one to the next intermediate receiver. With 2D Kalman filter, in addition to progressing in *space*, processing proceeds in *time*, namely from one to the next time step. We have found the capacity of the distributed and optimally-—that is, centrally—processed channels to be the same. In computing the capacity, we have made use of the direct relation between capacity and estimation theory. Under 1D and 2D Kalman filter, processing is linear. However, 2D Kalman filter mitigates a problem—of infinite memory for processing at each intermediate receiver—imposed by 1D Kalman filter. Under 2D Kalman filter, intermediate processing is done with finite memory.

Our treatment of distributed processing for point-to-point communication over parallel relays considers transmission capacity as the metric to contrast distributed against optimal (centralized) processing of observations. In order to derive capacity for communication under distributed processing, we have made use of the direct relation between *transmission capacity* and *estimation theory*.

When the capacity achieving input process (a wide sense stationary stochastic process) has infinite memory, a truncation strategy is needed. The truncation strategy will allocate some sensible finite amount of processing memory in the state vector of the state-space estimation model. By sensible, the strategy will reason-

ably narrow the error covariance gap between the two estimates, i.e., the estimate when the state vector has infinite memory and that when the respective memory in the state vector is truncated. One approach to truncating memory, when infinite, is a decomposition approach using the prolate spheroidal wave functions [12]. This problem is not elaborated in this chapter and is subject to further study.

An interesting extension to study is that of distributed processing on a channel that is not perfectly known at the intermediate receivers, namely a channel whereby the intermediate receivers perform channel estimation with some error. A cumulative error would build up in the case of such a channel. The results based on [19] may be extended to this case.

4.2 On the Capacity of a Serial Relay Channel

We have considered point-to-point communication over a series of AWGN channels. In such a series, one section of the channel is coupled with the next by a *relay function*. This communication set-up is particularly relevant in optical communications.

We have studied the maximum rate at which information can be reliably transmitted over a series of AWGN channels with memoryless relays. We have shown the following two results. First, assuming end-to-end noise power is finite and the serial relay channel is subject to energy constraints, we have shown that the capacity tends to infinity with the number of relay stages in the series. This result, though based on a set of assumptions that is not compatible at all with the serial channel model for optical communication purposes, points to the following insight. A 'folk theorem' stated that a hard-decision function is the best memoryless relay at low SNR and linear amplification function is the best at high SNR. Using the particular model used, we have argued that such is not always the case. Second, given a finite number of relay stages, finding maximum mutual information

subject to energy constraints alone is difficult. In addition to energy constraints, we have proposed entropy constraints at the output of individual relays. We have derived an explicit upper bound to mutual information, subject to these energy and entropy constraints. For a particular series of two AWGN channels, we have numerically demonstrated that lower bound attained by the pair of Gaussian input and linear relay function approaches the upper bound more closely than that attained by other choices of pair of input and relay function that we consider. In the context of a two-stage serial distribution network for optical communications, the noise power of the second section is likely to outgrow that of the first section. This is due to the fact that the major noise components are the noise from the relay—such as amplified stimulated emission (ASE) from erbium-doped fiber amplifier (EFDA)—and that from the receiver—such as the noise from the dark current generated by avalanche photodiode (APD). For optical communication purposes, this means that the choice of a code book, consisting of alphabets whose distributions are close to being Gaussian, at the source and a linear function at the relay outperforms other reasonable choices of code book and relay function.

Next, we have revisited the result from the model for infinite series of AWGN channels with memoryless relays, from a view that relay function is with memory, rather than memoryless—thus allowing coding/decoding at individual relays. We have shown two tradeoffs. First, given that information is sent at a fixed rate R and given the requirement that end-to-end error probability is bounded above by some arbitrary quantity, we have shown that the requisite length of memory N for decoding purpose at each stage is monotonically decreasing in the number of relay stages n. Second, given that information is sent at an achievable rate R, we have shown the rate at which mutual information per unit memory tends to R is higher as the number of relay stages n grows.

Appendix A

Proof of the Conditional Entropy-Power Inequality

Note that the steps of the proof are based on a proof of Shannon's entropy-power inequality [3]. The extension which is required to prove the desired conditional entropy-power inequality is almost trivial.

A.1 The Derivative of $h(X_f|W)$

We first define the conditional random variable X_f given W to be the sum of a conditional random variable X given W with probability density $p_{X|W}(x)$ plus an independent (with respect to X and W) zero-mean normal random variable with variance f. We show that the derivative of its conditional entropy $h(X_f|W = w)$ with respect to f is

$$\frac{1}{2}J_W(X_f) = \frac{1}{2}\int_{-\infty}^{\infty}\left(\frac{\partial p_{X_f|W}(x_f,f)}{\partial x_f}\right)^2 \frac{dx_f}{p_{X_f|W}(x_f,f)}\ , \tag{A.1}$$

where

$$p_{X_f|W}(x_f,f) = \frac{1}{\sqrt{2\pi f}}\int_{-\infty}^{\infty} p_{X|W}(x)\exp -\frac{(x_f-x)^2}{2f}\,dx \tag{A.2}$$

is the probability density of X_f given W.

Differentiating (A.2) inside the integral, we get the diffusion equation

$$\frac{\partial p_{X_f|W}(x_f, f)}{\partial f} = \frac{1}{2}\frac{\partial^2 p_{X_f|W}(x_f, f)}{\partial x_f^2} \tag{A.3}$$

From (A.2) we see that $p_{X_f|W} \leq 1/\sqrt{2\pi f}$, hence $h(X_f|W = w)$ either converges absolutely or is $+\infty$. For $f_2 > f_1$,

$$h(X_{f_1}|W = w) \overset{(a)}{\leq} h(X_{f_2}|W = w) \overset{(b)}{\leq} h(X_{f_1}|W = w) + \frac{1}{2}\log(f_2/f_1),$$

where

(a) follows since, for $f_2 > f_1$ and X_{f_2} being X_{f_1} plus some independent zero-mean normal random variable with variance $f_2 - f_1$, the mutual information $I(X_{f_2} - X_{f_1}; X_{f_2}|W = w) \geq 0$;

(b) follows since, for $f_2 > f_1$, the mutual information $I(X; X_{f_1}|W = w) \geq I(X; X_{f_2}|W = w)$.

Hence, $h(X_f|W = w) < \infty$ for all $f > 0$ or $h(X_f|W = w) = +\infty$ for all $f > 0$. We will show later that $h(X_f|W = w) < \infty$. Let us assume for now that $h(X_f|W =$

$w) < \infty$. Differentiating $h(X_f|W = w)$ inside the integral, we get

$$\begin{aligned}
\frac{dh(X_f|W=w)}{df} &= -\int_{-\infty}^{\infty} \frac{\partial p_{X_f|W}(x_f,f)}{\partial f} dx_f - \\
&\quad \int_{-\infty}^{\infty} \frac{\partial p_{X_f|W}(x_f,f)}{\partial f} \log p_{X_f|W}(x_f,f) dx_f \qquad \text{(A.4)} \\
&\overset{(a)}{=} 0 - \frac{1}{2}\int_{-\infty}^{\infty} \frac{\partial^2 p_{X_f|W}(x_f,f)}{\partial x_f^2} \log p_{X_f|W}(x_f,f) dx_f \\
&\overset{(b)}{=} -\frac{1}{2}\frac{\partial p_{X_f|W}(x_f,f)}{\partial x_f} \log p_{X_f|W}\Big|_{x_f=-\infty}^{\infty} \qquad \text{(A.5)} \\
&\quad + \frac{1}{2}\int_{-\infty}^{\infty} \left(\frac{\partial p_{X_f|W}(x_f,f)}{\partial x_f}\right)^2 \frac{dx_f}{p_{X_f|W}(x_f,f)} \\
&\overset{(c)}{=} 0 + \frac{1}{2}\int_{-\infty}^{\infty} \left(\frac{\partial p_{X_f|W}(x_f,f)}{\partial x_f}\right)^2 \frac{dx_f}{p_{X_f|W}(x_f,f)} \\
&\overset{(d)}{=} \frac{1}{2} J_W(X_f)\,, \qquad \text{(A.6)}
\end{aligned}$$

where

(a) follows from (A.3). The first term on the right hand side of (A.4) is 0 since

$$\begin{aligned}
-\int_{-\infty}^{\infty} \frac{\partial p_{X_f|W}(x_f,f)}{\partial f} dx_f &= \int_{-\infty}^{\infty} \frac{\partial^2 p_{X_f|W}(x_f,f)}{\partial x_f^2} dx_f \\
&= \frac{\partial p_{X_f|W}(x_f,f)}{\partial x_f}\Big|_{x_f=-\infty}^{\infty} \\
&= \left[-\frac{1}{\sqrt{2\pi f^3}}\int_{-\infty}^{\infty} p_{X|W}(x)(x_f - x)\exp -\frac{(x_f-x)^2}{2f}\, dx\right]\Big|_{x_f=-\infty}^{\infty} \\
&= 0;
\end{aligned}$$

(b) follows after performing integration by parts;

(c) follows from Lemma 1;

(d) follows from (A.1). Since the integrand of (A.1) is never negative, $J(X_f)$ either converges or is $+\infty$. Dividing (A.8) by $p_{X_f|W}$ and integrating over all x_f, we see that $J_W(X_f) \leq 1/f$. Hence, $J_W(X_f)$ is finite for all $f > 0$, converging uniformly for $0 < \epsilon \leq f$ and thus justifying (A.6).

Lemma 1 *For $f > 0$,*

$$-\frac{1}{2}\frac{\partial p_{X_f|W}(x_f,f)}{\partial x_f}\log p_{X_f|W}(x_f,f)\Big|_{x_f=-\infty}^{\infty} = 0 \tag{A.7}$$

Proof.

Consider the following string of inequalities.

$$\begin{aligned}\left(\frac{\partial p_{X_f|W}(x_f,f)}{\partial x_f}\right)^2 &= \left[\int_{-\infty}^{\infty}\frac{\sqrt{p_{X_f|W}(x_f,f)}\exp-\frac{(x_f-x)^2}{4f}}{\sqrt[4]{2\pi f}}\right.\\ &\quad\left.\cdot\ \frac{\sqrt{p_{X_f|W}(x_f,f)}(x_f-x)\exp-\frac{(x_f-x)^2}{4f}}{\sqrt[4]{2\pi f^5}}dx\right]^2\\ &\overset{(a)}{\leq} \frac{p_{X_f|W}(x_f,f)}{\sqrt{2\pi f^5}}\\ &\quad\int_{-\infty}^{\infty}p_{X|W}(x)(x_f-x)^2\exp-\frac{(x_f-x)^2}{2f}\,dx \qquad \text{(A.8)}\\ &\overset{(b)}{\leq} \sqrt{\frac{2}{\pi f^3}}\frac{p_{X_f|W}(x_f,f)}{e}, \qquad \text{(A.9)}\end{aligned}$$

where

(a) follows by applying Schwarz's inequality to

$$\frac{\partial p_{X_f|W}(x_f,f)}{\partial x_f} = -\frac{1}{\sqrt{2\pi f^3}}\int_{-\infty}^{\infty}p_{X|W}(x)(x_f-x)\exp-\frac{(x_f-x)^2}{2f}\,dx;$$

(b) follows by finding the maximum of $(x_f-x)^2\exp-(x_f-x)^2/2f$, i.e., a concave function. After taking the first derivative with respect to x_f-x and setting it to zero, we find the argument $(x_f-x)^*$ which sets $(x_f-x)^2\exp-(x_f-$

$x)^2/2f$ to zero. We then take its second derivative and verify the second derivative is negative given that $(x_f - x) = (x_f - x)^*$. It follows that $\max_{x_f - x} (x_f - x)^2 \exp -(x_f - x)^2/2f = 2f/e$, hence $\int_{-\infty}^{\infty} p_{X|W}(x)(x_f - x) \exp -\frac{(x_f-x)^2}{2f}\, dx \leq 2f/e$.

From (A.9) we see that $\frac{\partial p_{X_f|W}(x_f,f)}{\partial x_f}$ is, at most, of the order of $\sqrt{p_{X_f|W}(x_f,f)}$ as $p_{X_f|W}(x_f,f) \to 0$, and so $\frac{\partial p_{X_f|W}(x_f,f)}{\partial x_f} \log p_{X_f|W}(x_f,f) \to 0$ as $x_f \to \pm\infty$. □

A.2 The Convolution Inequality For J

Given a random variable W, let X, Y, and $Z = X + Y$ be some random variables with Y being independent of X and W. The conditional probability densities of X, Y, and Z are $p_{X|W=w}(x)$, $p_Y(y)$, and

$$p_{Z|W=w}(z) = \int_{\infty}^{\infty} p_{X|W=w}(x) p_Y(z-x) dx\,, \tag{A.10}$$

respectively. If these density functions are differentiable, then $dp_{Z|W=w}(z)/dz$ is

$$p'_{Z|W=w}(z) = \int_{\infty}^{\infty} p'_{X|W=w}(x) p_Y(z-x) dx\,. \tag{A.11}$$

Thus, if $p_{X|W=w}(x)$, $p_Y(y)$, and $p_{Z|W=w}(z)$ never vanish,

$$\begin{aligned} \frac{p'_{Z|W=w}(z)}{p_{Z|W=w}(z)} &= \int_{-\infty}^{\infty} \frac{p_{X|W=w}(x) p_Y(z-x)}{p_{Z|W=w}(z)} \frac{p'_{X|W=w}(x)}{p_{X|W=w}(x)} dx \\ &= E_{X|Z,W}\left[\frac{p'_{X|W=w}(x)}{p_{X|W=w}(x)} \middle| Z = z, W = w \right], \end{aligned}$$

that is, the conditional expectation of $p'_{X|W=w}(x)/p_{X|W=w}(x)$ for given values of Z and W. Likewise,

$$\frac{p'_{Z|W=w}(z)}{p_{Z|W=w}(z)} = E_{Y|Z,W}\left[\frac{p'_Y(y)}{p_Y(y)} \middle| Z = z \right].$$

Therefore, for any constants a and b,

$$E_{X,Y|Z,W}\left[a\frac{p'_{X|W=w}(x)}{p_{X|W=w}(x)}+b\frac{p'_Y(y)}{p_Y(y)}\middle| Z=z,W=w\right] = (a+b)\frac{p'_{Z|W=w}(z)}{p_{Z|W=w}(z)}.$$

Hence,

$$\begin{aligned}(a+b)^2\left[\frac{p'_{Z|W=w}(z)}{p_{Z|W=w}(z)}\right]^2 &= \left\{E_{X,Y|Z,W}\left[a\frac{p'_{X|W=w}(x)}{p_{X|W=w}(x)}+b\frac{p'_Y(y)}{p_Y(y)}\middle| Z=z,W=w\right]\right\}^2 \\ &\le E_{X,Y|Z,W}\left\{\left[a\frac{p'_{X|W=w}(x)}{p_{X|W=w}(x)}+b\frac{p'_Y(y)}{p_Y(y)}\right]^2\middle| Z=z,W=w\right\},\end{aligned} \tag{A.12}$$

since the second moment of any random variable is never less than the square of its mean, with equality only if

$$a\frac{p'_{X|W=w}(x)}{p_{X|W=w}(x)}+b\frac{p'_Y(y)}{p_Y(y)}=(a+b)\frac{p'_{Z|W=w}(z)}{p_{Z|W=w}(z)} \tag{A.13}$$

with probability one.

Averaging both sides of (A.12) over the conditional distribution of Z given $W=w$ gives us

$$\begin{aligned}(a+b)^2E_{Z|W}\left\{\left[\frac{p'_{Z|W=w}(z)}{p_{Z|W=w}(z)}\right]^2\middle| W=w\right\} &\le E_{X,Y|W}\left\{\left[a\frac{p'_{X|W=w}(x)}{p_{X|W=w}(x)}+b\frac{p'_Y(y)}{p_Y(y)}\right]^2\middle| W=w\right\} \\ &= a^2E_{X|W}\left\{\left[\frac{p'_{X|W=w}(x)}{p_{X|W=w}(x)}\right]^2\middle| W=w\right\} \\ &\quad +b^2E_Y\left\{\left[\frac{p'_Y(y)}{p_Y(y)}\right]^2\right\}\end{aligned} \tag{A.14}$$

or

$$(a+b)^2J_W(Z)\le a^2J_W(X)+b^2J_W(Y) \tag{A.15}$$

Averaging both sides of (A.15) over the distribution of W gives us

$$(a+b)^2E_W[J_W(Z)]\le a^2E_W[J_W(X)]+b^2E_W[J_W(Y)]. \tag{A.16}$$

Setting $a = 1/E_W[J_W(X)]$ and $b = a = 1/E_W[J_W(Y)]$, we obtain

$$\frac{1}{E_W[J_W(Z)]} \geq \frac{1}{E_W[J_W(X)]} + \frac{1}{E_W[J_W(Y)]}, \tag{A.17}$$

with equality if (A.13) holds. Let us suppose (A.13) holds. By substituting $x = z - y$ into (A.13) and integrating with respect to y, we get

$$-a \log p_{X|W=w}(z-y) + b \log p_{Y|W=w}(y) = (a+b)\frac{p'_{Z|W=w}(z)}{p_{Z|W=w}(z)} y + c(z) .$$

Setting $y = 0$ shows that the "constant" of integration $c(z)$ is differentiable, and it follows that $\frac{p'_{Z|W=w}(z)}{p_{Z|W=w}(z)} y$, too, is differentiable. Thus, differentiating with respect to z and setting $z = 0$, we have

$$-a\frac{p'_{X|W=w}(-y)}{p_{X|W=w}(-y)} = (a+b)\frac{p_{Z|W=w}(0)\, p''_{Z|W=w}(0) - p'^2_{Z|W=w}(0)}{p^2_{Z|W=w}(0)} y + c(0) , \tag{A.18}$$

from which we see that (A.18) is satisfied if $p_{X|W=w}(x)$ is a normal density function. Similarly, the condition (A.13) for equality in (A.17) implies that equality is satisfied if $p_Y(y)$ is a normal density function. Thus, (A.17) holds with equality when $p_{X|W=w}(x)$ and $p_Y(y)$ are normal density functions.

A.3 One-Dimensional Conditional Entropy-Power Inequality

Given $W = w$, we now define $X_{f(t)}$, $Y_{g(t)}$, and $Z_{j(t)}$ to be X, Y, and Z plus independent zero-mean normal random variables with variance $f(t)$, $g(t)$, and $j(t) = f(t) + g(t)$, respectively. We suppose that $f(0) = g(0) = j(0) = 0$, and we restrict t to non-negative values. When $f(t)$ and $g(t)$ are positive, the conditional probability density functions of $X_{f(t)}$, $Y_{g(t)}$, and $Z_{j(t)}$, given $W = w$ are everywhere differentiable and positive. Thus, differentiating

$$s(t) \triangleq \frac{\exp 2h(X_{f(t)}|W) + \exp 2h(Y_{g(t)}|W)}{\exp 2h(Z_{j(t)}|W)} \tag{A.19}$$

with respect to t with the help of (A.6), we have

$$\begin{aligned} e^{2h(Z_{j(t)}|W)}s'(t) &= e^{2h(X_{f(t)}|W)}f'(t)E_W[J_W(X_{f(t)})] + e^{2h(Y_{g(t)}|W)}g'(t)E_W[J_W(Y_{g(t)})] \\ &- \left[e^{2h(X_f(t)|W)} + e^{2h(Y_{g(t)}|W)}\right][f'(t) - g'(t)]E_W[J_W(Z_{j(t)})]\,. \quad \text{(A.20)} \end{aligned}$$

Substituting the upper bound on $E_W[J_W(Z_{j(t)})]$ given by (A.17), we find that

$$\begin{aligned} &e^{2h(Z_{j(t)}|W)}s'(t) \geq \\ &\frac{\left(e^{2h(X_{f(t)}|W)}E_W[J_W(X_{f(t)})] + e^{2h(Y_{g(t)}|W)}E_W[J_W(Y_{g(t)})]\right)}{E_W[J_W(X_{f(t)})] + E_W[J_W(Y_{g(t)})]} \cdot \\ &\left(f'(t)E_W[J_W(X_{f(t)})] + g'(t)E_W[J_W(Y_{g(t)})]\right). \quad \text{(A.21)} \end{aligned}$$

Hence, by choosing

$$f'(t) = \exp 2h(X_{f(t)}|W) \text{ and } g'(t) = \exp 2h(Y_{g(t)}|W)\,, \quad \text{(A.22)}$$

we ensure that $s'(t) \geq 0$, with equality if equality holds in (A.17). Thus, under (A.21), $s'(t) = 0$ if, conditional on $W = w$, $X_{f(t)}$ and $Y_{g(t)}$ are normal. This further implies $s(t)$ is a constant if, conditional on $W = w$, X and Y are normally distributed.

If the entropy integrals $h(X_{f(t)}|W)$, $h(Y_{g(t)}|W)$, and $h(Z_{j(t)}|W)$ converge uniformly near $t = 0$, then $s(t)$ is continuous at $t = 0$ and

$$s(0) = \frac{\exp 2h(X|W) + \exp 2h(Y|W)}{\exp 2h(Z|W)}$$

To evaluate $s(+\infty)$ we note that (A.22) implies $f(+\infty) = g(+\infty) = j(+\infty) = +\infty$, and we define $X_{F(t)}$ to be $X_{f(t)}/\sqrt{f}$. Then (A.2) gives us for its probability density

$$p_{X_{F(t)}|W}(x_F) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\sqrt{f(t)}p_{X|W}(\sqrt{f(t)}u)\exp -\frac{(x_F - u)^2}{2}\,du\,.$$

Given $W = w$, $h(X_{F(t)}|W = w) = h(X_{f(t)}|W = w) - \frac{1}{2}\log f(t)$. As $f(t)$ grows infinite, $\sqrt{f(t)}p_{X|W=w}(\sqrt{f(t)}u)$ becomes a Dirac delta function, and $p_{X_{F(t)}|W=w}$

approaches $\exp -\frac{(x_F-x)^2}{2}/\sqrt{2\pi}$. Thus, if $h(X_{F(t)}|W = w)$ converges uniformly as $f(t)$ grows infinite, it becomes $\frac{1}{2}\log 2\pi e$ in the limit, and $h(X_{f(t)}|W = w) - \frac{1}{2}\log 2\pi e f(t) \to 0$ as $t \to +\infty$. Similarly, $h(Y_{g(t)}|W = w) - \frac{1}{2}\log 2\pi e g(t) \to 0$ and $h(Z_{j(t)}|W = w) - \frac{1}{2}\log 2\pi e(f(t) + g(t)) \to 0$. From (A.19), then, we have $s(+\infty) = 1$, completing the proof for the one-dimensional conditional entropy-power inequality, i.e.,

$$e^{2h(Z|W)} \geq e^{2h(X|W)} + e^{2h(Y|W)} \tag{A.23}$$

Appendix B

Numerical Figures to Support Section 3.3.1

σ_1^2	λ	β_1	$I_{Lambert}(X_1;X_3)$
10	0.00322574	0.9545	0.0030
20	0.001519135	0.9411	0.0027
30	0.000937113	0.9176	0.0026
40	0.000650135	0.8918	0.0024
100	0.0001700164	0.7444	0.0018
200	0.00004609095	0.5668	0.0012
300	0.00001754179	0.4504	9.3739e-4
400	0.00000769528	0.3703	7.4728e-4

Table B.1: β_1 for computing $I^{UB}(X_1;Y_2)$ and $I_{Lambert}(X_1;X_3)$ given that $\sigma_2^2 = 100$

σ_1^2	λ	β_1	$I_{Lambert}(X_1;X_3)$
10	0.00241402	0.9720	0.0016
20	0.00119004	0.9719	0.0015
30	0.000763584	0.9628	0.0014
40	0.000549017	0.9512	0.0014
100	0.0001720957	0.8711	0.0011
200	0.00006045746	0.7476	8.9097e-4
300	0.00002935973	0.6480	7.3505e-4
400	0.000016360003	0.5684	6.2225e-4

Table B.2: β_1 for computing $I^{UB}(X_1;Y_2)$ and $I_{Lambert}(X_1;X_3)$ given that $\sigma_2^2 = 200$

σ_1^2	λ	β_1	$I_{Lambert}(X_1;X_3)$
10	0.00201106	0.9777	0.0011
20	0.001008853	0.9813	0.0010
30	0.0006573	0.9768	9.871e-4
40	0.0004792175	0.9702	9.5934e-4
100	0.0001610391	0.9186	8.3546e-4
200	0.00006225747	0.8288	6.9568e-4
300	0.00003297228	0.7487	5.9350e-4
400	0.00001995973	0.6795	5.2148e-4

Table B.3: β_1 for computing $I^{UB}(X_1;Y_2)$ and $I_{Lambert}(X_1;X_3)$ given that $\sigma_2^2 = 300$

σ_1^2	λ	β_1	$I_{Lambert}(X_1;X_3)$
10	0.00175972	0.9804	8.7077e-4
20	0.000891083	0.9858	7.6547e-4
30	0.000585433	0.9834	7.4977e-4
40	0.0004300915	0.9792	7.3458e-4
100	0.0001501585	0.9427	6.6003e-4
200	0.00006114028	0.8738	5.6842e-4
300	0.00003390846	0.8084	5.0072e-4
400	0.000021426755	0.7492	4.4708e-4

Table B.4: β_1 for computing $I^{UB}(X_1;Y_2)$ and $I_{Lambert}(X_1;X_3)$ given that $\sigma_2^2 = 400$

σ_1^2	α	$\exp(h(X_2) - h(Y_1))$	$I_{MMSE}(X_1;X_3)$
100	100.9968	0.9999	0.0016
200	200.9584	$\approx$1	0.0011
300	300.9989	$\approx$1	7.9334e-4
400	400.9992	$\approx$1	6.3742e-4

Table B.5: $I_{MMSE}(X_1;X_3)$ given that $\sigma_1^2 = 100$

σ_1^2	α	$\exp(h(X_2) - h(Y_1))$	$I_{MMSE}(X_1; X_3)$
100	100.9968	0.9999	0.0011
200	200.9584	$\approx$1	8.0445e-4
300	300.9989	$\approx$1	6.5076e-4
400	400.9992	$\approx$1	5.3143e-4

Table B.6: $I_{MMSE}(X_1; X_3)$ given that $\sigma_1^2 = 200$

σ_1^2	α	$\exp(h(X_2) - h(Y_1))$	$I_{MMSE}(X_1; X_3)$
100	100.9968	0.9999	0.00080566
200	200.9584	$\approx$1	6.4821e-4
300	300.9989	$\approx$1	5.4458e-4
400	400.9992	$\approx$1	4.8156e-4

Table B.7: $I_{MMSE}(X_1; X_3)$ given that $\sigma_1^2 = 300$

σ_1^2	α	$\exp(h(X_2) - h(Y_1))$	$I_{MMSE}(X_1; X_3)$
100	100.9968	0.9999	6.5253e-4
200	200.9584	$\approx$1	5.5189e-4
300	300.9989	$\approx$1	4.8298e-4
400	400.9992	$\approx$1	4.3974e-4

Table B.8: $I_{MMSE}(X_1; X_3)$ given that $\sigma_1^2 = 400$

List of Figures

List of Tables

Bibliography

[1] I. Abou-Faycal and M. Médard: *Uncoded Regeneration for Binary Signalling,* accepted to the Communication Theory Symposium of the IEEE Internation Conference on Communication (ICC 2004)

[2] S. Borade,L. Zheng, and R. Gallager: *Maximizing Degrees of Freedom in Wireless Networks,* 41st Allerton Annual Conference on Communications, Control, and Computing (October 2003)

[3] N.M. Blachman: *The Convolution Inequality for Entropy Powers,* IEEE Trans. Inf. Theory Vol. 11, No. 2 (Apr '65)

[4] R.E. Blahut: *Principles and Practice of Information Theory,* Addison-Wesley Publishing Company, 1987

[5] R.U. Nabar, O. Oyman, H. Bölcskei, and A.J. Paulraj: *Capacity Scaling Laws in MIMO Wireless Networks,* 41st Allerton Annual Conference on Communications, Control, and Computing (October 2003)

[6] R.E. Blahut: *Principles and Practice of Information Theory,* Addison-Wesley Publishing Company, 1987

[7] E. Desurvire: *Erbium-Doped Fiber Amplifiers: Principles and Applications* , John Wiley & Sons, 1994

[8] R.G. Gallager: *A Simple Derivation of the Coding Theorem and Some Application*, IEEE Trans. Inf. Theory Vol. 11, No. 1 (Jan '65)

[9] M. Gaspar, B. Rimoldi, and M. Vetterli: *To Code, Or Not To Code: On the Optimality of Symbol-by-Symbol Communication*, IEEE Trans. Inf. Theory Vol. 49, No. 5 (May '03)

[10] I.F. Akyildiz *et al.*: *A Survey on Sensor Networks*, IEEE Communication Magazine, pp. 102-114, Aug. 2002

[11] B.D.O. Anderson and J.B. Moore: *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ, 1979

[12] P. Bello: *Characterization of Randomly Time-Invariant Channels*, IEEE Trans. Commun. Syst. Vol. CS-11, pp. 361-393, Dec. 1963

[13] T. Berger, Z. Zhang, and H. Viswanathan: *The CEO Problem*, IEEE Trans. Inf. Theory Vol. 42, pp. 887-902, May 1996

[14] T.M. Cover and J.A. Thomas: *Elements of Information Theory*, John Wiley & Sons, Inc., 1991

[15] S.C. Draper: *Successive Structuring of Source Coding Algorithms for Data Fusion, Buffering, and Distribution in Networks*, PhD Thesis, MIT, 2002

[16] R.G. Gallager: *Information Theory and Reliable Communication*, John Wiley & Sons, Inc., 1968

[17] R.M. Gray: *On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices*, IEEE Trans. Inf. Theory, Vol. IT-18, pp. 725-730, 1972

[18] W. Hirt and J.L. Massey: *Capacity of the Discrete-Time Gaussian Channel with Intersymbol Interference*, IEEE Trans. Inf. Theory Vol. 34, No. 3, 1988

[19] M. Médard: *The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel*, IEEE Trans. Inf. Theory Vol. 46, pp.933-946, May 2000

[20] Y. Oohama: *The Rate-Distortion Function for the Quadratic Gaussian CEO Problem*, IEEE Trans. Inf. Theory Vol. 44, pp. 1057-1070, May 1998

[21] N. Peranginangin, M. Médard, R.G. Gallager: *Capacity of a Multi Output Channel with Distributed Processing*, Proc. International Sympsium on Information Theory (Yokohama, Japan), pp. 97, June 2003

[22] Y. Hong and A. Scaglione: *Opportunistic Large Arrays: Cooperative Transmission in Wireless Multihop Ad Hoc Networks to Reach Far Distances*, IEEE Trans. on Signal Proc. Vol. 51, pp. 2082-2092, Aug 2003

[23] S. Servetto, J. Barros: *Reachback Capacity with Non-Interfering Nodes*, Proc. International Symposium on Information Theory (Yokohama, Japan), pp. 366, June 2003

[24] Q. Zhao and L. Tong : *QoS Specific Medium Access Control for Wireless Sensor Networks with Fading*, Eighth International Workshop on Signal Processing for Space Communications (invited paper), July 2003

[25] D.P. Bertsekas and J.N. Tsitsiklis: *Introduction to Probability*, Athena Scientific, 2002

[26] B.S. Tsybakov: *Capacity of a Discrete-Time Gaussian channel with a Filter*, Probl. Peredach. Inform., Vol. 1, pp. 26-40, 1965

[27] H. Viswanathan and T. Berger: *The Quadratic Gaussian CEO Problem*, IEEE Trans. Inf. Theory Vol. 43, pp. 1549-1559, September 1997

[28] A.S. Willsky, G.W. Wornell, J.H. Shapiro: *Course Notes on Stochastic Processes, Detection, and Estimation*, Mass. Inst. Technol., Cambridge, MA, Sept. 1998 *On the Distribution and Moments of RC-Filtered Hard-Limited RC-Filtered White Noise*, IEEE Trans. Inf. Theory Vol. 19, No. 4 ('73)

[29] L.D. Davisson and P. Papantoni-Kazakos: *On the Distribution and Moments of RC-Filtered Hard-Limited RC-Filtered White Noise*, IEEE Trans. Inf. Theory Vol. 19, No. 4 ('73)

[30] L.D. Davisson, L.B. Milstein: *On the Performance of Digital Communication Systems With Bandpass LimitersPart I: One-Link System*, IEEE Trans. On Communication (Oct '72)

[31] L.D. Davisson, L.B. Milstein: *On the Performance of Digital Communication Systems With Bandpass LimitersPart II: Two-Link System*, IEEE Trans. On Communication (Oct '72)

[32] R. Gaudino,M. Shell,M. Len, G. Desa, and D.J. Blumenthal: *MOSAIC: A Multiwavelength Optical Subcarrier Multiplexed Controlled Network*, a Report, Optical Communication and Photonic Networks (OCPN) Laboratory, Georgia Institute of Technology ('97)

[33] P.A. Humblet and M. Azizoğlu: *On the Bit Error Rate of Lightwave Systems with Optical Amplifiers*, Journal of Lightwave Technology Vol. 9, No. 11 ('91)

[34] A.R. Kaye, D.A. George, and M.J. Eric: *Analysis and Compensation of Bandpass Nonlinearities for Communications*, IEEE Trans. On Communications (Oct '72)

[35] S.Y. Kwon, and R.S Simpson: *Effect of Hard Limiting on a Quadrature PSK Signal*, IEEE Trans. On Communications (Aug '73)

[36] R.G. Lyons: *The Effect of a Bandpass Nonlinearity on Signal Detectability*, IEEE Trans. On Communication Vol. 21, No. 1 (Jan '73)

[37] E. Masry: *The Recovery of Distorted Band-Limited Stochastic Processes*, IEEE Trans. Inf. Theory Vol. 19, No. 4 (July '73)

[38] J.E. Mazo, R.F. Pawula, and S.O. Rice: *On a Nonlinear Problem Involving RC Noise*, IEEE Trans. Inf. Theory Vol. 19, No. 4 ('73)

[39] H. Marko, R. Weis, and G. Binkert: *A Digital Hybrid Transmission System for 280 Mbits/s and 560 Mbits/s*, In European Conference on Optical Communication (ECOC), volume 4 ('01)

[40] B. Mikkelsen, S.L. Danielsen, C. Joergensen, R.J.S. Pedersen, H.N. Poulsen, and K.E. Stubkjaer : *All-optical Noise Reduction Capability of Interferometric Wavelength Converters*, Electronic Letters Vol. 32, No.6 (March '96)

[41] P. Ohlén and E. Berglind: *Noise Accumulation and BER Estimates in Concatenated Nonlinear Optoelectronic Repeaters*, IEEE Photonics Technology Letters Vol. 9, No. 7 ('97)

[42] R.F. Pawula and S.O. Rice: *On Filtered Binary Processes*, IEEE Trans. Inf. Theory Vol. 32, No.1 (Jan '86)

[43] W.S. Wong, S. Namiki, M. Margalit, E.P. Ippen, and H.A. Haus: *Optical Pulse Filtering with Dispersion-Imbalanced Nonlinear Loop Mirrors*, a Report, Research Laboratory of Electronics, Massachusetts Institute of Technology ('97)

[44] O. Zeitouni: *On the Filtering of Noise-Contaminated Signals Observed via Hard Limiters*, IEEE Trans. Inf. Theory Vol. 34, No. 5 (Sep '88)

[45] M. Gastpar and M. Vetterli: *To code, or Not to Code: Lossy Source-Channel Communication Revisited*, IEEE Trans. Inf. Theory Vol. 49, No. 5,(May '03)

[46] P. Gupta and P.R. Kumar: *The Capacity of Wireless Networks*, IEEE Trans. Inf. Theory Vol. 46, No. 2 (Mar '00)

[47] B.E. Schein: *Distributed Coordination in Network Information Theory*, Ph.D thesis, Massachusetts Institute of Technology,('01)

[48] M. Grossglauser and D.N. Tse: *Mobility Can Increase the Capacity of Wireless Networks*, IEEE INFOCOM 2001, Anchorage, Alaska (April '01)

[49] R. Barry and M. Médard: *BER Analysis of Low-Rate Communications through a Single Electro-Optic R2 Nonlinear Regenerator*, Lasers and Electro-Optics (CLEO) ('98)

[50] J.C Simon, L. Billes, A. Dupas, B. Kowalsky, M. Henry and B. Landousies: *All Optical Regeneration*, European Conference on Optical Communication (ECOC), volume 1 ('98)

[51] M. Matsumoto and O. Leclerc: *Analysis of 2R Optical Regenerators Utilizing Self-Phase Modulation in Highly Nonlinear Fibre*, Electronic Letters, volume 38 (June '02)

[52] J. De Merlier, G. Morthier, T. Van Caenegem, R. Baefs, I. Moerman and P. Van Daele: *Experimental Demonstration of 15 dB Extinction Ratio Improvement in a New 2R Optical Regenerator based on an MMI-SOA*, European Conference on Optical Communication (ECOC), volume 4 ('01)

[53] D. Chiaroni, B. Lavigne, A. Jourdan, L. Hamon, C. Janz and M. Renaud: *New 10 Gbit/s 3R NRZ Optical Regenerative Interface Based on Semiconductor Optical Amplifiers for All-Optical Networks*, European Conference on Optical Communication (ECOC), volume 5 ('97)

[54] A.E Kelly, I.D Phillips, R.J. Manning, A.D Ellis, D. Nesset, D.G Moodie, and R. Kashyap: *Regeneration of Optical Frequency-Shift-Keying Signal Using Self-Pulsating and Self-Oscilating Mechanisms in Laser Diodes*, Laser and Electro-Optics Society Annual Meeting (LEOS), volume 2 ('00)

[55] O. Leclerc, P. Brindel, D. Rouvillain, B. Dany, E. Pincemin, E. Desurvire, C. Duchet, A. Shen, F. Blache, F. Devaux, A. Coquelin, M. Goix, S. Bouchoule, and P. Nouchi: *Dense WDM (0.27 bit/s/Hz) 440 Gbit/s Dispersion-Managed Transmission over 10000 km with In-Line Optical Regeneration by Channel Pairs*, Electronic Letters, volume 36 (Feb '00)

[56] O. Leclerc, P. Brindel, D. Rouvillain, B. Dany, E. Pincemin, E. Desurvire, C. Duchet, E. Boucherez, and S. Bouchoule: *40 Gbit/s Polarization-Independent, Push-Pull Input Mach-Zender Modulator for All-Optical Regeneration*, Optical Fiber Communication Conference and Exhibit (OFC), volume Post-deadline supplement ('99)

[57] G.E. Falkovich, I. Kolokolov, V. Lebedev, V.K. Mezentsev, and S.K. Turitsyn: *Calculation of Non-Gaussian Intensity and Timing Jitter Statistics in Optical Regenerated Systems*, Optical Fiber Communication Conference and Exhibit (OFC), volume 3 ('01)

[58] W.B. Davenport: *Signal to Noise Ratios on Bandpass Filters*, Journal of Applied Physics, volume 24 ('53)

[59] P.C. Jain: *Limiting of Signals in Random Noise*, IEEE Trans. Inf. Theory Vol. IT-18 (May '72)

[60] J.J. Jones: *Hard-Limiting of Two Signals in Random Noise*, IEEE Trans. Inf. Theory Vol. IT-9 (May '63)

[61] W. Sollfrey: *Hard-Limiting of Three and Four Sinusoidal Signals*, IEEE Trans. Inf. Theory Vol. IT-15 (May '69)

[62] R.F. Pawula: *Level Crossings of Filtered Dichotomous Noise*, Physical Review A, Volume 37 (Mar '88)

[63] N.M. Blachman: *The Effect of Nonlinearity Upon Signals in the Presence of Noise,* IEEE Trans. On Comm. (Feb '73)

[64] D.A. Darling and A.J.F Siegert: *A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena - Part I,* IRE Transactions on Information Theory (Mar '57)

[65] A.J.F Siegert: *A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena - Part II,* IRE Transactions on Information Theory (Mar '57)

[66] A.J.F Siegert: *A Systematic Approach to a Class of Problems in the Theory of Noise and Other Random Phenomena - Part III,* IRE Transactions on Information Theory (Mar '58)