Doctoral Thesis

Capacity and information rates of discrete-time channels with memory

Author(s):
Hirt, Walter

Publication Date:
1988

Permanent Link:
https://doi.org/10.3929/ethz-a-000518487

Rights / License:
In Copyright - Non-Commercial Use Permitted
Capacity and Information Rates
of
Discrete-Time Channels with Memory

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH
for the degree of
Doctor of Technical Sciences

Presented by
WALTER HIRT
M.A.Sc. (El.-Eng.), University of Toronto
born December 3, 1947
citizen of Stilli, Aargau

Accepted on the recommendation of
Prof. Dr. J.L. Massey, examiner
Prof. Dr. J. Justesen, co-examiner

1988
Meinen Eltern gewidmet
ACKNOWLEDGEMENTS

I gratefully acknowledge Professor Dr. J.L. Massey for his always stimulating advice and support given to me during the course of this research.

I would like to thank Professor Dr. J. Justesen of the Technical University of Denmark for his interest in my work.

The management of the IBM Zurich Research Laboratory is especially acknowledged for giving me the opportunity to pursue this project.

I owe very special thanks to Ann for all the help and encouragement which I received from her as my wife, friend and colleague.
The equivalent discrete-time channel model of a digital communication system which makes use of channels with *intersymbol-interference* (ISI) is analyzed from an information-theoretical viewpoint. Channel capacity and cutoff-rate parameters are defined and computed (or bounded) for different channel models where each model consists of 1) the basic channel model and 2) the constraints on channel usage (input constraints). Two types of inputs are considered, input symbols which are constrained in their average energy (per symbol or per block) and independent binary symbols of equal probability. The results permit one to make some quantitative and qualitative statements about the effects of the channel's ISI on the various information rate parameters, as well as about the required (additional) complexity of coding systems for such channels.

A new approach is introduced to obtain the capacity of the basic *discrete-time Gaussian channel* (DTGC) with ISI where the inputs are subject to a *per symbol* average-energy constraint. A member of this class of channels is called simply a *DTGC*. The capacity of the DTGC is derived by means of a hypothetical channel model whose capacity is readily found by using the well-known theory of the *discrete Fourier transform* (DFT). This novel DFT approach allows one to prove that the capacity of the DTGC equals the limit as block length increases of the capacity of a channel which is defined by the basic discrete-time Gaussian channel and the usual block average-energy input constraint.

The basic DTGC with independent equally-likely binary input symbols is considered. The *symmetric information rate* (SIR) of this channel model is defined as the corresponding $N$-block average mutual information (per input symbol) in the limit of increasing block-length, $N$. The SIR is used to establish
a lower bound on the capacity of this channel. It is argued that it is not practical to compute the SIR exactly unless the channel is memoryless. The SIR is then approximated from statistical estimates that are obtained by Monte-Carlo integration of both the $N$-block average mutual information and an upper bound thereon.

The cutoff rate of the DTGC is defined as the corresponding $N$-block cutoff rate (per input symbol) in the limit of increasing block-length, $N$. A lower bound on the cutoff rate of the DTGC is obtained by adopting the DFT method which was introduced earlier. This lower bound is shown to be closely related to the capacity of the DTGC and is further compared with the zero-rate intercept of the straight-line portion of Gallager's random-coding exponent. For specific channels the cutoff rate of the DTGC and its capacity are shown to differ by at most $1/2$ bit per symbol at any given signal-to-noise ratio; it is conjectured that this result holds for all DTGC's.

The symmetric cutoff rate (SCR) of the basic DTGC with independent equally-likely binary input symbols is studied in detail. A computationally efficient procedure is developed to determine the SCR as the negative logarithm of the dominant eigenvalue of a certain (reduced) state transition matrix whose size depends on the length of the channel's memory. A comparison based on the SCR criterion between partial-response channels and so-called minimum-distance channels (MDC's) is presented. The numerical procedure by which the SCR is computed is further extended to include an efficient direct-search algorithm; given that the maximum memory length and the signal-to-noise ratio are both fixed, those unit-energy channel responses are identified which minimize the SCR. It is shown that the unit-sample responses of these minimum-SCR channels depend on the signal-to-noise ratio. The worst-case SCR is determined for a large range of ISI memory lengths and signal-to-noise
ratios. The worst-case degradation for medium-range rates in terms of the noise margin is shown to be on the order of 1 dB for each additional unit of ISI memory.

Some applications and properties of the SCR are presented that appear to be new. By using random-coding arguments, a very simple expression is derived that gives a close approximation to the (squared) minimum Euclidean distance, which determines the reliability of uncoded binary systems that use maximum-likelihood sequence estimation. This method does not require knowledge of the specific error-event that induces the minimum distance. By exploiting asymptotic properties of the newly defined minimum-SCR channels, a method is introduced which yields close approximations to the unit-sample responses of MDC's. Finally, a computationally simple, but tight, lower bound on the SCR is proposed that holds for all rates that are of practical interest for coded binary signaling.

Das Modell des zeitdiskreten Gaußschen Kanals mit statistisch unabhängigen, binären Eingangssymbolen gleicher Auftrittswahrscheinlichkeit wird betrachtet. Es wird gezeigt, dass sich die Transinformationsrate dieses Kanalmodells nur dann exakt bestimmen lässt, wenn der Kanal gedächtnisfrei ist. Durch die Anwendung statistischer Verfahren (Monte-Carlo Integration) wird die Transinformationsrate einiger interessanter Kanäle approximiert.

Es wird gezeigt, dass die cutoff rate (d.h. die Informationsrate welche im hypothetischen Grenzfall der fehlerfreien Übertragung bei unbeschränktem Codieraufwand der übertragbaren Datenrate entspricht) des zeitdiskreten Gaußschen Kanals mit Beschränkung der mittleren Signalenergie grundsätzlich mit Hilfe der vorgeschlagenen DFT-Methode bestimmt werden kann. Basierend auf dem neu eingeführten DFT-Kanalmodell wird eine untere Schranke für die cutoff rate definiert und es wird gezeigt, dass die gefundene Lösung in direktem Zusammenhang mit der Kapazität des vorgegebenen Kanalmodells steht. Die untere Schranke für die cutoff rate wird mit einem Parameter verglichen welcher von Gallaggers random-coding Exponenten abgeleitet ist. Es wird postuliert, dass die unbekannte cutoff rate des angenommenen Kanalmodells um höchstens 1/2 bit pro Symbol kleiner ist als die bekannte Kanalkapazität, und zwar unabhängig vom Signal-Rauschleistungsverhältnis.


# TABLE OF CONTENTS

1 Introduction ........................................... 1-1

2 Background and Definitions .......................... 2-1
   2.1 Channels with Memory ............................ 2-6
   2.2 Random-Coding Exponent for Finite-State Channels ....... 2-21
   2.3 Related Work .................................. 2-28

3 Capacity of the Discrete-Time Gaussian Channel with Intersymbol Interference ................. 3-1
   3.1 Definitions and Remarks .......................... 3-1
   3.2 Statement of Main Results ....................... 3-7
   3.3 Analysis of the NCGC ........................... 3-10
   3.4 Relations Between Channel Models ............... 3-16
   3.5 Derivation of \( C(E_d) \) and \( I^G(E_d) \) ............... 3-21
   3.6 Extension to Channels with Infinite Memory ......... 3-25
   3.7 Numerical Results ................................ 3-28
      Appendix 3.A
      Appendix 3.B

4 Information Rate with a Binary Input Constraint ................. 4-1
   4.1 Derivation of \( I^b_N(E_d) \) ....................... 4-5
   4.2 Properties of \( I^b(E_d) \) ....................... 4-7
   4.3 Generation of Numerical Results .................. 4-16
      Appendix 4.A
      Appendix 4.B
      Appendix 4.C
5 Lower Bound on the Cutoff Rate of the Discrete-Time Gaussian Channel with Intersymbol Interference
5.1 Background and Problem Definition
5.2 A Lower Bound on \( R_o(E_d) \)
5.3 Numerical Results
Appendix 5.A
Appendix 5.B

6 The Symmetric Cutoff Rate for Binary Inputs
6.1 Evaluation of \( R_o^b(E_d) \)
6.2 Bounds and Properties of \( R_o^b(E_d) \)
6.3 Examples and a Reduction Theorem
6.4 Numerical Results
Appendix 6.A

7 New Applications and Properties of the Symmetric Cutoff Rate
7.1 Approximation of \( d_{\text{min}}^2 \) by Means of \( R_o^b(E_d) \)
7.2 Approximating MDC's by Means of \( \overline{R}_o^b(E_d) \)
7.3 \( R_o^G(E_d) \) as a Lower Bound on \( R_o^b(E_d) \)

8 Summary and Conclusions
References
Curriculum Vitae
CHAPTER I

Introduction

The challenge to achieve the highest possible information rate at a specified reliability over existing or newly developed digital communication channels persists, both in data transmission and data storage applications [FOR-GAL-LAN-LON-QUR(84)], [MAL(85)], [WEI(87)]. The band-limiting nature of the physical channel in such systems leads to the inevitable intersymbol interference (ISI) effects at the channel output [LUC-SAL-WEL(68), Section 4.2]. To deal with ISI on a quantitative basis, mathematical channel models have been developed for digital data transmission applications that include the modulator-transmitter filter, the physical (waveform) channel with the (additive) noise source, and the receiver filter [PRO(83), Chapter 6]. Assuming that the channel is linear and time-invariant, these system models are given in the form of a discrete-time transversal filter with a finite number of constant coefficients and with independent, identically-distributed (i.i.d.) Gaussian noise samples added to the filter output. Under certain conditions, when considered from a communication system viewpoint, the digital magnetic recording channel can be identified as a linear system and thus modeled in the same way [HEY(77)], [MAL(87)].

Various systems which attempt to mitigate the generally undesired ISI effects have been proposed and developed in the past [PRO(75)]. A basic scheme that is often effective uses pulse-amplitude modulation (PAM) and partial-response techniques together with maximum-likelihood sequence estimation (MLSE) of the transmitted information digits [KOB(71)], [FOR(72)], [UNG(74)], [PRO(83), Chapter 6]. The reliability of such systems, usually measured in terms of probability of error, comes closest to the performance of a reference system that does not suffer from ISI. These advanced systems cope with ISI in an optimal
way, however the realization of their full performance potential is presently limited to channels with rather short memory because of practical constraints on receiver complexity measured either in terms of hardware or in the number of microcode instructions [FAL-MAG(73)], [EYU-QUR(88)].

As the information rate (storage density) in a given application is further increased, the effective noise interference from the physical channel may reach levels where the prescribed reliability can only be maintained if coding is added to the system [MAS(84-1)]. Reliability bounds on digital communication systems, where coded information sequences are transmitted over channels with known finite-length ISI and where the receiver performs combined maximum-likelihood demodulation and decoding, have been investigated by Omura [OMU(71)], Mackechnie [MAC(73)], and Acampora [ACA(76)]. These techniques and results were later extended and reported by Viterbi and Omura [VIT-OMU(79), Chapter 5]. These investigations have reconfirmed one of Shannon's fundamental results, namely that there exist coding schemes which allow transmission of information over noisy (band-limited) channels at a positive information rate with arbitrary reliability, provided that this rate does not exceed a certain number [SHA(48)], [SHA(49)].

Presently, there is a strong interest in effective practical codes for noisy channels with memory that can be modeled by linear discrete-time filtering of the input symbols [CAL-OZA-HEE(85)], [KET(86)], [WOL-UNG(86)], [CAL-HEE-LEE(86)]. For the code designer, the capacity and the maximum "practical" information rate, i.e., the cutoff rate [MAS(74)], of this class of channels are of interest. The exact origin of the ISI that characterizes the channel is unimportant in this context, what matters for the coding system is the equivalent discrete-time channel which is created by the overall transmission system [MAS(84-1)].

The equivalent discrete-time channel model for digital communication systems with ISI is analyzed in this thesis from an information-theoretical viewpoint.
The main objective is to determine how the two key parameters, channel capacity and cutoff rate, are influenced by ISI and by the signal-to-noise ratio of the channel. In general, a capacity or cutoff rate can be defined and computed provided that the channel model includes two specifications, namely 1) the basic channel model that specifies the conditional probability for the output given some input, and 2) the constraints on channel usage (input constraints). A further objective is the development of computationally efficient algorithms for the numerical evaluation of the quantities identified in this thesis. The contents of the remaining chapters are organized as follows:

Chapter 2: The theoretical and practical aspects of coded information transmission over channels with memory are briefly reviewed. The basic channel model used to represent linear communication channels that exhibit time-dispersive impulse responses is described. In the final section, a brief and selective overview of some lesser-known work related to this study is given.

Chapter 3: A new approach is introduced to obtain the capacity formula for the discrete-time Gaussian channel (DTGC) with ISI when the inputs are subject to a per symbol average-energy constraint. A member of this class of channels is simply referred to as a DTGC. The capacity of the DTGC is derived by means of a hypothetical channel model, called the \textit{N-circular Gaussian channel} (NCGC). The capacity of the NCGC is defined and readily found by using the well-known theory of the \textit{discrete Fourier transform} (DFT). The capacity of the NCGC is then related to the capacity of the DTGC. This approach allows one to prove that, in the limit of increasing block-length \( N \), the capacity of a different channel model, namely the basic discrete-time Gaussian channel with ISI and with the more-commonly-used \textit{per block} average-energy input constraint (\( N \)-block DTGC), is indeed the capacity of the DTGC.
Extension of the capacity result to channels with infinite memory and a presentation of numerical results complete the chapter.

**Chapter 4:** The basic discrete-time Gaussian channel with ISI and with independent, equally-likely, antipodal binary input symbols is considered. The *symmetric information rate* (SIR) of this channel model is defined as the corresponding $N$-block average mutual information (per input symbol) in the limit of increasing block-length $N$. The SIR establishes a lower bound on the capacity of this channel, except for the memoryless case (no ISI) in which the SIR equals capacity. It is argued that the SIR is numerically not computable in practice unless the channel is memoryless. It is then demonstrated that the SIR can be well approximated from statistical estimates of both the $N$-block average mutual information and an upper bound thereon. The principles of Monte-Carlo integration that apply to the present problem are summarized in the appendix.

**Chapter 5:** Analogous to the definition of the capacity introduced in Chapter 3, the cutoff rate of the DTGC is defined as the corresponding $N$-block cutoff rate (per input symbol) in the limit of increasing block-length $N$. Given some input probability density, an expression for the $N$-block cutoff rate is derived. It is shown that this expression is not maximized by a Gaussian density. A theorem is then proved that asserts that the cutoff rate of the DTGC may be obtained, at least in principle, from the cutoff rate of the NCGC in the limit of increasing block-length $N$. From the basic channel model of the NCGC, a lower bound on the cutoff rate of the DTGC is obtained by adopting the DFT method introduced in Chapter 3. The parametric expressions which determine this lower bound are shown to be closely related to corresponding expressions for the capacity of the DTGC. The lower bound for the cutoff rate of the DTGC is further compared with the zero-rate intercept of
the straight-line portion of Gallager's random-coding exponent. The analytical results of this chapter are also explored with respect to their spectral properties and their asymptotic behaviour. In particular, it is conjectured that the cutoff rate and the capacity of all DTGC's differ at most by $1/2$ bit per symbol at any signal-to-noise ratio.

**Chapter 6:** The basic discrete-time Gaussian channel with ISI and with independent, equally-likely, antipodal binary input symbols, as defined earlier in Chapter 4, is reconsidered. The symmetric cutoff rate (SCR) of this channel model is defined as the corresponding $N$-block cutoff rate (per input symbol) in the limit of increasing block-length $N$. It is shown that the SCR is given by the negative logarithm of the dominant eigenvalue of a certain state transition matrix whose size depends on the length of the channel's memory. A formal procedure is then developed to obtain an equivalent reduced state transition matrix that allows more efficient computation of the SCR. A comparison based on the SCR criterion between partial-response channels and so-called minimum-distance channels is presented in the numerical section. The numerical procedure by which the SCR is computed is further extended to include an efficient direct search algorithm. Given that the maximum memory length and the signal-to-noise ratio are both fixed, those channel responses are identified which minimize the SCR. It is demonstrated that the minimum-SCR channels and the minimum-distance channels exhibit similar symmetry properties. Furthermore, the worst-case SCR is determined for a large range of memory lengths and signal-to-noise ratios.

**Chapter 7:** The numerical methods and results described in Chapter 6 are extended and further interpreted. Some new applications and properties of the SCR are presented that have not been exploited previously. Through random-coding arguments, a simple expression is found which allows a
computationally efficient and tight approximation of the minimum Euclidean distance, which determines the performance of uncoded systems that use antipodal binary signaling and MLSE. This method does not require knowledge of the specific error-event that induces the minimum distance. Next, a method is introduced which yields good approximations to the unit-sample response of minimum-distance channels. This is achieved by exploiting some asymptotic properties of the newly defined minimum-SCR channels. Finally, a simple but tight lower bound for the SCR is proposed that, below some finite signal-to-noise ratio, holds for all rates of practical interest for coded binary signaling.

Chapter 8: The main results are summarized. The various rate parameters that are the subject of this thesis are compared for a representative set of sample channels. Conclusions are then drawn from the consolidated analytical and numerical results.
CHAPTER 2

Background and Definitions

The performance limits of coded digital communication systems as promised by Shannon's information theory are generally given in asymptotic form and their proofs are nonconstructive; thus, these absolute limits are rather difficult to approach in practice [BER-PE1-POP(87)], [WYN(81)]. When comparing different systems, however, relative gains\(^1\) that have been predicted by information-theoretical analysis can be closely realized in some cases [HAG(85)]. An obvious, but not always recognized, prerequisite for such an achievement is to adopt a suitable design model which does not preclude system realizations that approach (as close as is practical or desired) the theoretical optimum. The appropriate block diagram of a general digital communication system without feedback is shown in Figure 2.1. The foundations and the practical implications of this canonical view of digital communications are now well recognized, notably through the efforts of Massey [MAS(84-I)], [MAS(74)], Wyner [WYN(81)], Viterbi and Omura [VIT-OMU(79)], Berger [BER(71)], Gallager [GAL(68)], Wozencraft and Kennedy [WOZ-KEN(66)], and Wozencraft and Jacobs [WOZ-JAC(65)]. The focal point of the present study is introduced through a brief description of the system components shown in Figure 2.1 (see also [MAS(84-I)] for more detailed information).

Every information source can in principle be converted by the source encoder into a sequence of binary digits at the smallest possible information rate (meas-

\(^1\) Here, \textit{gain} means the factor (usually given in dB) by which the average transmitted power must be increased in a reference system so that it achieves the same performance as the competing system. The reference system is usually taken to be an \textit{uncoded} system with the same information rate, the same type of modulator, and a hard-decision demodulator [MAS(84-I)].
Fig. 2.1. Block diagram of a general digital communication system without feedback.
ured in bits per unit time), which is a source characteristic. Conversely, the source decoder should reproduce from these binary digits a replica of the source output. The secrecy encoder converts the source encoder output (plaintext) under the control of a secret key, $K$, into another sequence of binary information digits of the same or only slightly greater rate. The secrecy system should be designed such that its decoder reconstructs the plaintext only with knowledge of the secret key $K$. If there is no need for secrecy, the source coding system interfaces directly with the channel coding system. In any case, the combination of both the channel coding and the modulation systems are intended to reproduce the binary digits entering the channel encoder at the output of the channel decoder, by way of the waveform channel, with a specified reliability at the highest practicable rate and as efficiently as possible (e.g., with the smallest possible average power per information bit). A natural way to partition this task between channel coding and modulation is obtained if one insists that the modulator be memoryless and that the demodulator have no more memory than that required to deal with the memory of the waveform channel. This means that the modulator maps its current input symbol, $x_k$, directly into the corresponding waveform, $s_k(t)$, and that the channel encoder has the required memory to provide redundancy in the signal entering the waveform channel. This redundancy in the modulator input symbol sequence can be exploited by the channel decoder to make reliable decisions about the information bits, provided that the demodulator extracts from the received signal, $r(t)$, in each signaling interval sufficient likelihood information about the

---

1 A consequence of Shannon's theory is the separation theorem for source and channel coding [MAS(84-2)]. Note, however, that the possibility of joint source and channel coding is not precluded by adopting this canonic system model. The advantages and disadvantages of combining the functions of source coding and channel coding into a single device have been reported in [MAS(78)] and, more recently, in [AYA-GRA(87)].
transmitted waveforms. The term "soft decisions" is usually used to describe the output digits, \( \hat{y}_k \), of such a demodulator; they should be quantized to the smallest possible set of values that is sufficient to make the channel capacity (or the cutoff rate) of the resulting discrete channel (between modulator input and demodulator output) acceptably close to its maximum possible value. The use of a properly designed modulator and soft-decision demodulator is mandatory to obtain a coded digital communication system which achieves a large coding-demodulation gain relative to an uncoded but otherwise similar system [MAS(84-1)], [LEE(76)], [MAS(74)]. In other words, the channel coding system can only fulfill its purpose, viz., to prevent (rather than to correct or control) errors in the decoded (high-rate) information sequence, if the discrete channel's modulation-demodulation system is optimally designed according to information-theoretic criteria. This is in contrast to the design method used for uncoded systems where the symbol error probability in the hard-decision demodulator output is (approximately) minimized [LUC-SAL-WEL(68), Chapter 5].

The specific design of the modulation-demodulation system, and thus the capacity and the cutoff-rate parameters of the resulting discrete channel, depends on the constraints imposed on or by the waveform channel (e.g., constraints on bandwidth and on average- or peak signal power) and, if present,

1 This conceptual distinction between the demodulator and channel decoder functions does not preclude the possibility of combining the optimal demodulator and channel decoder functions in a single device [VIT-OMU(79), Sections 4.10 and 5.8].

2 Here, discrete means that both the inputs and outputs are time-discrete digits (occurring at regular time intervals) and amplitude-discrete digits (quantized to finite alphabets).

3 Note that soft-decision demodulation alone may contribute up to almost 2 dB (factor of \( \pi/2 \)) to the overall system gain when compared to hard-decision demodulation schemes [VIT-OMU(79), p. 101 and Fig. 2.19].
the type and the extent of the "memory" which arises from other impairments that are peculiar to the channel.

This study concentrates on some of the information-theoretical aspects of a specific class of discrete channels with memory. However, a large part of the literature on theoretical and applied work in the area of digital communications has been devoted to problems where the discrete channel is memoryless. The class of discrete memoryless channels (DMC's) may be defined as follows. Let the random vectors \( X = (X_0, X_1, \ldots, X_{N-1}) \) and \( \hat{Y} = (\hat{Y}_0, \hat{Y}_1, \ldots, \hat{Y}_{N-1}) \) be the ensembles of input and output sequences of length \( N \), respectively, for the discrete channel shown in Figure 2.1. Further, let \( x = (x_0, x_1, \ldots, x_{N-1}) \), \( \forall x_k \in A_x \), be an arbitrary sequence of \( N \) input digits and let \( \hat{y} = (\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_{N-1}) \), \( \forall y_k \in A_y \), be a corresponding output sequence with \( A_x \) and \( A_y \) being, respectively, the finite input and output alphabets. The discrete channel is memoryless in the sense that the conditional probability distributions of receiving \( \hat{y}_k \), given that the digits \( x_0, x_1, \ldots, x_k \) were sent and the digits \( \hat{y}_0, \hat{y}_1, \ldots, \hat{y}_{k-1} \) were received, are such that

\[
Q(\hat{y}_k|x_0, x_1, \ldots, x_k, \hat{y}_0, \hat{y}_1, \ldots, \hat{y}_{k-1}) = Q(\hat{y}_k|x_k), \tag{2.1a}
\]

for all \( k = 0, 1, \ldots \). Note that \( Q(\cdot|x) \) on the right-hand side of (2.1a) may depend on \( k \), i.e., the channel may be time-varying. The assumption that the discrete channel shown in Figure 2.1 is used without feedback (from its output to its input) implies that the inputs are selected in such a way that the choice of

\[\text{To simplify notation we use here, and subsequently, the convention that the subscripts may be dropped from probability distributions whenever the subscript is the capitalized version of the argument. For example, } Q(x_k) \text{ may be used for } Q_X(x_k) \text{ or } Q(y|x) \text{ may denote } Q_Y|X(y|x). \text{ However, if the random variable } X_k \equiv X, \forall k, \text{ then the } Q_X(\cdot) \text{ are invariant in } k \text{ and we write } Q_X(\cdot).\]
sending the digit $x_k$ may depend on the previously transmitted digits but not on the previously received digits, i.e.,

$$Q(x_k | x_0, x_1, \ldots, x_{k-1}, y_0, y_1, \ldots, y_{k-1}) = Q(x_k | x_0, x_1, \ldots, x_{k-1}), \quad (2.1b)$$

for all $k = 0, 1, \ldots$. It then follows [MAS(83), p. 4.6] from the multiplication rule for conditional probabilities that for a DMC used without feedback the conditional probability distributions of receiving the $N$-sequence $\hat{y}$, given that the $N$-sequence $x$ was sent, is

$$Q(\hat{y} | x) = \prod_{k=0}^{N-1} Q(y_k | x_k). \quad (2.1c)$$

Usually $Q(\cdot | \cdot)$ on the right-hand side of (2.1c) is restricted to be independent of $k$, i.e., the DMC is time-invariant. This type of DMC results, for example, when the waveform channel in Figure 2.1 is the stationary additive white Gaussian noise (AWGN) channel. The communication link between a platform in deep space and the earth is well modeled by this type of DMC; it was on the deep-space channel with a very limited power budget where sophisticated coding schemes were first systematically applied [McE(77), p. 267], [FOR(70)].

2.1 Channels with Memory

Real communication channels on earth and within the outer bounds of the ionosphere are more often degraded by a combination of additive, multiplicative and input-dependent interferences of time-varying nature. Whatever the underlying physical phenomena in a given situation (e.g., frequency-selective fading, burst or impulsive noise, scattering, dropouts), the demodulator output digits $\hat{y}_k$ in Figure 2.1 are in general no longer independent events given the input sequence. This means that (2.1a), the condition for a channel to be
memoryless, does not hold for these channels; the discrete channel is then said to have "memory". With a hard-decision demodulator, this means that the error-producing mechanisms are such that errors occur in clusters and/or that they are signal dependent. The characterization of channels with memory is in general more difficult than that of memoryless channels; ideally, such a characterization should 1) be mathematically tractable and 2) should represent accurately the pertinent properties of the channel's memory. A number of models for channels with memory have been proposed in the past, however some of them are at variance with either one or both of the requirements for a practical model [KAN-SAS(78)]. For present purposes, we distinguish between time-varying channels whose memory arises independently of the transmitted signal and channels with memory occurring from interference between transmitted signals.

A. Certain Time-Varying Channels

Given that the discrete channel in a digital communication system does have memory, it is often necessary (for good practical reasons) to render it memoryless with respect to the coding system [MAS(84-1)], [HAG(80)]. This can be achieved by applying interleaving techniques, where the deinterleaver tries to convert, for example, a bursty noise (which gives rise to the memory) into a seemingly less damaging type of random noise process. From an information-theoretical point of view, however, the very worst type of additive noise is stationary AWGN [GAL(68), Theorem 7.4.3], [CAH(71)]. Thus, if the channel memory arises from nonstationary (structured) noise, theory suggests that this memory should really be exploited rather than destroyed. However, the theory does not indicate how this memory should be used to realize its potential benefits. In practice, therefore, the issue is considerably more complex and interleaving is often the only practical alternative. The appropriate coding and interleaving strategies depend on factors such as the relation between
system-delay constraints and the length of noise bursts, the affordable encoding-decoding memory, and the ability of the demodulator to estimate the state of the channel. In any case, channel state information should appear in the output of the demodulator so that the decoder can take advantage of it.

When the channel modes are somehow "predictable", it seems plausible that the resulting "memory" should prove to be advantageous. Unless we specify precisely what is meant by "memory", not much more can be said. Wolfowitz [WOL(67)] is usually credited (e.g., [KAN-SAS(78)], [HAG(80)]) for an explicit proof of the thesis "memory increases capacity". It will be instructive to describe briefly his approach using our terminology.

Consider \(L\) DMC's having the same input alphabet, \(A_x\), and the same output alphabet, \(A^\wedge\). Let \(Q_p|_x(\cdot|x;s)\) for each \(x \in A_x\) be the conditional probability distributions over \(A^\wedge\) for the \(s\)-th one of the \(L\) DMC's, where \(s \in \{1, 2, \ldots, L\}\). Next, consider transmitting codewords \(x = (x_0, x_1, \ldots, x_{N-1})\), \(\forall x_k \in A_x\), of arbitrary length \(N\) over two different channels which are formed from this ensemble of \(L\) DMC's according to a probability (row) vector \(q = [q_1, q_2, \ldots, q_L]\). The first channel is formed by choosing the DMC to be used during each symbol interval at random, independently of all other choices, with probability \(q_s\) of choosing \(s \in \{1, 2, \ldots, L\}\). The choice of \(s\) is also independent of any input and output digits, and \(s\) is unknown to both encoder and decoder. The overall channel created by this first method is thus equivalent to a memoryless channel where the conditional probability distributions of receiving \(\hat{y}_k \in A^\wedge\), given that the encoded digit \(x_k \in A_x\) is sent, is

\[
\overline{Q}_p|_x(\hat{y}_k|x_k) = \sum_{s=1}^{L} q_s Q_p|_x(\hat{y}_k|x_k;s) \tag{2.2}
\]
for all \( k = 0, 1, \ldots, N - 1 \). The capacity of this channel will be called \( \bar{C} \).

With \( \hat{y}_k \) representing any digit in \( A_y \) at time instant \( k \), the probability of receiving the digit \( \hat{y}_k \), given that the codeword \( x \) is sent, is then given as

\[
\mathbb{P} \{ \hat{y}_k = y_k \mid x \} = \overline{Q}_{\hat{y} \mid x}(\hat{y}_k \mid x_k), \quad 0 \leq k \leq N - 1,
\] (2.3a)

so that the conditional probability of receiving from the demodulator the sequence \( \hat{y} \), given that \( x \) is sent, becomes

\[
\mathbb{P} \{ \hat{y} = y \mid x \} = \prod_{k=0}^{N-1} \overline{Q}_{\hat{y} \mid x}(\hat{y}_k \mid x_k),
\] (2.3b)

i.e., conditioned on \( x \), the components of any channel output sequence \( \hat{y} \) are independent.

The second channel differs from the first channel only in the way the individual DMC is chosen each time a symbol is being sent. For the same probability vector \( \mathbf{q} = [q_1, q_2, \ldots, q_L] \) used to create the first channel, we now choose \( \mathbf{M} = [m_{ij}] \) to be an \( L \times L \) stochastic matrix that defines an ergodic Markov chain which has \( \mathbf{q} \) as its unique stationary probability distribution, i.e., increasing powers of \( \mathbf{M} \) approach a matrix each of whose rows is \( \mathbf{q} \) [KEM-SNE(60), Theorem 4.1.6]. The DMC to transmit the first symbol of each codeword, \( x_0 \), is chosen according to the probability distribution \( \mathbf{q} \). If \( s_0 \) identifies the DMC chosen to transmit \( x_0 \), the probability that \( s_1 \) should be the identifier of the DMC to transmit the second symbol, \( x_1 \), is \( m_{s_0s_1} \). The

---

1 Note that \( \bar{C} \leq C_{\Sigma} \), where \( C_{\Sigma} \) is the capacity of the sum channel [GAL(68), p. 525 (Problem 4.18)]. Only in the special case where the capacity of the \( s \)-th DMC, \( C_s \), satisfies \( \bar{C} = C_s - \log(q_s) \), \( \forall s \in \{1, 2, \ldots, L\} \), will it be true that \( \bar{C} = C_{\Sigma} = \log_2 \sum_{s=1}^{L} 2^{C_s} \) bits per symbol.
probability that \( s_2 \) determines the DMC to transmit the \textit{third} symbol, \( x_2 \), is \( m_{s_1}s_2 \); and so on. Thus, the successive modes (states) during each symbol interval of the second channel are chosen according to the ergodic Markov process as specified through \( \mathbf{M} \). The choice of \( s \) is again independent of all input and output digits and it is also unknown to both encoder and decoder. Since \( q \) is the \textit{stationary measure} of \( \mathbf{M} \), it follows that the probability of receiving the digit \( \hat{y}_k \), given that the codeword \( x \) is sent, equals the right-hand side of (2.3a), i.e.,

\[
P^* \left\{ \hat{y}_k = \hat{y}_k \mid x \right\} = \bar{P} \left\{ \hat{y}_k = \hat{y}_k \mid x \right\}, \quad 0 \leq k \leq N - 1.
\]  

(2.4)

However, it is \textit{not} in general true for Channel 2 that the demodulator outputs are independently distributed; thus (2.3b) does not hold and the channel exhibits "memory" (arising from the underlying ergodic Markov process). On the other hand, both channels are directly comparable because of (2.4). Denote the capacity of Channel 2 by \( C^* \). Wolfowitz then proved that \( C^* \geq \overline{C} \) [WOL(67), Theorem 1]; more precisely, for the case where both encoder and decoder have no \textit{a priori} knowledge of the state process of the channel, he showed that the capacity of the channel with memory (Channel 2) is \textit{not less} than the capacity of the "comparable" memoryless channel (Channel 1). Later, Dobrushin and Pinsker [DOB-PIN(69)] confirmed this result and argued that it holds under more general conditions. In both cases, however, the proofs are nonconstructive, i.e., they do not indicate \textit{how} the channel memory should be exploited to benefit from the additional capacity that could potentially be realized from the channel's memory.

More constructive results are obtained under the assumption that the state process of a discrete channel with memory is observable by the receiver. As an example, consider the simple \textit{Rayleigh fading} channel with slow fading when it is used without feedback [HAG(80)]. Massey [MAS(84-1)] described a model of
a discrete channel with memory which is applicable to this type of channel and called it a state-observable channel with a freely-evolving state (SOCFES). With a SOCFES in Figure 2.1, the output digits of the demodulator at time instant \( k \) are interpreted as the pair of digits \( \hat{y}_k \equiv (\hat{y}_k, s_k) \) where \( \hat{y}_k \in A_\hat{y} \), is the (soft-decision) quasi-output and \( s_k \in A_s \) is the channel state. The latter may depend on the states prior to time instant \( k \), but evolves otherwise independently of all inputs and quasi-outputs. The channel is characterized for each codeword \( x \) of length \( N \) by the conditional probability distributions

\[
Q(\hat{y} | x) = Q(s_0, s_1, \ldots, s_{N-1}) \prod_{k=0}^{N-1} Q_{\hat{y}_{k} | x}(\hat{y}_k | x_k; s_k) \tag{2.5}
\]

over all \( N \)-sequences \( \hat{y} \) with components \( \hat{y}_k \equiv (\hat{y}_k, s_k) \in \{A_\hat{y}, x A_s\} \). When the state is given, a SOCFES behaves as a memoryless channel with respect to its input and the corresponding quasi-output. The memory in this channel arises from the dependence of the quasi-output sequence \( y' \) on the state sequence \( s = (s_0, s_1, \ldots, s_{N-1}), \forall s_k \in A_s \), which is determined by some independent mechanism. It is now clear how to make use of this type of channel memory: the (maximum-likelihood) decoder should use the state information provided by the demodulator to choose its decoding metric according to which DMC the SOCFES represents at any given time (since \( s \) evolves independent of \( x \), \( s \) by itself does not convey any information about the information source to the user).

What happens in the case where the demodulator denies the state information to the decoder has been shown by Hagenauer [HAG(80)]. He considered using a rate \( R = 1/2 \) binary convolutional code of constraint length 7 on the slowly fading Rayleigh-channel with full interleaving and coherent phase modulation. For a fixed bit error rate, the required signal-to-noise ratio per information bit \( (E_b / N_0) \) must be increased by a factor of up to 1.6 (2 dB) if the demodulator does not deliver the state information (fade depth) to the decoder.
The difference in $E_b/N_o$ based on the capacity criterion (assuming unlimited coding complexity) is about 0.7 dB. Similar results were obtained when studying the same channel with respect to the cutoff-rate criterion [SCH(84)]. Further, the use of coding and soft-decision decoding on fading channels can be considerably more effective than on the stationary AWGN channel [BUC(80)].

B. Channels with Intersymbol Interference Memory

A different class of discrete channels with memory arises from signaling over real communication channels when the waveform channel shown in Figure 2.1 is linear and produces intersymbol interference (ISI) in the received signal $r(t)$. The extent of this effect depends on the bandwidth constraints in the system and on the degree of linear distortion present in the transmitter (modulator) and in the transmission medium [LUC-SAL-WEL(68), Chapter 4]. This class of linear channels is the subject of the present study. The appropriate basic channel model used throughout the thesis will now be described.

Consider the analog model of a pulse amplitude modulated (PAM) digital communication system as shown in Figure 2.2(a). The received signal is

$$r(t) = u(t) + n(t)$$

$$= \sum_k x_k c(t - kT) + n(t),$$

where $c(t)$ is the overall channel impulse response, $x_k$ is the $k$-th transmitted digit (amplitude), $T$ is the symbol interval, and $n(t)$ is AWGN with mean zero power spectral density. The absolute difference in $E_b/N_o$ at bit error rate $10^{-5}$ between the performance curve with state information and the information-theoretical optimum (based on the capacity criterion) is almost 7 dB [HAG(80), Figure 8].

\[1 \] $E_b$ is the average signal energy per information bit and $N_o$ is the one-sided noise power spectral density. The absolute difference in $E_b/N_o$ at bit error rate $10^{-5}$ between the performance curve with state information and the information-theoretical optimum (based on the capacity criterion) is almost 7 dB [HAG(80), Figure 8].
Fig. 2.2. Discrete channel with ISI memory: (a) analog model of a PAM digital communication system with a linear time-dispersive channel, and (b) its equivalent discrete-time model.
and one-sided power spectral density \( N_o \). The output samples of the receiver filter, \( y_k \), are fed to a quantizer which delivers its outputs, \( \hat{y}_k \), to a decoder. The PAM system is said to exhibit ISI whenever the \textit{sampled} channel impulse response, \( c(mT) \), is nonzero for more than one integer \( m \).

The first question about the structure in Figure 2.2(a) is whether any useful "information" is lost by the sampling process. For no such loss to occur, regardless of the optimality criterion used to determine the receiver filter, it is necessary and sufficient that the sampled outputs of the receiver filter, \( y_k = y(kT) \), be a set of sufficient statistics for the estimation of the transmitted digits, \( x_k \), from the received signal, \( r(t) \) [WOZ-JAC(65), Chapter 4]. Ericson [ERI(71)] showed for fairly general criteria that the "optimal" receiver filter consists always of an ordinary matched filter with impulse response \( c(-t) \) followed by some \( T \)-spaced transversal filter. Forney [FOR(72)] has shown in general that an "information lossless" receiver filter can always be realized as the cascade of an ordinary matched filter with impulse response \( c(-t) \) and a certain \( T \)-spaced transversal filter such that the noise samples at its output are \textit{independent, identically-distributed} (i.i.d.) Gaussian random variables. Forney called this arrangement a \textit{whitened matched filter} (WMF) and showed that it is realizable (with finite delay) when \( c(t) \) is of finite duration. Andersen [AND(73)] extended and generalized Forney's results and proved that a WMF structure exists in all cases of practical interest. Andersen also introduced the designation \textit{sample-whitened matched filter} (SWMF) for this type of filter in order to emphasize that only the \textit{sampled} additive noise process in the output of such a filter has the properties of white noise.¹

¹ If, however, the output noise samples are passed through an ideal low-pass filter with cutoff frequency \( 1/(2T) \), then the continuous-time output process is low-pass white noise [AND(73)] (Andersen credits Massey for suggesting the clarifying designation \textit{SWMF} for the class of filters which have the properties of Forney's WMF).
Without loss of optimality, that part of the demodulator in Figure 2.1 that deals with the memory of the channel may be thought of as being combined with the decoder function [VIT-OMU(79), Sections 4.10 and 5.8], [BLA(83), Chapter 15], [OMU(71)]. By adopting this view, i.e., by assuming that the "demodulator" is simply the cascade of a SWMF, a symbol-rate sampler and some quantizer, the discrete channel in Figure 2.1 reduces to the discrete-time channel model of a PAM system shown in Figure 2.2(b). This channel model represents the discrete channel seen by the coding system; in particular, from the point of view of information transfer, it is equivalent to the continuous-time system shown in Figure 2.2(a). The discrete-time channel filter in Figure 2.2(b) is a $T$-spaced transversal filter with a finite number of coefficients, $h_m$, $0 \leq m \leq M$.\footnote{Although it is assumed that the waveform channel in Figure 2.2(a) is strictly band-limited, from a practical point of view its impulse response is finite in duration, i.e., the square-integrable impulse response satisfies $c(t) = 0$ for $t < 0$ and $t \geq (M + 1)T$. This assumption leads to a finite-length transversal filter [FOR(72)].} The unquantized channel output samples are given by the expression

$$y_k = v_k + w_k = \sum_{m=0}^{M} h_m x_{k-m} + w_k, \quad \forall k,$$

where the unit-sample response, $h^{M+1} = (h_0, h_1, ..., h_M)$, is square-summable, i.e., $\|h^{M+1}\|^2 < \infty$, and where the additive noise samples, $w_k$, are i.i.d. Gaussian random variables with mean zero and variance $N_0/2$. The same type of discrete-time model can be used to represent other modulation schemes as well as systems with time-varying channel impulse responses [PRO(75)]. For the present study, however, it is assumed that the modulation is PAM (the $x_k$
may be discrete or continuous real-valued digits) and that the unit-sample response of the discrete-time channel filter, $\mathbf{h}^{M+1}$, is real and time-invariant.

The coefficients of the discrete-time channel filter in Figure 2.2(b) are related to the original real-valued channel impulse response, $c(t)$, in the following way. Let $a_k$, $-M \leq k \leq M$, be the sample at time $t = kT$ of the real-valued autocorrelation function of $c(t)$, i.e., let

$$a_k = a(kT) = \int_{-\infty}^{\infty} c(\tau) c(\tau + kT) d\tau , \quad (2.8a)$$

which, as will be explained below, can also be expressed as

$$a_k = \sum_{m=0}^{M-k} h_m h_{m+k} , \quad 0 \leq k \leq M , \quad (2.8b)$$

with the coefficients $h_m$ being the components of $\mathbf{h}^{M+1}$. It is easily seen that $a_k = a_{-k}$, $\forall k$. Equation (2.8b) shows that the real-valued autocorrelation sequence can be written as a discrete convolution of the real-valued sequence $\{h_m: 0 \leq m \leq M\}$ and its time-reversed replica $\{h_m: 0 \leq m \leq M\}$. To see this let

$$A(z) = \sum_{k=-M}^{M} a_k z^{-k} \quad (2.9)$$

be the two-sided $z$-transform of the autocorrelation sample sequence $\{a_k: -M \leq k \leq M\}$. Since $a_k = a_{-k}$, $\forall k$, it follows that $A(z) = A(1/z)$; thus, if $\alpha$ is a root of $A(z)$ then $1/\alpha$ is also a root. Assume temporarily that $A(z)$ has no roots on the unit circle in the $z$-plane. Therefore, when $\alpha_r$, $r = 1, 2, \ldots, M$, are the $M$ roots of $A(z)$ which are outside the unit circle, then the
remaining $M$ roots $1/\alpha_r$, $r = 1, 2, \ldots, M$, are all inside the unit circle. $A(z)$ can thus be factored as

$$A(z) = y \ z^{-M} \prod_{r=1}^{M} (z - \alpha_r) \prod_{r=1}^{M} (1 - z \alpha_r)$$

(2.10a)

$$= H(z) H(1/z),$$

where $y$ is a real constant and the function [PAP(77), Section 7.2]

$$H(z) = \sqrt{y} \ z^{-M} \prod_{r=1}^{M} (z - \alpha_r)$$

(2.10b)

$$= \sum_{m=0}^{M} h_m z^{-k},$$

has all its roots outside the unit circle and the $h_m$, $0 \leq m \leq M$, are real.\(^1\) The appropriate noise-whitening filter which is connected in cascade with the ordinary matched filter has the $z$-transform transfer function $1/H(1/z)$. The latter specifies a stable recursive discrete-time filter because the roots of $H(1/z)$ were chosen to be inside the unit-circle [OPP-SCH(75), Section 4.3]. From (2.8) it follows directly that the energy in the unit-sample response of the transversal channel filter is

$$\|h^{M+1}\|^2 = \sum_{m=0}^{M} h_m^2 = a_0$$

(2.11)

$$= \|c(t)\|^2 = \int_{-\infty}^{\infty} c^2(\tau) \, d\tau.$$

It has been assumed in the procedure outlined above that $A(z)$ has no roots on the unit circle. The case where $A(z)$ does have $2K \leq 2M$ roots on the unit

\(^1\) The factorization in (2.10) is called the minimum-phase factorization. It leads to a unique $H(z)$, i.e., the energy in the first $m$ components of the corresponding channel unit-sample response, $h^{M+1}$, is maximized for every $m \leq M$, [PAP(77), Chapter 2].
circle can also be handled [FOR(72)]. In this case one writes

\[ A(z) = P(z) A'(z) \]

where \( P(z) = F(z) F(1/z) \) has the 2K roots on the unit circle and \( A'(z) = H'(z) H'(1/z) \) has the \( 2(M - K) \) roots which are not on the unit circle. Further, the K roots of \( F(z) \) are chosen such that there is no pair of reciprocal roots and \( H'(z) \) has the \( M - K \) roots of \( A(z) \) which are outside the unit circle. This factorization implies that the channel impulse response, \( c(t) \), can be written as the linear superposition

\[ c(t) = \sum_{i=0}^{K} f_i c'(t - iT) \quad (2.12) \]

where the coefficients \( f_i, i = 0, 1, \ldots, K \), correspond to the coefficients of \( F(z) \) and where the spectrum of \( c'(t) \) has no nulls within the Nyquist bandwidth.

The SWMF consists then of the cascade of an ordinary matched filter for the response \( c'(-t) \) and a noise-whitening filter with the z-transform transfer function \( 1/H'(1/z) \). The latter is again a stable recursive discrete-time filter because its singularities are all inside the unit circle in the z-plane. The unit-sample response of the equivalent discrete-time channel filter in Figure 2.2(b) is then obtained from the corresponding z-transform transfer function \( H(z) = F(z) H'(z) \), i.e., the sequence \( \{h_m: 0 \leq m \leq M\} \) is obtained by convolving the sequences \( \{f_i: 0 \leq i \leq K\} \) and \( \{h'_i: 0 \leq i \leq M - K\} \).

The channel model of the form shown in Figure 2.2(b) is the discrete channel which is of interest in the present study. In certain situations it is possible that this channel model is arrived at directly, for example, in case of the partial-response (PR) systems [KAB-PAS(75)] whose system polynomials have all roots on the unit circle (these roots correspond to nulls in the transfer function within the Nyquist bandwidth). However this model may arise, we will refer to it as the basic discrete-time Gaussian channel with ISI memory of (finite) length \( M \). Without loss of generality it can be assumed that \( h_0 h_M \neq 0 \); thus, the channel
is said to be memoryless (no ISI) when $M = 0$. For $M \geq 1$, the "state" of this channel at any given time instant, $\sigma_k$, can be taken as the $M$ most recent inputs, i.e., $\sigma_k = (x_{k-1}, x_{k-2}, \ldots, x_{k-M})$. For inputs from a finite alphabet of size $\chi$, the channel model shown in Figure 2.2(b) can thus be described as a finite-state machine (FSM) with $\chi^M$ possible states [FOR(72)], [OMU(71)].

The transfer function of the discrete-time channel filter is given by

$$H(\lambda) = \sum_{m=0}^{M} h_m e^{-jm\lambda}, \quad j = \sqrt{-1},$$

(2.13)

and is periodic in $\lambda$ with period $2\pi$, where $\lambda$ is the normalized frequency.

When comparing different channels, it will be assumed that their unit-sample response energies are identical, i.e., in (2.11) we shall let $\|h^{M+1}\|^2 = a_0 = 1$, for all finite $M = 0, 1, \ldots$. The underlying reason for this channel normalization is that in this way all channels achieve the same matched-filter bound performance, i.e., when only one isolated uncoded digit $x_k$ of (average) energy $E_s$ is transmitted through the channel in Figure 2.2(a), then the received (average) signal energy is $a_0 E_s = E_s$, independent of the particular ISI of the channel [LUC-SAL-WEL(68), pp. 101-102], [UNG(74)]. Furthermore, this channel normalization has been widely used in comparative studies with respect to a variety of performance criteria, such as error-rate performance of coded and uncoded systems [VIT-OMU(79), Sections 4.9 and 5.8], [PRO-KHA(86)]; convergence and performance of equalizers [PRO(83), Chapter 6], [CLA-SLA-RAJ(83)], [SAR(82)]; worst-case channel identification [MAG-PRO(73)], [AND-FOS(75)]; and various other criteria [BIG(84)], [SAS-CAS-NAM(82)], [HOF-MER-NOR(81)]. As a consequence of this channel normalization, it follows that $|H(\lambda)|$ exceeds unity for some $\lambda$ whenever ISI is present, i.e., $\max |H(\lambda)| > 1$, $|\lambda| \leq \pi$, for all $M > 0$. Figure 2.3 shows examples of the magnitude responses of four normalized channels which will be investigated sub-
Fig. 2.3. Examples of magnitude responses, $|H(\lambda)|$, $|\lambda| \leq \pi$, of normalized ISI channels with unit-sample responses of unit energy, i.e., $\|h^{M+1}\|^2 = 1$, where for

Channel 1: $h^1 = h_0 = 1$, (memoryless channel, $M = 0$)
Channel 2: $h^2 = (0.707, 0.707)$,
Channel 3: $h^3 = (0.500, 0.707, 0.500)$,
Channel 4: $h^7 = (0.19, 0.35, 0.46, 0.50, 0.46, 0.35, 0.19)$. 

Normalized Frequency, $|\lambda|/\pi$
sequently, including the memoryless channel \( (M = 0) \). Note that the unit-sample responses given in Figure 2.3 are not necessarily unique (minimum-phase), i.e., other responses of the same length may exist that yield identical magnitude responses \( |H(\lambda)|, |\lambda| \leq \pi \) (see [BER-RAJ-VDL(87)] for an example).

In the following chapters, it will be assumed that the quantizers shown in Figure 2.2 pass their inputs without modification to the decoder, i.e., that \( \hat{y}_k = y_k, \forall k \). Since the decoder then receives ideal soft decisions, the results obtained in the remaining chapters represent (upper) bounds for the same system when used with finite quantization.\(^1\) Finally, the signal-to-noise ratio is taken to be the quantity \( E_s/N_0 \), where \( E_s \) is the average energy used per coded input symbol, \( x_k \), and \( N_0 \) is the one-sided power spectral density of the AWGN. Thus, when \( R \) denotes the information rate in bits per symbol, the average signal energy per information bit is \( E_b = E_s/R \).

2.2 Random-Coding Exponent for Finite-State Channels

Consider the basic discrete-time Gaussian channel with ISI memory as shown in Figure 2.2(b) and described above. Let the input digits \( x_k \in A_x \) and the output digits \( \hat{y}_k \in A_y \), \( \forall k \), where \( A_x \) and \( A_y \) are, respectively, the finite input and output alphabets. Further, let \( \sigma_k = (x_{k-1}, x_{k-2}, \ldots, x_{k-M}) \) be the state of the channel at time instant \( k \); if \( A_x \) has size \( \chi \), then there are \( \chi^M \) possible states. When an input \( x_k \) is applied while the channel is in state \( \sigma_k \), then the channel produces a corresponding output digit, \( \hat{y}_k \), and moves to the next state, \( \sigma_{k+1} = (x_k, x_{k-1}, \ldots, x_{k+1-M}) \). Note that the state sequence \( (\sigma_1, \sigma_2, \ldots, \sigma_N) \), where \( \sigma_N = (x_{N-1}, x_{N-2}, \ldots, x_{N-M}) \) is the final state reached after trans-

---

\(^1\) With coded binary signaling and eight-level quantization, for example, the resulting loss in signal-to-noise ratio (for a fixed performance criterion) may be not more than 0.25 dB relative to the same system without quantization [VIT-OMU(79), Figure 4.12], [SAS-KAS-NAM(82)].
mission of the last digit of $x^N$, depends only on the input sequence $x^N$ and on
the initial state $\sigma_0$. Thus, given $x_k$, $k \in \{0, 1, \ldots, N-1\}$, the conditional
probability distributions over $A^\sigma$, $P(\hat{y}_k|x_k)$, depend on the input probability
assignment. Therefore, $P(\hat{y}_k|x_k)$ is generally undefined in terms of the channel
alone. The basic discrete-time Gaussian channel with ISI belongs thus to the
class of finite-state channels (FSC’s) as described in [GAL(68), Section 4.6]. In
particular, the present channel model represents an indecomposable FSC, i.e.,
each of its states can be reached from any other state by some input sequence
of finite length (here the length corresponds to the memory length, $M$)
[GAL(68), pp. 105-111]. Practically speaking, an indecomposable FSC is a
FSC for which the influence of the initial state, $\sigma_0$, vanishes with increasing
time.

Let $\Sigma$ be the set of the $\chi^M$ possible channel states. Gallager [GAL(68),
p. 100 and Theorem 4.6.4] has shown that the capacity of an indecomposable
FSC is independent of the initial state, i.e., for any $\sigma_0 \in \Sigma$, the capacity, $C$, is
given by

$$C = \lim_{N \to \infty} C_N,$$

(2.14a)

where

$$C_N = \max_{Q_X} \frac{1}{N} I(X^N; \hat{Y}^N|\sigma_0),$$

(2.14b)

and where $I(X^N; \hat{Y}^N|\sigma_0)$ is the average mutual information

\[1\] To be precise, $P_{\hat{Y}^N|X^N}(\hat{y}^N|x^N, \sigma_0)$ should be written as $P_{\hat{Y}^N|X^N}(\hat{y}^N|x^N, \sigma_0)$. To
simplify, however, the former notation will be used throughout, e.g., instead of writing
$Q_{X^N}(x^N)$ we will use the short notation $Q_X(x^N)$. This short notation will be
particularly useful when random sequences are written in the form $X^{[n, m]} = (X_n, X_{n+1}, \ldots, X_m)$. No ambiguity should arise with this convention since the
dimension(s) of the random vector(s) in the subscript of the probability distributions are
always identical with the dimension(s) of the vector(s) in the argument.
\[ I(X^N; \hat{Y}^N | \sigma_0) = \sum_{x^N} \sum_{\hat{y}^N} Q_X(x^N) P_{\hat{y}^N | x^N} (\hat{y}^N | x^N, \sigma_0) \times \]

\[ \times \log \left[ \frac{P_{\hat{y}^N | x^N} (\hat{y}^N | x^N, \sigma_0)}{\sum_{x^N} Q_X(x^N) P_{\hat{y}^N | x^N} (\hat{y}^N | x^N, \sigma_0)} \right] \]  \hspace{1cm} (2.14c)

The maximization in (2.14b) is taken over all probability assignments \( Q_X(\cdot) \) that satisfy some specified constraint on the input sequences \( x^N \). The initial state, \( \sigma_0 \), appears in (2.14) as a reminder that some \( \sigma_0 \) must be assumed in actual computations, e.g., without loss of generality, \( \sigma_0 \) may be the all-zero state. Thus, unless specified otherwise, it will be assumed hereafter that \( \sigma_0 = (0, 0, \ldots, 0) \) prior to transmission of the first digit of an input sequence. Consequently, \( \sigma_0 \) will be dropped in future expressions to simplify notation.

Gallager [GAL(68), Section 5.9] has proved a coding theorem for (indecomposable) FSC's that is applicable independent of the initial state. Let

\[ E_r(R) = \max_{0 \leq \rho \leq 1} \left[ F_{\infty}(\rho) - \rho R \right], \]  \hspace{1cm} (2.15a)

where

\[ F_{\infty}(\rho) = \lim_{N \to \infty} F_N(\rho), \]  \hspace{1cm} (2.15b)

\[ F_N(\rho) = -\frac{\rho \ln x^M}{N} + \max_{Q_X} E_{o,N}^*(\rho, Q_X), \]  \hspace{1cm} (2.15c)

and

\[ E_{o,N}^*(\rho, Q_X) = -\frac{1}{N} \ln \left[ \sum_{\hat{y}^N} \left( \sum_{x^N} Q_X(x^N) \left[ P_{\hat{y}^N | x^N} \left( \hat{y}^N | x^N \right) \right]^{1+\rho} \right)^{1+\rho} \right], \]  \hspace{1cm} (2.15d)
with the superscript $^*$ indicating that the right-hand side of (2.15d) is minimized over all $\sigma_0 \in \Sigma$. With the natural logarithm used in (2.15), the units of $E_r(R)$ are in nats per symbol (with the logarithm taken to the base 2, the units become bits per symbol).

**Theorem 2.1:** (after Gallager [GAL(68), Section 5.9]) For any $\varepsilon > 0$, there exists $N(\varepsilon)$ such that for each $N \geq N(\varepsilon)$ and each rate $R \geq 0$ there exists an $(N,R)$ block code for which, for all codewords $x_m^N$, $1 \leq m \leq \lceil e^{NR} \rceil$, and all initial states $\sigma_0$, the probability of deciding on a wrong codeword is upper bounded as

$$P_{e,m}(\sigma_0) \leq \exp\left[-N\left( E_r(R) - \varepsilon \right) \right].$$

(2.16)

Furthermore, for $0 \leq R < C$, where $C$ is the channel capacity given by (2.14), $E_r(R)$ is strictly positive, strictly decreasing in $R$, and convex-$\cup$.

For $0 \leq R < C$, the promise of Theorem 2.1 is that system reliability can be as high as is desired (assuming unlimited coding complexity) by choosing an appropriate $(N,R)$ code. The theorem is not as strong as the corresponding theorem for the memoryless channel [GAL(68), Theorem 5.6.2], since it is only valid for $N \geq N(\varepsilon)$. The quantity $E_r(R)$ is called the random-coding exponent; alternatively, the terms error exponent and Gallager exponent have been used. $E_r(R)$ represents, in general, a lower bound for the reliability function, $E(R)$, defined as

$$E(R) = \lim_{N \to \infty} \sup_{N} \frac{-\ln P_e(N,R)}{N},$$

(2.17)

---

1 For any positive integer $N$ and any positive number $R$, an $(N,R)$ block code is a code of block-length $N$ with $\lceil e^{NR} \rceil$ codewords (here $\lceil \cdot \rceil$ denotes the smallest integer greater than or equal to the argument).
where $P_e(N, R)$ is the minimum value of the probability that a wrong codeword is decoded over all $(N, R)$ codes for given $N$ and $R$ [GAL(68), p. 160]. $E(R)$ is in general not known exactly over the entire region of rates, $0 \leq R \leq C$. However, the characteristic properties of $E(R)$ are well established [GAL(65)].

The slope of $E_r(R)$ decreases from 0 at $R = C$ and it reaches the value $-1$ at $R = R_c$; $E_r(R)$ is linear in the range $0 \leq R \leq R_c$ with slope $-1$. Generally, $E(R)$ is uniquely known only for rates within the upper range $R_c \leq R \leq C$. For the range $0 \leq R < R_c$, $E(R)$ is bounded as $E_r(R) \leq E(R) \leq E^*_r(R)$, where $E^*_r(R)$ is the sphere-packing exponent [GAL(68), Section 5.8]. The typical form of $E_r(R)$, as defined by (2.15) for $0 \leq \rho \leq 1$, is illustrated in Figure 2.4. Its exact form depends on the signal-to-noise ratio, the specific input constraint (modulation) and the memory characteristics of the basic FSC. The capacity $C$ is that smallest rate $R$ where $E_r(R) = E(R) = E^*_r(R) = 0$. The random-coding exponent, $E_r(R)$, can be improved to yield a tighter upper bound on the probability of error for small rates $R$. This can be achieved by expurgating those codewords in the ensemble of codes for which the performance is poor [GAL(65)]. The tangent line for $E_r(R)$ drawn at the point $R = R_c$ has slope $-1$ and intercepts the rate axis at $R = R_o$. From (2.15), the parameter

\[ R_o = E_r(0) = F_{\infty}(1) = \lim_{N \to \infty} \max_{Q_X} E^*_{0,N}(1, Q_X) \]

\[ \text{Note that the slope of } E_r(R) \text{ is } -\rho, \text{ however, this slope is not necessarily continuous in } R. \text{ The reason is that the } Q_X \text{ which maximizes } E_r(R) \text{ may change with } R \text{ and this can cause discontinuities in the slope of } E_r(R). \text{ Furthermore, there exist channels where } E_r(R) \text{ consists only of the linear segment, i.e., its slope is } -1 \text{ over the entire range of rates where } P_{e,m}(\sigma_0) \to 0 \text{ is theoretically possible [GAL(65)]}. \text{ For a discussion on input constraints, see [GAL(68), Section 7.3] and [SHA(59)].} \]
Fig. 2.4. Typical form of the random-coding exponent of a FSC, $E_r(R)$, as defined by (2.15).

\[
\lim_{N \to \infty} -\frac{1}{N} \ln \left[ \min_{Q_x} \sum_{\hat{y}^N} \left( \sum_{x^N} Q_x(x^N) \sqrt{P_{\hat{y}|x}(\hat{y}^N|x^N)} \right)^2 \right]
\]

\[
(2.18)
\]

represents the zero-rate error exponent. The right-hand side of (2.16) can thus be further upper bounded as

\[
P_{e,m}(\sigma_0) \leq \exp\left[-N\{R_o - R - \varepsilon\}\right],
\]

\[
(2.19)
\]
where, as in (2.16), \( \varepsilon \) can be made arbitrarily small. The parameter \( R_o \) is more commonly called the \textit{cutoff rate} and is of significance for several reasons. For one, it determines the rate \( R \) above which the average number of computation steps per decoded digit for sequential decoding is unbounded [VIT(67)], [VIT-OMU(79), Chapter 6]. The parameter \( R_o \) appears also in code-ensemble performance bounds for the class of (time-varying) convolutional codes [VIT(71)], [VIT-OMU(79), Chapter 5]. Further, it can be noted from Figure 2.4 that the simpler bound (2.19) is fairly tight for \( R \approx R_c \), i.e., for rates \( R \) which are neither near zero nor near capacity. These observations have led some investigators to consider the cutoff-rate parameter, \( R_o \), as the principal design criterion for the modulation-demodulation scheme in a coded digital communication system [WOZ-KEN(66)], [MAS(74)], [LEE(76)], [MAS(84-1)].

In this context, the single number \( R_o \) is significant because 1) it determines an upper limit of a (practical) rate region over which operation with arbitrarily small reliability is possible, and 2) it provides a measure for the coding complexity (codeword length or code constraint length) which is required to achieve a specified decoding reliability for some \( R < R_o < C \). On the other hand, (2.19) does not show that \( P_{e,m}(\sigma_0) \to 0 \) is possible for \( R_o \leq R < C \).

Although \( R_o \) is generally simpler to compute than \( E_\varepsilon(R) \), the minimization in (2.18) is often difficult to perform. For this reason, and from other considerations, a lower bound on \( R_o \), called the \textit{symmetric cutoff rate}, \( \tilde{R}_o \), has been considered [LEE(76)], [WOZ-JAC(65), Chapter 5]. \( \tilde{R}_o \) is the value of the right-hand side of (2.18) when \( Q_X \) is such that the components of the random vector \( X \) of arbitrary length are \textit{independent} and \textit{uniformly} distributed over the input alphabet, \( A_x \). Similarly, by choosing some \( Q_X \) in (2.14) that satisfies these input constraints, one obtains a lower bound on \( C \). Such lower-bound quantities might be called \textit{feasible} or \textit{achievable} information rates, in the sense that \( P_{e,m}(\sigma_0) \to 0 \) is theoretically surely possible at these rates since they are below \( C \). When considering different channels within the same class, useful
results can still be obtained by comparing their relative performances based on such lower bounds on $R_p$ and $C$ (see [VIT-OMU(79), Section 5.8, Fig. 5.11] for an example).

2.3 Related Work

A brief and selective overview of some lesser-known work which relates to this background material will now be given. Some results on the evaluation of $E_r(R)$ for the Gaussian discrete-time channel with ISI have been reported in [SAS-KAS-NAM(82)]. Lower bounds on $E_r(R)$ were computed there for a channel with unit memory under the assumption that the inputs are i.i.d. binary digits. It was shown that this lower bound on $E_r(R)$ increases markedly when the demodulator delivers soft-decisions (up to 8 quantization levels) and the decoder accounts for the ISI memory. Savage [SAV(71)] introduced two measures to assess complexity in terms of computational work and decoding delay of various decoding rules (algorithms) for fixed reliability at some code rate $R$. Bounds on the size of these measures were derived that depend on bounds on $E(R)$.

An interesting application of the theory of FSC's has been reported by O'Neill and Lin [ONE-LIN(84)] who modeled a neuron encoder (a cell body which is part of a neuron) as an indecomposable FSC. They showed that the capacity-achieving input distribution is decidedly not Gaussian. For the class of FSC's where the states are the same as the channel outputs, simple necessary and sufficient conditions to determine when the capacity is positive have been derived by Wagner [WAG(71)] (the conditions for the general FSC are apparently unknown).

A novel class of channel models with memory, called block interference channels, were introduced by McEliece and Stark [McE-STA(84)] to study the effects of various kinds of interference phenomena. They point out that, with all other parameters held fixed, channel capacity is an increasing function of
channel memory, while the cutoff rate is generally a decreasing function. The conjecture was offered that, for these channels, $R_0$ is more properly an inverse measure of the coding delay rather than of the coding complexity required to achieve a given performance.

Signal design for non-white Gaussian noise channels based on the $R_0$ criterion has been considered by Bordelon [BOR(76)]. The cutoff rate of band-limited linear and nonlinear channels has been derived by Divsalar [DIV(78)] for the case where the decoder does not have exact knowledge of the channel parameters (mismatch conditions).

Saleh and Salz [SAL-SAL(87)] investigated $R_0$ for the peak-power-limited Gaussian channel. Their major result is that the $R_0$-achieving input distribution for the nondispersive channel with quadrature-amplitude modulation is discrete. For the peak-power-limited dispersive channel, the same conclusion was reached; however, only an asymptotic solution was given for large signal-to-noise ratios by showing that the inputs should be chosen from a uniform distribution. Similar results were obtained and reported earlier by Einarrson [EIN(79)] for the amplitude-limited nondispersive AWGN channel when considering signal design by optimization of $E_r(R)$. He also showed, for small signal-to-noise ratios, that binary signaling is optimal for one-dimensional modulation schemes and that amplitude limitation does not degrade the achievable $R_0$ compared to an average power limitation. The work by Smith [SMI(71)] on the capacity of amplitude- and variance constrained scalar Gaussian channels is closely related to the two previous studies. Smith, however, gave a rigorous proof that the amplitude-constrained (i.e., peak-power-limited) capacity of a memoryless Gaussian channel is achieved by a unique discrete random variable taken from a finite alphabet. In this context, the work of Färber [FAR(67-1)], [FAR(67-2)] and of Färber and Appel [FAR-APP(68)] published earlier (in German) are not well known in general. Besides reporting Smith's result, they also treated the case where the variance of the additive noise depends linearly
on the chosen amplitude of the transmitted input symbol. They found for this case that the capacity-achieving input distribution is also discrete but skewed. All problems involving amplitude-limited signal sets led to discrete input distributions and required nonlinear optimization procedures to generate numerical results.
CHAPTER 3

Capacity of the Discrete-Time Gaussian Channel

with Intersymbol Interference

Consider the class of discrete-time Gaussian channels with intersymbol interference (ISI) where the inputs are subject to a per symbol average-energy constraint. A member of this class of channels will be called a discrete-time Gaussian channel (DTGC). In this chapter, the capacity of the DTGC is of interest. In general, channel capacity can be defined and computed provided that the channel model includes two specifications, namely, 1) the basic channel model specifying the conditional probability for the output given some specific input, and 2) the constraints on channel usage (input constraints). In this chapter, the capacity of the DTGC is derived by means of a hypothetical channel model, called the N-circular Gaussian channel (NCGC), whose capacity is readily derived. The results obtained for the NCGC are used further to prove that in the limit of increasing block-length, \( N \), the capacity of the discrete-time Gaussian channel with ISI using a per block average-energy input constraint (\( N \)-block DTGC) is indeed also the capacity of the DTGC.

3.1. Definitions and Remarks

Three related channel models for channels with ISI are introduced and their respective capacities are defined. These channel models differ in one or both of the two defining specifications. Without loss of generality, only one-dimensional channel models are being considered, i.e., the channel responses and the input symbols are assumed to be real quantities; consequently, the channel outputs are also real.
A. Discrete-Time Gaussian Channel with ISI

The discrete-time model for the equivalent baseband channel of a PAM system with ISI, and with zero-mean additive white Gaussian noise (AWGN) having one-sided power spectral density, $N_0$, is the basic channel model of interest (see also Chapter 2). The input sequence, $\{x_k\}$, produces the noise corrupted output sequence, $\{y_k\}$, given by

$$y_k = \sum_{i=0}^{M} h_i x_{k-i} + w_k, \quad -\infty < k < \infty,$$  \hspace{1cm} (3.1)

where the finite-length sequence $(h_0, h_1, \ldots, h_M)$, with $h_0 h_M \neq 0$, is the unit-sample response of the equivalent channel filter. The transfer function of this filter is

$$H(\lambda) = \sum_{i=0}^{M} h_i e^{-j\lambda i}, \quad j = \sqrt{-1},$$  \hspace{1cm} (3.2)

and it is periodic in $\lambda$ with period $2\pi$. The noise samples, $w_k$, are independent and identically distributed (i.i.d.) Gaussian random variables with mean zero and variance $N_0/2$, i.e.,

$$\mathbb{E}[w_k] = 0, \quad \text{and} \quad \mathbb{E}[w_k w_l] = (N_0/2)\delta_{k-l},$$  \hspace{1cm} (3.3)

where $\mathbb{E}[.]$ means expectation, and where $\delta_0 = 1$, and $\delta_n = 0$, $n \neq 0$. With the time unit taken as the interval between input symbols, $M$ represents the ISI memory length of the channel. The channel is said to have ISI if, and only if, $M > 0$. Since $M$ is finite, the energy of the channel's unit-sample response is also finite. For $m \leq k < N$, we write (3.1) symbolically as

$$y_{[m, N-1]} = x_{[m-M, N-1]} \ast h_{[0, M]} + w_{[m, N-1]},$$  \hspace{1cm} (3.4)
where \( \ast \) denotes the linear convolution operator and where here, and hereafter, we use the sequence notation \( s_{[m,n]} = (s_m, s_{m+1}, \ldots, s_n) \). Note that the subsequence \( x_{[m-M,m-1]} \) of the input sequence \( x_{[m-M,N-1]} \) represents the initial contents of the channel memory, i.e., the channel state at time instant \( k = m \), when we begin to observe the output. It is required that the basic channel model (3.1) is used such that the inputs satisfy the constraint

\[
E[x_k^2] \leq E_s, \quad -\infty < k < \infty ,
\]

(3.5)

where \( E_s \) is the maximum allowed per symbol average energy. We call the channel defined by both, the basic channel model (3.1) and the input constraint (3.5), the discrete-time Gaussian channel with finite ISI memory \( M \), or simply, the DTGC. The DTGC appears to be the natural channel model for the energy-constrained Gaussian channel with ISI, although the difficulty of dealing with constraint (3.5) has led to the more general use of a different channel model, namely, the block-energy constrained channel described below (\( N \)-block DTGC).

It follows from the work of Gallager [GAL(68), Sections 4.6 and 5.9] (see also Chapter 2) that the appropriate definition for the capacity of the DTGC (in bits per channel input symbol) is

\[
C(E_s) = \lim_{N \to \infty} I_N(E_s)
\]

(3.6a)

with

\[
I_N(E_s) = \sup_{q_x} \frac{1}{N} I(x_{[0,N-1]} ; y_{[0,N-1]}) ,
\]

(3.6b)

where the supremum of the average mutual information, \( I(\cdot ; \cdot) \), is taken over all probability densities, \( q_x \), for the sequences \( x_{[0,N-1]} \) satisfying symbol-energy constraint (3.5), and where it is assumed that
\[ x^{[\![ -M, -1 ]\!]} = (0, 0, \ldots, 0). \]  

(3.6c)

The choice of \( x^{[\![ -M, -1 ]\!]} \) in (3.6c) is made for convenience, but has no influence on \( C(E_s) \) since \( M \) is finite. Definition (3.6) is appropriate in the sense that a coding theorem and its converse can be proved to show that \( C(E_s) \) is the upper limit of information rate (in bits per channel input symbol) such that arbitrarily reliable communication is possible over the DTGC.

B. N-Block DTGC

The second channel model also includes the basic channel model of (3.1), however, the constraint on the inputs is now

\[ \sum_{k=0}^{N-1} E[x_k^2] \leq NE_s, \]  

(3.7)

where \( N \) is the block length. Thus, \( NE_s \) is the maximum allowed per block average energy. We call the channel model as defined by (3.1) and (3.7) the \textit{N-block DTGC}. Note that for any \( N \), constraint (3.7) is weaker than constraint (3.5); constraint (3.5) always implies (3.7), but not conversely.

The capacity of the \textit{N-block DTGC} (in bits per channel input symbol) is defined as

\[ C_N(E_s) = \sup_{q_X} \frac{1}{N} I(\{X^{[\![ 0, N-1 ]\!]}; Y^{[\![ 0, N-1 ]\!]}), \]  

(3.8)

where the supremum is taken over all probability densities, \( q_X \), for the sequences \( X^{[\![ 0, N-1 ]\!]} \) satisfying block-energy constraint (3.7), and where \( X^{[\![ -M, -1 ]\!]} \) is the same as in (3.6c). It is conventional to define the quantity
\[ \hat{C}(E_d) = \lim_{N \to \infty} \hat{C}_N(E_d) \]  

(3.9)

to be the "capacity" of the energy-constrained (discrete-time) Gaussian channel with ISI [TSY(65)], [TSY(70)], [GRA(72)], [TOM-BER(73)]. However, it is important to note that \( \hat{C}(E_d) \) is not an actual capacity because the \( N \)-block DTGC is (by definition) a different channel for each \( N \). It seems intuitively obvious, however, that

\[ C(E_d) = \hat{C}(E_d) , \]  

(3.10)

but (to our knowledge) this has never previously been proved. All one could claim is that \( C(E_d) \leq \hat{C}(E_d) \), since the symbol-energy constraint (3.5) is stronger than the block-energy constraint (3.7). In order to prove the validity of (3.10), a new channel model is introduced whose capacity can be readily determined and used to relate \( C(E_d) \) and \( \hat{C}(E_d) \).

C. New Channel Model

We define a new channel model by modifying (3.1). The samples of the output sequence of the new basic channel model, \( \{\tilde{y}_k\} \), are determined by

\[ \tilde{y}_k = \sum_{i=0}^{N-1} \tilde{h}_i x((k-i)) + w_k , \quad 0 \leq k < N , \]  

(3.11)

where \(((\cdot))\) denotes addition modulo \( N \), and \( N > M \). Defining \( \tilde{h}^{[0,N-1]} = (h_0, h_1, \ldots, h_M, 0, 0, \ldots, 0) \) to be the original unit-sample response, \( h^{[0,M]} \), extended with \( (N-M-1) \) zeros, one can write (3.11) symbolically as

\[ \tilde{y}^{[0,N-1]} = x^{[0,N-1]} \otimes h^{[0,N-1]} + w^{[0,N-1]} , \]  

(3.12)
where $\otimes$ denotes the circular convolution operator. For the new channel model, the input constraint is

$$E[x_k^2] \leq E_s, \quad 0 \leq k < N,$$

(3.13)

where $E_s$ is again the maximum allowed per symbol average energy. We call the new channel model, defined by (3.11) and (3.13), the $N$-circular Gaussian channel (NCGC).

The capacity of the NCGC (in bits per channel input symbol) is defined as

$$\widetilde{C}_N(E_s) = \sup_{q_X} \frac{1}{N} I(x^{0, N-1}; \tilde{y}^{0, N-1}),$$

(3.14)

where the supremum is taken over all probability densities, $q_X$, for the sequences $x^{0, N-1}$ satisfying symbol-energy constraint (3.13). Note that there is no need to define an initializing input sequence since any output sequence, $\tilde{y}^{0, N-1}$, is completely determined from the input sequence, $x^{0, N-1}$, and the noise sequence, $w^{0, N-1}$. For the NCGC, we define the asymptotic capacity

$$\widetilde{C}(E_s) = \lim_{N \to \infty} \widetilde{C}_N(E_s),$$

(3.15)

which again is not itself a true capacity since the NCGC is also (by definition) a different channel for each $N$.

D. Remarks

The capacity of the $N$-block DTGC, $\hat{C}_N(E_s)$, as well as its limit as $N \to \infty$, $\hat{C}(E_s)$, have previously been derived by Tsybakov [TSY(65), TSY(70)] and others [TOM-BER(73), BRA-WYN(74)]. Tsybakov also treated the case where the length of the channel memory is unbounded ($M \to \infty$). In all of these cases, $\hat{C}_N(E_s)$ was obtained by solving an eigenvalue problem, and $\hat{C}(E_s)$ was
found by invoking asymptotic properties known from the theory of Toeplitz forms [GRE-SZE(58)], [GRA(72)].

This chapter introduces an approach to finding the capacity of the DTGC, $C(E_x)$, based on the discrete Fourier transform (DFT). It seemed worthwhile giving a rigorous derivation of capacity for this important channel model using only the well-known theory of the DFT rather than the more specialized theory of Toeplitz forms. Moreover, the new approach allows one to prove that the conventional capacity, $C(E_x)$, defined as the limit as $N \to \infty$ of the capacity of the $N$-block DTGC, $\hat{C}_N(E_x)$, is indeed also the capacity of the DTGC, $C(E_x)$.

Section 3.2 states the main results of this chapter, which are then proved in the following sections. Section 3.3 analyzes the new NCGC, and Section 3.4 proves fundamental relations between the channel models introduced above. Section 3.5 completes the proofs of the main results, Section 3.6 considers extension of these results to channels with infinite memory, and Section 3.7 contains numerical results.

3.2 Statement of Main Results

**Theorem 3.1:** The capacity of the NCGC (in bits/symbol for $\log = \log_2$) is given by

$$\tilde{C}_N(E_x) = \frac{1}{2N} \sum_{i=0}^{N-1} \log \left[ \max \left( \Theta |\tilde{H}_i|^2, 1 \right) \right],$$

(3.16a)

where $\tilde{H}[0,N-1]$ is the DFT of $\tilde{H}[0,N-1]$, i.e.,

$$\tilde{H}_i = \sum_{m=0}^{N-1} \tilde{h}_m e^{-j2\pi im/N}, \quad 0 \leq i < N,$$

(3.16b)

and where the parameter $\Theta$ is the solution of
Moreover, for the capacity-achieving $q_X$, the components of the input sequence, $x[0,N-1]$, are correlated Gaussian random variables with mean zero and covariances $\tilde{r}_n$, $0 \leq n < N$, given by

$$\tilde{r}_n = E[x_{k+n}x_k]$$

$$= \frac{1}{N} \sum_{i=0}^{N-1} \varepsilon_i \cos(2\pi ni/N), \quad 0 \leq k \leq k + n < N,$$

where the components of the spectral input energy sequence, $\varepsilon[0,N-1]$, satisfy

$$\varepsilon_i = \begin{cases} \frac{N_0}{2} & \left( \Theta - |\tilde{H}_i|^{-2} \right), \quad \Theta |\tilde{H}_i|^2 > 1 \\ 0, & \text{otherwise.} \end{cases}$$

In particular, capacity is achieved when equality holds in (3.13), i.e., when all inputs, $x_k$, $0 \leq k < N$, have the same average energy, $E[x_k^2] = \tilde{r}_0 = E_s$. Note that (3.17a) implies that the covariance sequence, $\tilde{r}[0,N-1]$, is the inverse DFT of the spectral energy sequence, $\varepsilon[0,N-1]$; conversely, $\varepsilon[0,N-1]$, is the DFT of $\tilde{r}[0,N-1]$.

**Corollary 3.1:** The DTGC, the $N$-block DTGC, and the NCGC, are asymptotically equivalent channel models, in the sense that

$$C(E_s) = \hat{C}(E_s) = \tilde{C}(E_s).$$
Theorem 3.2: The capacity of the DTGC (in bits/symbol for \( \log \equiv \log_2 \)) is given by

\[
C(E_s) = \frac{1}{2\pi} \int_{0}^{\pi} \log \left( \max \left( \Theta |H(\lambda)|^2, 1 \right) \right) d\lambda ,
\] (3.19a)

where \( H(\lambda) \) is the channel transfer function given in (3.2) and where the parameter \( \Theta \) is the solution of

\[
\int_{0}^{\pi} \max \left( \Theta - |H(\lambda)|^{-2}, 0 \right) d\lambda = 2\pi E_s / N_0 .
\] (3.19b)

Moreover, for the capacity-achieving \( q_X \), the inputs, \( x_k, -\infty < k < \infty \), are correlated Gaussian random variables with mean zero and covariances \( r_n, -\infty < n < \infty \), given by

\[
r_n = E[x_{k+n}x_k] = \frac{1}{\pi} \int_{0}^{\pi} S_x(\lambda) \cos(n\lambda) d\lambda ,
\] (3.20a)

where the input power spectral density satisfies

\[
S_x(\lambda) = \begin{cases} 
\frac{(N_0/2)\left[\Theta - |H(\lambda)|^{-2}\right]}{\Theta |H(\lambda)|^2 > 1, |\lambda| \leq \pi} \\
0 , & \text{otherwise}.
\end{cases}
\] (3.20b)

In particular, capacity is achieved when equality holds in (3.5), i.e., when all inputs, \( x_k, -\infty < k < \infty \), have the same average energy, \( E[x_k^2] = r_0 = E_s \).
3.3 Analysis of the NCGC

The proof of Theorem 3.1 will be given by first considering a channel with the basic channel model of the NCGC given in (3.11) and the block-energy input constraint given in (3.7). We are thus interested in the quantity

$$\tilde{I}_N(E_d) = \sup_{q_X} \frac{1}{N} I(X^{[0,N-1]}; Y^{[0,N-1]})$$

where the supremum is taken over all probability densities, $q_X$, for the sequences $x^{[0,N-1]}$, satisfying block-energy constraint (3.7). We shall then show that the optimizing $q_X$ in (3.21) also satisfies the stronger symbol-energy constraint (3.13) so that $\tilde{C}_N(E_d) = \tilde{I}_N(E_d)$.

A. Derivation of $\tilde{I}_N(E_d)$

The DFT of a sequence $b^{[0,N-1]} = (b_0, b_1, \ldots, b_{N-1})$ is the sequence $B^{[0,N-1]} = (B_0, B_1, \ldots, B_{N-1})$, defined by

$$B_i = \sum_{k=0}^{N-1} b_k \Omega_N^{-ik}, \quad 0 \leq i < N,$$

(3.22a)

where $\Omega_N = e^{j\frac{2\pi}{N}}$, $j = \sqrt{-1}$, [OPP-SCH(75), p. 100]. The inverse DFT is given by

$$b_k = \frac{1}{N} \sum_{i=0}^{N-1} B_i \Omega_N^{ki}, \quad 0 \leq k < N.$$

(3.22b)

Taking the DFT on both sides of (3.12) yields

$$\tilde{Y}_i = \tilde{H}_i X_i + W_i, \quad 0 \leq i < N,$$

(3.23)
as a result of the linearity and the circular convolution properties of the DFT
[OPP-SCH(75), p. 110]. In (3.23), \( \tilde{Y}_i \), \( \tilde{H}_i \), \( X_i \), and \( W_i \), are the components of
\( \text{DFT}\{y^{[0,N-1]}\} \), \( \text{DFT}\{h^{[0,N-1]}\} \), \( \text{DFT}\{x^{[0,N-1]}\} \), and \( \text{DFT}\{w^{[0,N-1]}\} \),
respectively. In the following, let \( L = \lfloor N/2 \rfloor \), where \( \lfloor \cdot \rfloor \) denotes the integer
part of a real number.

For any sequence \( \mathbf{b}^{[0,N-1]} = \text{DFT}\{\mathbf{b}^{[0,N-1]}\} \) where \( \mathbf{b}^{[0,N-1]} \) is real,
\( B_i = B_{N-i}^* \), \( 1 \leq i < N \), [OPP-SCH(75), p. 110] (here, * denotes complex
conjugate). Note that \( B_0 \) is real for \( N \) even or odd, and \( B_L \) is real for \( N \) even.
Therefore, knowledge of the components \( B_i \), \( 0 \leq i \leq L \), is sufficient to
reconstruct the real sequence \( \mathbf{b}^{[0,N-1]} \) and no information is lost by discarding
the components \( B_i \), \( L < i < N \). The remaining complex components, \( B_i \),
\( 0 \leq i \leq L \), may be further transformed according to

\[
B_i' = \begin{cases} 
B_i^R, & i = \begin{cases} 
0, & N \text{ odd} \\
0 \text{ or } L, & N \text{ even}
\end{cases} \\
\sqrt{2} B_i^R, & 1 \leq i \leq \begin{cases} 
L, & N \text{ odd} \\
L - 1, & N \text{ even}
\end{cases} \\
\sqrt{2} B_{N-i}^L, & L < i < N,
\end{cases}
\]  

(3.24a)

where \( B_i^R = \text{Re}(B_i) \) and \( B_i^L = \text{Im}(B_i) \). The obvious inverse of (3.24a) is

\[
B_i = \begin{cases} 
B_i', & i = \begin{cases} 
0, & N \text{ odd} \\
0 \text{ or } L, & N \text{ even}
\end{cases} \\
(B_i' + jB_{N-i}^*)/\sqrt{2}, & 1 \leq i \leq \begin{cases} 
L, & N \text{ odd} \\
L - 1, & N \text{ even}
\end{cases} \\
B_{N-i}^*, & L < i < N.
\end{cases}
\]  

(3.24b)

To obtain an equivalent representation of (3.11), we divide both sides of (3.23)
by the complex constant \( \tilde{H}_i \) (assuming temporarily that \( \tilde{H}_i \neq 0 \), all \( i \)) and
transform the resulting first \((L + 1)\) components with (3.24a). The equivalent form of (3.11) in the transform domain is now

\[
Y_i' = X_i' + V_i', \quad 0 \leq i < N, \tag{3.25}
\]

where the \(X_i'\) and \(V_i'\) were obtained from (3.24a) using the identities \(B_i = X_i\), and \(B_i = W_i|\tilde{H}_i\), respectively, and where \(\tilde{H}_i \neq 0, \quad 0 \leq i \leq L\). Later, it will become clear that this restriction on the \(\tilde{H}_i\) is not necessary.

It will be useful to combine the transforms defined in (3.22) and (3.24). The relations between the real time-domain variables \(b_k, \quad 0 \leq k < N\), and the real transform-domain variables \(B_i', \quad 0 \leq i < N\), may be written as the transform pair

\[
\begin{align*}
B_i' &= \begin{cases} 
     d_i \sum_{k=0}^{N-1} b_k \cos(2\pi ik/N), & 0 \leq i \leq L, \\
     d_i \sum_{k=1}^{N-1} b_k \sin(2\pi ik/N), & L < i < N,
\end{cases} \tag{3.26a} \\

b_k &= \frac{1}{N} \left( \sum_{i=0}^{L} d_i B_i' \cos(2\pi ki/N) + \sum_{i=L+1}^{N-1} d_i B_i' \sin(2\pi ki/N) \right), \quad 0 \leq k < N, \tag{3.26b}
\end{align*}
\]

where

\[
d_i = \begin{cases} 
     1, & i = \begin{cases} 
        0, & N \text{ odd} \\
        0 \text{ or } L, & N \text{ even}
\end{cases} \\
     \sqrt{2}, & \text{otherwise}.
\end{cases} \tag{3.26c}
\]

\[\text{1 The ensemble of any class of transform-domain random vectors } B_{[0, N-1]} = (R_0', R_1', ..., R_{N-1}') \text{ will be denoted subsequently by } B_{[0, N-1]}'.\]
Transform (3.26) is a form of the real discrete Fourier transform (RDFT) [ERS(85)]. Proofs of the following lemmas for transforms (3.24a), (3.26a), and (3.26b), are given in Appendix 3.A.

**Lemma 3.1:** Let \( U^{[0, N-1]} = \text{DFT}\{u^{[0, N-1]}\} \), where the components \( u_k \), \( 0 \leq k < N \), are real i.i.d. Gaussian random variables with mean zero and variance \( \sigma^2 \), and let \( B_l = C_l U_l \) be the components of the sequence \( B^{[0, N-1]} \), where the \( C_l \) are complex constants, and \( C_l = C_{N-l}^* \), \( 1 \leq i < N \). Then, application of transform (3.24a) to the complex subsequence \( B^{[0, L]} \) yields the real sequence \( B^{[0, N-1]} \), whose components \( B_i \) are independent Gaussian random variables with mean zero and variances \( N\sigma^2|C_i|^2 \), \( 0 \leq i < N \).

**Lemma 3.2:** Let \( b^{[0, N-1]} \) be a sequence whose components \( b_k \), \( 0 \leq k < N \), are real i.i.d. Gaussian random variables with mean zero and variance \( \sigma^2 \). Then, the components \( B_i \), \( 0 \leq i < N \), of the transform-domain sequence, \( B^{[0, N-1]} \), as obtained from (3.26a), are also real i.i.d. Gaussian random variables with mean zero and variance \( N\sigma^2 \).

**Lemma 3.3:** Let \( B^{[0, N-1]} \) be a sequence whose components \( B_i \), \( 0 \leq i < N \), are real independent Gaussian random variables with mean zero and variances \( N\sigma_i^2 \), and where \( \sigma_i = \sigma_{N-i} \), \( 1 \leq i < N \). Then, the components \( b_k \), \( 0 \leq k < N \), of the time-domain sequence \( b^{[0, N-1]} \), as obtained from (3.26b), are real correlated Gaussian random variables with mean zero and covariances \( \tilde{r}_n \), \( 0 \leq n < N \), given as

\[
\tilde{r}_n = \mathbb{E}[b_{k+n}b_k] = \frac{1}{N} \sum_{i=0}^{N-1} \sigma_i^2 \cos\left(\frac{2\pi ni}{N}\right), \quad 0 \leq k \leq k + n < N.
\]
Using Parseval's relation for the DFT [OPP-SCH(75), p. 125], one finds that the original block-energy constraint (3.7) becomes in the transform domain

\[
\sum_{i=0}^{N-1} E[X_i^2] \leq N^2 E_s. \tag{3.27}
\]

By Lemma 3.1, the \( V_i \) in (3.25) are statistically independent Gaussian random variables with mean zero and variance

\[
\sigma_i^2 = N(N_0/2)|\widetilde{H}_i|^{-2}, \quad 0 \leq i < N. \tag{3.28}
\]

Thus, it follows from (3.25) that the equivalent transform-domain channel model for the NCGC is a set of \( N \) parallel (discrete-time) memoryless additive Gaussian noise channels, where the channel inputs, \( X'_i, 0 \leq i < N, \) satisfy (3.27). This equivalence implies that

\[
\sup_{Q_X} I(X[0,N-1], \widetilde{Y}[0,N-1]) = \sup_{Q'_X} I(X'[0,N-1], \widetilde{Y}'[0,N-1]), \tag{3.29}
\]

where \( Q'_X \) is the class of probability densities for all \( X'[0,N-1] \) satisfying block-energy constraint (3.27). To write (3.29), 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 [GAL(68), p. 30]. Application of a theorem due to Gallager [GAL(68), Theorem 7.5.1] to this equivalent transform-domain channel model provides the solution for \( \widetilde{I}_N(E_s) \) in the form of the parametric expression

\[
\widetilde{I}_N(E_s) = \frac{1}{2N} \sum_{i=0}^{N-1} \log \left[ \max \left( \Theta |\widetilde{H}_i|^2, 1 \right) \right], \tag{3.30a}
\]

where the parameter \( \Theta \) is the solution of
\[
\sum_{i=0}^{N-1} \max \left( \Theta - |\tilde{H}_i|^{-2}, 0 \right) = 2NE_s/N_0. \tag{3.30b}
\]

\section*{B. Properties of the NCGC}

Solution (3.30) was obtained under the assumption that \(\tilde{H}_i \neq 0\), \(0 \leq i < N\). However, (3.30a) indicates that the \(i\)-th component channel does not contribute to \(\hat{l}_N(E_s)\) whenever \(\Theta |\tilde{H}_i|^2 \leq 1\); a condition which certainly holds when \(\tilde{H}_i = 0\). This implies that any optimal transmission scheme will not make use of those component channels for which \(\tilde{H}_i = 0\). Therefore, to include this case in the solution, the sum in (3.30b) is taken only over those \(i\) where \(\tilde{H}_i \neq 0\).

Theorem 7.5.1 in [GAL(68)] implies that \(I(X^{[0,N-1]}, Y^{[0,N-1]})\) in (3.29) is optimized by choosing \(Q_X\), such that the transform-domain inputs \(X_i^t\), \(0 \leq i < N\), are statistically independent Gaussian random variables with mean zero and variances \(E[X_i^t]^2 = N\varepsilon_i\), where the \(\varepsilon_i\) are given in (3.17b), i.e., (3.27) and thus also (3.7) holds with equality. Note that \(N\varepsilon_i\) is the average input energy that must be used in the \(i\)-th component channel. Since \(\tilde{H}_i = \tilde{H}_{N-i}^t\), \(1 \leq i < N\), it follows that \(\varepsilon_i = \varepsilon_{N-i}\), \(1 \leq i < N\), as can be seen from (3.17b).

Invoking Lemma 3.3 leads to the result that the optimizing \(q_X\) in (3.21) is such that the time-domain inputs \(x_k\), \(0 \leq k < N\), are statistically correlated Gaussian random variables with mean zero and covariances \(\tilde{r}_n\), \(0 \leq n < N\), as given in (3.17a). Moreover, from (3.17a) and with equality in (3.27), it follows immediately that \(\tilde{r}_0 = E[x_k^2] = E_s\), \(0 \leq k < N\). One may conclude that the optimizing \(q_X\) in (3.21) also satisfies the stronger constraint (3.13) with equality; thus, this \(q_X\) is also the optimizing \(q_X\) in (3.14), with the implication that \(\tilde{C}_N(E_s) = \hat{l}_N(E_s)\) for this \(q_X\). This completes the proof of Theorem 3.1.
3.4 Relations Between Channel Models

This section proves Corollary 3.1. Further, the case where the input symbols are chosen to be i.i.d. random variables is considered and it is shown that the resulting upper bound on the information rate is readily obtained using the same approach as for capacity.

In Section 3.1, it was stated that for the NCGC, in contrast to both the DTGC and the N-block DTGC, there is no need for an initializing input sequence. At the discrete-time instant \( k = 0 \) (when we begin to observe the output), the initial state of the basic channel model of the NCGC is represented by the input subsequence \( x^{[N-M,N-1]} \). This means that the linear convolution in (3.1) can be made into a circular convolution by choosing \( x_k = x_{N+k}, \, -M \leq k < 0 \). Conversely, by letting in (3.11) \( x_k = 0, \, N-M \leq k < N \), the circular convolution becomes a linear one. In the following, we shall frequently use these facts to establish fundamental relations between the three channel models introduced in Section 3.1.

A. Proof of Corollary 3.1

**Theorem 3.3:** The capacities of the NCGC and channels of the \( N \)-block DTGC type are related by

\[
\left[ 1 - \frac{M}{N} \right] \hat{C}_{N-M} \left( \frac{N}{N-M} E_s \right) \leq \tilde{C}_N (E_s) \leq \left[ 1 + \frac{M}{N} \right] \hat{C}_{N+M} (E_s). \quad (3.31)
\]

**Proof:** The lower bound in (3.31) is proved first. Defining

\[
\beta^m_n = \sum_{k=m}^{n} E[x_k^2], \quad n \geq m, \quad (3.32a)
\]

and using the result \( \tilde{C}_N (E_s) = \tilde{I}_N (E_s) \) obtained in Section 3.3, one obtains from definition (3.21)
\[
N \hat{C}_N(E_s) \geq \sup_{\left\{ \begin{array}{l}
\beta_0^{N-1} \leq NE_s \\
x_k = 0, \ N-M \leq k < N
\end{array} \right\}} I(X^{[0,N-1]}; \hat{Y}^{[0,N-1]})
\]

\[
= \sup_{\left\{ \begin{array}{l}
\beta_0^{N-M-1} \leq NE_s \\
x_k = 0, \ N-M \leq k < N
\end{array} \right\}} I(X^{[0,N-M-1]}; \hat{Y}^{[0,N-1]})
\]

\[
\geq \sup_{\left\{ \begin{array}{l}
\beta_0^{N-M-1} \leq NE_s \\
x_k = 0, \ N-M \leq k < N
\end{array} \right\}} I(X^{[0,N-M-1]}; \hat{Y}^{[0,N-M-1]})
\]

\[
= [N-M] \hat{C}_{N-M}(\frac{N}{N-M}E_s)
\]

where the suprema are over all probability densities, \(q_x\), satisfying the indicated constraints. The first inequality holds because an additional input constraint can only decrease average mutual information, and the first equality holds because circular convolution becomes linear convolution for this input constraint. The second inequality holds since information can only be lost if the received sample vector is truncated [GAL(68), pp. 16-27]. The last equality follows from definition (3.8).

The upper bound in (3.31) is proved by assuming (for notational convenience) that the first input digit is transmitted at time instant \(k = -M\), i.e., \(m = -M\) in (3.4). Thus,

\[
[N + M] \hat{C}_{N+M}(E_s) \geq [N + M] I_{N+M}(E_s)
\]

\[
\geq \sup_{\left\{ \begin{array}{l}
\mathbb{E}[x_k^2] \leq E_s, \ -M \leq k < N \\
x_k = x_{N+k}, \ -M \leq k < 0
\end{array} \right\}} I(X^{[-M,N-1]}; \hat{Y}^{[-M,N-1]})
\]
$$\geq \sup_{\{B[x_k^2] \leq E, 0 \leq k < N\}} I(x^{[0,N-1]}, \tilde{y}^{[0,N-1]})$$

$$= N\tilde{C}_N(E)$$  \quad (3.32c)$$

where the first two inequalities hold because stronger input constraints can only decrease average mutual information. The third inequality holds because of truncation of the input and output sequences, and because the constraint $x_k = x_{N+k}$, $-M \leq k < 0$, implies that circular and linear convolution coincide. The last equality follows from definition (3.14). From Theorem 3.3 and definition (3.9) one obtains immediately the following.

**Corollary 3.2:** The quantity $\hat{C}(E)$, defined as the limit as $N \to \infty$ of the capacity of the $N$-block DTGC, $\hat{C}_N(E)$, may be obtained as

$$\hat{C}(E) = \lim_{N \to \infty} \tilde{C}_N(E).$$  \quad (3.33)$$

**Lemma 3.4:** The optimized average block mutual information of the DTGC defined in (3.6b), and the capacity of the NCGC, $\tilde{C}_N(E)$, are related by

$$\left[1 - \frac{M}{N}\right] I_{N-M}\left(\frac{N}{N-M}E\right) \leq \tilde{C}_N(E) \leq \left[1 + \frac{M}{N}\right] I_{N+M}(E).$$  \quad (3.34)$$

The lower bound in (3.34) follows from (3.32b) and the obvious inequality $\hat{C}_N(E) \geq I_N(E)$; the upper bound is implicit in (3.32c). From Lemma 3.4 and definition (3.6) one obtains the following.

**Corollary 3.3:** The capacity of the DTGC, $C(E)$, may be obtained as
Corollary 3.1 in Section 3.2 now follows from Corollaries 3.2 and 3.3, and from definition (3.15).

B. Upper Bound on the Information Rate for Independent Average-Energy Constrained Inputs

The method introduced above to determine the capacity of the DTGC, $C(E_s)$, may be applied more generally. Further interesting and useful relations between channel models of the DTGC and the NCGC type may be derived. They yield an upper bound for (achievable) information rates in the case where the input symbols are chosen independently and are constrained in their average energy.

Let the inputs $x_k$, $0 \leq k < N$, be i.i.d. random variables, with mean zero and variance not larger than $E_s$, that have otherwise an arbitrary probability measure, $q_X$ (density in the case of continuous valued variables, distribution in the case of discrete variables). Call the resulting channel models DTGC-$q$ and NCGC-$q$, in order to distinguish them from the previously defined DTGC and NCGC, respectively. For the DTGC-$q$, define the information rate (in bits per channel input symbol)

$$I^q(E_s) = \lim_{N \to \infty} I_N^q(E_s) ,$$

(3.36a)

where the average block mutual information

$$I_N^q(E_s) = \frac{1}{N} I(X^0[N-1] ; Y^0[N-1]) ,$$

(3.36b)

with all $X^{[-M,-1]}$ as in (3.6c). Similarly, for the NCGC-$q$ define
3-20

\[ \tilde{I}_N^q(E_d) = \frac{1}{N} I(X^{[0,N-1]}; \tilde{Y}^{[0,N-1]}). \]  

**Theorem 3.4:** The average block mutual informations of the DTGC-q and the NCGC-q, as defined above, are related by

\[ \left[ 1 - \frac{M}{N} \right] I_{N-M}^q(E_d) \leq \tilde{I}_N^q(E_d) \leq \left[ 1 + \frac{M}{N} \right] I_{N+M}^q(E_d), \]  

(3.38a)

where

\[ \tilde{I}_N^q(E_d) \leq I_N^G(E_d) = \frac{1}{2N} \sum_{i=0}^{N-1} \log \left[ 1 + 2 \left( E_d / N_0 \right) \left| \tilde{H}_i \right|^2 \right]. \]  

(3.38b)

Furthermore, equality holds in (3.38b) if and only if \( q_X \) is the Gaussian density for i.i.d. random variables with mean zero and variance \( E_d \).

**Corollary 3.4:** The information rate of the DTGC-q, \( I^q(E_d) \), is obtained as

\[ I^q(E_d) = \lim_{N \to \infty} \tilde{I}_N^q(E_d), \]  

(3.39a)

and it is bounded above as

\[ I^q(E_d) \leq I^G(E_d) = \lim_{N \to \infty} \tilde{I}_N^G(E_d) \leq C(E_d), \]  

(3.39b)

where \( I^q(E_d) = I^G(E_d) \) if and only if \( q_X \) is the Gaussian density for i.i.d. random variables with mean zero and variance \( E_d \), as \( N \to \infty \).

Theorem 3.4 is proved in Appendix 3.B and Corollary 3.4 follows from (3.38a) and definition (3.36a). The last inequality in (3.39b) is obvious as it was assumed that the inputs to the DTGC-q are independent; for the DTGC, no such assumption was made. In fact, the capacity, \( C(E_d) \), is generally achieved
for correlated Gaussian inputs (Theorem 3.2), so that \( I^G(E_s) = C(E_s) \) if and only if the channel is memoryless \((M = 0)\). In the case where \( I^q(E_s) = I^G(E_s) \), i.e., when \( q_X \) is the Gaussian probability density with i.i.d. components, we call the DTGC-\( q \) (NCGC-\( q \)) more precisely the DTGC-G (NCGC-G). The expression for \( I^G(E_s) \) is derived in the following section.

### 3.5 Derivation of \( C(E_s) \) and \( I^G(E_s) \)

This section proves Theorem 3.2 (Section 3.2) and evaluates \( I^G(E_s) \) in (3.39). In Section 3.1, we introduced the hypothetical NCGC described within an \( N \)-dimensional space. The relations derived in Section 3.4 indicate, however, that the circularity restriction imposed becomes less and less important as \( N \) is increased. Thus, the infinite-dimensional generalization of the results obtained for the NCGC (NCGC-G) will yield the corresponding results for the DTGC (DTGC-G). This means that the NCGC (NCGC-G) and the DTGC (DTGC-G) become asymptotically equivalent as \( N \to \infty \).

From (3.16b) and the expression for the channel transfer function in (3.2), it follows that \( |\tilde{H}_i|^2 = |H(\lambda_i)|^2 \), \( 0 \leq i < N \), where \( \lambda_i = i\Delta\lambda_N \), \( 0 \leq i \leq L \), \( \lambda_i = (i - N)\Delta\lambda_N \), \( L < i < N \), and \( \Delta\lambda_N = 2\pi/N \). We shall make use of the following simple property of Riemann integrals.

**Lemma 3.5:** Let \( G(\cdot) \) be a continuous real-valued function. Then,

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{i=0}^{N-1} G(|\tilde{H}_i|^2) = \lim_{N \to \infty} \frac{1}{2\pi} \sum_{i=0}^{N-1} G(|H(\lambda_i)|^2) \Delta\lambda_N
\]

\[
= \frac{1}{2\pi} \int_{-\pi}^{\pi} G(|H(\lambda)|^2) d\lambda ,
\]

where \( H(\lambda) \) is defined in (3.2).
The first part of Theorem 3.2 is now proved by Corollary 3.3 and application of Lemma 3.5 to the parametric expressions in (3.16). Using the fact that 
\[ |H(\lambda)| = |H(-\lambda)|, \ |\lambda| \leq \pi, \]
leads to the final result in (3.19).

Assuming that \( E_\delta \) is finite, it can be shown that the real covariance sequence \( r_0(N-1) \) in (3.17) is positive definite [DOO(53), p. 473], thus confirming that the components of the input sequence, \( x_0(N-1) \), belong to a (wide-sense) stationary random process. The space over which this process is defined is \( N \)-dimensional. We may now generalize and consider the case where \( N \rightarrow \infty \). The covariances \( r_n = r_{-n}, \ 0 \leq n < \infty \), exist as the Fourier-Stieltjes coefficients of the spectral distribution function \( F(\lambda), \ |\lambda| \leq \pi, \) in the form

\[
r_n = E[x_{k+n}x_k] = \int_{-\pi}^{\pi} e^{jkn\lambda} dF(\lambda), \ -\infty < k < \infty, \quad (3.41)
\]

provided that \( F(\lambda) \geq 0 \) is absolutely continuous and non-decreasing in the interval \( |\lambda| \leq \pi, F(-\pi) = 0 \), and \( F(n) = r_0 < \infty \) [DOO(53), pp. 474-476], [HAN(67), pp. 8-18]. On the other hand, from (3.17) and Lemma 3.5, with \( \varepsilon_i = \varepsilon_{N-i}, 1 \leq i < N \), one finds

\[
\lim_{N \rightarrow \infty} \tilde{r}_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(\lambda)e^{j\pi\lambda} d\lambda \quad (3.42)
\]

where \( S_X(\lambda) \) is given in (3.20b). In order for the relation

\[
r_n = \lim_{N \rightarrow \infty} \tilde{r}_n \quad (3.43)
\]

to hold, it follows by comparison of (3.41) and (3.42), that \( F(\lambda) \) must be of the form

\[
F(\lambda) = \frac{1}{2\pi} \int_{-\pi}^{\pi} S_X(\nu)d\nu, \ |\lambda| \leq \pi. \quad (3.44)
\]
The right side of (3.44) represents $F(\lambda)$ as required by (3.41). In particular, $F(\lambda) \geq 0$, $F(-\pi) = 0$, and $F(\lambda)$ is nondecreasing since $S_x(\lambda) \geq 0$. Furthermore, using $\tilde{r}_0 = E_s$ (Theorem 3.1), it follows from (3.42) and (3.44) that

$$F(\pi) = \lim_{N \to \infty} \tilde{r}_0 = E_s.$$  \hspace{1cm} (3.45)

Therefore, (3.43) is valid, and (3.20a) is obtained from (3.42) and from the fact that $S_x(\lambda)$ is an even function. It follows that the capacity-achieving input process of the DTGC, $\{x_k\}$, $-\infty < k < \infty$, is zero-mean Gaussian having a continuous spectral distribution function $F(\lambda)$ with corresponding spectral density $S_x(\lambda)$. As a consequence, capacity is achieved when all inputs $x_k$, $-\infty < k < \infty$, have the same average energy, $E[x_k^2] = E_s$.

From (3.20b), we note that (3.19) has the well-known water-filling interpretation of capacity [GAL(68), p. 389]. For $C(E_s)$ to be finite, a constraint must be placed on $H(\lambda)$. Assuming $0 < \Theta < \infty$, and using $\ln[\max(z,1)] \leq \ln(1+z) \leq z$, $z \geq 0$, one obtains from (3.19a) the sufficient condition

$$\int_0^\pi |H(\lambda)|^2 d\lambda < \infty,$$  \hspace{1cm} (3.46)

which is equivalent to stating that the unit-sample response of the channel filter, $h[0,M]$, has finite energy. This completes the proof of Theorem 3.2.

Finally, invoking Lemma 3.5 yields from (3.38b) and Corollary 3.4 an expression for $I^G(E_s)$ in the form

$$I^G(E_s) = \frac{1}{2\pi} \int_0^\pi \log \left[ 1 + 2(E_s/N_0)|H(\lambda)|^2 \right] d\lambda.$$  \hspace{1cm} (3.47)
Corollary 3.5: The information rate of the DTGC-G, $I^G(E_s)$, is upper bounded as

$$I^G(E_s) \leq \frac{1}{2} \log \left[ 1 + 2 \frac{(E_s/N_o) \|h^{[0, M]}\|^2}{(E_s/N_o) \|h^{[0, M]}\|^2} \right], \quad (3.48a)$$

where the squared norm of $h^{[0, M]}$, $\|h^{[0, M]}\|^2$, represents the energy of the channel's unit-sample response. Equality holds in (3.48a) if and only if the channel is memoryless ($M = 0$). Furthermore, the bound in (3.48a) becomes tight for small values of $E_s/N_o$; in particular,

$$I^G(E_s) \approx \frac{1}{\ln 2} \frac{(E_s/N_o) \|h^{[0, M]}\|^2}{(E_s/N_o) \|h^{[0, M]}\|^2}, \quad E_s/N_o \ll 1, \quad (3.48b)$$

where the units are bits/symbol.

Proof: The squared magnitude of $H(\lambda)$ may be written as

$$|H(\lambda)|^2 = t_0 + 2 \sum_{k=1}^{M} t_k \cos(k\lambda), \quad (3.49a)$$

where the channel's autocorrelation coefficients are

$$t_k = \sum_{i=0}^{M} h_i h_{i-k}, \quad -M \leq k \leq M, \quad (3.49b)$$

and $t_k = t_{-k}$. Thus, from (3.47) and (3.49a) one obtains

$$I^G(E_s) = \frac{1}{2\pi} \int_0^\pi \log \left[ 1 + 2(E_s/N_o) \left\{ t_0 + 2 \sum_{k=1}^{M} t_k \cos(k\lambda) \right\} \right] d\lambda$$
where $T_0 = \left[1 + 2(E_s/N_o)t_0\right]$. By using the simple inequality $\ln(1 + x) \leq x$, $x > -1$, the last integral can be shown to be upper bounded by zero. Therefore, (3.48a) follows since $t_0 = \|h[0,M]\|^2$; similarly, (3.48b) follows. Assuming that $E_s/N_o > 0$, the statement on the equality in (3.48a) follows from the last line in (3.50) since the integral is identically zero if and only if $M = 0$. This completes the proof of Corollary 3.5.

### 3.6 Extension to Channels with Infinite Memory

The results obtained for channels with finite memory, $M$, generalize to channels with infinite memory. Consider the discrete-time Gaussian channel where the input sequence, $\{x_k\}$, produces the noise corrupted output sequence, $\{y_k\}$, given by

$$y_k = \sum_{i=0}^{\infty} a_i x_{k-i} + w_k, \quad -\infty < k < \infty,$$

(3.51)

where the infinite-length sequence $(a_0, a_1, ...)$, with $a_0 \neq 0$, is the causal unit-sample response of the equivalent channel filter. Assume further that

$$\sum_{i=0}^{\infty} |a_i| < \infty,$$

(3.52a)
so that the transfer function of this filter,

\[ A(\lambda) = \sum_{i=0}^{\infty} a_i e^{-ji\lambda}, \quad j = \sqrt{-1}, \quad (3.52b) \]

exists and is continuous. The noise samples in (3.51), \( w_k \), are i.i.d. Gaussian random variables with mean zero and variance \( \frac{N_0}{2} \), as defined in (3.3). The basic channel model specified in (3.51) is used such that the inputs satisfy the per symbol average energy constraint (3.5). We call this channel model the discrete-time Gaussian channel with infinite memory, or simply, the DTGC\(^\infty\), to distinguish it from the DTGC as defined in Section 3.1.

Let

\[ h_i = \begin{cases} a_i, & 0 \leq i \leq M \\ 0, & \text{otherwise} \end{cases} \quad (3.53) \]

where \( M \) is finite, and consider a channel having the causal unit-sample response \((h_0, h_1, ..., h_M)\). Define the capacity

\[ C(E_s, M) = C(E_s), \quad (3.54) \]

where \( C(E_s) \) is given in (3.19). Similarly, define the transfer function

\[ H(\lambda, M) = H(\lambda), \quad (3.55) \]

where \( H(\lambda) \) is given in (3.2). From (3.55), (3.53), and (3.52b), it follows that

\[ \lim_{M \to \infty} H(\lambda, M) = A(\lambda). \quad (3.56) \]

The capacity of the DTGC\(^\infty\), \( C^\infty(E_s) \), is then obtained as
where the limit of $C(E_s, M)$ exists provided that (3.46) holds in the limit of infinite $M$. In the present case, this condition holds true since (3.52a) and (3.53) assert that the $h_i$ are absolutely summable.\footnote{The infinite-length unit-sample response $(a_0, a_1, \ldots)$ has finite energy as a consequence of assumption (3.52a) since}

\[
\sum_{i=0}^{\infty} a_i^2 \leq \left[ \sum_{i=0}^{\infty} |a_i| \right]^2 < \infty,
\]

which in turn implies that

\[
\int_0^{\pi} |A(\lambda)|^2 d\lambda < \infty.
\]

It follows that $C^\infty(E_s)$ is given by (3.19) when $H(\lambda)$ is replaced by $A(\lambda)$. Similarly, the expression for $I^G(E_s)$ in (3.47) extends also to channels with infinite ISI memory. $A(\lambda)$ allows the representation of shift-invariant linear (causal) discrete-time systems. Such systems are described by a rational transfer function, i.e.,

\[
A(\lambda) = \frac{\sum_{m=0}^{M'} b_m e^{-jm\lambda}}{1 - \sum_{n=1}^{M''} c_n e^{-jn\lambda}},
\]

where $M'$ and $M''$ are finite [OPP-SCH(75), Chapter 4]. For $M'' > 0$, the unit-sample response of such a system is of infinite duration.
3.7 Numerical Results

The analytical results obtained in this chapter will be explored numerically. A simple example is considered in detail to illustrate these results and their interpretation.

Comparison between numerical results for different channels requires introduction of some normalization rule with respect to the channel responses. In Chapter 2, it was argued that a suitable normalization for the DTGC (DTGC-G) is the requirement that all channels have the same unit-sample response energy, \( \parallel h_{[0,M]} \parallel^2 \). Here, it will be assumed that all channels satisfy the constraint

\[
\parallel h_{[0,M]} \parallel^2 = 1. \tag{3.59}
\]

The channel responses used to generate the numerical results are defined in Figure 3.1, where the two channels with ISI memory \( M = 1 \) are equivalent with respect to capacity and achievable information rate. A similar dual relationship exists for any channel with \( M > 1 \); as a rule, the equivalent channel is obtained by changing the sign of every other component of \( h_{[0,M]} \) (see also Chapter 6, Section 6.4.2). In addition to the channels listed in Figure 3.1, the following channel will be examined in more detail.

**Example 3.1:** Consider the channel with \( M = 1 \), where \( h_{[0,1]} = (h_0, h_1) = (h_0, \alpha h_0) = h_0(1, \alpha) \), \(-\infty < \alpha < \infty\), and where \( \parallel h_{[0,1]} \parallel^2 = 1 \), i.e., \( h_0 = \sqrt{1/(1 + \alpha^2)} \). From (3.49), one obtains immediately

\[
|H(\lambda)|^2 = 1 + K_\alpha \cos \lambda, \tag{3.60a}
\]

where

\[
K_\alpha = \frac{2\alpha}{1 + \alpha^2}, \quad |K_\alpha| \leq 1. \tag{3.60b}
\]
Fig. 3.1. Definition of the channel unit-sample responses used to generate numerical results, $h^{[0,M]} = (h_0, h_1, ..., h_M)$, $\|h^{[0,M]}\|^2 = 1$, with ISI memory $M = 0, 1, 2, \text{ and } 6$. 
A. Capacity of the DTGC

Figure 3.2 indicates that for a sufficiently large block length, \( N \), the capacity of the NCGC, \( \tilde{C}_N(E_d) \), approximates closely the capacity of the DTGC, \( C(E_d) \). Therefore, if \( N \) is chosen large enough, e.g., \( N = 100 \) in Figure 3.2, it will be sufficient to evaluate \( \tilde{C}_N(E_d) \) given in (3.16), instead of using numerical integration to obtain \( C(E_d) \), as required by (3.19). In addition, the use of a fast Fourier transform (FFT) algorithm for the computation of the DFT of \( \hat{h}[0,N-1] \) in (3.16b) may further improve the computational efficiency [OPP-SCH(75), Chapter 6].

The capacity, \( C(E_d) \), of the channels defined in Figure 3.1 is shown in Figure 3.3. A striking feature of these results is that \( C(E_d) \) is not strictly upper bounded by the capacity of the memoryless channel (\( M = 0 \)). Rather, the capacity curve for the latter is crossed over by the capacity curves of all other channels with \( M > 0 \). This is a direct consequence of the chosen channel response normalization (3.59). The results which are shown in Figure 3.3 can be understood better through the following example.

Example 3.1 (continued): Assume that \( \alpha > 0 \), i.e., \( 0 < K_\alpha \leq 1 \) in (3.60a). In this case, \(|H(\lambda)|^{-2}, |\lambda| \leq \pi\), is monotonically increasing in \(|\lambda|\), so that the integrals in (3.19) must be evaluated in the interval \( 0 \leq \lambda \leq \lambda_\Theta \), where

\[
\lambda_\Theta = \begin{cases} 
0, & \Theta \leq \frac{1}{1 + K_\alpha} \\
\cos^{-1} \left( \frac{1 - \Theta}{\Theta K_\alpha} \right), & \frac{1}{1 + K_\alpha} < \Theta < \frac{1}{1 - K_\alpha} \\
\pi, & \Theta \geq \frac{1}{1 - K_\alpha}.
\end{cases}
\]

(3.61)

Here, \( \lambda_\Theta \) may be called the input signal bandwidth, or the channel utilization bandwidth. In terms of \( \Theta \) and \( \lambda_\Theta \), one obtains from (3.19a)
Fig. 3.2. Approximation of the capacity of the DTGC, \( C(E_s) \), by means of the capacity of the NCDC, \( \tilde{C}_N(E_s) \), for the channel \( h_{[0,1]} = (1, \pm 1)/\sqrt{2} \). For \( \Theta = 3.69 \), \( C(E_s) = 0.7825 \) bits/symbol, and \( E_s/N_0 = 1 \) (≈ 0 dB).
Fig. 3.3. Capacity of the DTGC, $C(E_s)$, as a function of signal-to-noise ratio, $E_s/N_o$, for the channels defined in Figure 3.1.
and the required signal-to-noise ratio, \( E_s/N_o \), is obtained from (3.19b) as

\[
E_s/N_o = \frac{\lambda_\Theta}{2\pi} \left\{ \Theta - \frac{(2/\lambda_\Theta)}{\sqrt{1 - K_\alpha^2}} \tan^{-1} \left( \frac{1 - K_\alpha}{1 + K_\alpha \tan(\lambda_\Theta/2)} \right), \quad 0 < K_\alpha < 1 \right\} \\
\Theta - (1/\lambda_\Theta) \tan(\lambda_\Theta/2), \quad K_\alpha = 1,
\]

where the transformation \( z = \tan(\lambda/2) \) was used to solve the particular integral in (3.19b) for the case \( 0 < K_\alpha < 1 \). The capacity-achieving input power spectral density in (3.20b), \( S_x(\lambda) \), becomes

\[
S_x(\lambda) = \frac{N_0}{2} \left[ \Theta - \frac{1}{1 + K_\alpha \cos \lambda} \right], \quad 0 \leq |\lambda| \leq \lambda_\Theta, \quad 0 < K_\alpha \leq 1.
\]

where \( \lambda_\Theta \) is such that the average energy per input symbol satisfies

\[
E_s = \frac{1}{\pi} \int_0^{\lambda_\Theta} S_x(\lambda) d\lambda.
\]

Figure 3.4 shows the even-symmetric function \( |H(\lambda)|^{-2} \) for the positive frequency interval, \( 0 \leq \lambda/\pi \leq 1 \), for the case \( K_\alpha = 1 \) (\( \alpha = 1 \)). Also indicated are the channel utilization bandwidth that corresponds to \( E_s/N_o = 1 \) \( (\Theta = 3.69) \), \( \lambda_\Theta = 0.76 \pi \), and the input power spectral density \( S_x(\lambda) \), which suggests the water-filling interpretation of the capacity result [GAL(68), p. 389].

The normalized input power spectral density for the case \( K_\alpha = 1 \), \( S_x(\lambda)/S_x(0) \), is shown in Figure 3.5 for different values of \( E_s/N_o \). It does
Fig. 3.4. The even-symmetric function $|H(\lambda)|^{-2}$, plotted over the positive frequency interval, $0 \leq \lambda/\pi \leq 1$, for the case $K_\alpha = 1$ ($\alpha = 1$) in Example 3.1. $\lambda_\Theta = 0.76\pi$ ($\Theta = 3.69$) corresponds to $E_s/N_o = 1$, and $S_x(\lambda)$ indicates the capacity-achieving input power spectral density as expressed by Eq. (3.62c).
Fig. 3.5. The normalized input power spectral density achieving capacity, \( S_x(\lambda)/S_x(0) \), for the case \( K_\alpha = 1 (\alpha = 1) \) in Example 3.1, plotted for different signal-to-noise ratios: \( E_s/N_0 = -25 \text{ dB} \) (\( \lambda_\Theta/\pi = 0.20 \)), \( E_s/N_0 = -10 \text{ dB} \) (\( \lambda_\Theta/\pi = 0.51 \)), \( E_s/N_0 = 0 \text{ dB} \) (\( \lambda_\Theta/\pi = 0.76 \)), \( E_s/N_0 = 10 \text{ dB} \) (\( \lambda_\Theta/\pi = 0.91 \)), \( E_s/N_0 = 30 \text{ dB} \) (\( \lambda_\Theta/\pi = 0.99 \)).
emphasize that the shape of \( S_x(\lambda), 0 \leq |\lambda| \leq \lambda_\Theta \), is distinct for each value of \( E_s/N_0 \). In particular, as \( E_s/N_0 \to \infty \), (3.62c) and (3.62d) indicate that \( S_x(\lambda) \to E_s \), since \( \Theta \to \infty \), and thus \( \lambda_\Theta \to \pi \). This means that
\[
C(E_s) \approx \frac{1}{2\pi} \int_0^\pi \log \left[ 1 + 2 \left( E_s/N_0 \right) (1 + K_\alpha \cos \lambda) \right] d\lambda , \quad E_s/N_0 >> 1 ,
\]
(3.63)
as can be seen from (3.62c) and (3.62a), i.e., \( C(E_s) \) approaches \( I^G(E_s) \), given in (3.47), asymptotically from above as \( E_s/N_0 \to \infty \). On the other hand, \( C(E_s) \) is upper bounded by the capacity of a hypothetical memoryless channel where the average symbol energy equals \( \max_\lambda |H(\lambda)|^2 E_s \), \( 0 \leq |\lambda| \leq \pi \), i.e., for the present example
\[
C(E_s) \leq \frac{1}{2} \log \left[ 1 + 2 \left( 1 + K_\alpha \right) E_s/N_0 \right] , \quad 0 < K_\alpha \leq 1.
\]
(3.64)
The bound in (3.64) is tight as \( E_s/N_0 \to 0 \). This may be verified as follows. From (3.62c) and (3.62d), one obtains
\[
\Theta \geq \frac{1}{1 + K_\alpha} + \frac{2\pi}{\lambda_\Theta} \left( E_s/N_0 \right) , \quad 0 < K_\alpha \leq 1 .
\]
(3.65)
Substituting (3.65) into (3.62a) yields the lower bound
\[
C(E_s) \geq \overline{C}(E_s) \approx \frac{1}{2\pi} \int_0^{\lambda_\Theta} \log \left[ \frac{1 + K_\alpha \cos \lambda}{1 + K_\alpha} + \frac{2\pi}{\lambda_\Theta} \left( 1 + K_\alpha \cos \lambda \right) E_s/N_0 \right] d\lambda ,
\]
\[
0 < K_\alpha \leq 1 .
\]
(3.66)
Since \( E_s/N_0 \to 0 \) implies \( \lambda_\Theta \to 0 \), the lower bound in (3.66) may be approximated as
where we have used the fact that \((\frac{E_s}{N_0})/\lambda_{\Theta} \to 0\) for \(E_s/N_0 \to 0\), as can be seen from (3.62b). Thus, the lower bound, \(\bar{C}(E_s)\), approaches asymptotically the upper bound in (3.64).

The asymptotic behaviour of \(C(E_s)\) is illustrated in Figure 3.6, where we have introduced the minimum signal-to-noise ratio per information bit, \(E_b/N_o\), that is required for (error-free) communication at rate \(R = C(E_bR)\), and \(R^G = I^G(E_bR^G)\), respectively. The numerical example in Figure 3.6 is given for \(K_a = 1\), i.e., when \(R \to 0\), the minimum value of \(E_b/N_o\) decreases to \(-4.6\, dB\), from the minimum value of \(-1.6\, dB\) required for the memoryless channel, as well as for communicating with i.i.d. Gaussian symbols.

### B. Information Rate \(I^G(E_s)\)

The information rate of the DTGC-G, \(I^G(E_s)\), is shown in Figure 3.7 for the channels defined in Figure 3.1. The tightness of the upper bound on \(I^G(E_s)\) for small values of \(E_s/N_o\), given in (3.48), is evident from these results. Comparison with Figure 3.3 confirms also that \(I^G(E_s)\) provides the lower asymptote for \(C(E_s)\) when \(E_s/N_o\) is allowed to increase without bound. The results of this section may be summarized as follows.

**Corollary 3.6:** The capacity, \(C(E_s)\), of a DTGC with unit-sample response \(h[0,M] = (h_0, h_1, ..., h_M)\), is bounded as
Fig. 3.6. Codrates $R = C(E_b R)$ and $R^G = I^G(E_b R^G)$, for the case $K_\alpha = 1$ ($\alpha = 1$) in Example 3.1, as a function of minimum required signal-to-noise ratio per information bit, $E_b / N_0$. The curve for the memoryless channel is shown for reference.
Fig. 3.7. Achievable information rate of the DTGC-G, $I^G(E_s)$, as a function of signal-to-noise ratio, $E_s/N_o$, for the channels defined in Figure 3.1.
where $H(\lambda)$ is defined in (3.2), and units are bits/symbol. The lower bound in (3.68) is asymptotically tight when $E_s/N_0 \to \infty$, and the upper bound in (3.68) is asymptotically tight when $E_s/N_0 \to 0$. Moreover, if all DTGC’s (including the memoryless channel) satisfy constraint (3.59), then the upper bound in (3.68) for DTGC’s with ISI memory $M \geq 1$ always exceeds the upper bound for the memoryless channel ($M = 0$), given as $(1/\ln 2) E_s/N_0$ (bits/symbol).
Proof of Lemma 3.1: We want to show that the real components $B_i^r$, $0 \leq i < N$, are Gaussian random variables with

$$
E[B_i^r] = 0, \quad \text{and} \quad E[B_i^r B_k^r] = N\sigma^2 |C_i|^2 \delta_{i-k}, \quad 0 \leq k < N. \quad (3.A.1)
$$

It follows from (3.22a) that the complex $B_i = C_i U_i$, $0 \leq i < N$, are a weighted sum of the $u_k$, $0 \leq k < N$, which are (by definition) real i.i.d. zero-mean Gaussian random variables. Therefore, the $B_i$ obtained from (3.24a) are also zero-mean Gaussian. In order to prove the covariances in (3.A.1), it is required to evaluate the expectations $E[B_i^R B_k^R]$, $E[B_i^R B_k^I]$, and $E[B_i^I B_k^I]$, $0 \leq i \leq L$, $0 \leq k \leq L$. First, form the products $B_i^R B_k^R$, $B_i^R B_k^I$, and $B_i^I B_k^I$, by substituting $B_i^* = (B_i^R + B_i^I)/2$, and $B_i^I = -j(B_i^R - B_i^I)/2$, to obtain

$$
B_i^R B_k^R = \left( B_i B_k^* + B_i^* B_k + B_i^I B_k^I + \text{c.c.} \right)/4, \quad (3.A.2a)
$$

$$
B_i^R B_k^I = -j \left( B_i B_k - B_i^* B_k^* + B_i^I B_k^I - B_i^I B_k^* \right)/4, \quad (3.A.2b)
$$

$$
B_i^I B_k^I = - \left( B_i B_k - B_i^* B_k^* - B_i^I B_k^I + B_i^I B_k^* \right)/4. \quad (3.A.2c)
$$

Since $E[B_i^* B_k^*] = (E[B_i B_k])^*$, and $E[B_i^* B_k] = (E[B_i B_k])^*$, the expectations of the products in (3.A.2) can be expressed in terms of $E[B_i B_k]$, and $E[B_i^* B_k^*]$,

$$
E[B_i B_k] = (C_i C_k) E[U_i U_k]
$$

$$
= N\sigma^2 |C_i|^2 \delta_{i-k}, \quad i = \begin{cases} 0, & N \text{ odd} \\ 0 \text{ or } L, & N \text{ even} \end{cases}
$$

and
\[ E[B_i B_k^*] = (C_i C_k^*) E[U_i U_k^*] \]
\[ = N \sigma^2 |C_i|^2 \delta_{i-k}, \quad 0 \leq i \leq L. \]

In (3.A.3), we have used \( E[u_m u_n] = \sigma^2 \delta_{m-n} \), to obtain

\[ E[U_i U_k] = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E[u_m u_n] \Omega_N^{-im-kn} \]
\[ = \sigma^2 \sum_{m=0}^{N-1} \Omega_N^{-m(i+k)} = \begin{cases} N \sigma^2, & i + k = 0 \mod N \\ 0, & \text{otherwise} \end{cases} \]

and

\[ E[U_i U_k^*] = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E[u_m u_n] \Omega_N^{-im+kn} \]
\[ = \sigma^2 \sum_{m=0}^{N-1} \Omega_N^{-m(i-k)} = N \sigma^2 \delta_{i-k}, \quad 0 \leq i \leq L. \]

Equation (3.A.4) is a consequence of the orthogonality relationships between powers of the exponential \( \Omega_N \) [OPP-SCH(75), p. 88]. Taking expectation on both sides of (3.A.2), and using (3.A.3), gives

\[ E[B_i^R B_k^R] = \begin{cases} N \sigma^2 |C_i|^2 \delta_{i-k}, & i = \begin{cases} 0, & N \text{ odd} \\ 0 \text{ or } L, & N \text{ even} \end{cases} \\ (N/2) \sigma^2 |C_i|^2 \delta_{i-k}, & 1 \leq i \leq \begin{cases} L, & N \text{ odd} \\ L - 1, & N \text{ even} \end{cases} \end{cases} \]

\[ E[B_i^I B_k^I] = (N/2) \sigma^2 |C_i|^2 \delta_{i-k}, \quad 1 \leq i \leq \begin{cases} L, & N \text{ odd} \\ L - 1, & N \text{ even} \end{cases} \]

\[ E[B_i^L B_k^L] = \begin{cases} N \sigma^2 |C_i|^2 \delta_{i-k}, & 0 \leq i \leq L \end{cases} \]
\[ \mathbb{E}[B^R_i B^L_k] = 0, \quad 0 \leq i \leq L, \quad 0 \leq k \leq L. \quad (3.A.5c) \]

From (3.24a) and (3.A.5), it follows that the covariances \( \mathbb{E}[B^R_i B^L_k] \) are indeed as given in (3.A.1). The \( B^R_i, \quad 0 \leq i < N \), are therefore uncorrelated, and since they are Gaussian, they are also independent.

**Proof of Lemma 3.2:** Since transform (3.26) is equivalent to the combination of transforms (3.22) and (3.24), Lemma 3.2 is a special case of Lemma 3.1. The proof of Lemma 3.2 is obtained by replacing \( u[0,N-1] \) and \( u'[0,N-1] \), in the proof of Lemma 3.1 by \( b[0,N-1] \) and \( B[0,N-1] \), respectively, and letting \( C_i = 1, \quad 0 \leq i < N \).

**Proof of Lemma 3.3:** By definition, the \( B^R_i, \quad 0 \leq i < N \), are statistically independent Gaussian random variables with

\[ \mathbb{E}[B^R_i] = 0, \quad \text{and} \quad \mathbb{E}[B^R_i B^L_k] = N \sigma_i^2 \delta_{i,-k}. \quad (3.A.6) \]

According to (3.26b), the \( b_k, \quad 0 \leq k < N \), are obtained from a weighted sum of zero-mean Gaussian random variables; therefore, the \( b_k \) are also zero-mean Gaussian. The covariances are obtained from (3.26b) as

\[
\mathbb{E}[b_k b_m] = \frac{1}{N^2} \left\{ \sum_{i=0}^{L} \sum_{n=0}^{L} d_i d_n \mathbb{E}[B^R_i B^L_n] \cos(2\pi ki/N) \cos(2\pi mn/N) \\
+ \sum_{i=0}^{L} \sum_{n=L+1}^{N-1} d_i d_n \mathbb{E}[B^R_i B^L_n] \cos(2\pi ki/N) \sin(2\pi mn/N) \right\}
\]
\[
\begin{align*}
&+ \sum_{i=L+1}^{N-1} \sum_{n=0}^{L} d_i d_n \mathbb{E}[B_i B_n'] \sin(2\pi ki/N) \cos(2\pi mn/N) \\
&+ \sum_{i=L+1}^{N-1} \sum_{n=L+1}^{N-1} d_i d_n \mathbb{E}[B_i B_n'] \sin(2\pi ki/N) \sin(2\pi mn/N) \\
= \frac{1}{N} \left\{ \sum_{i=0}^{L} d_i^2 \sigma_i^2 \cos(2\pi ki/N) \cos(2\pi mi/N) \right. \\
&\left. + \sum_{i=L+1}^{N-1} d_i^2 \sigma_i^2 \sin(2\pi ki/N) \sin(2\pi mi/N) \right\}.
\end{align*}
\]

For \( N \) odd, one finds from (3.A.7) and (3.26c), with \( \sigma_i = \sigma_{N-i} \), \( 1 \leq i < N \),

\[
\mathbb{E}[b_k b_m] = \frac{1}{N} \left\{ \sigma_0^2 + 2 \sum_{i=1}^{L} \sigma_i^2 \left[ \cos(2\pi ki/N) \cos(2\pi mi/N) + \sin(2\pi ki/N) \sin(2\pi mi/N) \right] \right\}
\]

\[
= \frac{1}{N} \left\{ \sigma_0^2 + 2 \sum_{i=1}^{L} \sigma_i^2 \cos[2\pi (k - m)i/N] \right\}
\]

\[
= \frac{1}{N} \sum_{i=0}^{N-1} \sigma_i^2 \cos[2\pi (k - m)i/N], \ 0 \leq k < N, \ 0 \leq m < N.
\]

(3.A.8)

For \( N \) even, the last equality is obtained similarly. The covariance matrix is completely specified by \( N \) different covariances since \( \mathbb{E}[b_k b_m] \) depends only on the absolute (time) difference \( |k - m| \) (the covariance matrix is circular, i.e., its first row completely specifies it). Letting \( m = k + n \) in (3.A.8), and defining the covariances \( \tilde{r}_n = \mathbb{E}[b_{k+n} b_k] \), \( 0 \leq n < N \), completes the proof of Lemma 3.3.
APPENDIX 3.B

Proof of Theorem 3.4: From definitions (3.37) and (3.36b), one obtains the lower bound of (3.38a) by using arguments very similar to those used in the development of (3.31) of Theorem 3.3, i.e.,

\begin{align}
N \tilde{I}_N^q(E_d) &\geq \{ I(X[0,N-1]; \gamma[0,N-1]): x_k = 0, \ N - M \leq k < N \} \\
&= I(X[0,N-M-1]; \gamma[0,N-1]) \\
&\geq I(X[0,N-M-1]; Y[0,N-M-1]) \\
&= [N - M] \tilde{I}_{N-M}^q(E_d).
\end{align}

(3.B.1a)

The upper bound of (3.38a) is proved similarly by using (3.4) with \( m = -M \).

Thus,

\begin{align}
[N + M] \tilde{I}_{N+M}^q(E_d) &\geq \{ I(X[-M,N-1]; \gamma[-M,N-1]): x_k = x_{N+k}, \ -M \leq k < 0 \} \\
&\geq \{ I(X[0,N-1]; \gamma[0,N-1]): x_k = x_{N+k}, \ -M \leq k < 0 \} \\
&= I(X[0,N-1]; \gamma[0,N-1]) \\
&= N \tilde{I}_N^q(E_d).
\end{align}

(3.B.1b)

and the equivalent transform-domain channel model for the NCGC-q is a set of \( N \) parallel (discrete-time) memoryless additive Gaussian noise channels, each having uncorrelated zero-mean inputs of variance not larger than \( NE_d \). This follows from the proof of Lemma 3.1 in Appendix 3.A by letting \( \sigma^2 \leq E_d \) and not assuming that \( q_x \) is Gaussian. Thus, while the transform-domain inputs \( X'_i, 0 \leq i < N, \) are uncorrelated, they are in general not independent unless
3.B-2

$q_X$ is Gaussian. In any case, energy constraint (3.27) is satisfied and the additive noise in the $i$-th channel is independent zero-mean Gaussian having variance $N(N_0/2)|\tilde{H}_i|^{-2}$, as previously shown in proving (3.28). It follows that (see, e.g., [GAL(68), p. 345])

$$I(X^i[0,N-1], Y^i[0,N-1]) \leq \sum_{i=0}^{N-1} I(X^i_i; Y^i_i), \quad (3.B.2a)$$

where equality holds for independent $X^i_i$. Let $E[X^2_i] = N\varepsilon_i$, $0 \leq i < N$, so that $\varepsilon_i \leq E_s$, $0 \leq i < N$. It follows that [GAL(68), Theorem 7.4.2]

$$\sum_{i=0}^{N-1} I(X^i_i; Y^i_i) \leq \frac{1}{2} \sum_{i=0}^{N-1} \log \left[ 1 + \frac{2\varepsilon_i}{N_0|\tilde{H}_i|^2} \right], \quad (3.B.2b)$$

with equality if and only if the inputs $X^i_i$ are Gaussian. The right side of (3.B.2b) is a convex-$\cap$ function over the $\varepsilon_i$ and it is maximized when $\varepsilon_i = E_s$, $0 \leq i < N$, i.e., all inputs must use the maximum allowed average energy.

Combining (3.B.2a) and (3.B.2b), and using definition (3.37), yields (3.38b). Clearly, component channels where $\tilde{H}_i = 0$ do not contribute to $\tilde{H}^G_N(E_s)$; the input energy for these component channels is thus lost.
CHAPTER 4

Information Rate with a Binary Input Constraint

We consider the discrete-time channel model shown in Figure 4.1. The input symbols, \( x_k \), are independently selected with equal probability from a binary alphabet, i.e., \( x_k \in \{-\sqrt{E_s}, +\sqrt{E_s}\} \), where \( E_s \) is the energy per input symbol. The channel is characterized by the unit-sample response \( h^{M+1} = (h_0, h_1, \ldots, h_M) \), where \( h_0 h_M \neq 0 \), and \( M \) is the finite channel memory. The noise samples, \( w_k \), are independent and identically distributed (i.i.d.) Gaussian random variables with mean zero and variance \( N_0/2 \). This channel model operates as follows. Assume that the inputs \( x_k, k < 0 \), have been transmitted and that we now wish to transmit the \( N \)-block symbol sequence \( x^N = (x_0, x_1, \ldots, x_{N-1}) \), \( N \geq 1 \). Define the channel state at time instant \( k \), \( s_k \), as

\[
s_k = (x_{k-1}, x_{k-2}, \ldots, x_{k-M}) \tag{4.1}
\]

so that \( s_0 = (x_{-1}, x_{-2}, \ldots, x_{-M}) \). The samples of the channel output sequence, \( \{y_k: 0 \leq k < \infty\} \), are given as

\[
y_k = \sum_{i=0}^{M} h_i x_{k-i} + w_k \tag{4.2}
\]

\[
= \sum_{i=0}^{\min(M,k)} h_i x_{k-i} + \sum_{i=\min(M,k)+1}^{M} h_i x_{k-i} + w_k.
\]

For \( 0 \leq k < M \), the last sum in (4.2) is the contribution to \( y_k \) of the input symbols that determine the initial state, \( s_0 \); this sum vanishes for \( M \leq k < \infty \). Clearly, the input symbols that represent \( s_0 \) do not add any
Fig. 4.1. Discrete-time Gaussian channel with a binary input constraint.
information about the independent input block, \( x^N \), to that given by the output samples with \( k \geq 0 \). Without loss of generality, one may thus define \( s_0 = (0, 0, \ldots, 0) \), or equivalently, \( x_k = 0, k < 0 \). Because of the finite memory \( M \) of the channel and the fact that the input symbols are independent, the output samples for \( k \geq N + M \) also carry no information about \( x^N \). Therefore, with respect to information about the input block \( x^N \), the relevant output samples of the channel in Figure 4.1 are

\[
y_k = v_k + w_k, \quad 0 \leq k < N + M, \tag{4.3a}
\]

where, with \( P = \max(0, k - N + 1) \), and \( Q = \min(M, k) \), the signal sample \( v_k \) is given by

\[
v_k = \sum_{i=P}^{Q} h_i x_{k-i}. \tag{4.3b}
\]

For this channel model, we define the symmetric information rate

\[
i^b(E_s) = \lim_{N \to \infty} I^b_N(E_s), \tag{4.4a}
\]

where the superscript \( b \) denotes i.i.d., equally-likely binary inputs, and where the quantity \( I^b_N(E_s) \) is the \( N \)-block average mutual information

\[
i^b_N(E_s) = \frac{1}{N} I(X^N; Y^N). \tag{4.4b}
\]

---

1 See Chapter 2 or [LEU(76)] for an explanation of the term symmetric.
According to Gallager [GAL(68), Section 2.4], the average mutual information, \( I(X^N;Y^N) \), between the ensemble of binary input sequences, \( X^N \), and the ensemble of amplitude-continuous output sequences, \( Y^N \), is given by

\[
I(X^N;Y^N) = \sum_{i=0}^{K-1} \int Q_X(x_i^N) p_{Y|X}(y^N|x_i^N) \log \frac{p_{Y|X}(y^N|x_i^N)}{p_Y(y^N)} dy^N, \quad K = 2^N. \tag{4.5a}
\]

Each dimension of the \( N \)-dimensional integral in (4.5a) is evaluated over the interval \((-\infty, +\infty)\), and the unconditional joint probability density is given by

\[
p_Y(y^N) = \sum_{k=0}^{K-1} Q_X(x_k^N) p_{Y|X}(y^N|x_k^N). \tag{4.5b}
\]

For the present i.i.d. case,

\[
Q_X(x_i^N) = \frac{1}{K} = 2^{-N}, \quad 0 < i < K, \tag{4.6}
\]

and the conditional joint probability density of the output block is

\[
p_{Y|X}(y^N|x_i^N) = p_{Y|V}(y^N|v_i^N) = p_w(y^N - v_i^N), \quad 0 < i < K, \tag{4.7a}
\]

where the joint probability density of the noise sample sequence, \( w^N = (w_0, w_1, ..., w_{N-1}) \), is given by

\[
p_w(w^N) = \prod_{k=0}^{N-1} \frac{1}{\sqrt{\pi N_0}} \exp\left(-\frac{w_k^2}{N_0}\right). \tag{4.7b}
\]

A justification of definition (4.4) is given in Appendix 4.A, which shows that \( I_N^b(E_d) \leq I^h(E_d), \quad N \geq 0 \), so that \( I_N^b(E_d) \) is always a lower bound for \( I^h(E_d) \). Note that definition (4.4) ignores those output samples \( y_k, \quad N < k < N + M \),
which contain additional information about the last $M$ symbols of the input $x^N$. Since $M$ is finite, the effect of this truncation vanishes as $N \to \infty$ (see also Corollary 4.4). In (4.7a), the output sample sequence, $y^N_i = (v_0^i, v_1^i, \ldots, v_{N-1}^i)$, is the uniquely determined response of the channel in Figure 4.1 in the absence of noise, given that the input sequence $x^N_i$ is sent and that the initial channel state $s_0 = 0$.

In the following, an expression is derived for $I^b_N(E_s)$. It follows from this result that there exists, unfortunately, no simple "closed-form" solution for $I^b(E_s)$. As a consequence, Monte-Carlo techniques will be used to estimate $I^b_N(E_s)$, and thus to estimate $I^b(E_s)$.

4.1 Derivation of $I^b_N(E_s)$

An expression for $I^b_N(E_s)$, as defined in (4.4), is now obtained as follows. First, let

$$d^N_{mn} = \frac{1}{\sqrt{E_s}} (v^N_m - v^N_n), \quad 0 \leq m, n < K,$$

be the normalized difference in the $N$-dimensional Euclidean space, $\mathbb{R}^N$, between two arbitrary signal sequences, $v^N_m$ and $v^N_n$. Furthermore, define the inner product between any two sequences $a^N$ and $b^N$ as

$$\langle a^N, b^N \rangle = \sum_{k=0}^{N-1} a_k^* b_k ,$$

where $^*$ denotes complex-conjugate, and define the square of the norm of $a^N$ as

$$\|a^N\|^2 = \langle a^N, a^N \rangle = \sum_{k=0}^{N-1} |a_k|^2.$$
Theorem 4.1: An expression for the $N$-block average mutual information, $I_N^b(E_o)$, as defined in (4.4b), is given by

$$I_N^b(E_o) = 1 - \frac{1}{N} \int p_z(z^N) g(z^N) dz^N,$$  \hspace{1cm} (4.10a)

where each dimension of the $N$-dimensional integral is evaluated over the interval $(-\infty, +\infty)$ and where the units are bits/symbol. The dummy variable in (4.10a), $z^N$, is a sequence of i.i.d. Gaussian random variables with mean zero and unit variance, i.e.,

$$p_{z^N}(z^N) = \frac{1}{(2\pi)^{N/2}} \exp\left(-\frac{\|z^N\|^2}{2}\right), \hspace{1cm} (4.10b)$$

and the function

$$g(z^N) = \frac{1}{K} \sum_{i=0}^{K-1} \log_2 \left[ \sum_{k=0}^{K-1} \exp\left\{ -\sqrt{2E_s/N_o} \langle z^N, d_{ik}^N \rangle - (E_s/N_o) \|d_{ik}^N\|^2 \right\} \right], \hspace{1cm} (4.10c)$$

where $K = 2^N$.

Proof: Substitution of (4.6) and (4.7a) into (4.5) yields (4.5a) in the form

$$I(X^N; Y^N) = \frac{\log K}{K} \sum_{i=0}^{K-1} \int p_W(y^N - v_i^N) dy^N \hspace{1cm} (4.11)$$

The first term in (4.11) evaluates simply to $\log K = N \log 2$. Performing a transformation of variables in the integral of the second term of (4.11) with
\[ z^N = (y^N - v_t^N) / \sqrt{N_o/2} \] so that \[ dy^N = (N_o/2)^N dz^N, \] and using (4.7b) and (4.9b) together with the identity

\[ \|a^N + b^N\|^2 = \|a^N\|^2 + 2\langle a^N, b^N \rangle + \|b^N\|^2, \]

gives

\[ I(X^N; Y^N) = N \log 2 - \frac{1}{K} \sum_{i=0}^{K-1} \int p_Z(z^N) \times \]

\[ \times \log \left[ \sum_{k=0}^{K-1} \exp \left\{ -\sqrt{2E_s/N_o} \langle z^N, d_{ik}^N \rangle - (E_s/N_o) \|d_{ik}\|^2 \right\} \right] dz^N, \]

(4.12)

where \( p_Z(z^N) \) is given in (4.10b), and \( d_{ik}^N \) is defined in (4.8). Interchanging the order of summation and integration in (4.12), defining \( g(z^N) \) as in (4.10c), and taking the logarithm with the base 2, yields \( I^b_N(E_s) \) in the form of (4.10a). This completes the proof of Theorem 4.1.

4.2 Properties of \( I^b(E_s) \)

The symmetric information rate, \( I^b(E_s) \), serves as a lower bound for the corresponding asymptotic capacity of the channel in Figure 4.1, \( C^b(E_s) \), i.e.,

\[ I^b(E_s) \leq C^b(E_s) = \lim_{N \to \infty} C^b_N(E_s), \]

(4.13a)

where the \( N \)-block capacity is defined by

\[ C^b_N(E_s) = \frac{1}{N} \sup_{Q_X} I(X^N; Y^N), \]

(4.13b)

1 Equation (4.10c) confirms our assertion that \( I^b_N(E_s) \) is independent of the initial channel state, \( s_0 \), because \( g(z^N) \) depends only on all possible differences between the noiseless channel output sequences, \( v_m^N \) and \( v_n^N \), \( 0 \leq m, n < K \).
with the supremum taken over all joint probability distributions, $Q_X(.)$, for the binary sequence ensemble, $X^N$. The inequality in (4.13a) is a direct consequence of definition (4.4) which assumes $Q_X(.)$ as in (4.6).

**Corollary 4.1:** Equality holds in (4.13a) if and only if the channel is memoryless ($M = 0$). The optimizing $Q_X(.)$ in (4.13b) is then given by (4.6) for all $N \geq 1$, and the capacity becomes

$$
C^b(E_s)
= 1 - \int_{-\infty}^{+\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \log_2 \left[ 1 + e^{-\sqrt{8E_s/N_0} z - 4E_s/N_0} \right] dz,
$$

where the units are bits/symbol.

**Proof:** We first prove the identity

$$
C_N^b(E_s)_{M=0, h_0=1} = C_l^b(E_s)_{M=0, h_0=1}, \quad N \geq 1,
$$

which implies, because of definition (4.13a), that

$$
C^b(E_s)_{M=0, h_0=1} = C_l^b(E_s)_{M=0, h_0=1}.
$$

Let $X^N$ be the input ensemble achieving $C_N^b(E_s)$, i.e., $I(X^N;Y^N) = NC_N^b(E_s)$, where $Y^N$ is the corresponding output ensemble. For any memoryless channel,

$$
I(X^N;Y^N) \leq \sum_{i=0}^{N-1} I(X_i;Y_i), \quad N \geq 1,
$$
where equality holds if the inputs $x_i \in X_i$, $0 \leq i < N$, are statistically independent [FAN(61), pp. 125/142]. By definition (4.13b), the right side of (4.16) is at most equal to $NC^b(E_d)$, with equality if and only if each $Q_{x_i}()$, $0 \leq i < N$, achieves the maximum of $I(X_i;Y_i)$, i.e., $C^b_N(E_d) \leq C^b(E_d)$. For the converse, let $X$ be the (one-dimensional) ensemble achieving $C^b_1(E_d)$, i.e., $Q_X()$ is such that $I(X;Y) = C^b_1(E_d)$, where $Y$ is the corresponding (one-dimensional) output ensemble. Now, let $X^N$ be the input ensemble where the components $x_i$, $0 \leq i < N$, are i.i.d. random variables with $Q_X()$ achieving $C^b_1(E_d)$, and let $Y^N$ be the corresponding output ensemble. Also, for any $X^N$ whose components are independent,

$$I(X^N;Y^N) \geq \sum_{i=0}^{N-1} I(X_i;Y_i), \quad N \geq 1,$$

where equality holds if the channel is memoryless [McE(77), pp. 31-43]. Thus, by definition (4.13b), the left side of (4.17) is at most equal to $NC^b_N(E_d)$, with equality if and only if $[Q_X()]^N$ achieves the maximum of $I(X^N;Y^N)$, i.e., $C^b_N(E_d) \geq C^b_1(E_d)$. From the above, it then follows that $C^b_N(E_d) = C^b_1(E_d)$ if and only if the channel is memoryless and the inputs are i.i.d. random variables (the channel is assumed to be stationary), i.e., in (4.13b),

$$Q_X(x_i^N) = {\left[Q_X(x)\right]}^N, \quad 0 \leq i < K,$$

which establishes (4.15a). For output-symmetric memoryless channels with binary inputs, $Q_X(+\sqrt{E_x}) = Q_X(-\sqrt{E_x}) = 1/2$ are the input probabilities achieving $C^b(E_d)$ [VIT-OMU(79), p. 141]. Therefore, we conclude that, for this channel, $C^b(E_d)$ is achieved for all $N \geq 1$ with $Q_X()$ as specified in (4.6). Hence, (4.13a) holds with equality for the memoryless channel, i.e., $I^b(E_d) =$
$C^b(E_s) = I_1^b(E_s)$, and (4.14) follows when evaluating (4.10) for $N = 1$, $M = 0$, and $h_0 = 1$. This completes the proof of Corollary 4.1.

Corollary 4.2: For small values of $E_s/N_o$ and provided that the energy of the unit-sample response of the channel is finite, i.e., $\|h^{M+1}\|^2 < \infty$, the symmetric information rate defined in (4.4), $I^b(E_s)$, is approximately given by

$$I^b(E_s) \approx \frac{1}{\ln 2} \left( E_s/N_o \right) \|h^{M+1}\|^2$$

(4.19)

$$\approx 1.443 \left( E_s/N_o \right) \|h^{M+1}\|^2, \quad E_s/N_o << 1,$$

where the units are bits/symbol.

Proof: The proof is straightforward but lengthy (some derivations are given in Appendix 4.B). Let

$$f_i(\alpha, z^N) = \sum_{k=0}^{K-1} \exp \left( -\sqrt{2} \alpha \langle z^N, d_{ik}^N \rangle - \alpha^2 \|d_{ik}^N\|^2 \right)$$

(4.20a)

where

$$\alpha^2 = E_s/N_o,$$

(4.20b)

and expand the function

$$F_i(\alpha, z^N) = \ln \left[ f_i(\alpha, z^N) \right]$$

(4.20c)

in a Taylor series about $\alpha = 0$ [THO(65), pp. 785-800]. Thus,

---

1 As pointed out by Gallager [GAI(68) p. 76], one should not conclude from this result that statistical dependence between input symbols should be avoided when designing reliable communication systems. Actual coding schemes usually introduce (or require) some statistical dependence between transmitted symbols to be effective.
\[ F_i(\alpha, z^N) = \ln[f_i(0, z^N)] + \left[ \frac{f_i''(0, z^N)}{f_i(0, z^N)} \right] \alpha + \left[ \frac{f_i''(0, z^N)}{f_i(0, z^N)} - \left[ \frac{f_i'(0, z^N)}{f_i(0, z^N)} \right]^2 \right] \frac{\alpha^2}{2} + R_i(\alpha, z^N), \]  

where

\[ f_i(0, z^N) = K = 2^N, \]

\[ f_i'(0, z^N) = -\sqrt{2} \sum_{k=0}^{K-1} \langle z^N, d_{ik}^N \rangle, \]

\[ f_i''(0, z^N) = 2 \sum_{k=0}^{K-1} \left[ \langle z^N, d_{ik}^N \rangle^2 - \|d_{ik}^N\|^2 \right], \]

and the Lagrange remainder

\[ R_i(\alpha, z^N) = F_i'''(\theta, z^N) \frac{\alpha^3}{6}, \]

for some number \( \theta, 0 < \theta < \alpha \). Substituting (4.21) into (4.10) of Theorem 4.1 and integrating as shown in Appendix 4.B, one obtains

\[ l_N^b(\alpha) = \frac{1}{N} \left[ \frac{\alpha^2}{\ln 2} \sum_{m=0}^{N-1} \bar{\delta}_m + c_\alpha \alpha^3 \right], \]

where \( \bar{\delta}_m \) is the expectation of the product \( d_{ik}^m d_{ir}^m \) over \( 0 \leq i, k, r < K \), i.e.,

\[ F_i(\alpha, z^N) \) is continuous on \( 0 \leq \alpha < \infty \), and all its derivatives with respect to \( \alpha \) exist and are also continuous on \( 0 \leq \alpha < \infty \).
\[
\tilde{\delta}_m = \frac{1}{K^3} \sum_{i=0}^{K-1} \sum_{k=0}^{K-1} \sum_{r=0}^{K-1} d^i_{mk} d^k_{mr}, \quad (4.22b)
\]

with
\[
d^i_{mk} = \frac{1}{\sqrt{E_s}} \sum_{t=p}^{Q} h_t (x^i_{m-t} - x^j_{m-t}) \quad (4.22c)
\]
and
\[
P = \max(0, m - N + 1), \quad Q = \min(M, m). \quad (4.22d)
\]

It is shown in Appendix 4.B that
\[
\tilde{\delta}_m = \sum_{n=0}^{Q} h_n^2, \quad 0 \leq m < N. \quad (4.23)
\]

Substituting (4.23) into (4.22a) and using the fact that the term \(c \alpha^3\) becomes insignificant as \(\alpha \to 0\), one obtains the approximation
\[
I_N^b(\alpha) \simeq \frac{\alpha^2}{\ln 2} \left\{ \begin{array}{ll}
\sum_{m=0}^{N-1} \left( 1 - \frac{m}{N} \right) h_m^2, & N \leq M \\
 \left( 1 - \frac{M}{N} \right) \|h_{M+1}\|^2 + \frac{1}{N} \sum_{m=0}^{M-1} (M - m) h_m^2, & N > M.
\end{array} \right. \quad (4.24)
\]

Finally, (4.19) is obtained with \(E_s/N_o\) replacing \(\alpha^2\), and by taking the limit as \(N \to \infty\). This completes the proof of Corollary 4.2.

**Corollary 4.3:** Under the conditions of Corollary 4.2, the symmetric information rate defined by (4.4) (in units of bits/symbol), \(I^b(E_s)\), is upper bounded as
\[ I^b(E_s) \leq I^G(E_s) \]

\[ = \frac{1}{2\pi} \int_0^\pi \log_2 \left[ 1 + 2 \left( \frac{E_s}{N_0} \right) |H(\lambda)|^2 \right] d\lambda \]

\[ \leq \frac{1}{2} \log_2 \left[ 1 + 2 \left( \frac{E_s}{N_0} \right) \|h^{M+1}\|^2 \right] \]

\[ \leq \frac{1}{\ln 2} \left( \frac{E_s}{N_0} \right) \|h^{M+1}\|^2. \]  \hspace{1cm} (4.25)

The second inequality holds with equality if and only if \(|H(\lambda)| = \|h^{M+1}\|^2 = h_0^2, \ |\lambda| \leq \pi, i.e., in the case where the channel is memoryless (M = 0). Furthermore, the last bound in (4.25) is asymptotically tight in the limit as \(E_s/N_0 \to 0\).

**Proof:** We make use of the results obtained in Chapter 3. Thus, invoking Corollary 3.4 and using (3.47) give directly the first two lines in (4.25). The next inequality follows since the logarithm is (strictly) convex-f over the interval \(0 \leq E_s/N_0 < \infty\), and because of the identity

\[ \frac{1}{\pi} \int_0^\pi |H(\lambda)|^2 d\lambda = \|h^{M+1}\|^2. \]  \hspace{1cm} (4.26)

The last inequality follows from the simple inequality \(\ln(1 + x) \leq x, x > -1\). The tightness of this bound for small values of \(E_s/N_0\) is asserted by Corollary 4.2. This completes the proof of Corollary 4.3.

An upper bound for \(I^b_N(E_s)\) is now introduced which will be useful when estimating \(I^b_N(E_s)\) and \(I^b(E_s)\).

---

1 Attempts to bound \(I^b_N(E_s)\), and thus \(I^b(E_s)\), by using (4.10) were not successful.

2 Hence Jensen's inequality implies \(E[\log|g(\lambda)|] \leq \log[E[|g(\lambda)|]]\) with equality if and only if \(g(\lambda)\) is a constant (alternatively, the calculus of variations may be used).
Corollary 4.4: For $N \geq 1$, define

$$\hat{I}_N^b(E) = \frac{1}{N} I(X^N; Y^{N+M})$$

as the average mutual information per input symbol between the ensemble of binary input sequences, $X^N$, and the ensemble of continuous output sequences, $Y^{N+M}$, where any realization $y^{N+M} \in Y^{N+M}$ has components as defined in (4.3). Thus, $I(X^N; Y^{N+M})$ is obtained from (4.5) by replacing $y^N$ with $y^{N+M}$, i.e., the integral in (4.5a) becomes $(N + M)$-dimensional. Thus,

$$I_N^b(E) \leq \hat{I}_N^b(E), \quad N \geq 1,$$

where $I_N^b(E)$ is given by (4.10) of Theorem 4.1, and $\hat{I}_N^b(E)$ is also given by (4.10) with $z^N$ replaced by $z^{N+M}$ and $\delta_{ik}^N$ replaced by $\delta_{ik}^{N+M} = (v_i^{N+M} - v_k^{N+M})/\sqrt{E}$, $0 \leq i, k < K$, and with the components of $v_r^{N+M}$, $0 \leq r < K$, given as in (4.3b). Furthermore, the upper bound in (4.27b) is asymptotically tight in the limit as $N \rightarrow \infty$, i.e., the asymptotic information rate defined in (4.4), $I^b(E)$, is equivalently obtained as

$$I^b(E) = \lim_{N \rightarrow \infty} \hat{I}_N^b(E).$$

Finally, for $N = 1$, the upper bound defined in (4.27a) becomes

$$\hat{I}_1^b(E) = 1 - \int_{-\infty}^{+\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \log_2 \left[ 1 + e^{-\sqrt{8 E_4 l N_0} \|h^M+1\| z - 4(E_4 l N_0) \|h^M+1\|^2} \right] dz,$$

where

$$\hat{I}_1^b(E) \bigg|_{\|h^M+1\|^2 = 1} = C^b(E), \bigg|_{M=0} \bigg|_{h_0=1}.$$
Proof: Let \( Y^{N+M} = Y^N \hat{Y}^M \), where \( Y^N = (Y_0, Y_1, \ldots, Y_{N-1}) \), and \( \hat{Y}^M = (Y_N, Y_{N+1}, \ldots, Y_{N+M-1}) \), \( N \geq 1 \). Then,

\[
I_N^b(E_3) = \frac{1}{N} I(X^N; Y^N \hat{Y}^M) = \frac{1}{N} I(X^N; Y^N) + \frac{1}{N} I(X^N; \hat{Y}^M | Y^N)
\]

(4.30a)

since (conditional) average mutual information is nonnegative; this proves (4.27b). To prove (4.28), let \( N > M \) and define \( X^N = X^{N-M} \hat{X}^M \), where \( X^{N-M} = (X_0, X_1, \ldots, X_{N-M-1}) \) and \( \hat{X}^M = (X_{N-M}, X_{N-M+1}, \ldots, X_{N-1}) \). The last term in the second line of (4.30a) may now be expanded and upper bounded as

\[
\frac{1}{N} I(X^N; \hat{Y}^M | Y^N) = \frac{1}{N} I(X^{N-M} \hat{X}^M; \hat{Y}^M | Y^N)
\]

(4.30b)

where we have used the facts that \( X^{N-M} \) and \( \hat{Y}^M \) are statistically independent, i.e., \( I(X^{N-M}; \hat{Y}^M | Y^N \hat{X}^M) = 0 \) and that \( I(\hat{X}^M; \hat{Y}^M | Y^N) \) is upper bounded by \( M \log 2 \) since there are \( 2^M \) elements in the ensemble of binary sequences \( \hat{X}^M \).

Thus, substitution of (4.30b) into (4.30a) yields

\[
I_N^b(E_3) \leq I_N^b(E_3) \leq I_N^b(E_3) + (M/N) \log 2 ,
\]

and (4.28) follows in the limit as \( N \to \infty \). After some obvious simplifications, one obtains from (4.10) directly the expression
\[ i_1^b(E_3) = 1 - \int p_Z(z^{M+1}) \times \]
\[ \times \log_2 \left[ 1 + e^{-\sqrt{8E_z/N_0} \langle z^{M+1}, h^{M+1} \rangle - 4(E_z/N_0) \|h^{M+1}\|^2} \right] dz^{M+1}. \]

(4.31)

where the \((M + 1)\)-dimensional integral is invariant to the direction of \(h^{M+1}\), because \(z^{M+1}\) has a spherically symmetric probability density, \(p_Z(z^{M+1})\), as can be seen from (4.10b). Therefore, one may replace \(h^{M+1}\) in (4.31) by any vector, e.g., \(\hat{h}^{M+1}\), as long as \(\|\hat{h}^{M+1}\| = \|h^{M+1}\|\). In other words, the inner products \(\langle z^{M+1}, \hat{h}^{M+1} \rangle\) and \(\langle z^{M+1}, h^{M+1} \rangle\) span the same space. Choosing \(\hat{h}_0 = \|h^{M+1}\|\) and \(\hat{h}_k = 0, 1 \leq k \leq M\), gives \(\langle z^{M+1}, \hat{h}^{M+1} \rangle = \|h^{M+1}\|z_0\), and integrating and renaming the dummy variable, \(z_0\), yields (4.29a). Finally, letting \(\|h^{M+1}\| = 1\) in (4.29a) leads to (4.14), and (4.29b), respectively. This completes the proof of Corollary 4.4.

### 4.3 Generation of Numerical Results

The expression in (4.14) for the capacity of the memoryless Gaussian channel with binary inputs is well known from the solution of [GAL(68), Probl. 4.22]. Equivalent expressions in integral form can be found in [McE(77), Probl. 4.14], and [VIT-OMU(79), p. 153]. In each case, the particular integral must be evaluated numerically with respect to the dummy variable (\(z\) in the present case). The generalized \(N\)-dimensional expression for \(i_N^b(E_3)\) in (4.10) of Theorem 4.1 includes the scalar solution (4.14) as a special case. This implies, unfortunately, that \(i_N^b(E_3)\) should also be evaluated using numerical integration methods. Therefore, analytical evaluation of \(i_N^b(E_3)\) in the limit as \(N \to \infty\) will generally not be possible and \(i^b(E_3)\), as defined in (4.4), can thus only be approximated (lower bounded) by evaluating \(i_N^b(E_3)\) for increasing but finite \(N\).
Even for moderately large $N$, however, it will be difficult to compute $I_N^b(E_3)$ with (standard) numerical integration methods. In such a situation, one often resorts to statistical estimation of the $N$-dimensional integrals [DAV-RAB(67), Chapter 5], [STR(71), Chapter 6]. In the following, this technique is pursued further to obtain statistical estimates (rather than true approximations) for $I_N^b(E_3)$ and $\hat{I}_N^b(E_3)$. Based on these estimates, approximate results for $I_N^b(E_3)$ can then be obtained.

There exists a large body of literature on the application of Monte Carlo methods (MCM’s) to the problem of estimating multi-dimensional integrals. References, and a brief summary of definitions and principles of the MCM as they apply to the present problem, are given in Appendix 4.C. Subsequently, frequent reference will be made to Appendix 4.C. The material presented in this appendix reflects the methods that were chosen after an extended search for a computationally efficient type of MCM. Some aspects of this selection process are included in the following.

A. Estimation of $I_N^b(E_3)$ and $\hat{I}_N^b(E_3)$

Using Corollary 4.4, one obtains

$$I_N^b(E_3) = 1 - \frac{1}{N} \int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} p_Z(z^{N+M}) g(z^{N+M}) \, dz^{N+M} ,$$

(4.32a)

where each dimension of the $(N+M)$-dimensional integral is evaluated over the interval $(-\infty, +\infty)$ and where the units are bits/symbol. The dummy variable in (4.32a), $z^{N+M}$, is a sequence of i.i.d. Gaussian random variables with mean zero and unit variance, i.e.,

$$p_Z(z^{N+M}) = \frac{1}{(2\pi)^{(N+M)/2}} \exp\left(-\|z^{N+M}\|^2/2\right) ,$$

(4.32b)

and the function
\[ g(z^{N+M}) = \frac{1}{K} \sum_{i=0}^{K-1} g_i(z^{N+M}), \quad (4.32c) \]

where, with \( K = 2^N \),

\[ g_i(z^{N+M}) = \log_2 \left[ \sum_{k=0}^{K-1} \exp \left\{ -\sqrt{2E_s/N_o} \langle z^{N+M}, d_{ik}^{N+M} \rangle - (E_s/N_o) \| d_{ik}^{N+M} \|^2 \right\} \right]. \quad (4.32d) \]

Ideally, one would like to estimate \( \hat{I}_N^b(E_s) \) and \( I_N^b(E_s) \) for sufficiently large \( N \), such that \( \hat{I}_N^b(E_s) - I_N^b(E_s) \leq \delta \), where \( \delta \) is an arbitrarily small positive number. Note that \( I_N^b(E_s) \) may be obtained directly from (4.32) by setting the last \( M \) components of \( d_{ik}^{N+M} \) in (4.32d) to zero. Therefore, \( I_N^b(E_s) \) and \( \hat{I}_N^b(E_s) \) can be estimated simultaneously, and the following development for \( \hat{I}_N^b(E_s) \) holds similarly for \( I_N^b(E_s) \).

According to Appendix 4.C, the integral in (4.32a) represents the expectation of \( g(z^{N+M}) \), i.e.,

\[ \hat{I}_N^b(E_s) = 1 - \frac{1}{N} \mathbb{E} \left[ g(z^{N+M}) \right] = 1 - \frac{1}{N} \hat{J}_g, \quad (4.33) \]

and

\[ \hat{\xi}_L = \frac{1}{L} \sum_{s=1}^{L} g(z_s^{N+M}) \quad (4.34a) \]

is an unbiased estimator of \( \hat{J}_g \). The sample variance of \( g(\cdot) \) in (4.34a) is thus given as

\[ \hat{s}_{g,L}^2 = \frac{1}{L-1} \left[ \sum_{s=1}^{L} g^2(z_s^{N+M}) - L \hat{\xi}_L^2 \right], \quad (4.34b) \]
and the $z_s^{N+M}$, $1 \leq s \leq L$, are generated according to the multivariate Gaussian density (4.32b). Usually, i.i.d. random numbers having a Gaussian (normal) density, denoted by $N(0,1)$ for mean zero and unit variance, are constructed from i.i.d. random numbers with a uniform density, denoted by $U(0,1)$ for the interval $[0,1]$ (see, e.g., [LAW-KEL(82), Chapters 6/7]). It could thus be advantageous to use directly the $U(0,1)$ density when estimating $\hat{J}_g$ in (4.33). In addition, the transformed integrand may have a reduced (sample) variance. It will be instructive to determine whether any net benefit results when using directly the $U(0,1)$ density.

Write $J$ in (4.33) as

$$J_g = \int_0^\infty \int_0^\infty \cdots \int_0^\infty \frac{1}{(2\pi)^{(N+M)/2}} e^{-\left(z_0^2 + z_1^2 + \cdots + z_{N+M-1}^2\right)/2} \times$$

$$\times g_S(z_0, z_1, \ldots, z_{N+M-1}) dz_0 dz_1 \cdots dz_{N+M-1}, \quad (4.35a)$$

where

$$g_S(z_0, z_1, \ldots, z_{N+M-1}) =$$

$$g(z_0, z_1, \ldots, z_{N+M-1}) + g(-z_0, z_1, \ldots, z_{N+M-1}) +$$

$$g(z_0, -z_1, \ldots, z_{N+M-1}) + g(-z_0, -z_1, \ldots, z_{N+M-1}) +$$

$$\cdots + \cdots +$$

$$\cdots + \cdots +$$

$$+ g(z_0, -z_1, \ldots, -z_{N+M-1}) + g(-z_0, -z_1, \ldots, -z_{N+M-1}). \quad (4.35b)$$

Note that $g_S(\cdot, \cdot, \ldots, \cdot)$ in (4.35b) requires $2^{N+M}$ functional evaluations of $g(\cdot, \cdot, \ldots, \cdot)$. Define now the transformation $u_i = e^{-z_i^2/2}$, $0 \leq i < N+M$, so that $z_i = \sqrt{-2 \ln u_i}$, $dz_i = -du_i/\left(u_i \sqrt{-2 \ln u_i}\right)$, and so that the new dummy variables $u_i \rightarrow 1$ (0) as $z_i \rightarrow 0$ ($\infty$). Then, one may write $J_g$ as
\[
\hat{J}_g = \int_0^1 G(u^{N+M}) du^{N+M},
\]

(4.36a)

where

\[
G(u^{N+M}) = \left[ \prod_{i=0}^{N+M-1} \frac{1}{\sqrt{-4\pi \ln u_i}} \right] \times \\
\times g_\Sigma(\sqrt{-2 \ln u_0}, \sqrt{-2 \ln u_1}, ..., \sqrt{-2 \ln u_{N+M-1}}),
\]

(4.36b)

and \(g_\Sigma(\cdot, \cdot, ..., \cdot)\) is defined in (4.35b). One may now use

\[
\hat{G}_L = \frac{1}{L} \sum_{s=1}^{L} G(u_s^{N+M})
\]

(4.37a)

as an unbiased estimator of \(\hat{J}_g\) in (4.33), where the sample variance is

\[
\hat{\sigma}^2_{G,L} = \frac{1}{L-1} \left[ \sum_{s=1}^{L} G^2(u_s^{N+M}) - L \hat{G}_L^2 \right].
\]

(4.37b)

The sample sequences \(u_s^{N+M}, 1 \leq s \leq L\), have i.i.d. components which are generated according to the U(0,1) density. A variance reduction, which could result from the use of the U(0,1) instead of the N(0,1) density, will be at least partly offset by the additional workload that is required to compute \(\hat{G}_L\). The additional work involves the summation in (4.35b) and the evaluation of the logarithm and square-root for each \(u_i\) in (4.36b). Whether this increased workload results in a substantial reduction of the (sample) variance may be determined when comparing the two MCM's by means of the relative efficiency, \(\eta\) (see Appendix 4.C for the definition of \(\eta\)).
Later, and in the following example, it will be useful to identify a specific binary sequence, \( x^N = (x_0, x_1, ..., x_{N-1}) \), \( x_k = \pm \sqrt{E_s} \), among the set of \( K = 2^N \) possible sequences, by the index

\[
i = \sum_{k=0}^{N-1} (1 - x_k \sqrt{E_s}) 2^{k-1}.
\] (4.38)

For instance, the sequence \( x^3 = (\pm \sqrt{E_s}, \sqrt{E_s}, -\sqrt{E_s}) \) will be identified by the index \( i = 2 \times 2^1 + 0 \times 2^0 + 2 \times 2^1 = 5 \).

**Example 4.1:** Consider the memoryless channel where \( M = 0 \) and \( h_0 = 1 \). For this channel, \( C^b(E_s) \) is given by (4.14). With \( N = 1 \) and \( z \equiv z_0 \), (4.35b), (4.32c/d), and (4.8) yield

\[
g_2(z) = g(z) + g(-z)
\]

\[
= \frac{1}{2} \left[ g_0(z) + g_1(z) \right] + \frac{1}{2} \left[ g_0(-z) + g_1(-z) \right] \quad (4.39a)
\]

\[
= g_0(z) + g_0(-z),
\]

where

\[
g_0(z) = \log_2 \left[ 1 + e^{-\sqrt{8E_s / N_0} z - 4E_s / N_0} \right] \quad (4.39b)
\]

and where the last equality in (4.39a) follows from \( g_1(z) = g_0(-z) \). Thus,

\[
C^b(E_s) \bigg|_{M=0} = 1 - \int_0^1 G(u) du \quad ,
\] (4.40a)

where

\[
G(u) = \frac{1}{\sqrt{-4 \pi \ln u}} \left[ g_0(\sqrt{-2 \ln u}) + g_0(-\sqrt{-2 \ln u}) \right].
\] (4.40b)
The integral in (4.14) may be estimated (method 1) by means of

\[ \tilde{g}_L = \frac{1}{L} \sum_{s=1}^{L} g_0(z_s) , \]  

(4.41)

where \( g_0(\cdot) \) is given in (4.39b), and where the \( z_s \) are N(0,1) random numbers.

The second estimator (method 2) is given as

\[ \bar{g}_L = \frac{1}{L} \sum_{s=1}^{L} G(u_s) , \]  

(4.42)

where \( G(\cdot) \) is given in (4.40b) and where the \( u_s \) are U(0,1) random numbers.

Estimates for \( C^b(E_s) \) are then given by \( (1 - \tilde{g}_L) \) and \( (1 - \bar{g}_L) \), respectively.

For control purposes, \( C^b(E_s) \) is also computed by standard numerical integration.

Figure 4.2(a) shows the efficiency \( \eta \) of method 2 relative to method 1 as a function of \( E_s/N_0 \). For some range of \( E_s/N_0 \), \( \eta \) can be improved by using the simple method of *antithetic variates* [KLE(74), Section III.6], [LAW-KEL(82), Section 11.3]. With this method, one computes for each \( u_s \) the functions \( G(u_s) \) and \( G(1 - u_s) \) and replaces \( G(u_s) \) in (4.42) by \( [G(u_s) + G(1 - u_s)]/2 \).

The pair \( \{u, (1 - u)\} \) is called a pair of antithetic variates; if \( u \) is a U(0,1) variate, then \( (1 - u) \) is as well. This method, denoted \( U(0,1)/A \) in Figure 4.2, works best for monotonically increasing or decreasing integrands, i.e., small (large) values of \( G(u) \) tend to be compensated by large (small) values of \( G(1 - u) \). Note that the \( U(0,1) \) methods perform better than the \( N(0,1) \) method over only the range of \( E_s/N_0 \) where the capacity is in the range \( 0.25 < C^b(E_s) \leq 1 \) bits/symbol.

Figure 4.2(c) demonstrates the good correlation in this example between the absolute error, \( |C^b(E_s) - (1 - \bar{g}_L)| \), and the standard deviation of the capacity.
Fig. 4.2. Comparison of MCM's (Example 4.1):
(a) Relative efficiency, $\eta$, as a function of $E_s / N_0$, for the memoryless channel ($M = 0$, $h_0 = 1$) when using random numbers with uniform density, U(0,1), and pairs of antithetic random numbers with uniform density, U(0,1)/A. The reference method uses random numbers with a Gaussian density, N(0,1). $L = 1000$ sample functions were averaged for each method.
(b) Capacity for this channel, $C^b(E_s)$, obtained by numerical integration (see Figure 4.2(c) for the error performance of the U(0,1)/A and N(0,1) methods).
Fig. 4.2. (continued)

(c) Standard deviation (A) of the capacity estimate \(1 - \tilde{G}_L\), \(s_{G,L}/\sqrt{L}\), and absolute error (B) of the capacity estimate, \(|C_l^b(E_s) - (1 - \tilde{G}_L)|\), as a function of \(E_s/N_o\) when using \(L = 1000\) pairs of \(U(0,1)\)/A random numbers. Also shown here is the absolute error (C) of the capacity estimate \(1 - \tilde{G}_L\), \(|C_l^b(E_s) - (1 - \tilde{G}_L)|\), for the reference method using \(N(0,1)\) random numbers. To obtain a fair comparison (unit work ratio), \(L = 3800\) sample functions were used for the latter method. (Units of standard deviation and absolute errors are \(\text{bits/symbol}\).)
estimate, \( s_{G,L} \sqrt{L} \), when using the \( U(0,1)/A \) method. Also shown is the absolute error for the \( N(0,1) \) method, \( |C^b(E_s) - (1 - \bar{g}_l)| \). To obtain a fair comparison, the results in Figure 4.2(c) were generated using a unit work ratio, i.e., both methods were allowed the same amount of computer run time. For \( E_s/N_0 \) below \(-7 \) dB, the \( N(0,1) \) method performs clearly better, while for larger values of \( E_s/N_0 \) its performance is not significantly worse in this example.

Example 4.1 indicates that the benefits of using \( U(0,1) \) random numbers, compared to using \( N(0,1) \) random numbers, are limited or non-existent when the allowed workload (computer run time) is kept the same for each method. For the general case to be considered here, the number of functional evaluations required to compute \( g_\Sigma(\cdot, \cdot, \cdot, \cdot) \) in (4.36b) increases exponentially with \( N \) and \( M \). On the other hand, the variance reduction which may be achieved (depending on the signal-to-noise ratio, \( E_s/N_0 \)) when using the \( U(0,1) \) method, vanishes quickly as \( N \) increases. Therefore, it was decided to adopt the \( N(0,1) \) method to estimate \( I^b_N(E_s) \) and \( \hat{I}^b_N(E_s) \) and, if necessary, to combine it with the control-variate MCM (CV-MCM) as described in Appendix 4.C.

First, reconsider the function \( g(\cdot) \) specified by (4.32c/d). In the general case where ISI is present \((M > 0)\), the vertices of the received (noiseless) signal sequences, \( v^N_{r+M} \), \( 0 \leq r < K \), form a skewed \((N + M)\)-dimensional cube in signal space. For this reason, averaging over the index \( i \) in (4.32c) is necessary. As indicated in Example 4.1, however, a simplification may be introduced when the set of all input sequences, \( \{x^N_i : 0 \leq i < K \} \), is partitioned into two complementary sets, \( \{x^N_i : 0 \leq i < K/2 \} \), and \( \{x^N_i : K/2 \leq i < K \} \), where the identifying index, \( i \), is defined as in (4.38). Define

\[
g'(z^{N+M}) = \sum_{i=0}^{(K/2)-1} g_i(z^{N+M}), \quad (4.43a)
\]

and
where \( g_i(\cdot), 0 \leq i < K, \) is defined in (4.32d). It then follows that

\[
g''(z^{N+M}) = g'(-z^{N+M})
\]

because for every \( d^{N+M}_{ik}, 0 \leq i < K/2, 0 \leq k < K, \) in \( g_i(\cdot) \) of (4.43a), there exists in \( g_{K-1-i}(\cdot) \) of (4.43b) a corresponding \( d^{N+M}_{K-1-i,k-k} = -d^{N+M}_{ik}. \) Since \( p_z(\cdot) \) in (4.32b) is an isotropic function, one obtains from (4.32a) the identity \( E[g(z^{N+M})] = E[g'(z^{N+M})]. \) Therefore, (4.33) may be expressed as

\[
\hat{J}_N^b(E_s) = 1 - \frac{1}{N} E[g'(z^{N+M})]
\]

where \( g'(\cdot) \) is defined by (4.43a). The basic MCM or the CV-MCM (see also Appendix 4.C) may now be used to estimate \( \hat{J}_N^b \) in (4.45).

For notational simplicity, it will be assumed in the remainder of this chapter that the unit-sample response of the channel, \( h^{M+1} = (h_0, h_1, \ldots, h_M), \) is normalized such that its energy \( \|h^{M+1}\|^2 = 1, \) and \( h_0 > 0. \) Define for the memoryless channel (\( M = 0, h_0 = 1 \)) the integrand

\[
g^*(z^{N+M}) = g'(z^{N+M}) \bigg|_{M=0, h_0=1}
\]

and the corresponding capacity

\[
C^b_s(E_s) = C^b_s(E_s) \bigg|_{M=0, h_0=1}
\]

where \( g''(\cdot), 0 \leq i < K, \) is defined in (4.32d). It then follows that

\[
g''(z^{N+M}) = g'(-z^{N+M})
\]

because for every \( d^{N+M}_{ik}, 0 \leq i < K/2, 0 \leq k < K, \) in \( g_i(\cdot) \) of (4.43a), there exists in \( g_{K-1-i}(\cdot) \) of (4.43b) a corresponding \( d^{N+M}_{K-1-i,k-k} = -d^{N+M}_{ik}. \) Since \( p_z(\cdot) \) in (4.32b) is an isotropic function, one obtains from (4.32a) the identity \( E[g(z^{N+M})] = E[g'(z^{N+M})]. \) Therefore, (4.33) may be expressed as

\[
\hat{J}_N^b(E_s) = 1 - \frac{1}{N} E[g'(z^{N+M})]
\]

where \( g'(\cdot) \) is defined by (4.43a). The basic MCM or the CV-MCM (see also Appendix 4.C) may now be used to estimate \( \hat{J}_N^b \) in (4.45).

For notational simplicity, it will be assumed in the remainder of this chapter that the unit-sample response of the channel, \( h^{M+1} = (h_0, h_1, \ldots, h_M), \) is normalized such that its energy \( \|h^{M+1}\|^2 = 1, \) and \( h_0 > 0. \) Define for the memoryless channel (\( M = 0, h_0 = 1 \)) the integrand

\[
g^*(z^{N+M}) = g'(z^{N+M}) \bigg|_{M=0, h_0=1}
\]

and the corresponding capacity

\[
C^b_s(E_s) = C^b_s(E_s) \bigg|_{M=0, h_0=1}
\]
and let the difference integrand be defined as
\[ D(z^{N+M}) = g'(z^{N+M}) - \kappa g^*(z^{N+M}). \quad (4.47) \]

The integrand for the memoryless channel is thus taken as the control variate. Substitution of (4.47) into (4.45) yields
\[ \hat{i}_N^b(E_d) = (1 - \kappa) + \kappa C^b_{\kappa}(E_d) - \frac{1}{N} \mathbb{E}[D(z^{N+M})] \]
\[ = (1 - \kappa) + \kappa C^b_{\kappa}(E_d) - \frac{1}{N} \hat{j}_D. \quad (4.48) \]

By defining the so-called error sequence as
\[ e^N_{ik} = \frac{1}{\sqrt{E_s}} (x^N_i - x^N_k), \quad (4.49a) \]
the difference integrand defined in (4.47) becomes after some manipulations
\[ D(z^{N+M}) = \frac{2}{K} \sum_{i=0}^{(K/2) - 1} D_i(z^{N+M}), \quad (4.49b) \]
where
\[ D_i(z^{N+M}) = \log_2 \left[ \frac{\sum_{k=0}^{K-1} \exp \left\{ -\frac{2E_s}{N_o} \langle z^{N+M}, d_i^{N+M} \rangle - (E_s/N_o) \| d_i^{N+M} \|^2 \} \right\}^{\kappa} }{\sum_{k=0}^{K-1} \exp \left\{ -\frac{2E_s}{N_o} \langle z^N, e_i^N \rangle - (E_s/N_o) \| e_i^N \|^2 \} \right\}^{\kappa} } \right], \quad (4.49c) \]
and where it is understood that \( z^N \) represents the first \( N \) components of \( z^{N+M} \), i.e., \( z^{N+M} = (z^N, z_N, z_{N+1}, ..., z_{N+M-1}) \). The sample mean
\[ \hat{D}_L = \frac{1}{L} \sum_{s=1}^{L} D(z_s^{N+M}) , \quad (4.50a) \]

will be taken as the unbiased estimator of \( \hat{I}_D \) in (4.48). The sample variance is given by

\[ \hat{s}_{DL}^2 = \frac{1}{L-1} \left[ \sum_{s=1}^{L} D^2(z_s^{N+M}) - \frac{1}{L} \hat{D}_L^2 \right] , \quad (4.50b) \]

and \( \hat{s}_{DL} / (N/\sqrt{L}) \) is used to test the stability of the estimate of \( \hat{I}_N^b(E_s) \), which is defined as

\[ \overline{\hat{I}_N^b(E_s)} = (1 - \kappa) + \kappa C_{\cdot}^b(E_s) - \frac{1}{N} \hat{D}_L \]

\[ \approx \hat{I}_N^b(E_s) . \quad (4.51) \]

The optimal value for the parameter \( \kappa \) is determined by observing the stability criterion, \( \hat{s}_{DL} / (N/\sqrt{L}) \). The estimate of \( I_N^b(E_s) \), \( \overline{I_N^b(E_s)} \), is defined similarly.

For later use, define also the average quantity

\[ \overline{I_N^b(E_s)} = \frac{1}{2} \left[ \overline{I_N^b(E_s)} + \overline{\hat{I}_N^b(E_s)} \right] . \quad (4.52) \]

It can be seen from (4.48) to (4.51) that the CV-MCM reduces to the basic MCM when \( \kappa = 0 \).

Experiments have shown that the optimal value for the parameter \( \kappa \) obtained when estimating \( \hat{I}_N^b(E_s) \) is generally different from the one found when estimating \( I_N^b(E_s) \); however, this difference becomes insignificant as \( N \) increases. Furthermore, the optimal value of \( \kappa \) is essentially independent of the signal-
to-noise ratio, $E_s/N_0$. The choice $\kappa = 1$ was found to be a useful compromise. The CV-MCM requires about 25% more computing time than the basic MCM ($\tau_1/\tau_2 \approx 0.8$) because of the additional summation in the denominator of (4.49c). For $M = 1$ and $E_s/N_0 \equiv 0 \, \text{dB}$, for example, the estimate of the variance ratio $\sigma_1^2/\sigma_2^2$ decreases to about four with increasing $N$, when the optimal value for $\kappa$ is being used. Thus, the relative efficiency, $\eta$, of the CV-MCM with respect to the basic MCM equals about three in this case. This order of improvement in computational efficiency is typical for small values of $M$ and $N \leq 10$. Since it may take up to several hours to obtain stable estimates when $N$ is large, such relatively small savings are still significant. In the case where $N$ is sufficiently larger than $M$, a further reduction in required computing time is possible by using only a subset of the $K/2$ input sequences, $x_i^N$, when computing the average of the $D((z)_{(z)}$ in (4.49b). This subset may be determined by uniform random sampling of the index $i$ from the interval $[0, 1, \ldots, (K/2) - 1]$. To avoid biased estimates, the size of the chosen subset of input sequences should be as large as is practical.

**Example 4.2:** Consider the channel with ISI memory $M = 1$, where the unit-sample response $h^2 = (h_0, h_1) = (h_0, \alpha h_0) = h_0(1, \alpha), -\infty < \alpha < \infty$, has unit energy, i.e., $h_0 = 1/\sqrt{1 + \alpha^2}$. Figure 4.3 shows the estimate of $I_N^b(E_s)$, $I_N^b(E_s)$, and the estimate of $I_N^b(E_s)$, $I_N^b(E_s)$, as a function of the block length, $N$, for the case $\alpha = 1$. Also shown are the capacity values for the memoryless channel, $C^b(E_s)$, and the average quantity $I_N^b(E_s)$ which was defined in (4.52). The latter appears to reach a steady value for smaller values of $N$ than any of the two individual estimates. Figure 4.3 indicates that large block lengths are required to achieve $I_N^b(E_s) \simeq I_N^b(E_s)$; generally, it is required that $N >> M$.

Figure 4.4 shows $I_N^b(E_s)$, $I_N^b(E_s)$, and their average, $I_N^b(E_s)$, as a function of the signal-to-noise ratio, $E_s/N_0$, for $N = 6$. For reference purposes, $C^b(E_s)$ is also shown. Note that $I_N^b(E_s)$ indicates a loss of about 1 dB in signal-to-noise
Fig. 4.3. (Example 4.2) Convergence of $I_N^b(E_s)$, $i_N^b(E_s)$, and $i_N^b(E_s)$, as a function of the block length, $N$. The channel has unit-sample response $h^2 = (1, 1)/\sqrt{2}$, i.e., its ISI memory $M = 1$. $L = 1000$ Gaussian sample vectors were used for each estimate. The estimated standard deviation (stability criterion) is typically 1% to 4% of the estimate.
Fig. 4.4. (Example 4.2) $I_N^b(E_s)$, $\overline{I}_N^b(E_s)$, and $\tilde{I}_N^b(E_s)$, as a function of the signal-to-noise ratio, $E_s/N_o$, for block length $N = 6$. The channel has unit-sample response $h^2 = (1, 1)/\sqrt{2}$, i.e., its ISI memory $M = 1$. The capacity of the memoryless channel, $C_b(E_s)$, is shown for reference.
ratio when compared with $C^b_<(E_s)$. Figure 4.5 presents the quantities $\bar{I}_b^b(E_s)$, $\bar{I}_b^b(E_s)$, and $I_b^b(E_s)$, for $0 \leq \alpha \leq 1$, and $E_s/N_0 \equiv 0$ dB. Note that the difference between the two estimates increases as the amount of ISI (i.e., $\alpha$) increases.

5. Practical Approximation of $I^b(E_s)$

It was argued at the beginning of this section that the sought quantity, $I^b(E_s)$, cannot be computed in general with presently known methods. However, the results obtained in Example 4.2 suggest that $I_b^b(E_s)$, the average of $I_b^b(E_s)$ and $\bar{I}_b^b(E_s)$ as defined in (4.52), could provide a useful approximation for $I^b(E_s)$ in the case where $N$ is sufficiently larger than the ISI memory, $M$. It is thus proposed to consider $I_b^b(E_s)$ as a practical approximation for $I^b(E_s)$, i.e., to assume that

$$I_b^b(E_s) \approx I^b(E_s), \quad N \gg M. \quad (4.53)$$

The motivation to consider this approximation stems from the fact that $I^b(E_s)$ may also be written in the form

$$I^b(E_s) = \lim_{N \to \infty} \left[ \varepsilon I_b^b(E_s) + (1-\varepsilon)\bar{I}_b^b(E_s) \right], \quad 0 \leq \varepsilon \leq 1, \quad (4.54)$$

as can be seen from definition (4.4) and Corollary 4.4.

Figure 4.6 shows $I_b^b(E_s)$ as determined for the channels which were defined in Figure 3.1 of Chapter 3. The approximations, $\bar{I}_b^b(E_s)$, for Channel 2 and Channel 3 were obtained with $N = 6$. For Channel 4, $N = 10$ was used and a subset of input sequences was sampled when computing the average in (4.49b). The capacity of the memoryless channel, $C(E_s) = \log_2 \sqrt{1 + 2E_s/N_0}$ (bits/symbol), is shown for reference. In addition, the information rate for i.i.d. Gaussian inputs, $I^G(E_s)$, which has been derived in Chapter 3, is plotted for
Fig. 4.5. (Example 4.2) $\bar{I}_N^B(E_d), \bar{I}_N^B(E_d)$, and $\bar{I}_N^B(E_d)$, for block length $N = 6$, as a function of the ISI parameter, $\alpha$, i.e., the channel has ISI memory $M = 1$ and unit-sample response $h^2 = (1, \alpha)/\sqrt{1 + \alpha^2}$.
Fig. 4.6. $I_N^b(E_s)$ as a function of the signal-to-noise ratio, $E_s/N_o$, for the channels defined in Figure 3.1 (Chapter 3). For Channel 1, Channel 2, and Channel 3, the block length $N = 6$, and for Channel 4, $N = 10$. For reference purposes, the capacity of the memoryless channel (Channel 1), $C(E_s) = \log_2 \sqrt{1 + 2E_s/N_o}$ (bits/symbol), and the information rate of Channel 4 for i.i.d. Gaussian inputs, $I^G(E_s)$, are also shown.
Channel 4. Clearly, $I_N^b(E_3) \leq I^G(E_3)$, in this case. Comparison of Figure 4.5 and Figure 3.7 reveals that this inequality holds for all channels considered here, including the memoryless channel for which $I^G(E_3) = C(E_3)$. According to Corollary 4.3, $I^G(E_3)$ must provide an upper bound for $I^b(E_3)$; the fact that this has been confirmed experimentally as shown in Figure 4.6, where $I_N^b(E_3)$ replaces $I^b(E_3)$, supports the usefulness of $I_N^b(E_3)$ as a practical approximation for $I^b(E_3)$. 
Justification of Definition (4.4):

To show that the definition

\[ I^b(E_2) = \sup_N I^b_N(E_3) \quad (4.A.1) \]

is equivalent to definition (4.4), i.e., that

\[ \lim_{N \to \infty} I^b_N(E_2) = \sup_N I^b_N(E_3) , \quad (4.A.2) \]

let \( 0 < n < N \), and define the input ensembles \( X^N = (X_0, X_1, \ldots, X_{N-1}) \), \( X^n = (X_0, X_1, \ldots, X_{n-1}) \), and \( \hat{X}^{N-n} = (X_n, X_{n+1}, \ldots, X_{N-1}) \). Further, define the output ensembles which correspond to these input ensembles as \( Y^N = (Y_0, Y_1, \ldots, Y_{N-1}) \), \( Y^n = (Y_0, Y_1, \ldots, Y_{n-1}) \), and \( \hat{Y}^{N-n} = (Y_n, Y_{n+1}, \ldots, Y_{N-1}) \). Then, by definition (4.4b),

\[ N I^b_N(E_2) = I(X^n; Y^n) = I(X^n \hat{X}^{N-n}; Y^n \hat{Y}^{N-n}) \]

\[ = I(X^n; Y^n \hat{Y}^{N-n}) + I(\hat{X}^{N-n}; Y^n \hat{Y}^{N-n} | X^n) , \quad (4.A.3) \]

where the first term in the last equality is lower bounded by \( n I^b_n(E_2) \). The second term in (4.A.3) may be expanded and lower bounded as

\[ I(\hat{X}^{N-n}; Y^n \hat{Y}^{N-n} | X^n) = I(\hat{X}^{N-n}; Y^n \hat{Y}^{N-n} X^n) - I(\hat{X}^{N-n}; X^n) \geq I(\hat{X}^{N-n}; \hat{Y}^{N-n}) = (N-n) I^b_{N-n}(E_3) , \quad (4.A.4) \]

where \( I(\hat{X}^{N-n}; X^n) = 0 \), since the inputs are statistically independent. The last equality follows from definition (4.4b) and the fact that \( I(\hat{X}^{N-n}; \hat{Y}^{N-n}) \) is
independent of the channel state at time instant \( n \). Combination of (4.A.3) and (4.A.4) yields

\[
 l^b_N(E_s) \geq \frac{n}{N} l^b_n(E_s) + \frac{N-n}{N} l^b_{N-n}(E_s), \quad 0 < n < N. \quad (4.A.5)
\]

Application to (4.A.5) of the following result, which was proved by Gallager [GAL(68), pp. 112-113], asserts the equivalence expressed in (4.A.2).

**Lemma 4.A.1:** Let \( a_N, N = 1, 2, \ldots \), be a bounded sequence of numbers, i.e., define

\[
 \bar{a} = \sup_N a_N < \infty.
\]

Then, provided that for all \( N \geq 1 \) and all \( N > n \)

\[
 a_N \geq \frac{n}{N} a_n + \frac{N-n}{N} a_{N-n},
\]

it follows that

\[
 \lim_{N \to \infty} a_N = \bar{a}.
\]

Clearly, the conditions of Lemma 4.A.1 are satisfied since \( l^b_N(E_s) \leq \log 2 \), i.e., \( l^b_N(E_s) \) is a bounded sequence of numbers. Note that for \( N = 2n \), (4.A.5) yields the inequality \( l^b_{2n}(E_s) \geq l^b_n(E_s), n \geq 1 \).
APPENDIX 4.B

Derivation of (4.22a):

With the use of (4.20), equation (4.10c) of Theorem 4.1 may be written as

\[ g(z^N) = \frac{1}{K \ln 2} \sum_{i=0}^{K-1} F_i(\alpha, z^N) , \]  

(4.B.1)

so that (4.10a) becomes

\[ I_N^b(\alpha) = 1 - \frac{1}{NK \ln 2} \sum_{i=0}^{K-1} \int p_z(z^N) F_i(\alpha, z^N) dz^N , \]

(4.B.2)

where the order of summation and integration have been interchanged. Substituting (4.21) into (4.B.2), writing the inner product \( \langle z^N, d_{ik}^N \rangle \) as the sum of the component products, \( z_m d_{ik}^m \), \( 0 \leq m < N \), and interchanging the order of summation and integration, we can reduce (4.B.2) to

\[
I_N^b(\alpha) = \frac{\sqrt{2} \alpha}{NK^2 \ln 2} \sum_{i=0}^{K-1} \sum_{k=0}^{K-1} \sum_{m=0}^{N-1} d_{ik}^m I_m \\
- \frac{\alpha^2}{NK^2 \ln 2} \sum_{i=0}^{K-1} \sum_{k=0}^{K-1} \left[ \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} d_{ik}^m d_{in}^{jr} J_{mn} - \|d_{ik}^N\|^2 \right] \\
+ \frac{\alpha^2}{NK^3 \ln 2} \sum_{i=0}^{K-1} \sum_{k=0}^{K-1} \sum_{r=0}^{K-1} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} d_{ik}^m d_{in}^{jr} J_{mn} \\
- \frac{1}{NK \ln 2} \sum_{i=0}^{K-1} L_i(\alpha) ,
\]

(4.B.3a)

where
I_m = \int p_{z}(z^N) z_m dz^N , \quad (4.B.3b)

J_{mn} = \int p_{z}(z^N) z_m z_n dz^N , \quad (4.B.3c)

and

L_i(\alpha) = \int p_{z}(z^N) R_i(\alpha, z^N) dz^N . \quad (4.B.3d)

The components of the noise vector, \( z^N \), are i.i.d. (Gaussian) random variables with mean zero and unit variance. It follows immediately that

\[ I_m = 0 , \quad 0 \leq m < N \quad (4.B.4a) \]

and

\[ J_{mn} = \begin{cases} 1 , & m = n \\ 0 , & m \neq n \end{cases} , \quad 0 \leq m, n < N . \quad (4.B.4b) \]

By (4.21e), the integral in (4.B.3d), which represents the mean of the remainder \( R_i(\alpha, z^N) \) with respect to \( z^N \), reduces to a term that is proportional to \( \alpha^3 \), i.e.,

\[ L_i(\alpha) = \alpha^3 c_i(\alpha) , \quad 0 \leq i < K , \quad (4.B.4c) \]

where \( c_i(\alpha) \) is unknown but finite. Defining

\[ c_\alpha = \frac{-1}{NK \ln 2} \sum_{i=0}^{K-1} c_i(\alpha) , \quad (4.B.5) \]

substituting (4.4) into (4.B.3a), and defining \( \tilde{\delta}_m \) as in (4.22b), one obtains the result in (4.22a).
Derivation of (4.23):

One may write (4.22b) as

\[ \overline{\delta}_m = E[d_{ik}^m d_{ir}^m], \quad 0 \leq i, k, r < K, \quad (4.B.6) \]

where \( E[\cdot] \) means expectation over the mutually independent, equally likely superscripts \( i, k \) and \( r \). Substituting (4.22c) into (4.B.6) gives

\[ \overline{\delta}_m = \frac{1}{E_s} \sum_{n=0}^{Q} \sum_{j=0}^{Q} h_n h_j \varepsilon_{nj}, \quad 0 \leq m < N, \quad (4.B.7a) \]

where

\[ \varepsilon_{nj} = E[(x_{m-n}^i - x_{m-n}^k)(x_{m-j}^i - x_{m-j}^r)] \]

\[ = E[x_{m-n}^i x_{m-j}^i] - E[x_{m-n}^k x_{m-j}^i] - E[x_{m-n}^i x_{m-j}^r] + E[x_{m-n}^k x_{m-j}^r]. \quad (4.B.7b) \]

The components of \( x_i^N, x_k^l = \pm \sqrt{E_s}, \quad 0 \leq k < N, \) are mutually independent and equally likely. For \( 0 \leq i < K, \) \( x_k^l \) is \( K/2 \) times positive and \( K/2 \) times negative. It follows immediately that

\[ E[x_k^i x_n^i] = \begin{cases} E_s, & i=j, k=n \\ 0, & \text{otherwise} \end{cases} \quad (4.B.8) \]

Applying (4.B.8) to (4.B.7b) gives

\[ \varepsilon_{nj} = \begin{cases} E_s, & n=j \\ 0, & \text{otherwise} \end{cases} \quad (4.B.9) \]

and (4.23) follows when using (4.B.9) in (4.B.7a).
APPENDIX 4.C

On Monte-Carlo Integration:

We summarize briefly the principles of the Monte-Carlo method (MCM) as they apply to the problem of estimating multi-dimensional integrals of the kind considered in this chapter. There exists a large body of literature on MCM's, notably [HAM-HAN(64)], [DAV-RAB(67)], [FRE-GRE(71)], [STR(71)], [KLE(74)], [LAW-KEL(82)], and the references listed therein.

The idea of using MCM's to estimate multiple integrals is based on the following observation. Suppose that one wishes to compute a number, \( J \), that is specified by a multi-dimensional integral which has no "closed-form" solution, and which does not lend itself to direct numerical evaluation (e.g., the number of dimensions of the integrand is very large). Further, suppose that this integral expression is, or has been transformed, such that the number \( J \) represents the expected value of the outcome of a certain stochastic process. The expectation of this outcome may then be estimated by sampling from its population; the average of these samples is an estimate (rather than a true approximation) for the number \( J \). The confidence interval of this estimate shortens with an increasing number of samples which contribute to the average. The rate of confidence improvement depends generally on the type of MCM chosen.

A. The Basic MCM

The MCM for estimating the \( N \)-dimensional integral of the form

\[
J_g = \int_{\mathbb{R}^N} p_x(x^N) g(x^N) dx^N
\]

(4.C.1a)

\[
= E[g(x^N)],
\]

where \( p_x(\cdot) \) is a probability density, i.e.,
is called the basic MCM. Although not necessary, it is usually assumed that the region of integration, \( \mathbb{R}^N \), is finite; this can always be achieved with an appropriate transformation of variables. It is assumed, however, that \( J_g \) exists, i.e., \( \mathbb{E}[|g(\cdot)|] \) is finite. Given some integer, \( L \), one samples \( L \) independent points, \( x_i^N, 1 \leq i \leq L \), according to \( p_X(\cdot) \). The sample mean

\[
\bar{g}_L = \frac{1}{L} \sum_{i=1}^{L} g(x_i^N)
\]

(4.C.2)
is then an unbiased estimator of \( J_g \), i.e., \( \mathbb{E}[\bar{g}_L] = J_g \). Equivalently, according to the strong law of large numbers, \( \lim_{L \to \infty} \bar{g}_L = J_g \), with probability one (see, e.g., [DAV-RAB(67), p. 143]). The (basic) MCM assumes the availability of some random number generator to produce the i.i.d. sample points, \( x_i^N, 1 \leq i \leq L \), in (4.C.2). In practice, so-called pseudorandom numbers are usually used. Such numbers are generated by a deterministic mechanism (e.g., by using a recurrence relation) which has been shown to satisfy certain statistical tests for randomness [HAM-HAN(64), Chapter 3], [LAW-KEL(82), Chapter 6], [LEW-GOO-MIL(69)].

An important question concerns the error introduced when using \( \bar{g}_L \) to estimate \( J_g \). From a statistical viewpoint, a measure of the error may be deduced from the Central Limit Theorem (CLT). Let

\[
\sigma_g^2 = \mathbb{E}[g^2(x^N)] - \left\{ \mathbb{E}[g(x^N)] \right\}^2
\]

(4.C.3)
be the variance of $g(\cdot)$ in (4.C.1a). For the scalar case ($N = 1$), it follows from the CLT that, for sufficiently large $L$ [LAW-KEL(82), Chapter 4] \footnote{For $N > 1$, the multi-dimensional CLT applies [HII(80)].},

$$\Pr\left( |\bar{g}_L - g| \leq \lambda \sigma_g / \sqrt{L} \right) = \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-y^2/2} dy. \quad (4.C.4)$$

The error bound in (4.C.4) (e.g., for $\lambda = 1.96$, the confidence level is 0.95) is proportional to the ratio $\sigma_g / \sqrt{L}$. This rate of convergence is typical for the basic MCM. The MCM has the advantage, however, that convergence is independent of the dimension, $N$, except that $\sigma_g$ may increase with $N$ [DAV-RAB(67), pp. 144-145]. For the present purpose, the ratio $\sigma_g / \sqrt{L}$ will be taken as a measure for the stability of $\bar{g}_L$. The standard deviation, $\sigma_g$, is not known in practice and has to be replaced by its estimate, $\hat{s}_{g,L}$. This estimate is obtained as the square-root of the sample variance

$$s^2_{g,L} = \frac{1}{L-1} \left[ \sum_{i=1}^{L} g^2(x_i^N) - L \bar{g}^2_L \right], \quad (4.C.5)$$

which is an unbiased estimator of $\sigma_g^2$, i.e., $\mathbb{E}[s^2_{g,L}] = \sigma_g^2$.

B. Variance Reduction

Often, the basic MCM is not satisfactory because the stability criterion, $\sigma_g / \sqrt{L}$ (respectively its estimate, $s_{g,L} / \sqrt{L}$), cannot be reduced to acceptable levels compared to $\bar{g}_L$ without using an excessive sample size. While the basic MCM is always useful in pilot studies for the purpose of exploring initial results, one should (and often has to) care about computational economy when considering production runs. It is thus worthwhile to investigate whether the original problem may be modified such that the variance, $\sigma_g^2$ (or its estimate,
can be reduced substantially. A variety of procedures called variance reduction techniques (VRT's) are known [HAM-HAN(64), Chapter 5], [LAW-KEL(82), Chapter 11]. The question which VRT to use in a particular problem has rarely a simple and unique answer. For a MCM using a specific VRT (method 2), the relative efficiency, \( \eta \), with respect to a reference (basic) MCM (method 1) is defined as [HAM-HAN(64), p. 51]

\[
\eta = \frac{\tau_1 \sigma_1^2}{\tau_2 \sigma_2^2},
\]

where \( \tau_1 \) and \( \tau_2 \) are units of computing time for the two methods, and \( \sigma_1^2 \) and \( \sigma_2^2 \) are the respective variances of the resulting estimates of \( J \). Given a class of problems, the most suitable VRT should be identified by comparing \( \eta \) (with the variances in (4.C.6) replaced by their respective estimates). In doing so, one must be careful to choose a typical problem, e.g., if the actual problem is multi-dimensional, results obtained for the corresponding scalar problem may not be conclusive. In (4.C.6), \( \eta \) is the product of two terms, the variance ratio, \( \sigma_1^2/\sigma_2^2 \), and the work ratio, \( \tau_1/\tau_2 \). Most VRT's achieve a variance reduction at the cost of increased workload (e.g., the number of functional evaluations doubles). A VRT is only effective if the additional work is more than compensated by a (sizable) reduction of the variance.

Obviously, whenever it is possible to perform part of the integration by analytical or simple numerical means, one should do so. An effective class of VRT's is based on the idea that the integrand can be approximated by a specific function which can be integrated analytically or numerically. The (basic) MCM is then only used to estimate the integral of the difference between the original integrand and its approximation. For multi-dimensional integrands, however, it is often difficult to identify suitable approximating functions.

Suppose that one can find a function \( g^*(x^N) \) such that
\[ |g(x^N) - \kappa g^*(x^N)| \leq \Delta, \quad \forall x^N \in \mathbb{R}^N, \]  

(4.C.7a)

where \(\kappa\) is a constant to be specified, \(\Delta \geq 0\), and the expectation of \(g^*(x^N)\),

\[
\int_{\mathbb{R}^N} p_x(x^N) g^*(x^N) dx^N = J_{g^*}, \tag{4.C.7b}
\]

is known or easily determined with high accuracy. Letting

\[
d(x^N) = g(x^N) - \kappa g^*(x^N), \tag{4.C.8}
\]

it follows that

\[
J_g = \kappa J_{g^*} + J_d, \tag{4.C.9a}
\]

where

\[
J_d = \int_{\mathbb{R}^N} p_x(x^N) d(x^N) dx^N \tag{4.C.9b}
\]

is the difference integral to be estimated. The variance in this case becomes

\[
\sigma_d^2 = J_d^2 - J_d^2 = \sigma_g^2 + \kappa^2 \sigma_{g^*}^2 - 2 \kappa \text{Cov}(g, g^*) \tag{4.C.10}
\]

\[
\leq \Delta^2,
\]

where \(\text{Cov}(\cdot, \cdot)\) denotes the covariance. Note that \(\sigma_d^2 < \sigma_g^2\) if and only if \(2 \kappa \text{Cov}(g, g^*) > \kappa^2 \sigma_{g^*}^2\), and \(\sigma_d^2\) is minimized with

\[
\kappa = \kappa_{\text{opt}} = \frac{\text{Cov}(g, g^*)}{\sigma_{g^*}^2}. \tag{4.C.11}
\]

The value of \(\kappa_{\text{opt}}\) is not known in practice, but it could be estimated using the MCM (see, e.g., [KLE(74), Section III.4] for a discussion of this method and its
Consequences). Alternatively, a suitable value for \( \kappa \) may be found empirically. For simplicity, one often chooses \( \kappa = 1 \) for \( \text{Cov}(\cdot, \cdot) > 0 \), and \( \kappa = -1 \) for \( \text{Cov}(\cdot, \cdot) < 0 \).

It follows from the above that the new sample mean given as

\[
\bar{g}_{d,L} = \kappa g^* + \frac{1}{L} \sum_{i=1}^{L} d(x_i^N) \quad (4.C.12a)
\]

is also an unbiased estimator of \( J_g \). Similarly, the sample variance given as

\[
s_{d,L}^2 = \frac{1}{L - 1} \left[ \sum_{i=1}^{L} d^2(x_i^N) - L(\bar{g}_{d,L} - \kappa g^*)^2 \right] \quad (4.C.12b)
\]

is an unbiased estimator of \( \sigma_d^2 \). The approximating function, \( g^*(x^N) \), is called the control variate for \( g(x^N) \) in the sense that \( g^*(x^N) \) is correlated with \( g(x^N) \) and that its expectation is known [HAM-HAN(64), p. 59]. This class of VRT's, as described above by (4.C.7) to (4.C.12) in its basic form, is called the control-variate MCM (CV-MCM). The CV-MCM is one of the most effective VRT's. Similar methods exist where the expectation of the control variate need not be known explicitly [HEI(80)].
CHAPTER 5

Lower Bound on the Cutoff Rate of the Discrete-Time Gaussian Channel
with Intersymbol Interference

The class of discrete-time Gaussian channels with intersymbol interference (ISI) is considered where the inputs are subject to a per symbol average-energy constraint. This channel model has been mathematically defined in Chapter 3 and a member of this class of channels has been called a discrete-time Gaussian channel (DTGC). In this chapter, the cutoff rate of the DTGC and a lower bound thereof are of interest. As in the case of channel capacity, the cutoff rate (or bounds thereof) can be defined and computed provided that the channel model includes 1) the basic channel model specifying the conditional probability for the output given some specific input, and 2) the constraints on channel usage (input constraints). In Chapter 3, the capacity of the DTGC has been obtained by means of a hypothetical channel model, called the N-circular Gaussian channel (NCGC); here, a similar approach will be developed to determine a lower bound on the cutoff rate of the DTGC. Consequently, the notation, definitions and results of Chapter 3 will be used whenever appropriate.

5.1 Background and Problem Definition

Analogous to the definition for channel capacity used in Chapter 3, the cutoff rate of the DTGC (in bits per channel input symbol) is defined as

$$ R_0(E_s) = \lim_{N \to \infty} R_{0,N}(E_s), \quad (5.1a) $$

with

$$ R_{0,N}(E_s) = \max_{\mathbf{x}} \frac{1}{N} R(\mathbf{X}[0,N]; \mathbf{Y}[0,N-1]), \quad (5.1b) $$
where the block cutoff rate

\[ R(X[0, N-1]; Y[0, N-1]) = \]

\[ -\log_2 \left[ \int \left( \sqrt{P_{Y|X}(y[0, N-1]|x[0, N-1])} q_X(x[0, N-1]) \right)^2 dy[0, N-1] \right], \]

(5.1c)

and where each dimension of the two \( N \)-dimensional integrals is evaluated over the interval \(( -\infty, +\infty)\). The maximization in (5.1b) is taken over all probability densities, \( q_X \), for the sequences \( x[0, N-1] \), satisfying the symbol-energy constraint (3.5). It is also assumed that (3.6c) holds, i.e., that the initial content of the channel memory is zero. This convention has no influence on \( R_o(E_s) \) since the channel memory, \( M \), is finite. Note that the maximization in (5.1b) is equivalent to minimizing the integral inside the logarithm of (5.1c).

Lemma 5.1: The block cutoff rate of the DTGC, as defined by (5.1c), is given by

\[ R(X[0, N-1]; Y[0, N-1]) = -\log_2 \left[ \Gamma_N(q_X) \right], \]

(5.2a)

with

\[ \Gamma_N(q_X) = \int \int q_X(x[0, N-1]) q_X(x'[0, N-1]) \times \]

\[ \times \exp \left\{ -\frac{1}{4N_o} \|x[0, N-1] - x'[0, N-1]\|_N^2 \right\} dx[0, N-1] dx'[0, N-1], \]

(5.2b)

where each dimension of the two \( N \)-dimensional integrals is evaluated over the interval \(( -\infty, +\infty)\), where \( \| \cdot \| \) denotes Euclidean norm, and where the \( N \times N \)
matrix $H_N$ in (5.2b) is determined by the coefficients of the channel's unit-sample response, $h_{[0,M]} = (h_0, h_1, ..., h_M)$, i.e.,

$$H_N = \begin{bmatrix} h_0 & h_1 & \cdots & h_M \\ h_0 & h_1 & \cdots & h_M \\ & h_0 & \cdots & h_M \\ & & \ddots & \vdots \\ & & & h_M \\ & & & & h_0 \\ & & & & & h_0 \end{bmatrix}.$$

**Proof:** It follows from (3.1) and (3.6c) that the conditional probability density

$$p_{Y|X}(y_{[0,N-1]}|x_{[0,N-1]}) = p_w(y_{[0,N-1]} - x_{[0,N-1]}H_N), \quad (5.3a)$$

where $H_N$ is given by (5.2c) and where

$$p_w(w_{[0,N-1]}) = \frac{1}{(\pi N_0)^{N/2}} \exp \left( - \frac{\|w_{[0,N-1]}\|^2}{N_0} \right) \quad (5.3b)$$

is the $N$-dimensional probability density for the i.i.d. Gaussian noise samples, $w_k, 0 \leq k < N$, having mean zero and variance $N_0/2$. Substituting (5.3) into (5.1c), interchanging the order of integration, completing the square of the resulting exponent, and integrating with respect to the sequence $y_{[0,N-1]}$, one obtains (5.2).
It follows from (5.1b) and Lemma 5.1 that

\[ R_{\alpha,N}(E_0) = -\frac{1}{N} \log_2 \left[ \min_{q_X} \Gamma_N(q_X) \right] = -\frac{1}{N} \log_2 \Gamma_N^*, \quad (5.4) \]

i.e., \( \Gamma_N^* \) denotes the minimum value of \( \Gamma_N(q_X) \) over \( q_X \) for all \( x^{[0,N-1]} \) satisfying the symbol-energy constraint (3.5) and subject to

\[ q_X(x^{[0,N-1]}) \geq 0 \quad \text{and} \quad \int q_X(x^{[0,N-1]}) \, dx^{[0,N-1]} = 1. \quad (5.5) \]

By application of the calculus of variations to this minimization problem, one obtains the integral equation

\[ \int q_X(x^{[0,N-1]}) \exp \left\{ -\frac{1}{4N^2_o} \| x^{[0,N-1]} - x^{[0,N-1]} \| H_N \|^2 \right\} \, dx^{[0,N-1]} = \lambda \quad (5.6) \]

as a necessary condition for \( q_X \) to minimize \( \Gamma_N(q_X) \) (see also [SAL-SAL(87)] where a related problem is treated). In (5.6), the Lagrange multiplier, \( \lambda \), is a constant for all \( x^{[0,N-1]} \) which belong to the admissible region defined by constraint (3.5). Moreover, \( \lambda = \Gamma_N^* \), as can be seen by substituting (5.6) into (5.2b).

The DTGC, the channel model that describes discrete-time Gaussian channels with ISI, includes the memoryless channel \( (M = 0) \) as a special case. Therefore, if the minimization problem in (5.4) could be solved for the general DTGC, then the solution for the memoryless channel would also be known. If no solution can be found for the memoryless channel, then of course the solution for the general DTGC will also remain inaccessible. Thus, it is instructive first to consider problem (5.4) for the special case of the memoryless channel.
Corollary 5.1: For the memoryless channel \( (M = 0) \), the input symbols, \( x_k \), \( 0 \leq k < N \), that minimize \( \Gamma_N(q_X) \) should be generated independently, i.e.,

\[
\Gamma_N(q_X) \Big|_{M=0} = [\gamma(q_X)]^N,
\]

(5.7a)

where

\[
\gamma(q_X) = \int \int q_X(x) q_X(x') \exp \left\{ -\frac{h_0^2}{4N_0} (x - x')^2 \right\} dx \, dx'.
\]

(5.7b)

Furthermore, \( R_o(E_s) = -\log_2 \gamma^* \), where \( \gamma^* \) is the minimum of \( \gamma(q_X) \) over all probability densities \( q_X \), such that \( \mathbb{E}[x^2] \leq E_s \) and

\[
\int q_X(x) \exp \left\{ -\frac{h_0^2}{4N_0} (x - x')^2 \right\} dx = \lambda
\]

(5.8)

for all \( x' \) which satisfy \( \mathbb{E}[x'^2] \leq E_s \).

Proof: A similar problem considered by Saleh and Salz [SAL-SAL(87)] is equivalent to the present one, except that they consider transmission of symbols which are constrained in peak-power, i.e., satisfying \( x_k^2 \leq E_s \), \( 0 \leq k < N \). This stronger constraint implies the present one which only limits the average energy of each symbol; for both constraints, however, the admissible region for each symbol \( x_k \) is independent of the remaining symbols. Thus, the proof given in the appendix of [SAL-SAL(87)] for Proposition 1 applies directly to the present case and leads to (5.7). Condition (5.8) is then obtained as a special case of (5.6) when letting \( N = 1 \).

When the memoryless Gaussian channel is used with a peak-power constraint (i.e., \( x_k^2 \leq E_s \), \( 0 \leq k < N \)), the optimizing \( q_X(\cdot) \) in (5.7) describes a discrete distribution [EIN(79)], [SAL-SAL(87)]. The optimal discrete \( q_X(\cdot) \) in this case
is a function of the signal-to-noise ratio, $E_s/N_0$, and can be determined numerically. For any other, generally weaker, input constraint, i.e., any form of average-energy constraint, the optimizing $q_x(\cdot)$ is not known. In particular, it is shown in Appendix 5.A that a Gaussian probability density function does not generally achieve $R_o(E_s)$ as defined in (5.1) for the DTGC. Chapter 3 has shown, however, that the capacity of the DTGC is achieved with Gaussian inputs even if the average block-energy constraint (3.7) is replaced by the stronger average symbol-energy constraint (3.5). These observations lead us to redefine the problem as follows.

Define the cutoff rate of the NCGC (in bits per channel input symbol) as

$$\tilde{R}_{o,N}(E_s) = \max_{q_x} \frac{1}{N} R(x^{[0,N-1]}; y^{[0,N-1]})$$

(5.9)

where $R(\cdot; \cdot)$ is given by (5.1c) with the ensemble $Y^{[0,N-1]}$ replaced by $\tilde{Y}^{[0,N-1]}$, and where the maximization in (5.9) is taken over all probability densities, $q_x$, for the sequences $x^{[0,N-1]}$, satisfying symbol-energy constraint (3.5). As shown in Appendix 5.B, the following results can now be proved.

**Theorem 5.1:** The quantity $R_{o,N}(E_s)$ as defined by (5.1b) for the DTGC, and the cutoff rate of the NCGC, $\tilde{R}_{o,N}(E_s)$, are related by

$$\left[1 - \frac{M}{N}\right] R_{o,N-M}(E_s) \leq \tilde{R}_{o,N}(E_s) \leq \left[1 + \frac{M}{N}\right] R_{o,N+M}(E_s), \quad N > M.$$ 

(5.10)

**Corollary 5.2:** By Theorem 5.1, the cutoff rate of the DTGC defined in (5.1) may also be obtained as

$$R_o(E_s) = \lim_{N \to \infty} \tilde{R}_{o,N}(E_s).$$

(5.11)
The result of Corollary 5.2 implies that it is sufficient to find a solution for $\tilde{R}_{o,N}(E_s)$ in order to determine the quantity of interest, $R_{o}(E_s)$. However, a solution for $\tilde{R}_{o,N}(E_s)$ is still difficult to obtain and a lower bound for $\tilde{R}_{o,N}(E_s)$ will be proposed in Section 5.2. A corresponding lower bound for $R_{o}(E_s)$ can then be obtained and compared with the capacity result which was derived in Chapter 3. The former is further compared with a cutoff rate parameter obtained from the (zero-rate) random-coding exponent that applies when each input sequence (codeword) is constrained in energy [GAL(68), Chapter 7]. Finally, the chapter concludes with numerical results in Section 5.3.

5.2 A Lower Bound on $R_{o}(E_s)$

Consider first the channel consisting of the basic channel model of the NCGC defined by (3.11), and the average block-energy constraint defined by (3.7). For this channel model, we define the quantity (in bits per channel input symbol)

$$
\tilde{R}_{o,N}'(E_s) = \max_{q_{x}'} \frac{1}{N} R(X^{[0,N-1]}; Y^{[0,N-1]}),
$$

(5.12)

where $R(\cdot; \cdot)$ is also given by (5.1c), but where the maximization is now taken over all probability densities, $q_{x}^{'}$, for the sequences $x^{[0,N-1]}$, satisfying the block-energy constraint (3.7). Clearly,

$$
\tilde{R}_{o,N}(E_s) \leq \tilde{R}_{o,N}'(E_s),
$$

(5.13)

since the symbol-energy constraint (3.5) is generally stronger than the block-energy constraint (3.7). Therefore, any lower bound on $\tilde{R}_{o,N}(E_s)$ is also a lower bound on $\tilde{R}_{o,N}'(E_s)$. Furthermore, equality holds in (5.13) when the symbol-energy constraint (3.5) and the block-energy constraint (3.7) are equivalent. In view of the capacity results obtained in Chapter 3, we conjecture at this point (and prove later) that equality holds in (5.13).
The well-known theorem of reversibility [WOC-JAC(65), p. 222] states that any succession of reversible transformations of the sequences \(x^{[0,N-1]}\) and \(\tilde{y}^{[0,N-1]}\) in (5.12) preserves the likelihood ratio. On the other hand, if the likelihood ratio is indeed preserved, the cutoff rate as well as the capacity of the channel are preserved [MAS(74)], [MAS(80)]. Therefore, \(\tilde{R}_{o,N}(E_S)\) as defined in (5.12) may be equivalently written as

\[
\tilde{R}_{o,N}(E_S) = \max_{Q_{X'}} \frac{1}{N} R(X^{[0,N-1]}; Y^{[0,N-1]}),
\]

(5.14)

where the maximization is over the class of probability densities \(Q_{X'}\) for all \(X'^{[0,N-1]}\) satisfying the average block-energy constraint (3.27), and where \(X^{[0,N-1]}\) and \(Y^{[0,N-1]}\) are the input- and output sequence ensembles, respectively, of the transform-domain channel model defined by (3.25). Figure 5.1 depicts this equivalent channel model in the form of \(N\) parallel, time-discrete memoryless channels. The additive noise in the \(i\)-th channel is zero-mean Gaussian with variance \(\sigma_i^2\) as given by (3.28). Furthermore, the \(N\) channels are independent in the sense that

\[
P_{X,Y}(X^{[0,N-1]}, Y^{[0,N-1]}) = \prod_{i=0}^{N-1} P_{X,Y}(X_i^{[0,N-1]}, Y_i^{[0,N-1]}).
\]

(5.15)

Therefore, the problem of determining \(\tilde{R}_{o,N}(E_S)\) reduces to finding the cutoff rate of \(N\) parallel and independent time-discrete memoryless Gaussian channels subject to the constraint

\[
\int Q_{X'}(X^{[0,N-1]}) \|X^{[0,N-1]}\|^2 dX^{[0,N-1]} \leq N^2 E_S,
\]

(5.16)
Fig. 5.1. Equivalent (basic) transform-domain channel model.
i.e., the average length of the input vector \( X^{[0,N-1]} \) is confined to an \( N \)-dimensional hypersphere of radius \( N \sqrt{E_s} \). In (5.14), no further assumptions about \( Q_X \) have been made. In particular, the optimizing \( Q_X \) may require that the individual channels in Figure 5.1 are not used independently. Consider the special case where the \( N \) channels in Figure 5.1 are identical, i.e., where the equivalent time-domain channel is memoryless \( (M = 0) \). In this case, \( N = 1 \) may be assumed (Corollary 5.1) so that the transform-domain channel and the time-domain channel are the same channel. As indicated in the previous section, the optimizing input distribution remains unknown even for this simple situation. One may therefore conclude that the transform-domain channel does not facilitate the solution of problem (5.1). However, considering the capacity results obtained in Chapter 3, the transform-domain channel model suggests how one may obtain a lower bound on \( \tilde{R}_{o,N} (E_s) \).

Assume that the individual channels in Figure 5.1 are used independently, i.e., that the input probability density has the form

\[
Q_X (X^{[0,N-1]}) = \prod_{i=0}^{N-1} Q_X (X_i^t) .
\]  
(5.17a)

Further, let the \( X_i^t \) be Gaussian random variables with mean zero and variance \( \mathbb{E}[X_i^t] = Ne_i, \ 0 \leq i < N \), i.e., let

\[
Q_X (X_i^t) = \frac{1}{\sqrt{2\pi Ne_i}} \exp \left\{ - \frac{X_i^2}{2Ne_i} \right\} ,
\]  
(5.17b)

where, according to (3.27), the average input symbol energies, \( e_i \), satisfy the block-energy constraint

\[
\sum_{i=0}^{N-1} e_i \leq NE_s.
\]  
(5.18)
A lower bound on $\tilde{R}'_{o,N}(E_d)$ is then

$$\tilde{R}'_{o,N}(E_d) = \max_{\varepsilon_i} \frac{1}{N} R(X_i^{[0,N-1]}; Y_i^{[0,N-1]}),$$

(5.19)

where the input symbols are distributed according to (5.17), and where the maximization is over all $\varepsilon_i$ satisfying block-energy constraint (5.18).

Lemma 5.2: The lower bound on $\tilde{R}'_{o,N}(E_d)$ as defined in (5.19) is given by

$$\tilde{R}'_{o,N}(E_d) = \max_{\varepsilon_i} \frac{1}{N} \sum_{i=0}^{N-1} \frac{1}{2} \log_2 \left[ 1 + \frac{\varepsilon_i}{N_o} |\tilde{H}_i|^2 \right],$$

(5.20)

where the $\tilde{H}_i$ are defined by (3.16b).

Proof: Since all of the $N$ individual channels in Figure 5.1 are statistically independent under the present assumptions, it follows from (5.1c) that

$$R(X_i^{[0,N-1]}; Y_i^{[0,N-1]}) = -\sum_{i=0}^{N-1} \log_2 \gamma_i,$$

(5.21a)

where

$$\gamma_i = \int_{-\infty}^{\infty} \left( \int_{-\infty}^{\infty} \sqrt{p_{Y_i^i|X_i^i}(X_i^i; Y_i^i)} Q_{X_i^i} dX_i^i \right)^2 dY_i^i$$

(5.21b)

with

$$p_{Y_i^i|X_i^i}(Y_i^i; X_i^i) = \frac{1}{\sqrt{2\pi \sigma_i^2}} \exp \left[ -\frac{(Y_i^i - X_i^i)^2}{2 \sigma_i^2} \right].$$

(5.21c)

Substitution of (5.17b) and (5.21c) into (5.21b) yields, after some manipulations (completion of the square of the resulting exponent and double integration),
Finally, after substitution of the right-hand side of (3.28) for $\sigma_i^2$ into (5.22) and using the result in (5.21a), one obtains expression (5.20) of Lemma 5.2.

The maximization in (5.20) is over the region specified by (5.18) and the condition $\epsilon_i \geq 0$, $0 \leq i < N$. In this region, the sum on the right-hand side of (5.20) is a convex function of the spectral energy sequence, $\epsilon_i^{[0, N-1]}$, and its maximum occurs for equality in (5.18). Thus, Gallager's Theorem 7.5.1 [GAL(68), p. 344] provides also the solution for the present maximization problem. Direct substitution of the respective quantities yields the necessary and sufficient conditions to achieve the maximum in (5.20) as

$$
\epsilon_i = \begin{cases} 
B - N_0 |\tilde{H}_i|^{-2} & , B > N_0 |\tilde{H}_i|^{-2} \\
0 & , B \leq N_0 |\tilde{H}_i|^{-2} 
\end{cases},
$$

(5.23a)

where $B$ is chosen such that

$$
\sum_{i=0}^{N-1} \epsilon_i = NE_x.
$$

(5.23b)

**Theorem 5.2:** The lower bound for $\tilde{R}_{o,N}(E_x)$, as defined in (5.19), is given by the parametric expression

$$
\tilde{R}_{o,N}^{G^*}(E_x) = \frac{1}{2N} \sum_{i=0}^{N-1} \log_2 \left[ \max \left( \Phi |\tilde{H}_i|^2 , 1 \right) \right],
$$

(5.24a)

where the $\tilde{H}_i$ are given by (3.16b), and where the parameter $\Phi$ is the solution of the equation

$$
\gamma_i = \left( \frac{2\sigma_i^2}{2\sigma_i^2 + NE_i} \right)^{\frac{1}{2}}.
$$

(5.22)
Moreover, the components $x_k$ of the equivalent time-domain input sequence, $x[0, N-1]$, are correlated Gaussian random variables with mean zero and covariances $\tilde{r}_n$, $0 \leq n < N$, given by
\[
\tilde{r}_n = E[x_{k+n}x_k]
\]
\[
= \frac{1}{N} \sum_{i=0}^{N-1} \epsilon_i \cos \left(\frac{2\pi i n}{N}\right), \quad 0 \leq k \leq k+n < N,
\]
where the components of the spectral input energy sequence, $\epsilon[0, N-1]$, satisfy
\[
\epsilon_i = \begin{cases} 
N_0 \left[\Phi - |\tilde{H}_i|^{-2}\right], & \Phi |\tilde{H}_i|^2 > 1 \\
0, & \text{otherwise}
\end{cases}
\]
In particular, $\tilde{R}_o^G(\epsilon_s)$ is achieved when the time-domain input symbols, $x_k$, have the same average energy, i.e., when $E[x_k^2] = \tilde{r}_0 = E_s$, $0 \leq k < N$.

Proof: Using (5.23a) in (5.20) and letting $\Phi = B/N_0$ yields directly the right-hand side of (5.24a). Equation (5.24b) is then obtained by combining (5.23a) and (5.23b). Lemma 3.3 implies (5.25a), and (5.25b) follows from (5.23a) with $B = \Phi N_0$. This completes the proof of Theorem 5.2.

The result of Theorem 5.2 is closely related to the capacity result obtained for the NCGC in Chapter 3 (see Theorem 3.1). In particular, the result for $\tilde{R}_o^G(\epsilon_s)$ has a similar water-filling interpretation with respect to the spectral distribution of the input energy. Moreover, the result of Theorem 5.2, which
has been derived under an average block-energy constraint, is also valid for the NCGC, since this channel also satisfies the average symbol-energy constraint of the NCGC. The main result of this section may now be stated as follows.

**Corollary 5.3:** The cutoff rate of the DTGC as defined by (5.1), \( R_o(E_s) \), is lower bounded as

\[
R_o(E_s) \geq R_o^{G^*}(E_s) ,
\]

where

\[
R_o^{G^*}(E_s) = \frac{1}{2\pi} \int_0^\pi \log_2 \left[ \max \left( \Phi |H(\lambda)|^2, 1 \right) \right] d\lambda ,
\]

and where \( H(\lambda) \) is the channel transfer function given in (3.2). The parameter \( \Phi \) is the solution of the equation

\[
\int_0^\pi \max \left( \Phi - |H(\lambda)|^{-2}, 0 \right) d\lambda = \pi E_s / N_o .
\]

Moreover, the bound-achieving input symbols, \( x_k, -\infty < k < \infty \), are correlated Gaussian random variables with mean zero and covariances \( r_n \), \( -\infty < n < \infty \), given by

\[
r_n = E[x_{k+n}x_k] \]

\[
= \frac{1}{\pi} \int_0^\pi S_x(\lambda) \cos(n\lambda) d\lambda ,
\]

where the input power spectral density satisfies

\[
S_x(\lambda) = \begin{cases} 
N_o \left[ \Phi - |H(\lambda)|^{-2} \right] , & \text{if } \Phi |H(\lambda)|^2 > 1 , \ |\lambda| \leq \pi \\
0 , & \text{otherwise} .
\end{cases}
\]
In particular, $R^G_o(E_s)$ is achieved when the input symbols have the same average energy, i.e., when $E[x_k^2] = r_0 = E_s$, $-\infty < k < \infty$.

**Proof:** By Theorem 5.2, $\tilde{R}^{G*}_{0,N}(E_s)$ is also a lower bound on the cutoff rate of the $NCGC$, $\tilde{R}_{o,N}(E_s)$, i.e.,

$$\tilde{R}_{o,N}(E_s) \geq \tilde{R}^{G*}_{o,N}(E_s) , \quad N > M .$$

Defining

$$R^G_o(E_s) = \lim_{N \to \infty} \tilde{R}^{G*}_{o,N}(E_s) ,$$

and invoking Corollary 5.2 gives directly (5.26a). The proof of the remaining results follows that given in Section 3.5 of Chapter 3.

Comparison of Corollary 5.3 and Theorem 3.2 shows that the lower bound on the cutoff rate of the DTGC, $R^G_o(E_s)$, and on the capacity of the DTGC, $C(E_s)$, are directly related as

$$R^G_o(E_s) = C(E_s/2) , \quad 0 \leq E_s/N_o < \infty .$$

Thus, the curve for $R^G_o(E_s)$ plotted as a function of $E_s/N_o$ is simply the curve for $C(E_s)$ shifted by $10 \log_{10}(2) \approx 3$ dB to the right; otherwise, both $R^G_o(E_s)$ and $C(E_s)$ exhibit the same characteristics.

Consider now the special case where the time-domain input symbols, $x_k$, $0 \leq k < N$, are i.i.d. Gaussian random variables with zero mean and variance $E[x_k^2] = E_s$. This channel model has been called the DTGC-G in Chapter 3 (Section 3.4). By Lemma 3.2, the inputs of the equivalent transform-domain channel are then also i.i.d. Gaussian random variables with zero mean and variance $E[x_i^2] = NE_s$. With respect to Lemma 5.2, it is now easy to verify
that this situation corresponds to $e_i = E_s$, $0 \leq i < N$. Thus, defining the cutoff rate of the DTGC-G as

$$R_o^G(E_s) = \lim_{N \to \infty} \left[ \tilde{R}_{0,N}^G(E_s) \bigg| e_i = E_s \right],$$

(5.31)

and using the same arguments which led to Corollary 3.6, one obtains the following result.

**Corollary 5.4:** The lower bound on the cutoff rate of the DTGC as defined by (5.29) (in units of bits/symbol), $R_o^{G*}(E_s)$, satisfies

$$\max |H(X)| \leq R_o^{G*}(E_s) \leq \frac{\max |H(\lambda)|^2}{2 \ln 2} \frac{E_s}{N_0},$$

(5.32a)

where $H(\lambda)$ is defined in (3.2), and

$$R_o^G(E_s) = \frac{1}{2\pi} \int_0^\pi \log_2 \left[ 1 + \frac{E_s}{N_0} |H(\lambda)|^2 \right] d\lambda$$

(5.32b)

$$= \lambda^G(E_s/2) \leq \frac{1}{2 \ln 2} \frac{E_s}{N_0} \|H[0,M]\|^2,$$

with $\lambda^G(\cdot)$ being the information rate of the DTGC-G (see also Section 3.4 of Chapter 3). The lower bound in (5.32a) is asymptotically tight in the limit as $E_s/N_0 \to \infty$, and the upper bound is asymptotically tight when $E_s/N_0 \to 0$. Moreover, if all DTGC’s (including the memoryless channel) satisfy the constraint (3.59), the upper bound in (5.32a) for DTGC’s with ISI memory $M > 1$ always exceeds the upper bound for the memoryless channel ($M = 0$), namely, $(E_s/N_0)/(2 \ln 2)$ bits/symbol.

The lower bound on the cutoff rate of the DTGC, $R_o^{G*}(E_s)$, may now be compared with a cutoff rate parameter that is obtained from the (zero-rate)
random-coding exponent and that applies when each input sequence (codeword) is constrained in allowed energy. Consider the basic channel model of the DTGC as defined by (3.1), where each independently chosen input sequence, \( x[0,N-1] \), satisfies

\[
\sum_{k=0}^{N-1} x_k^2 \leq NE_s .
\] (5.33)

This channel model will be denoted by \( DTGC^* \), where the asterisk is a reminder that the average symbol-energy constraint (3.5) has been replaced by the constraint (5.33). Similarly, the \( N \)-circular extension of the DTGC* model will be denoted by \( NCGC^* \). The equivalent transform-domain channel model of the \( NCGC^* \) is then given by the basic channel model of Figure 5.1, where each input vector, \( X'[0,N-1] \), satisfies

\[
\sum_{i=0}^{N-1} X_i^2 \leq N^2 E_s .
\] (5.34)

Assume now that the probability density function for these input vectors (i.e., for the ensemble of codes) is chosen as

\[
Q_{X'}^*(X'[0,N-1]) = \frac{1}{\mu} \phi(X'[0,N-1]) \prod_{i=0}^{N-1} Q_{X_i'}(X_i') ,
\] (5.35a)

where \( Q_{X_i'}(\cdot) \) is the scalar Gaussian density given by (5.17b) for equality in (5.18). The function \( \phi(\cdot) \) in (5.35a) ensures that the constraint (5.34) is satisfied and is defined as

\[
\phi(X'[0,N-1]) = \begin{cases} 
1 , & E_s - \Delta < N^{-2} \|X'[0,N-1]\|^2 \leq E_s \\
0 , & \text{otherwise ,}
\end{cases}
\] (5.35b)
where $A > 0$ and where $\mu$ is chosen such that $Q_X^*(\cdot)$ integrates to unity.  

The results obtained by Gallager [GAL(68), Chapter 7] for the random-coding exponent of parallel additive Gaussian noise channels apply directly to the present transform-domain channel model. To obtain expressions which can be integrated, Gallager [GAL(68), p. 346] replaced the function $\phi(\cdot)$ in (5.35a) by the upper bound

$$\phi(x^*[0,N-1]) \leq \exp \left[ s\Delta + \frac{s}{N^2} \sum_{i=0}^{N-1} (X_i^2 - N\epsilon_i) \right],$$  

(5.35c)

where $s \geq 0$. The random-coding exponent of this channel, denoted here by $\tilde{E}_{r,N}(R', Q')$, is then given as

$$\tilde{E}_{r,N}(R', Q') = \begin{cases} 
\max_{0 \leq \rho \leq 1} \tilde{E}_{r,N}(\rho, R', Q') , & R_c' \leq R' \leq C' \\
\tilde{E}_{r,N}(1, R_c', Q') + R_c' - R', & 0 \leq R' < R_c' ,
\end{cases}$$  

(5.36)

where $R'$ is the total rate used over the $N$ channels, where $Q'$ is the (chosen) input probability density specified by (5.35), and where $C'$ is the total capacity of the $N$ channels. $R_c'$ in (5.35c) is the rate $R'$ for which $\rho = 1$ maximizes $\tilde{E}_{r,N}(\rho, R', Q')$ over $0 \leq \rho \leq 1$. This latter quantity is determined (in bits per $N$-symbol vector) by the parametric relations

$$\tilde{E}_{r,N}(\rho, R', Q') = \frac{\rho N (E_2/N_0)}{(1 + \rho) \Psi \ln 2} - \frac{1}{2} \log_2 \left[ 1 + \rho - \frac{\rho}{\Psi |\tilde{H}_i|^2} \right],$$  

(5.37a)

---

1 In (5.35a), $\mu$ is the probability that the independently-chosen transform-domain components of $X^*[0,N-1]$ satisfy $F_i - \Delta < N^{-2} \|X^*[0,N-1]\|^2 \leq F_i$. 

---
\[ R' = \sum_{i=0}^{N-1} \frac{1}{2} \log_2 \left( \Psi | \tilde{H}_i |^2 \right), \quad (5.37b) \]

\[ \Psi | \tilde{H}_i |^2 > 1 \]

and

\[ \sum_{i=0}^{N-1} \frac{(1 + \rho)^2 \Psi \left( \Psi - | \tilde{H}_i |^{-2} \right)}{(1 + \rho) \Psi - \rho | \tilde{H}_i |^{-2}} = 2N \left( \frac{E_s}{N_o} \right), \quad (5.37c) \]

\[ \Psi | \tilde{H}_i |^2 > 1 \]

where \( R' = C' \) for \( \rho = 0 \). Thus the capacity of the NCGC*, here denoted by \( \tilde{C}_N^*(E_s) \), is given by \( C'/N \). Moreover, \( \tilde{C}_N^*(E_s) = \tilde{C}_N(E_s) \), as can be seen by letting \( \rho = 0 \) and \( \Psi = \Theta \) when comparing (5.37b) and (5.37c) with (3.16a) and (3.16c), respectively. It appears that exclusion of the subset of all input sequences (codewords) which have energy greater than \( NE \) does not reduce the capacity. Actually, although these input sequences are excluded from the ensemble defined by (5.35a) and (5.35b), they have been essentially reintroduced when \( \phi(\cdot) \) in (5.35a) was replaced by the upper bound in (5.35c). In the case where \( \rho = 0 \), the solution for \( \tilde{E}_{r,N}(\rho, R', Q') \) requires that \( s = 0 \) in (5.35c), i.e., apart from the constant factor \( \mu^{-1} \), the resulting upper bound on \( Q_{X}^*(\cdot) \) in (5.35a) is identical to \( Q_{X}^*(\cdot) \) in (5.17) for equality in (5.18).

The cutoff rate of the NCGC*, here denoted by \( \tilde{R}_{0,N}^*(E_s) \), is now obtained from (5.36) and (5.37) as

\[ \tilde{R}_{0,N}^*(E_s) = \frac{1}{N} \left[ \tilde{E}_{r,N}(1, R', Q') + R_c' \right] \]

\[ = \frac{E_s}{2N_0} + \frac{1}{2N} \sum_{i=0}^{N-1} \log_2 \left[ \frac{\Psi | \tilde{H}_i |^2}{2 - 1/(\Psi | \tilde{H}_i |^2)} \right], \quad (5.38a) \]
where the parameter $\Psi$ is uniquely determined by the equation

$$
\sum_{i=0}^{N-1} \frac{2 \left( \Psi - |\tilde{H}_i|^{-2} \right)}{2 - 1/\left( \Psi |\tilde{H}_i|^2 \right)} = N \left( E_s/N_o \right). 
$$

(5.38b)

$\Psi |\tilde{H}_i|^2 > 1$

Denoting the cutoff rate of the DTGC* by $R^*_o(E_s)$, we obtain the following result from Theorem 5.1 and Lemma 3.5.

**Corollary 5.5:** The cutoff rate of the DTGC*, $R^*_o(E_s)$, is given as

$$
R^*_o(E_s) = \lim_{N \to \infty} \tilde{R}^*_{o,N}(E_s)
$$

$$
= \frac{E_s/N_o}{2 \Psi \ln 2} + \frac{1}{2\pi} \int_0^\pi \log_2 \left[ \frac{\Psi |H(\lambda)|^2}{2 - 1/\left( \Psi |H(\lambda)|^2 \right)} \right] d\lambda, 
$$

(5.39a)

where $H(\lambda)$ is the channel transfer function in (3.2) and where the parameter $\Psi$ is the unique solution of the integral equation

$$
\int_0^\pi \frac{2 \left( \Psi - |H(\lambda)|^{-2} \right)}{2 - 1/\left( \Psi |H(\lambda)|^2 \right)} d\lambda = \pi \left( E_s/N_o \right). 
$$

(5.39b)

$\Psi |H(\lambda)|^2 > 1$

Moreover, the input symbols, $x_k$, $-\infty < k < \infty$, are appropriately chosen correlated Gaussian random variables having the spectral density

$$
S^*_x(\lambda) = \begin{cases} 
N_o \left[ \frac{2 \left( \Psi - |H(\lambda)|^{-2} \right)}{2 - 1/\left( \Psi |H(\lambda)|^2 \right)} \right], & \Psi |H(\lambda)|^2 > 1, \ |\lambda| \leq \pi \\
0, & \text{otherwise .}
\end{cases}
$$

(5.39c)
Consider now the special case where the DTGC* is memoryless \((M = 0)\) and has the unit-sample response \(h_0 = 1\). The transform-domain channel model of the corresponding NCGC* consists then of \(N\) identical channels with \(|\tilde{H}_i|^2 = 1\), \(0 \leq i < N\), so that \(|H(\lambda)|^2 = 1\) for \(|\lambda| \leq \pi\) in (5.39). For \(\alpha = E_s / N_0\), (5.39b) yields the condition

\[
\Psi_0^2 = (1 + \alpha) \Psi_0 - \frac{\alpha}{2} , \quad \Psi_0 \geq 1 ,
\]

and (5.39a) yields

\[
R_o^*(E_s) \bigg|_{M=0 \atop h_0=1} = \frac{\alpha}{2 \Psi_0 \ln 2} + \frac{1}{2} \log_2 \left[ \frac{\Psi_0^2}{2 \Psi_0 - 1} \right] ,
\]

where the solution of (5.40a) is

\[
\Psi_0 = \frac{1}{2} \left( 1 + \alpha + \sqrt{1 + \alpha^2} \right)
\]

\[
= \frac{\alpha}{1 + \alpha - \sqrt{1 + \alpha^2}} , \quad \alpha = E_s / N_0 .
\]

By substituting the second expression for \(\Psi_0^2\) in (5.41) into the non-logarithmic term of (5.40b), replacing \(\Psi_0^2\) in (5.40b) by the right-hand side of (5.40a), and by using the first expression for \(\Psi_0\) in (5.41) in the resulting argument of the logarithm, one obtains \(R_o^*(E_s)\) for the memoryless channel in the form

\[
R_o^*(E_s) \bigg|_{M=0 \atop h_0=1} = \frac{1}{2 \ln 2} \left[ 1 + (E_s / N_0) - \sqrt{1 + (E_s / N_0)^2} \right] 
\]

\[
+ \frac{1}{2} \log_2 \left[ \frac{1}{2} \left( 1 + \sqrt{1 + (E_s / N_0)^2} \right) \right] ,
\]
where the units are bits/symbol. This result coincides with the expression for $R^*_o$ as given in [WOZ-JAC(65), Eq. (5.44b)]; it is also consistent with the straight-line portion of the error exponent $E(R)$ as given in [GAL(65), Eqs. (126), (138), and (139)].

For the general DTGC*, it is clear that $R^*_o(E_s) \leq C(E_s), 0 \leq E_s/N_o < \infty$. In the case where $E_s/N_o \to 0$, it follows from (5.39b) that $\Psi^{-1} \to \max_{\lambda} |H(\lambda)|^2, |\lambda| \leq \pi$. Letting temporarily $x = \Psi |H(\lambda)|^2$ in (5.39a) and using the inequality $\ln[x^2/(2x - 1)] \leq (x - 1)^2/(2x - 1), x > 1/2$, (equality holds for $x = 1$) one can show that the integral in (5.39a) vanishes as $E_s/N_o \to 0$ and that

$$R^*_o(E_s) \approx \frac{\max_{\lambda} |H(\lambda)|^2}{2 \ln 2} \left(\frac{E_s}{N_o}\right)$$

$$\approx \frac{1}{2} C(E_s), \ E_s/N_o \ll 1,$$

where the second line follows from the upper bound in (3.68) of Corollary 3.6. The result that $R^*_o(E_s) \approx C(E_s)/2, E_s/N_o \ll 1$, is consistent with the zero-rate random-coding exponent for the class of the very noisy channels [GAL(65)]. The asymptotic relation in (5.43) is implicit in the following result, which relates the quantities $R^G_o(E_s), R^*_o(E_s), \text{ and } C(E_s)$.

**Corollary 5.6:** The lower bound on the cutoff rate of the DTGC, $R^G_o(E_s)$, the cutoff rate of the DTGC*, $R^*_o(E_s)$, and the capacity of the DTGC, $C(E_s)$, are related as

---

1. The original derivation of $E(R)$ for the memoryless channel with constraint (5.33) applied to each codeword was provided by Shannon [SHA(59)] (see also Gallager's remarks in [GAL(65), p. 16, Note 9]).

2. The class of very noisy channels is characterized by the property that the probability of receiving a given output sequence is almost independent of the input sequence (see also [VIT(67)]) and [VIT-OMU(79), Section 3.4]).
\[ \frac{1}{2} C(E_s) \leq R^G_\alpha(E_s) \leq R^*_\alpha(E_s) \leq C(E_s), \quad 0 \leq E_s/N_\alpha < \infty. \] (5.44)

Moreover, the lower bound in (5.44) is achieved asymptotically by both parameters, \( R^G_\alpha(E_s) \) and \( R^*_\alpha(E_s) \), as \( E_s/N_\alpha \to 0 \), and the upper bound is achieved as \( E_s/N_\alpha \to \infty \).

**Proof:** Because of (5.30), it is sufficient to show that \( C(E_s) \leq 2C(E_s/2) \) to prove the first inequality in (5.44). The capacity of the DTGC, \( C(E_s) \), may be written in the form

\[ C(E_s) = \frac{1}{2\pi} \int_0^{\pi} \log \left[ 1 + \frac{2S_X(\lambda)}{N_\alpha} |H(\lambda)|^2 \right] d\lambda, \] (5.45a)

where \( S_X(\cdot) \) is given by (3.20b) such that

\[ \frac{1}{\pi} \int_0^{\pi} S_X(\lambda) \, d\lambda = E_s \] (5.45b)

for some given value of \( \Theta \) in (3.20b). Using the inequality \( \log(1+x)^{1/2} \leq \log(1+x/2), \, x \geq 0 \), and defining

\[ S'_X(\lambda) = \frac{1}{2} S_X(\lambda), \] (5.46a)

one obtains from (5.45a) the inequality

\[ C(E_s) \leq \frac{1}{\pi} \int_0^{\pi} \log \left[ 1 + \frac{2S'_X(\lambda)}{N_\alpha} |H(\lambda)|^2 \right] d\lambda \]

\[ = 2C(E'_s), \] (5.46b)
where
\[ \frac{1}{\pi} \int_0^\pi S_x(\lambda) \, d\lambda = E'_s. \]  
(5.46c)

Substitution of (5.46a) into (5.46c) and comparison with (5.45b) show that
\[ E'_s = E_s / 2, \]  
and hence that \( C(E_s) / 2 < C(E'_s) = R_g^*(E_s). \)

The second inequality in (5.44) may be explained from the input constraints
code ensembles) that have been assumed when defining \( \tilde{R}_{o,N}^*(E_s) \) and
\( R_{o,N}^*(E_s), \) respectively. First, an argument will be given that shows that the
converse of the second inequality in (5.44) is not possible. By Theorem 5.2,
\( \tilde{R}_{o,N}^*(E_s) \) is obtained under the input constraint
\[ E\left[ \|X^{[0,N-1]}\|^2 \right] = E_s, \]  
(5.47a)

where \( X^{[0,N-1]} \) is the normalized input vector
\[ X^{[0,N-1]} = \frac{1}{N} X^{[0,N-1]} \]  
(5.47b)

and consists of independent Gaussian components as specified by (5.17). The
variance of the squared norm of the normalized input vector satisfies
\[ \text{Var}\left( \|X^{[0,N-1]}\|^2 \right) = \sum_{i=0}^{N-1} \text{Var}\left( X_i^2 \right) = \sum_{i=0}^{N-1} \left[ E[X_i^4] - (E[X_i^2])^2 \right] \]  
(5.48)
\[ = \frac{2}{N^2} \sum_{i=0}^{N-1} \epsilon_i^2 \leq \frac{2}{N} \hat{\epsilon}^2, \]

where \( \hat{\epsilon} = \max \{ \epsilon_i : 0 \leq i < N \} \) with \( \epsilon_i \) as given by (5.25b). The first
equality in (5.48) holds because the \( X_i^4, 0 \leq i < N, \) are statistically
independent random variables, and the second equality holds by definition. The
third equality in (5.48) follows from the facts that, for a zero-mean Gaussian random variable, \( y \), the fourth moment is \( E[y^4] = 3(E[y^2])^2 \) [PAP(81), pp. 147-148] and that \( E[X_i^2] = \varepsilon_i/N, \ 0 \leq i < N \). It follows that the expected value of the squared norm of the normalized input vector, \( E[\|X'[0,N-1]\|^2] \), is independent of \( N \), while the variance of \( \|X'[0,N-1]\|^2 \) tends to zero as \( N \to \infty \). By applying Chebyshev's inequality [PAP(81), pp. 149-151], one finds that the probability that \( \|X'[0,N-1]\|^2 \) differs from its expected value of \( E_s \) by less than \( \delta \) can be lower bounded as

\[
\Pr \left( \|X'[0,N-1]\|^2 - E_s < \delta \right) \geq 1 - \frac{2\delta^2}{N\delta^2}
\]

for any \( \delta > 0 \). Since \( \delta \) may be arbitrarily small, it follows that the random variable \( \|X'[0,N-1]\|^2 \) almost surely takes the value \( E_s \) as \( N \to \infty \). Thus \( X'[0,N-1] \) falls almost surely on the surface of a sphere with radius \( \sqrt{E_s} \) as \( N \to \infty \). This result, which is related to the sphere hardening phenomenon [WOZ-JAC(65), Section 5.5], implies that the expectation operator in (5.47a) becomes irrelevant in the limit as \( N \to \infty \). \footnote{Since \( \lim_{N \to \infty} \text{Var}(\|X'[0,N-1]\|^2) = \lim_{N \to \infty} E\left[(\|X'[0,N-1]\|^2 - E_s)^2\right] = 0 \), it follows that \( \lim_{N \to \infty} \|X'[0,N-1]\|^2 = E_s \) with probability 1.}

As \( N \to \infty \), it may be deduced from the above that the average block-energy constraint (5.18) together with the input probability density for \( X'[0,N-1] \) as defined in (5.17), \( Q_X(\cdot) \), become equivalent to definition (5.35) for the NCGC*. Thus, one may exclude the case \( R_o^{G^*}(E_s) > R_o^*(E_s) \) since, in the limit as \( N \to \infty \), the class of input sequences which satisfy constraint (5.34) for the NCGC* converges towards the class of input sequences which satisfy the apparently weaker constraint (5.18) for the NCGC. The converse, i.e., \( R_o^{G^*}(E_s) \leq R_o^*(E_s) \), may be proved by examining the upper bound on \( \phi(\cdot) \) in
which finally determines \( R^*_0(\Delta_0) \). Clearly, this bound admits codewords which are not in the ensemble specified by (5.35a) and (5.35b); thus, the resulting upper bound of \( Q^*_x(\cdot) \) violates constraint (5.34). In particular, for \( s = 0 \), the right-hand side of

\[
Q^*_x(x^*_0[N-1]) \leq \frac{e^{sA}}{\mu} \prod_{i=0}^{N-1} Q^*_x(x^*_i) \exp\left[ sN^{-2} (x^*_i^2 - \bar{\epsilon}_i) \right]
\]

becomes identical to \( Q^*_x(\cdot) \) in (5.17) when equality holds in (5.18), apart from the factor \( \mu^{-1} \) which is unimportant with respect to the random-coding exponent. The achievable exponent is thus the same under the two constraints [GAL(68), pp. 327-328]. In this situation (\( s = 0 \)) where the random-coding exponent is dominated by those (few) codewords whose normalized squared norm is substantially less than the value of \( E_\Delta \), one finds that \( R^*_0(E_\Delta) = R^*_0(E_\Delta), s = 0 \). By choosing \( s > 0 \) in (5.50), one sees that the influence of these poor input sequences (codewords) on the exponent is reduced; in fact, Gallager's results imply that \( s = N\rho_0[(1 + \rho)^2\Psi\Theta]^{-1} \) satisfies the solution for \( E_{rN}(\rho, \Theta, \Psi') \) in (5.37) [GAL(68), Section 7.5]. Since \( R^*_0(E_\Delta) \) represents the zero-rate intercept of this now improved random-coding exponent, it can be concluded that \( R^*_0(E_\Delta) > R^*_0(G_\Theta(E_\Delta)), s > 0 \).

The third inequality in (5.44) is now obvious, and the fact that \( R^*_0(E_\Delta), R^*_0(E_\Delta) \) and \( C(E_\Delta) \) increase without bound as \( E_\Delta/N_0 \to \infty \) follows by comparing (5.26b), (5.39a) and (3.19a) as \( \Phi \to \infty, \Psi \to \infty \) and \( \Theta \to \infty \), respectively. \(^1\)

Finally, the asymptotic tightness of the lower bound in (5.44) for \( E_\Delta/N_0 \ll 1 \) is implied by Corollary 5.4, Eq. (5.43), and Corollary 3.6. This completes the proof of Corollary 5.6.

\(^1\) Note from (5.39b) that in the first term of (5.39a) \( \lim_{\Psi \to \infty} \left[ (E_\Delta/N_0)^{1/\Psi} \right] = 1. \)
5.3 Numerical Results

The analytical results obtained in this chapter will be explored numerically. As in Chapter 3, it is assumed that all channels exhibit the unit-sample response energy 1 for all finite ISI memory lengths, i.e., that \( \| h^{[0,M]} \|^2 = 1, \ M \geq 0 \). The specific channel responses used to generate the numerical results are defined in Figure 3.1 (see also the remarks at the beginning of Section 3.7 in Chapter 3).

Figure 5.2 illustrates the result (5.44) of Corollary 5.6 for a channel with unit-sample response \( h^{[0,1]} = (1, \pm 1)/\sqrt{2} \). The double-logarithmic representation shows clearly the asymptotic behavior of the various quantities for small and large values of \( E_s/N_o \). In absolute terms, the difference between the rates \( R_o^G(E_s) \) and \( R_o^*(E_s) \) appears not to be significant. Figure 5.3, where \( R_o^G(E_s) \) replaces the lower bound \( C(E_s)/2 \) in the previous figure, suggests that this difference converges towards a fixed value as \( E_s/N_o \to \infty \); similarly, the difference between \( C(E_s) \) and \( R_o^*(E_s) \) appears to converge towards a constant for very large values of \( E_s/N_o \). The asymptotic relations between \( R_o^G(E_s) \), \( R_o^*(E_s) \), and \( C(E_s) \), will be investigated in more detail later. As predicted in Corollary 5.4, Figure 5.3 confirms that \( R_o^G(E_s) \) provides a lower asymptotic bound on \( R_o^G(E_s) \) in the limit as \( E_s/N_o \to \infty \).

Let \( S_x^C(\lambda), S_x^G(\lambda), \) and \( S_x^*(\lambda) \), be the input power spectral densities which achieve \( C(E_s), R_o^G(E_s), \) and \( R_o^*(E_s) \), respectively. The normalized input power spectral densities are shown in Figure 5.4 for the channel with unit-sample response \( h^{[0,1]} = (1, 1)/\sqrt{2} \), and for different values of \( E_s/N_o \). The changing characteristics of \( S_x^*(\lambda) \) with respect to \( S_x^C(\lambda) \) and \( S_x^G(\lambda) \) are of particular interest. An essentially flat input power spectral density is approached by all three (normalized) spectral densities in Figure 5.4 as \( E_s/N_o \to \infty \). This means that, for large values of \( E_s/N_o \), uncorrelated input symbols are nearly optimal with respect to capacity and cutoff rate (lower bound).

Note that \( R_o^G(E_s) \) in (5.26) may be written in the form
Fig. 5.2. The result of Corollary 5.6 illustrated for a channel with unit-sample response \( h^{[0,1]} = (1, \pm 1)/\sqrt{2} \): \( C(E_s) \), \( R_o^*(E_s) \), \( R_o^{G*}(E_s) \), and \( C(E_s)/2 \), as a function of the signal-to-noise ratio, \( E_s/N_o \).
Fig. 5.3. $C(E_s), R^*_o(E_s), R^{G*}_o(E_s),$ and $R^G_o(E_s),$ as a function of the signal-to-noise ratio, $E_s/N_o,$ for a channel with unit-sample response $h^{[0,1]} = (1, \pm 1)/\sqrt{2}.$
Fig. 5.4. Normalized input power spectral densities $S_x^C(\lambda)$, $S_x^*(\lambda)$, and $S_x^{G*}(\lambda)$, which achieve $C(E_s)$, $R_o^*(E_s)$, and $R_o^{G*}(E_s)$, respectively, for a channel with unit-sample response $h[0, 1] = (1, 1)/\sqrt{2}$, and at different values of $E_s/N_0$. 
\[ R_o^{G*}(E_s) = \frac{1}{2\pi} \int_0^{\pi} \log_2 \left[ 1 + \frac{S_x^{G*}(\lambda)}{N_0} |H(\lambda)|^2 \right] d\lambda, \quad (5.51a) \]

where \( S_x^{G*}(\lambda) \) is given by the right-hand side of (5.27b) normalized to

\[ \frac{1}{\pi} \int_0^{\pi} S_x^{G*}(\lambda) d\lambda = E_s \quad (5.51b) \]

for some value of \( \Phi \) in (5.27b). It is interesting to evaluate the right-hand side of (5.51a) when \( S_x^{G*}(\lambda) \) is replaced by the capacity-achieving input spectral density, \( S_x^C(\lambda) \), for the same average symbol energy, \( E_s \). Thus, we define the quantity

\[ R_o^C(E_s) = \frac{1}{2\pi} \int_0^{\pi} \log_2 \left[ 1 + \frac{S_x^C(\lambda)}{N_0} |H(\lambda)|^2 \right] d\lambda, \quad (5.52a) \]

where \( S_x^C(\lambda) \) equals the right-hand side of (3.20b) normalized to

\[ \frac{1}{\pi} \int_0^{\pi} S_x^C(\lambda) d\lambda = E_s. \quad (5.52b) \]

Substitution of \( S_x^C(\lambda) \) into (5.52a) yields

\[ R_o^C(E_s) = \frac{1}{2\pi} \int_0^{\pi} \log_2 \left[ \max \left( \frac{1 + \Theta |H(\lambda)|^2}{2}, 1 \right) \right] d\lambda, \quad (5.53) \]

where \( \Theta \) satisfies (5.52b). Clearly, \( R_o^C(E_s) \leq R_o^{G*}(E_s) \), \( 0 \leq E_s/N_0 < \infty \), with equality for finite, non-zero values of \( E_s/N_0 \) if and only if the channel is memoryless. For the channel specified in Figure 5.4 with \( E_s/N_0 = 1 \) (\( \equiv \) 0 dB), one obtains \( R_o^C(E_s) = 0.5350 < R_o^{G*}(E_s) = 0.5389 \) (bits/symbol); the difference
of less than 1% is insignificant. It was found for this channel that the rate ratio \( \frac{R^C_g(E_s)}{R^G_0(E_s)} \) never drops below 0.99 for any value of \( E_s/N_o \). Similar results were obtained for all channels specified in Figure 3.1, i.e., for all of these channels it was observed that

\[
0.99 \leq \frac{R^C_g(E_s)}{R^G_0(E_s)} \leq 1, \quad 0 \leq E_s/N_o < \infty ,
\]

where the lower bound is reached for intermediate values of \( E_s/N_o \) and where the upper bound is approached as \( E_s/N_o \to 0 \) or \( \infty \). \(^1\)

The lower bound on the cutoff rate of the DTGC, \( R^G_0(E_s) \), and the cutoff rate of the DTGC\(^*\), \( R^*_0(E_s) \), are shown in Figure 5.5 for the channels with unit-sample responses defined in Figure 3.1. As expected, the respective curves for the memoryless channel (\( M = 0 \), Channel 1) are crossed-over by the curves for \( R^G_0(E_s) \) and for \( R^*_0(E_s) \) of all other channels with \( M > 0 \). This result is again a direct consequence of the chosen channel response normalization (3.59), as seen previously in the case of the capacity of the DTGC, \( C(E_s) \). For completeness, Figure 5.6 shows also the lower bound on \( R^G_0(E_s) \) given in (5.32), \( R^G_0(E_s) = I_d(E_s/2) \). Using the terminology introduced in Chapter 3, \( R^G_0(E_s) \) has been called the cutoff rate of the DTGC-G (see also Eq. (5.31) and Corollary 5.4). Note that \( R^G_0(E_s) \) of the memoryless channel provides a strict upper bound on \( R^G_0(E_s) \) of any other DTGC-G with an energy-normalized unit-sample response.

Finally, the asymptotic relations between \( R^G_0(E_s) \), \( R^*_0(E_s) \), and \( C(E_s) \) will be discussed in some more details. Figure 5.7 shows the absolute difference \( C(E_s) - R^*_0(E_s) \) and the rate ratio \( C(E_s)/R^*_0(E_s) \) as functions of the signal-

\(^1\) The simple example given by Bordelon [BOR(76), pp. 20-23], in a related study of the cutoff rate (\( R_o \)) for non-white Gaussian noise channels, indicates similarly small differences (< 1%) when evaluating the lower bound expression for \( R_o \) by using the capacity-achieving energy distribution instead of the bound-achieving distribution.
Fig. 5.5. (a) Lower bound of the cutoff rate of the DTGC, $R_o^{G^*}(E_s)$, and (b) cutoff rate of the DTGC*, $R_o^*(E_s)$, as a function of the signal-to-noise ratio, $E_s/N_o$. The unit-sample responses of the channels are defined in Figure 3.1.
Fig. 5.6. Cutoff rate of the DTGC-G, $R^G_\phi(E_s)$, as a function of the signal-to-noise ratio, $E_s/N_0$, for the channels with unit-sample responses as defined in Figure 3.1.
Fig. 5.7. Comparison of $C(E_s)$ and $R_o^*(E_s)$ for the channels with unit-sample responses as defined in Figure 3.1: (a) rate difference $[C(E_s) - R_o^*(E_s)]$, and (b) rate ratio $C(E_s)/R_o^*(E_s)$, as a function of the signal-to-noise ratio, $E_s/N_0$. 
to-noise ratio, $E_s/N_0$, for the class of channels whose unit-sample responses are specified in Figure 3.1. Similarly, and for the same class of channels, Figure 5.8 shows the absolute difference $\left[ R_o^*(E_s) - R_o^{G*}(E_s) \right]$ and the rate ratio $R_o^*(E_s)/R_o^{G*}(E_s)$. The figures indicate clearly the asymptotic relations between the respective quantities as $E_s/N_0 \to 0$ or $\infty$. In particular, Figures 5.7(b) and 5.8(b) confirm the tightness of the upper and lower bounds in (5.44) of Corollary 5.6. Figure 5.8(b) indicates that $R_o^*(E_s)$ exceeds $R_o^{G*}(E_s)$ by about 13% at most, but at different values of $E_s/N_0$ for each channel. The corresponding absolute values of the rates $R_o^*(E_s)$ and $R_o^{G*}(E_s)$ may be determined from Figure 5.5.

Figure 5.7(a) suggests that, for some fixed value of $E_s/N_0$, the absolute difference between $C(E_s)$ and $R_o^*(E_s)$ is upper bounded by a finite constant which is independent of the specific unit-sample response of the channel. Likewise, Figure 5.8(a) indicates a similar behavior for the absolute difference between $R_o^*(E_s)$ and $R_o^{G*}(E_s)$. In both cases, the upper bound on the difference is reached asymptotically in the limit as $E_s/N_0 \to \infty$. The exact numerical values of these upper bounds may be derived as follows. Define the region of integration in (3.19b), (5.26c) and (5.39b) as

$$\Lambda_\mu = \left\{ \lambda : \mu |H(\lambda)|^2 > 1, \ 0 \leq \lambda \leq \pi \right\},$$

and call the resulting signal-to-noise ratios $(E_s/N_0)_\mu$, where $\mu$ stands for either $\Theta$, $\Phi$, or $\Psi$. Consider first the difference $\left[ C(E_s) - R_o^*(E_s) \right]$ for some fixed value of $E_s/N_0$, i.e., let $(E_s/N_0)_\Theta = (E_s/N_0)_\psi$. It can then be shown that the last equality requires that

$$1 \leq \Theta/\Psi \leq 2,$$
Fig. 5.8. Comparison of $R_o^*(E_s)$ and $R_o^{G*}(E_s)$ for the channels with unit-sample responses as defined in Figure 3.1: (a) rate difference $[R_o^*(E_s) - R_o^{G*}(E_s)]$, and (b) rate ratio $R_o^*(E_s)/R_o^{G*}(E_s)$, as a function of the signal-to-noise ratio, $E_s/N_0$. 
where $\Theta/\Psi$ increases continuously from one to two as $E_s/N_0$ increases from zero to infinity.  

Let $\Theta = \kappa \Psi$, where $\kappa$ is such that $(E_s/N_0)_\Theta = (E_s/N_0)_\Psi$. Assume that $|H(\lambda)|^2$ has only a finite number of zeros within the interval $0 \leq |\lambda| \leq \pi$, and define the limit region

$$\Lambda_\infty = \lim_{\mu \to \infty} \Lambda_\mu = \{ \lambda: 0 \leq \lambda \leq \pi \}.$$  \hspace{1cm} (5.57)

One then obtains

$$\left[ C(E_s) - R_o^*(E_s) \right] = \frac{\log_2(\kappa)}{2\pi} \int_{\Lambda_\Theta} d\lambda$$

$$+ \frac{1}{2\pi} \left[ \int_{\Lambda_\Theta} \log_2\left( \frac{\Psi |H(\lambda)|^2}{|H(\lambda)|^2} \right) d\lambda - \int_{\Lambda_\Psi} \log_2\left( \frac{\Psi |H(\lambda)|^2}{|H(\lambda)|^2} \right) d\lambda \right]$$

$$+ \frac{1}{2\pi} \int_{\Lambda_\Psi} \log_2\left( 2 - \frac{1}{\Psi |H(\lambda)|^2} \right) d\lambda - \frac{(E_s/N_0)_\Psi}{2 \Psi \ln 2} , \hspace{0.5cm} \Theta = \kappa \Psi .$$  \hspace{1cm} (5.58)

Taking the limit as $\Psi \to \infty$ on both sides of (5.58) yields

$$\lim_{E_s/N_0 \to \infty} \left[ C(E_s) - R_o^*(E_s) \right] = 1 - \frac{1}{2 \ln 2}$$

$$\simeq 0.27865 \hspace{0.5cm} \text{bits/symbol} ,$$  \hspace{1cm} (5.59)

where we have used the facts that $\kappa \to 2$, $\Lambda_\Theta \to \Lambda_\infty$, $\Lambda_\Psi \to \Lambda_\infty$, and

1 The lower bound in (5.56) follows directly from (3.19b) and (5.39b). The fact that $\Theta/\Psi$ is increasing with $E_s/N_0$ follows by comparing the asymptotes of $(E_s/N_0)_\Theta$ and $(E_s/N_0)_\Psi$ for fixed values of $\Theta/\Psi$.

2 Since we assume that all channels of interest have finite-length unit-sample responses, this condition is always satisfied.
\[ \lim_{\Psi \to \infty} \frac{(E_s/N_0) \Psi}{\Psi} = 1, \quad (5.60) \]

as can be seen from (5.39b). Similarly, one obtains with \( \Theta = \kappa' \Phi \), where \( \kappa' \) is such that \((E_s/N_0)_\Theta = (E_s/N_0)_\Phi\),

\[ \lim_{E_s/N_0 \to \infty} \left[ C(E_s) - R_0^{G^*}(E_s) \right] = \frac{1}{2} \text{ (bits/symbol)} \quad (5.61) \]

and the combination of this result with (5.59) yields

\[ \lim_{E_s/N_0 \to \infty} \left[ R_0^*(E_s) - R_0^{G^*}(E_s) \right] = \frac{1}{2} \left( \frac{1}{\ln 2} - 1 \right) \quad (5.62) \]

\[ \approx 0.22135 \text{ (bits/symbol)}. \]

Based on the numerical results shown in Figures 5.7(a) and 5.8(a), one is led to conjecture that the differences between the respective quantities are monotonically increasing functions of \( E_s/N_0 \). Under this hypothesis, the following conjecture can be stated.

**Conjecture 5.1:** The upper bounds on the differences between the quantities \( C(E_s), R_0^*(E_s), \) and \( R_0^{G^*}(E_s) \), i.e., on the quantities \( \left[ C(E_s) - R_0^*(E_s) \right], \)

\( \left[ R_0^*(E_s) - R_0^{G^*}(E_s) \right], \) and \( \left[ C(E_s) - R_0^{G^*}(E_s) \right], \) are finite constants which are independent of the specific (finite-length) unit-sample response of the channel. In particular, these differences are bounded as

\[ 0 \leq \left[ C(E_s) - R_0^*(E_s) \right] \leq 1 - \frac{1}{2 \ln 2} \text{ (bits/symbol)}, \quad (5.63) \]

\[ 0 \leq \left[ R_0^*(E_s) - R_0^{G^*}(E_s) \right] \leq \frac{1}{2} \left( \frac{1}{\ln 2} - 1 \right) \text{ (bits/symbol)}. \quad (5.64) \]
and

\[ 0 \leq \left[ C(E_s) - R_o^{G^*}(E_s) \right] \leq \frac{1}{2} \text{ (bits/symbol)}. \quad (5.65) \]

The lower bounds are tight as \( E_s/N_o \to 0 \), and the upper bounds are reached asymptotically as \( E_s/N_o \to \infty \). Furthermore, the capacity of the DTGC, \( C(E_s) \), and the cutoff rate of the DTGC as defined by (5.1), \( R_o(E_s) \), differ at most by 1/2 bit per symbol, i.e.,

\[ C(E_s) - \frac{1}{2} \leq R_o(E_s) \leq C(E_s), \quad 0 \leq E_s/N_o < \infty, \quad (5.66) \]

where the units are bits per symbol.
It is shown that the Gaussian probability density function does not generally achieve \( R_0(E_s) \) for the DTGC as defined in (5.1). The memoryless \((M = 0)\) channel is considered first. The result is then shown to hold for the general DTGC as well.

Let \( a = h_0^2/(4N_o) \) so that condition (5.8) becomes

\[
\int_{-\infty}^{\infty} q_X(x) \exp\left[ -a(x - x')^2 \right] dx - \lambda = 0 \tag{5.A.1}
\]

for all \( x' \) which satisfy \( \mathbb{E}[x'^2] \leq E_s \). Assume now that the input symbols are distributed according to the scalar Gaussian density

\[
q_X(x) = \frac{1}{\sqrt{2\pi \sigma^2}} \exp\left[ -\frac{x^2}{2\sigma^2} \right], \tag{5.A.2}
\]

where \( 0 < \mathbb{E}[x^2] = \sigma^2 \leq E_s \). Substitution of (5.A.2) into (5.A.1), completion of the square in the resulting exponent, and evaluation of the integral, yield the condition

\[
\frac{1}{\sqrt{1 + \frac{h_0^2}{2N_o}(\sigma^2/N_o)}} \exp\left[ -\frac{(h_0^2/2)^2}{1 + \frac{h_0^2}{2}(\sigma^2/N_o)} \cdot (x'^2/N_o) \right] - \lambda = 0. \tag{5.A.3}
\]

Since \( \lambda \) is a constant for some given value of \( E_s/N_o \), it follows that (5.A.3) cannot be satisfied for all \( x' \) which belong to the admissible region defined by (5.A.2). This result implies that the Gaussian probability density function does not satisfy the necessary condition (5.8) for the minimization of \( \gamma(q_X) \); thus, \( R_0(E_s) \) is not achieved for this type of density function.
Consider now the general DTGC where the optimizing $q_X$ is required to satisfy condition (5.6). For simplicity, the dimension indices in the vector and matrix notation will be omitted in the following. It is assumed that the input symbols are distributed according to the $N$-dimensional Gaussian density

$$q_X(a) = \frac{1}{(2\pi)^{N/2} \|\Sigma\|^{1/2}} \exp\left[-\frac{1}{2} a \Sigma^{-1} a^T\right], \quad (5.A.4)$$

where $\Sigma = [\sigma_{ij}^2]$, $0 \leq i, j < N$, is the $N \times N$ covariance matrix such that $0 < \mathbb{E}[a_i^2] = \sigma_{ii}^2 \leq E_s$, $0 \leq i < N$, $\|\cdot\|$ denotes the determinant of the matrix, and $T$ means transpose. Substituting (5.A.4) into (5.6) and applying the rules of matrix algebra when completing the square of the resulting exponent, one obtains after integration the condition

$$\frac{\|Z\|^{1/2}}{\|\Sigma\|^{1/2}} \exp\left[-\frac{1}{2} x^T Q^{-1} x \right] - \lambda = 0, \quad (5.A.5a)$$

where, with the definition

$$A = A^T = \frac{1}{4N_0} H_N H_N^T, \quad (5.A.5b)$$

the inverse of matrix $Z$ is given as

$$Z^{-1} = 2A + \Sigma^{-1}, \quad (5.A.5c)$$

and

$$Q^{-1} = 2(A - 2AZA). \quad (5.A.5d)$$

Matrix $H_N$ in (5.A.5b) is defined by (5.2c). Since $\lambda$ is a constant for some given value of $E_s/N_0$, it follows that condition (5.A.5) cannot be satisfied for all $x' \equiv x[0, N-1]$ which belong to the admissible region defined by (5.A.4). Thus, $R_o(E_s)$ as defined by (5.1) is not achieved when $q_X$ is Gaussian. Note that condition (5.A.5) reduces to (5.A.3) for $N = 1$. 
Proof of Theorem 5.1:

Theorem 5.1 is proved with the help of the following two lemmas.

**Lemma 5.B.1:** Assume a DTGC which outputs the sequence

\[ y[0,N+M-1] = x[0,N-1] \hat{H}_N + w[0,N+M-1], \quad (5.1) \]

where the input sequence \( x[0,N-1] \) is chosen according to a specific probability density \( q_x \). The noise sample sequence, \( w[0,N+M-1] \), consists of i.i.d. Gaussian samples with mean zero and variance \( N_0/2 \). The \( N \times (N+M) \) matrix

\[
\hat{H}_N = \begin{bmatrix}
H_N & 0 \\
\vdots & \ddots \\
0 & \cdots & G_M
\end{bmatrix}, \quad (5.2a)
\]

where the \( N \times N \) matrix \( H_N \) is defined in (5.2c) and the \( M \times M \) matrix \( G_M \) is given as

\[
G_M = \begin{bmatrix}
h_M & 0 & \cdots & 0 \\
\cdot & h_M & \cdots & \cdot \\
\cdot & \cdots & \cdots & \cdot \\
h_1 & h_2 & \cdots & h_M
\end{bmatrix}, \quad (5.2b)
\]

The block cutoff rate for this modified DTGC is then obtained as
\[ R(X^{[0,N-1]}; Y^{[0,N+M-1]}) = -\log_2 \left[ \hat{\Gamma}_N(q_X) \right], \quad (5.B.3) \]

where \( \hat{\Gamma}_N(\cdot) \) is given by the right-hand side of (5.2b) with \( H_N \) replaced by \( \hat{H}_N \).

Furthermore, the block cutoff rate of the DTGC as determined by (5.2) in Lemma 5.1, \( R(X^{[0,N-1]}; Y^{[0,N-1]}) \), is upper bounded as

\[ R(X^{[0,N-1]}; Y^{[0,N-1]}) \leq R(X^{[0,N-1]}; Y^{[0,N+M-1]}) \], \quad (5.B.4) \]

where equality holds only if the channel has no ISI memory (\( M = 0 \)).

**Proof:** Replace the output ensemble \( Y^{[0,N-1]} \) in (5.1c) by \( Y^{[0,N+M-1]} \) and denote the resulting argument of the logarithm by \( \hat{\Gamma}_N(q_X) \). Following the proof for Lemma 5.1, one can verify that \( \hat{\Gamma}_N(q_X) \) is given by the right-hand side of (5.2b) with \( H_N \) replaced by \( \hat{H}_N \). Inequality (5.B.4) follows since

\[
\| (x^{[0,N-1]} - x^{[0,N-1]}) \|_{\hat{H}_N}^2 = \\
\| (x^{[0,N-1]} - x^{[0,N-1]}) \|_{H_N}^2 + \| (x^{[N-M,N-1]} - x^{[N-M,N-1]}) G_M \|^2 \\
\geq \| (x^{[0,N-1]} - x^{[0,N-1]}) \|_{H_N}^2. \tag{5.B.5}
\]

The term containing matrix \( G_M \) vanishes when \( M = 0 \), i.e., (5.B.5) holds with equality for all \( x^{[0,N-1]} \neq x^{[0,N-1]} \) only if the channel is memoryless. It then follows from (5.2b) and (5.B.5) that

\[ \hat{\Gamma}_N(q_X) \leq \Gamma_N(q_X), \quad (5.B.6) \]

and one obtains (5.B.4) from (5.B.3) and (5.2a). This proves Lemma 5.B.1.
Consider next the \(NCGC\), where \(N > M\), and in (5.9) let
\[
R(X^{[0,N-1]}, Y^{[0,N-1]}) = -\log_2 \left( \hat{\Gamma}_N(q_X) \right). 
\] (5.B.7)

By Lemma 5.1, \(\hat{\Gamma}_N(\cdot)\) is then also given by the right-hand side of (5.2b), but with the \(N \times N\) matrix \(H_N\) being replaced by the \(N \times N\) circular matrix
\[
\hat{H}_N = \begin{bmatrix}
h_0 & h_1 & \cdots & h_M & O \\
h_0 & h_1 & \cdots & h_{M-1} & h_M \\
h_0 & \cdots & h_{M-1} & \cdots & \cdots \\
h_M & h_0 & \cdots & \cdots & O \\
O & O & \cdots & O & O \\
\end{bmatrix}. 
\] (5.B.8)

For the DTGC, assume (for notational convenience) that the first input symbol is transmitted at time instant \(k = -M\). Define the \((N+M) \times (N+M)\) matrix
\[
H_{-M,N} = \begin{bmatrix}
Q_M & G_M & O \\
O & H_N & O \\
\end{bmatrix} = \begin{bmatrix}
Q_M \\
O \\
\end{bmatrix} \begin{bmatrix}
\bar{H}_N \\
O \\
\end{bmatrix}, 
\] (5.B.9a)

where the \(N \times N\) matrix \(H_N\) is defined in (5.2c), the \(M \times M\) matrix \(G_M\) is defined in (5.B.2b), and the \(M \times M\) matrix \(Q_M\) is
The resulting matrix $\bar{H}_N$ in (5.9a) has thus size $(N + M) \times N$. By Lemma 5.1, it then follows that the block cutoff rate for this DTGC is given by

$$R(X[-M,N-1]; Y[-M,N-1]) = -\log_2 \left[ \Gamma_{-M,N}(q_X) \right] \quad (5.10a)$$

with

$$\Gamma_{-M,N}(q_X) = \int \int q_X(x[-M,N-1]) q_X(x'[-M,N-1]) \times$$

$$\times \exp \left\{ -\frac{1}{4N_0} \| (x[-M,N-1] - x'[-M,N-1]) H_{-M,N} \|^2 \right\} \, dx[-M,N-1] \, dx'[-M,N-1].$$

(5.10b)

One may thus write

$$R_{0,N+M}(E_3) = \max_{q_X} \frac{1}{(N + M)} R(X[-M,N-1]; Y[-M,N-1]),$$

(5.10c)

where it is assumed that $x_k = 0$, $-2M \leq k < M$, but where the notation on the left-hand side is used in the sense of definition (5.1b). This equivalence is possible since the channel is assumed to be stationary.

Consider now a modified version of the DTGC just described. Assume that its inputs satisfy the additional constraint $x_k = x_{N+k}$, $-M \leq k < 0$, i.e., that $x[-M,-1] = x[N-M,N-1]$. However, the subsequence $y[-M,-1]$, which now
contains additional information about the input subsequence \( x^{[N-M,N-1]} \), is not made available at the channel output. For this modified DTGC, let \( \tilde{q}_X \) denote the class of probability densities for the sequence \( x^{[0,N-1]} \) that satisfy the symbol-energy constraint (3.5), and denote by \( \tilde{q}'_X \) a specific member of this class. Clearly, for any \( \tilde{q}'_X \), the block cutoff rate of this channel model is always upper bounded by the \textit{maximized} block cutoff rate of the original DTGC as determined by (5.B.10). Furthermore, the additional input constraint that \( x_k = x_{N+k} \), \(-M \leq k < 0\), implies that

\[
y^{[0,N-1]} = x^{[-M,N-1]} \tilde{H}_N + w^{[0,N-1]}
\]

(5.B.11)

as can be seen from (5.B.8) and (5.B.9a) by noting that the matrix \( G_M \) in (5.B.9a) represents the lower-left triangular submatrix in matrix \( \tilde{H}_N \). In other words, with the additional constraint, linear convolution coincides with circular convolution for the time instants \( k = 0, 1, \ldots, N-1 \). Thus, the following result has been proved.

\textbf{Lemma 5.B.2:} The maximized block cutoff rate as defined in (5.B.10) of the DTGC is lower bounded as

\[
\left[ N + M \right] R_{o,N+M}(E_s) = \max_{\tilde{q}_X} R(x^{[-M,N-1]}, y^{[-M,N-1]}) \\
\geq \max_{\tilde{q}_X} \left\{ \right. \\
\left\{ x_k = x_{N+k}, -M \leq k < 0 \right\} \\
\max_{\tilde{q}_X} R(x^{[0,N-1]}, \tilde{Y}^{[0,N-1]}) \\
= \max_{\tilde{q}_X} R(x^{[0,N-1]}, \tilde{Y}^{[0,N-1]})
\]
Theorem 5.1 may now be proved as follows.

Proof of Theorem 5.1: The lower bound in (5.10) is proved first. For the DTGC, let \( q_X' \) represent the class of probability densities for the sequence \( x^{[0, N-M-1]} \) that satisfy the symbol-energy constraint (3.5), as required by definition (5.1b) when \( N \) is replaced by \( (N - M) \). Further, let \( q_X^* \) be the specific probability density in this class which achieves \( R_{o,N-M}(E_3) \), i.e., let

\[
R(X^{[0,N-M-1]}; Y^{[0,N-M-1]}) |_{q_X^*} = [N - M] R_{o,N-M}(E_3) \tag{5.B.12}
\]

For the NCNGC, let \( q_X \) represents the class of probability densities that satisfy the symbol-energy constraint (3.5), as required by definition (5.9). Then, from definition (5.9), one finds that

\[
N R_{o,N}(E_3) \geq \max_{\{ q_X : x_k = 0, \ N-M \leq k < N \}} R(X^{[0,N-1]}; \tilde{Y}^{[0,N-1]}) \tag{5.B.13}
\]

\[
= \max_{q_X} R(X^{[0,N-M-1]}; Y^{[0,N-1]}) \geq R(X^{[0,N-M-1]}; Y^{[0,N-1]}) |_{q_X^*} \geq R(X^{[0,N-M-1]}; Y^{[0,N-M-1]}) |_{q_X^*} = [N - M] R_{o,N-M}(E_3), \tag{5.B.14}
\]
where the first inequality holds because of the additional input constraint and the following equality holds since circular convolution coincides with linear convolution for this constraint. The second inequality in (5.B.14) follows because \( q'_X \) may not be the maximizing density in the previous line. The last inequality is implied by Lemma 5.B.1 upon replacing \( N \) by \( (N - M) \) in all expressions of the lemma. Finally, the last line in (5.B.14) holds by definition (5.B.13). This proves the lower bound in (5.10).

To prove the upper bound in (5.10), suppose that \( q^* \) is the specific probability density function for the sequence \( x^{[0,N-1]} \) that achieves \( \tilde{R}_{o,N}(E_s) \) according to definition (5.9), i.e., that

\[
R(X^{[0,N-1]}, Y^{[0,N-1]}) \bigg|_{q_X^*} = N \tilde{R}_{o,N}(E_s). \tag{5.B.15}
\]

Consider now the DTGC which transmits the first input symbol at time instant \( k = -M \) as was described above. From Lemma 5.B.2 and definition (5.B.15), one concludes immediately that

\[
\left[ N + M \right] R_{o,N+M}(E_s) \geq R(X^{[0,N-1]}, Y^{[0,N-1]}) \bigg|_{q_X^* = q_X^*} = N \tilde{R}_{o,N}(E_s). \tag{5.B.16}
\]

This proves the upper bound in (5.10) and completes the proof of Theorem 5.1.
CHAPTER 6

The Symmetric Cutoff Rate for Binary Inputs

We consider the discrete-time Gaussian channel with ISI and i.i.d. binary inputs of equal probability as described in Chapter 4 (Figure 4.1). For this channel model, we wish to derive and compute the so-called symmetric cutoff rate, \( R^b_0(E_s) \), defined as

\[
R^b_0(E_s) = \lim_{N \to \infty} R^b_{0,N}(E_s), \quad (6.1a)
\]

where the superscript \( b \) stands for i.i.d., equally-likely binary inputs, and where

\[
R^b_{0,N}(E_s) = -\frac{1}{N} \log \left[ \int \left( \sum_{i=0}^{K-1} \sqrt{p_{Y|X}(y^N|x_i^N)} Q_x(x_i^N) \right)^2 dy^N \right], \quad K = 2^N. \quad (6.1b)
\]

Each dimension of the \( N \)-dimensional integral in (6.1b) is evaluated over \((-\infty, +\infty)\), and \( p_{Y|X}(\cdot|\cdot) \) and \( Q_x(\cdot) \) are given in (4.6) and (4.7), respectively.

**Lemma 6.1:** An expression for \( R^b_{0,N}(E_s) \), as defined in (6.1b), is given by

\[
R^b_{0,N}(E_s) = -\frac{1}{N} \log \left[ \frac{1}{K^2} \sum_{i=0}^{K-1} \sum_{k=0}^{K-1} \exp \left\{ -\frac{(E_s / N_o) \| d_{ik}^N \|^2 / 4} \right\} \right], \quad (6.2)
\]

where the Euclidean distance vector, \( d_{ik}^N \), is defined in (4.8).

**Proof:** The argument of the logarithm in (6.1b) becomes

\[\cdots\]

\[\cdots\]

1 See Chapter 2 or [IEEE(76)] for an explanation of the term symmetric.
after interchanging the order of summation and integration and upon substitution of (4.6) and (4.7a). Using (4.7b) and the notation introduced in (4.9) together with the identity

\[ \|a^N - b_i^N\|^2 + \|a^N - b_k^N\|^2 = 2\|a^N - (b_i^N + b_k^N)/2\|^2 + \|b_i^N - b_k^N\|^2/2 \]

and definition (4.8), one finds that the integral evaluates to the exponential under the double sum in (6.2). This completes the proof of Lemma 6.1.

In the following, a method to evaluate \( R_o^b(E_g) \) is developed which is based in part on an approach used by Omura [VIT-OMU(79), Section 5.8] for evaluating the ensemble performance for the class of time-varying convolutional codes used on the Gaussian ISI channel. \(^1\) Here, a computationally efficient procedure is developed by making use of certain symmetry properties of the finite-state channel model. Simple bounds for \( R_o^b(E_g) \) are derived and examples are given to illustrate the method. The algorithms are then applied to compute \( R_o^b(E_g) \) for various types of ISI channels; in particular, channels that minimize \( R_o^b(E_g) \) are defined and identified.

6.1 Evaluation of \( R_o^b(E_g) \)

Let \( \gamma_N \) be the argument of the logarithm in (6.2) of Lemma 6.1 and note that we may express it as an expectation of the exponential, i.e.,

\[ \frac{1}{K^2} \sum_{i=0}^{K-1} \sum_{k=0}^{K-1} \int \sqrt{p_w(y^N - \gamma_i^N)p_w(y^N - \gamma_k^N)} \, dy^N \]
\[ \gamma_N = \mathbb{E} \left[ \exp \left\{ - \left( \frac{E_s}{N_o} \right) \left\| \mathbf{d}_{mn}^N \right\|^2 / 4 \right\} \right], \quad 0 \leq m, n < K, \quad (6.3) \]

where the \( \mathbf{d}_{mn}^N \) are all equally likely with probability \( K^{-2} \). Defining

\[ e_{km}^{mn} = (x_k^m - x_k^n) / (2 \sqrt{E_s}) \]

\[ = \begin{cases} 
1, & x_k^m = -x_k^n = + \sqrt{E_s} \\
0, & x_k^m = x_k^n = \pm \sqrt{E_s} \\
-1, & x_k^m = -x_k^n = - \sqrt{E_s} 
\end{cases} \quad (6.4) \]

and using (4.3b) and (4.8) yields the components of \( \mathbf{d}_{mn}^N \) in the form

\[ d_{km}^{mn} = 2 \sum_{i=0}^{Q} h_i e_{k-i}^{mn} , \quad 0 \leq k < N. \quad (6.5a) \]

The squared norm of \( \mathbf{d}_{mn}^N \) is given by

\[ \left\| \mathbf{d}_{mn}^N \right\|^2 = 4 \sum_{k=0}^{N-1} \left[ h_0 e_k^{mn} + \sum_{i=1}^{Q} h_i e_{k-i}^{mn} \right]^2, \quad (6.5b) \]

where \( Q = \min(M, k) \) and where the second sum in (6.5b) vanishes for \( Q = 0 \). The error components in (6.4), \( e_k \in \{1, 0, -1\} \), are independent of each other. The expectation in (6.3) can now be taken over the \( 3^N \) error sequences, \( \mathbf{e}_i^N = (e_0^i, e_1^i, \ldots, e_{N-1}^i) \), \( 0 \leq i < 3^N \), each sequence having probability

\[ Q_{F}(\mathbf{e}^N) = \prod_{k=0}^{N-1} q(e_k), \quad (6.6a) \]

where
Define the state sequence

$$\sigma_k = (e_{k-1}, e_{k-2}, \ldots, e_{k-M}), \quad 0 \leq k < N,$$

and let

$$F(e_k, \sigma_k) = q(e_k) \exp \left\{ -\left( \frac{E_z}{N_o} \right) \left[ h_0 e_k + \sum_{i=1}^{Q} h_i e_{k-i} \right] \right\}, \quad 0 \leq k < N.$$  \hfill (6.8)

The $F(e_k, \sigma_k)$ may be thought of as being generated by the finite-state machine (FSM) in Figure 6.1. The expectation in (6.3) can now be written as

$$\gamma_N = \sum_{i=0}^{J-1} \prod_{k=0}^{N-1} F(e_i^k, \sigma_k),$$  \hfill (6.9)

where $J = 3^N$ and $e_i^k$ is the $k$-th component of $e_i^N$, $0 \leq i < J$. The products in (6.9), which are formed from $N$ outputs of the FSM in Figure 6.1, are obtained by applying the ensemble of sequences $E = \{ e_i^N : 0 \leq i < J \}$ to its input, with the machine starting from state $\sigma_0 = (0, 0, \ldots, 0)$ for each sequence. Let the $L = 3^M$ distinct states of the machine in Figure 6.1 be $\Sigma_j$, $-P \leq j \leq P$, where $P = (L-1)/2$ and $\Sigma_0 = (0, 0, \ldots, 0)$. Then, given the present state, $\sigma_k$, and the present input, $e_k$, define a shift operation which produces the next state, $\sigma_{k+1}$, in the manner

$$\sigma_{k+1} = \gamma(e_k, \sigma_k) = (e_k, e_{k-1}, \ldots, e_{k+1-M}), \quad 0 \leq k < N.$$  \hfill (6.10)
Fig. 6.1: The finite-state machine (FSM) generating $F(e_k, \sigma_k)$, $0 \leq k < N$, as defined by Equation (6.8). For each input sequence, $e^N = (e_0, e_1, \ldots, e_{N-1})$, the starting state $\sigma_0 = (0, 0, \ldots, 0)$. 
The FSM in Figure 6.1 generates a regular Markov chain. It is possible to drive the machine, with nonzero probability, into any state, $\Sigma_j$, $-P \leq j \leq P$, with at most $N = M$ inputs, no matter what the starting state, $\sigma_0$. Without loss of generality, one may assume that $N \geq M$ in (6.9). The ensemble of error sequences defined above, $E$, can be partitioned into $L = 2P + 1$ subsets $E_j = \{e_i^N : i \ni \sigma_N = \Sigma_j\}$, $-P \leq j \leq P$, i.e., all of the $J/L = 3^{N-M}$ sequences, $e_i^N \in E_j$, drive the FSM into the same terminating state, $\sigma_N = \Sigma_j$, after acceptance of the last input, $e_{N-1}^i$. Equation (6.9) may then be written in the form

$$\gamma_N = \sum_{j=-P}^{P} \sum_{i=0}^{J/L-1} \left[ F(e_i^1, \sigma_0 = \Sigma_0) \prod_{k=1}^{N-1} F(e_k^i, \sigma_k) \right]. \tag{6.11}$$

Define the $L \times L$ matrix ($L = 3^M = 2P + 1$)

$$A = [a_{mn}] , \quad -P \leq m, n \leq P , \tag{6.12a}$$

with elements

$$a_{mn} = \begin{cases} F(e, \Sigma_n) , & \text{if } \Sigma_m = g(e, \Sigma_n) \text{ for some } e \in \{-1, 0, 1\} \\ 0 , & \text{otherwise} \end{cases} \tag{6.12b}$$

Matrix $A$ is the state transition matrix of the machine in Figure 6.1; according to the underlying state diagram, it has only three nonzero elements in each row and each column. Consider now the $N$-th power of matrix $A$, i.e., consider the $L \times L$ matrix

$$A^N = [a_{ts}^N] , \quad -P \leq t, s \leq P , \tag{6.13a}$$

where the elements are given as
The matrix elements, \( a_{ts}^N \) in (6.13), represent the sum of all the products obtained from \( N \geq M \) outputs of the machine in Figure 6.1 that are generated by those \( J/L \) input sequences, \( e_i^N \), which drive the machine into the same terminating state, \( \sigma_N = \Sigma_I \), given that, for each input sequence, the starting state is \( \sigma_0 = \Sigma_s \). Define the \( L \)-component row vector

\[
u = [1, 1, \ldots, 1],
\]

and the \( L \)-component column vector

\[
i_0 = [0, 0, \ldots, 0, 1, 0, \ldots, 0]^T,
\]

where \( T \) means transpose and the position of the unit element corresponds to the column index \( s = 0 \) of matrix \( A^N \), i.e., the column which corresponds to the starting state, \( \Sigma_0 \). It then follows from (6.13) and (6.14) that (6.11) can be written in the form

\[
\gamma_N = \sum_{j=-p}^{P} a_{j0}^N = \nu A^N i_0.
\]

Lemma 6.2: The \( L \times L \) matrix \( A \), as defined in (6.12), is a real and irreducible (also called regular) matrix with nonnegative elements, i.e., \( a_{mn} \geq 0 \), \(-P \leq m, n \leq P\). In particular, \( a_{00} = F(0, \Sigma_0) = q(0) = 1/2 \).
Proof: Assuming that $E_j/N_0 < \infty$, one obtains immediately from (6.12b), (6.8) and (6.6b) that $a_{mn} \geq 0$, $-P \leq m,n \leq P$. Matrix $A$ is irreducible if and only if, for some $N$, the $N$-th power of $A$, $A^N$, has only positive elements, i.e., in (6.13b) one must have $a_{ts}^N > 0$, $-P \leq t,s \leq P$; otherwise, the matrix is called reducible (see [KEM-SNE(60), Theorem 4.1.2] or [GAN(59), p. 63]). Thus, $A$ is irreducible if and only if it is possible to reach each distinct state of the FSM in Figure 6.1, $-P \leq j \leq P$, from any state in some finite number of $N$ steps (including the case $\sigma_N = \sigma_0$). From the underlying state transition diagram, it follows that each distinct state can be reached in at most $N = M$ steps starting from any state (the inputs, $e_k \in \{-1,0,1\}$, $0 \leq k < N$, are independent of each other and of the present state, $\sigma_k$, and they occur with nonzero probability). Therefore, in (6.13b), $a_{ts}^N > 0$, $-P \leq t,s \leq P$, for all $N \geq M$, i.e., the underlying state transition diagram represents a strongly connected graph. The fact that $\alpha_{00} = 1/2$ follows from (6.8) and (6.6). This completes the proof of Lemma 6.2.

With Lemma 6.2, the Perron-Frobenius theorem [GAN(59), Section 111.2], [LAN(69), Chapter 9] may now be applied to matrix $A$. First, note that $A$ is primitive since $a_{ts}^M > 0$, $-P \leq t,s \leq P$ [GAN(59), p. 97, Theorem 8]. According to the Perron-Frobenius theorem, any $L \times L$ matrix that satisfies the conditions of Lemma 6.2 has a real and positive eigenvalue, $\Gamma$, that has multiplicity one and that exceeds the moduli of all other eigenvalues. $\Gamma$ is called the dominant eigenvalue of $A$. The eigenvector (left or right) of the matrix corresponding to this dominant eigenvalue has only real and positive components. Let $c = [c_p, c_{p+1}, \ldots, c_P]$, $c_i > 0$, $-P \leq i \leq P$, be the left (row) eigenvector which is associated with $\Gamma$, i.e.,

$$cA = \Gamma c.$$  \hspace{1cm} (6.16)
Lemma 6.3: Let

$$\rho = \frac{c_{\text{max}}}{c_{\text{min}}} ,$$  \hspace{1cm} (6.17a)

where $c_{\text{max}}$ ($c_{\text{min}}$) is the largest (smallest) component of the positive left eigenvector, $c$, associated with the dominant eigenvalue, $\Gamma$, of matrix $A$. Then, $\rho \geq 1$ and, for any $N \geq M$,

$$\frac{1}{\rho} \Gamma^{-N} \leq \gamma_N \leq \rho \Gamma^{-N} ,$$  \hspace{1cm} (6.17b)

or equivalently,

$$- \frac{1}{N} \log \rho - \log \Gamma \leq R_{0,N}^b(E_2) \leq \frac{1}{N} \log \rho - \log \Gamma .$$  \hspace{1cm} (6.17c)

Proof: With (6.16) one obtains the relation

$$c A^N = \Gamma^N c .$$  \hspace{1cm} (6.18)

Upper bounding the $i$-th (unity) component of $u$ in (6.14a) by $c_i / c_{\text{min}}$, $-P \leq i \leq P$, yields

$$u A^N i_0 \leq \frac{1}{c_{\text{min}}} c A^N i_0 = \frac{\Gamma^N}{c_{\text{min}}} \langle c , i_0 \rangle \leq \rho \Gamma^{-N} .$$  \hspace{1cm} (6.19a)

The last inequality holds because the inner product $\langle c , i_0 \rangle = c_0 \leq c_{\text{max}}$, and $\rho$ is defined in (6.17a). Similarly, lower bounding the $i$-th component of $u$ by $c_i / c_{\text{max}}$, $-P \leq i \leq P$, one obtains

$$u A^N i_0 \geq \frac{1}{c_{\text{max}}} c A^N i_0 = \frac{\Gamma^N}{c_{\text{max}}} \langle c , i_0 \rangle \geq \frac{1}{\rho} \Gamma^{-N} ,$$  \hspace{1cm} (6.19b)
and combining (6.15) and (6.19) gives (6.17b). The equivalent form in (6.17c) follows from (6.2) in Lemma 6.1 and definition (6.3). This completes the proof of Lemma 6.3.

**Theorem 6.1:** The symmetric cutoff rate parameter as defined in (6.1), $R_o^b(E_s)$, is given (in bits/symbol) by

$$R_o^b(E_s) = - \log_2 \Gamma,$$  \hspace{1cm} (6.20)

where $\Gamma$ is the dominant (largest) eigenvalue of the state transition matrix $A$ defined in (6.12).

**Proof:** The validity of (6.20) is asserted by Lemmas 6.2 and 6.3, and definition (6.1).

6.2 Bounds and Properties of $R_o^b(E_s)$

The symmetric cutoff rate in (6.20), $R_o^b(E_s)$, is upper bounded by the so-called asymptotic cutoff rate, $\hat{R}_o^b(E_s)$, i.e.,

$$R_o^b(E_s) \leq \hat{R}_o^b(E_s) = \lim_{N \to \infty} \hat{R}_{o,N}^b(E_s)$$ \hspace{1cm} (6.21a)

where $\hat{R}_{o,N}^b(E_s)$ is defined as

$$\hat{R}_{o,N}^b(E_s) = - \frac{1}{N} \log \left[ \min_{Q'_X} \int \left( \sum_{i=0}^{K-1} \sqrt{p_{Y|X}(y_N|x_i^N)} \ Q'_X(x_i^N) \right)^2 dy^N \right]$$ \hspace{1cm} (6.21b)

with the minimization taken over all joint probability distributions, $Q'_X$, for the ensemble of binary sequences, $X^N$. Inequality (6.21a) is a direct consequence of definition (6.1).
Corollary 6.1: For the memoryless channel (no ISI, i.e., \( M = 0 \), \( h_0 = 1 \)), the input symbols achieving \( \bar{R}_0^b(E_3) \) are i.i.d. binary random variables, so that \(^1\)

\[
\sqrt{p_{Y|X}(y^N|x^N_i)} \ Q'_X(x^N_i) = \prod_{k=0}^{N-1} \sqrt{p_{Y|X}(y^N_k|x^N_k)} \ Q'_X(x^N_k), \quad 0 \leq i < K.
\]

(6.22a)

In this case, \( Q'_X(x^N_k) = 1/2, \ x_k = \pm \sqrt{E_s} \), for all \( i \) and \( k \), and equality holds in (6.21a), i.e.,

\[
R_o^b(E_3) \bigg|_{M=0} \bigg| h_0=1 = R_o^b(E_3) \bigg|_{M=0} \bigg| h_0=1 = 1 - \log_2 (1 + e^{-E_s/N_0}) , \quad (6.22b)
\]

where the units are bits/symbol.

Proof: It must be shown that

\[
\bar{R}_o^b(N,E_3) \bigg|_{M=0} \bigg| h_0=1 = \bar{R}_o^b(N,E_3) \bigg|_{M=0} \bigg| h_0=1 , \quad N \geq 1 .
\]

(6.23)

Let \( y'_N \) be the argument of the logarithm in (6.21b). Substituting the right side of (6.22a) and applying the method of rearranging the resulting expression as in [GAL(68), p.133], one obtains

\[
y'_N = \min_{Q'_X} \left[ \int_{-\infty}^{+\infty} \left\{ \sqrt{p_{Y|X}(y|\sqrt{E_s})} \ Q'_X(\sqrt{E_s}) + \sqrt{p_{Y|X}(y|\sqrt{E_s})} \ Q'_X(-\sqrt{E_s}) \right\}^2 dy \right]^N .
\]

(6.24)

\(^1\) It is assumed that \( Q'_X(x^N_i) = Q'_X(x^N_0)Q'_X(x^N_1)\ldots Q'_X(x^N_{N-1}) \) for the memoryless channel (see [GAL(68), p. 133]). On the other hand, one may argue that \( Q'_X(x^N_i) = \prod_{k=0}^{N-1} Q'_X(x^N_k), \quad 0 \leq i < K, \) is capacity-achieving for the memoryless channel when \( Q'_X(x^N_i) = 1/2, \ x_k = \pm \sqrt{E_s}, \) for all \( i \) and \( k \) (see also Corollary 4.1 in Chapter 4).
Letting $Q'_x(\sqrt{E_s}) = q$ and $Q'_x(-\sqrt{E_s}) = 1 - q$, one sees that (6.24) becomes

$$\gamma'_N = \min_q \gamma'(q)^N, \quad 0 \leq q \leq 1,$$

(6.25)

where $\gamma'(q)$ stands for the expression in the square brackets of (6.24). It can now be shown that the second derivative of $\gamma'(q)$ is positive, i.e., $\gamma'(q)$ is strictly convex (in the interval $0 \leq q \leq 1$ unless the channel has zero capacity [MAS(83), p. 5.30]. It follows that $q = 1/2$ minimizes $\gamma'(q)$ in (6.25), and $Q'_x(\sqrt{E_s}) = Q'_x(-\sqrt{E_s}) = 1/2$ is the unique minimizing distribution. Thus,

$$\gamma'_N = [\gamma_1]^N, \quad N \geq 1,$$

(6.26)

where $\gamma_1$ is obtained from (6.3) with $N = 1$, and (6.23) follows. Finally, evaluation of (6.3), with $N = 1$ and $M = 0$ gives

$$\gamma_1 = \frac{1 + e^{-(E_s/N_0)h_0^2}}{2}.$$

(6.27)

With $h_0 = 1$, this leads to (6.22b), which proves Corollary 6.1.

**Corollary 6.2:** Provided that the energy of the unit-sample response of the channel is finite, i.e., that $\|h^{M+1}\|^2 < \infty$, the symmetric cutoff rate defined in (6.1), $R^{b}_o(E_s)$, is upper bounded (in bits/symbol) as

$$R^{b}_o(E_s) \leq \frac{1}{2 \ln 2} (E_s/N_0) \|h^{M+1}\|^2.$$

(6.28)

Furthermore, this bound is asymptotically tight for small values of $E_s/N_o$, i.e.,

$$R^{b}_o(E_s) \approx \frac{1}{2 \ln 2} (E_s/N_o) \|h^{M+1}\|^2 \approx 0.7213 (E_s/N_o) \|h^{M+1}\|^2, \quad E_s/N_o << 1.$$  

(6.29)
Proof: From (6.2) and (6.3), one obtains

\[ R_{0,N}(E_o) = -\frac{1}{N} \log \left( E \left[ \exp \left\{ -\left( E_s/N_o \right) \|d_{mn}\|^2 / 4 \right\} \right] \right), \tag{6.30} \]

where the expectation is taken over all subscripts \(0 \leq m, n < K\). Since the exponential in (6.30) is strictly convex, for all \(0 \leq m, n < K\), the Jensen inequality (see, e.g., [McE(77), p. 277]) may be applied, yielding (with log = \(\log_2\) units are bits per symbol)

\[ R_{0,N}(E_s) \leq -\frac{1}{N} \log_2 \left( \exp \left[ E \left\{ -\left( E_s/N_o \right) \|d_{mn}\|^2 / 4 \right\} \right] \right) \]

\[ \leq \frac{1}{4N \ln 2} \left( E_s/N_o \right) E \left[ \|d_{mn}\|^2 \right]. \tag{6.31} \]

From (6.5b), one obtains

\[ E\left[ \|d_{mn}\|^2 \right] = 4 \sum_{k=0}^{N-1} \sum_{i=0}^{Q} \sum_{r=0}^{Q} h_i h_r E\left[ e_{k-i}^{mn} e_{k-r}^{mn} \right], \tag{6.32} \]

where, from (6.6b),

\[ E\left[ e_{k-i}^{mn} e_{k-r}^{mn} \right] = \begin{cases} 1/2, & i = r \\ 0, & i \neq r. \end{cases} \tag{6.33} \]

Substituting (6.33) into (6.32) gives

\[ E\left[ \|d_{mn}\|^2 \right] = 2 \sum_{k=0}^{N-1} \sum_{i=0}^{Q} h_i^2, \tag{6.34} \]

and thus \(^1\)

\(^1\) Note also that \(E\left[ \|d_{mn}\|^2 \right] \leq 2N \|h^{M+1}\|^2\).
Finally, taking the logarithm to the base 2 and, on both sides of (6.35), the limit as \( N \to \infty \), one obtains (6.28). The tightness of the upper bound in (6.28) for \( E_s / N_o \ll 1 \) is implied by the fact that (6.31) holds with equality as \( E_s / N_o \to 0 \). This follows when expanding the log-function in (6.30) in a Taylor series about \( E_s / N_o = 0 \); retaining only the linear term leads directly to the second line of (6.31). This completes the proof of Corollary 6.2.

The simple upper bound on \( R_o^b(E_s) \) in (6.28) coincides with the one obtained for \( R_o^G(E_s) \) in Chapter 5. In view of Corollary 4.3 (Chapter 4), which asserts that \( I^b(E_s) \leq I^G(E_s) \), for all values of \( E_s / N_o \), and in particular as \( E_s / N_o \to 0 \), one might expect that \( R_o^b(E_s) \leq R_o^G(E_s) \) should hold similarly. However, it shall be demonstrated in Chapter 7 (Section 7.3) that this latter relation does not hold true for values of \( E_s / N_o \) which are smaller than some finite, positive number, \( \delta \). In fact, it will be shown that \( R_o^G(E_s) \) provides a lower bound for \( R_o^b(E_s) \) whenever \( 0 \leq E_s / N_o \leq \delta \).

### 6.3 Examples and a Reduction Theorem

Any numbering scheme may be used to enumerate the \( L = 2P + 1 = 3^M \), \( M > 0 \), distinct states \( \Sigma_j, -P \leq j \leq P \), of the machine in Figure 6.1. It will be convenient to let

\[
\Sigma_j = (t_0, t_1, \ldots, t_{M-1}), \quad t_k \in \{-1, 0, 1\}, \quad 0 \leq k < M, \quad (6.36a)
\]
and to assign the subscript \( j \in \{-P, -P + 1, \ldots, P\} \) such that \( \Sigma_{-j} = -\Sigma_j \), i.e.,

\[
j = \sum_{k=0}^{M-1} t_k 3^{M-k-1}.
\]

(6.36b)

With convention (6.36), the notation \( F(e, \Sigma_j) \) used in (6.12) may be replaced by the equivalent notation \( F(e, j) \).

**Example 6.1:** Consider the case \( M = 1 \), where \( h^2 = (h_0, h_1) = (h_0, \alpha h_0) = h_0(1, \alpha) \), \(-\infty < \alpha < \infty\), and where it is assumed that \( \|h^2\|^2 = E_h < \infty \), i.e.,

\( h_0 = \sqrt{E_h/(1 + \alpha^2)} \). Then, \( L = 3 \), \( P = 1 \) and, from (6.12), the state transition matrix is

\[
A = \begin{bmatrix}
F(-1, -1) & F(-1, 0) & F(-1, 1) \\
F(0, -1) & F(0, 0) & F(0, 1) \\
F(1, -1) & F(1, 0) & F(1, 1)
\end{bmatrix} = \begin{bmatrix}
a_0 & a_3 & a_2 \\
a_1 & a_4 & a_1 \\
a_2 & a_3 & a_0
\end{bmatrix} ,
\]

(6.37a)

where the entries of the second matrix are obtained from (6.8) as

\[
a_0 = F(-1, -1) = F(1, 1) = \frac{1}{4} e^{-(E_s/N_0) h_0^2 (1+\alpha)^2}
\]

\[
a_1 = F(0, -1) = F(0, 1) = \frac{1}{2} e^{-(E_s/N_0) h_0^2 \alpha^2}
\]

\[
a_2 = F(1, -1) = F(-1, 1) = \frac{1}{4} e^{-(E_s/N_0) h_0^2 (1-\alpha)^2}
\]

\[
a_3 = F(-1, 0) = F(1, 0) = \frac{1}{4} e^{-(E_s/N_0) h_0^2}
\]

\[
a_4 = F(0, 0) = \frac{1}{2} .
\]

(6.37b)

The state diagram for this example is shown in Figure 6.2.
Fig. 6.2. State diagram for Example 6.1 ($M = 1$).
Example 6.2: Consider now \( M = 2 \), where the channel response \( h^3 = (h_0, h_1, h_2) = (h_0, \alpha h_0, \beta h_0) = h_0(1, \alpha, \beta) \), \(-\infty < \alpha, \beta < \infty\), and where it is assumed that \( \|h^3\|^2 = E_h < \infty \), i.e., \( h_0 = \sqrt{E_h/(1 + \alpha^2 + \beta^2)} \). Thus, \( L = 9 \), \( P = 4 \), and from (6.12),

\[
A = \begin{bmatrix}
    a_0 & a_3 & a_6 & 0 & 0 & 0 & 0 & 0 & 0 \\
    0 & 0 & 0 & a_9 & a_{12} & a_{11} & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 & 0 & a_8 & a_5 & a_2 & 0 \\
    a_1 & a_4 & a_7 & 0 & 0 & 0 & 0 & 0 & 0 \\
    0 & 0 & 0 & a_{10} & a_{13} & a_{10} & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 & 0 & a_7 & a_4 & a_1 & 0 \\
    a_2 & a_5 & a_8 & 0 & 0 & 0 & 0 & 0 & 0 \\
    0 & 0 & 0 & a_{11} & a_{12} & a_9 & 0 & 0 & 0 \\
    0 & 0 & 0 & 0 & 0 & a_6 & a_3 & a_0 & 0 
\end{bmatrix}, \quad (6.38a)
\]

where the nonzero entries are given by (6.8) as

\[
\begin{align*}
a_0 &= F(-1, -4) = F(1, 4) = \frac{1}{4} e^{-(E_s |N_0) h_0^2 (1+\alpha+\beta)^2} \\
a_1 &= F(0, -4) = F(0, 4) = \frac{1}{2} e^{-(E_s |N_0) h_0^2 (\alpha+\beta)^2} \\
a_2 &= F(1, -4) = F(-1, 4) = \frac{1}{4} e^{-(E_s |N_0) h_0^2 (1-\alpha-\beta)^2} \\
a_3 &= F(-1, -3) = F(1, 3) = \frac{1}{4} e^{-(E_s |N_0) h_0^2 (1+\alpha)^2} \\
a_4 &= F(0, -3) = F(0, 3) = \frac{1}{2} e^{-(E_s |N_0) h_0^2 \alpha^2} \\
a_5 &= F(1, -3) = F(-1, 3) = \frac{1}{4} e^{-(E_s |N_0) h_0^2 (1-\alpha)^2} \\
a_6 &= F(-1, -2) = F(1, 2) = \frac{1}{4} e^{-(E_s |N_0) h_0^2 (1+\alpha-\beta)^2} \\
a_7 &= F(0, -2) = F(0, 2) = \frac{1}{2} e^{-(E_s |N_0) h_0^2 (\alpha-\beta)^2} \\
a_8 &= F(1, -2) = F(-1, 2) = \frac{1}{4} e^{-(E_s |N_0) h_0^2 (1-\alpha+\beta)^2} 
\end{align*}
\]
\[ a_9 = F(-1,-1) = F(1,1) = \frac{1}{4} e^{-\frac{1}{2}(E_s/N_0) \beta^2} \]
\[ a_{10} = F(0,-1) = F(0,1) = \frac{1}{2} e^{-\frac{1}{2}(E_s/N_0) \beta^2} \]
\[ a_{11} = F(1,-1) = F(-1,1) = \frac{1}{4} e^{-\frac{1}{2}(E_s/N_0) \beta^2} \]
\[ a_{12} = F(-1,0) = F(1,0) = \frac{1}{4} e^{-\frac{1}{2}(E_s/N_0) \beta^2} \]
\[ a_{13} = F(0,0) = \frac{1}{2} . \]

The state diagram for this example is shown in Figure 6.3.

According to Lemma 6.2, the \( M \)-th power of matrix \( A \), \( A^M \), should have only positive entries, i.e., \( a_{ts}^M > 0 \), \(-P \leq t, s \leq P \). For \( M=1 \) (Example 6.1) this condition is trivially satisfied, as can be seen from (6.37). For \( M=2 \) (Example 6.2) one obtains

\[ A^2 = \begin{bmatrix}
    a_0^2 & a_0a_3 & a_0a_6 & a_3a_9 & a_3a_{12} & a_3a_{11} & a_6a_8 & a_5a_6 & a_6a_2 \\
    a_1a_9 & a_4a_9 & a_7a_9 & a_{10}a_{12} & a_{12}a_{13} & a_{10}a_{12} & a_7a_{11} & a_4a_{11} & a_{11}a_{11} \\
    a_2a_8 & a_5a_8 & a_2^2 & a_5a_{11} & a_5a_{12} & a_5a_9 & a_2a_6 & a_2a_3 & a_0a_2 \\
    a_0a_1 & a_1a_3 & a_1a_6 & a_4a_9 & a_4a_{12} & a_4a_{11} & a_7a_8 & a_5a_7 & a_2a_7 \\
    a_{11}a_9 & a_{10}a_{10} & a_7a_{10} & a_{10}a_{13} & a_{13}^2 & a_{10}a_{13} & a_7a_{10} & a_4a_{10} & a_{11}a_{10} \\
    a_2a_7 & a_5a_7 & a_7a_8 & a_4a_{11} & a_4a_{12} & a_4a_9 & a_1a_6 & a_1a_3 & a_0a_1 \\
    a_0a_2 & a_2a_3 & a_2a_6 & a_5a_9 & a_5a_{12} & a_5a_{11} & a_2^2 & a_5a_8 & a_2a_8 \\
    a_{11}a_{11} & a_{10}a_{11} & a_7a_{11} & a_{10}a_{12} & a_{12}a_{13} & a_{10}a_{12} & a_7a_9 & a_4a_9 & a_{11}a_9 \\
    a_6a_2 & a_5a_6 & a_6a_8 & a_3a_{11} & a_3a_{12} & a_3a_9 & a_0a_6 & a_0a_3 & a_0^2 \\
\end{bmatrix} \]

(6.39)
Fig. 6.3. State diagram for Example 6.2 ($M = 2$).
so that the condition for $A$ to be irreducible is also satisfied. Alternatively, this fact may be verified with the state diagram in Figure 6.3.

As a consequence of convention (6.36) for enumerating the $L$ distinct states, $\Sigma_j, -P \leq j \leq P$, it follows from (6.8) that $F(e, j) = F(-e, -j)$. This leads to a symmetry in matrix $A$ with respect to its center element $a_{00} = 1/2$, i.e., $A$ has the property that $a_{mn} = a_{-m,-n}, -P \leq m, n \leq P$. Matrices with this property are called centrosymmetric [ANDR(73)]. The same symmetry is maintained for all powers of $A$, i.e., matrix $A^N$ has the property that $a_{ts}^N = a_{-t,-s}^N, -P \leq t, s \leq P$, as illustrated in (6.39) for $N=2$. The centrosymmetry of matrix $A$ is reflected in the corresponding state diagram (Figures 6.2 and 6.3).

It implies that the set of all $N$-step paths starting with state $\Sigma_0$ and ending in state $\Sigma_j, j = 1, 2, \ldots, P$, is equivalent to the set of all $N$-step paths starting with state $\Sigma_0$ and ending in state $\Sigma_{-j}$. In the following, this property is exploited to reduce the dimensionality in the computation of $R^0_\alpha(E_\delta)$.

A formal method, deduced from the theory of finite Markov chains [KEM-SNE(60), Section 6.3], leads to an equivalent reduced state description of the machine in Figure 6.1. Define the reduced state space $\{\tilde{\Sigma}_{\pm j}: 0 \leq j \leq P\}$ with $P + 1 = (3^M + 1)/2$ elements where

$$\tilde{\Sigma}_{\pm j} \equiv \begin{cases} \Sigma_0, & j = 0 \\ (\Sigma_j, \Sigma_{-j}), & 1 \leq j \leq P \end{cases},$$

(6.40)

i.e., the nonzero states, $\Sigma_j$ and $\Sigma_{-j}, 1 \leq j \leq P$, are combined to form a single state, $\tilde{\Sigma}_{\pm j}$. Further, define the $(P + 1) \times (2P + 1)$ matrix

$$G = [g_{ik}], \quad 0 \leq i \leq P, \quad -P \leq k \leq P,$$

(6.41a)

where

$$g_{ik} = \begin{cases} 1, & i = |k| \\ 0, & \text{otherwise} \end{cases},$$

(6.41b)
and the \((2P + 1) \times (P + 1)\) matrix

\[
V = \begin{bmatrix} v_{ik} \end{bmatrix}, \quad -P \leq i \leq P, \quad 0 \leq k \leq P, \tag{6.42a}
\]

where

\[
v_{ik} = \begin{cases} 
1 & , \quad i = k = 0 \\
1/2 & , \quad k = |i| \neq 0 \\
0 & , \quad \text{otherwise}.
\end{cases} \tag{6.42b}
\]

**Lemma 6.4:** If \(A\) is a \((2P + 1) \times (2P + 1)\) matrix as defined in (6.12) and \(G\) and \(V\) are matrices as defined in (6.41) and (6.42), respectively, then

\[
VGAV = AV. \tag{6.43}
\]

**Proof:** Let \((vg)_{ik}, -P \leq i, k \leq P\), be the elements of matrix \(VG\). Then, from (6.41) and (6.42), it follows that

\[
(vg)_{ik} = g_{|i|,k} = \begin{cases} 
1 & , \quad i = k = 0 \\
1/2 & , \quad |i| = |k| \neq 0 \\
0 & , \quad \text{otherwise}.
\end{cases} \tag{6.44}
\]

Let \((av)_{ik}, -P \leq i \leq P, 0 \leq k \leq P\), be the elements of matrix \(AV\). Then, from (6.42) and the centrosymmetry of matrix \(A\), it follows that

\[
(av)_{ik} = (av)_{-i,k} = \begin{cases} 
a_{ik} & , \quad k = 0 \\
(a_{ik} + a_{i,-k})/2 & , \quad \text{otherwise}.
\end{cases} \tag{6.45}
\]

Thus, matrix \(AV\) has row symmetry with respect to the center row \((i = 0)\). It follows from (6.44) that multiplication of matrix \(VG\) with matrix \(AV\) simply reproduces \(AV\), i.e., for a matrix of the type of \(AV\), matrix \(VG\) takes the role of the identity matrix. This completes the proof of Lemma 6.4.
Theorem 6.2: If $G$, $A$, and $V$ are matrices as specified in Lemma 6.4, then the $(P+1) \times (P+1)$ matrix

$$
\tilde{A} = GAV = \begin{bmatrix} \tilde{a}_{ik} \end{bmatrix}, \quad 0 \leq i, k \leq P,
$$

(6.46a)

where

$$
\tilde{a}_{ik} = \begin{cases} 
a_{0k}, & i = 0 
a_{ik} + a_{-i,k}, & \text{otherwise},
\end{cases}
$$

(6.46b)

is the equivalent reduced state transition matrix for the machine in Figure 6.1. Matrix $\tilde{A}$ has the same dominant eigenvalue, $\Gamma$, as matrix $A$. Therefore, $R^b_o(E)$ in (6.20) of Theorem 6.1 may be obtained by computing $\Gamma$ from the $(P+1) \times (P+1)$ matrix $\tilde{A}$, instead of the $(2P+1) \times (2P+1)$ matrix $A$.

Proof: It follows from (6.46) and Lemma 6.2 that matrix $\tilde{A}$ has also real nonnegative elements and is irreducible. Thus, by the Perron-Frobenius theorem, $\tilde{A}$ must have a dominant eigenvalue, $\tilde{\Gamma}$. Let

$$
\tilde{\gamma}_N = \tilde{u} \tilde{A}^N \tilde{i}_0,
$$

(6.47a)

where

$$
\tilde{u} = [1, 1, \ldots, 1]
$$

(6.47b)

is a $(P+1)$-component row vector and

$$
\tilde{i}_0 = [1, 0, 0, \ldots, 0]^T
$$

(6.47c)
is a \((P + 1)\)-component column vector. To prove that \(\tilde{\Gamma} = \Gamma\), it must be shown according to Lemma 6.3, that \(\tilde{\gamma}_N = \gamma_N\). From Lemma 6.4 one obtains

\[
\tilde{A}^N = (GAV)^N = GA^N V. \tag{6.48}
\]

Note further that \(\tilde{u}G = u\), and \(V\tilde{i}_0 = i_0\), where \(u\) and \(i_0\) are defined in (6.14). Then, substituting (6.48) into (6.47a) gives

\[
\tilde{\gamma}_N = \tilde{u}GA^N V\tilde{i}_0 = uA^N i_0 = \gamma_N, \tag{6.49}
\]

where the last equality follows from (6.15). Finally, (6.46b) may be verified using (6.41) and (6.45). This completes the proof of Theorem 6.2.

Example 6.1 (continued): With (6.37a), and using (6.41) and (6.42) for \(P = 1\), (6.46a) yields

\[
\tilde{A} = \begin{bmatrix}
0 & 1 & 0 \\
1 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
\begin{bmatrix} a_0 & a_3 & a_2 \\ a_1 & a_4 & a_1 \\ a_2 & a_3 & a_0 \end{bmatrix}
\begin{bmatrix} 0 & 1/2 \\ 1 & 0 \\ 0 & 1/2
\end{bmatrix}
\end{bmatrix}
\]

\[
= \begin{bmatrix}
0 & 1 & 0 \\
1 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
a_3 & (a_0 + a_2)/2 \\
\begin{bmatrix} a_4 & a_1 \\ a_3 & (a_0 + a_2)/2 \end{bmatrix}
\end{bmatrix}
= \begin{bmatrix}
a_4 & a_1 \\
2a_3 & (a_0 + a_2)
\end{bmatrix}. \tag{6.50}
\]

The last equality may be directly verified by using (6.46b). It can be seen from the second equality in (6.50) that matrix \(G\) simply removes the duplication of rows in matrix \(AV\). Thus, \(G\) could be replaced by
where $\epsilon$ is any number in the range $0 \leq \epsilon \leq 2$. The reduced state diagram for Example 6.1 is shown in Figure 6.4.

**Example 6.2 (continued):** Equation (6.38a), and direct usage of (6.46b) with $P = 4$, yield

$$
\begin{bmatrix}
a_{13} & a_{10} & 0 & 0 & 0 \\
0 & 0 & a_7 & a_4 & a_1 \\
0 & 0 & a_8 & a_5 & a_2 \\
2a_{12} & (a_9 + a_{11}) & 0 & 0 & 0 \\
0 & 0 & a_6 & a_3 & a_0
\end{bmatrix}
$$

(6.51)

The reduced state diagram for Example 6.2 is shown in Figure 6.5.

6.4 Numerical Results

This section introduces first an iterative method to determine the dominant eigenvalue, $\Gamma$, of the $(P + 1) \times (P + 1)$ matrix $\tilde{A}$ satisfying the conditions of Theorem 6.2. Next, the properties of the symmetric cutoff rate, $R^b_0(E_s)$, are investigated for a certain class of partial-response channels [KAB-PAS(75)]. These results are then compared with the ones obtained for the so-called minimum-distance channels originally introduced by Magee and Proakis [MAG-PRO(73)], i.e., the channels with fixed ISI memory, $M$, that exhibit the lowest attainable squared minimum distance, $d_{\min}^2$. Thirdly, the unit-sample responses of those channels which minimize $R^b_0(E_s)$ are defined and determined. These newly characterized channels are then related to the well-known minimum-distance channels.
Fig. 6.4. Reduced state diagram for Example 6.1 ($M = 1$).
Fig. 6.5. Reduced state diagram for Example 6.2 ($M = 2$).
A. Numerical Evaluation of $\Gamma$

For binary input symbols, it is expected that $0 \leq R^b_o(E_s) \leq \log_2$, i.e., that

$$\frac{1}{2} \leq \Gamma \leq 1.$$  \hfill (6.52)

In (6.52), $\Gamma \to 1/2$ as $E_s/N_0 \to \infty$, and $\Gamma \to 1$ as $E_s/N_0 \to 0$; this follows directly from (6.2) since $\|d^N_{i,k}\| > 0$ for $i \neq k$, and $\|d^N_{i,k}\| = 0$ for $i = k$. The following observations on the properties of the eigenvectors of the matrix $\tilde{A}$ associated with $\Gamma$ will be useful when $\Gamma$ is computed near the bounds indicated in (6.52).

The upper bound in (6.52) may also be verified by considering the form of $\tilde{A}$ as $E_s/N_0 \to 0$. For $E_s/N_0 = 0$, the transpose of $\tilde{A}$, $\tilde{A}^T$, reduces to a stochastic matrix, i.e., the sum of the (nonnegative) elements of each column of $\tilde{A}$ is unity. This can be verified with (6.50) and (6.51) in Examples 6.1 and 6.2, respectively. It is well known that the dominant eigenvalue of a stochastic matrix is one, and that its associated right eigenvector is positive, with all elements also being unity [GAN(59), pp. 99-105], [LAN(69), Theorem 9.5.1]. Thus, the left eigenvector of $\tilde{A}$ associated with $\Gamma$ approaches the all-one row vector as $\Gamma \to 1$. Verification of the lower bound in (6.52) directly from $\tilde{A}$ is less trivial. In the case of $E_s/N_0 \to \infty$, the form of $\tilde{A}$ depends on the unit-sample response of the channel, i.e., some matrix elements other than $a_{00} = 1/2$ may be nonzero (they converge towards 1/4 or 1/2). However, for $N = M$ and $E_s/N_0 \to \infty$, it can be observed in (6.13b) that at least one product term $F(e^i_k, \sigma_k) \to 0$, $0 < k < M$, whenever $F(e^i_0, \sigma_0 = \Sigma_i) \to 1/4$ or $1/2$ for all $i \ni \sigma_M = \Sigma_i$, $|t| \neq |s|$. This implies that the $M$-th power of the matrix $\tilde{A}$, $\tilde{A}^M$, becomes a diagonal matrix as $E_s/N_0 \to \infty$, with the largest element being $a_{00}^M = (1/2)^M$. Since $\tilde{A}^M$ has the dominant eigenvalue $\Gamma^M$, it follows that $\Gamma \to 1/2$ as $E_s/N_0 \to \infty$. Furthermore, since these (degenerate) diagonal matrices are symmetric, the left and right eigenvectors associated with
\( \Gamma = 1/2 \) are identical, with the leading element being one and all others being zero. Therefore, the left and right eigenvectors of \( \tilde{A} \) approach the vectors 
\[ [1, 0, 0, \ldots, 0] \quad \text{and} \quad [1, 0, 0, \ldots, 0]^T, \]
respectively, as \( \Gamma \to 1/2 \). The diagonal property of \( \tilde{A}^M \), as described above, may be verified, using (6.50) and (6.37b) in Example 6.1 or in Example 6.2, by constructing \( \tilde{A}^M \) from (6.39) or (6.51) and using (6.38b).

The \((2P+1) \times (2P+1)\) matrix \( A \) has \( N_A = 6P+3 \) nonzero elements; the \((P+1) \times (P+1)\) matrix \( \tilde{A} \) has \( N_\tilde{A} = 3P+1 \) nonzero elements. The fractions of nonzero elements in the matrices \( A \) and \( \tilde{A} \) are \( \eta_A = N_A/(2P+1)^2 \) and \( \eta_{\tilde{A}} = N_{\tilde{A}}/(P+1)^2 \), respectively. Table 6.1 characterizes the two matrices in terms of parameters \( P, N_A, \eta_A, N_\tilde{A}, \) and \( \eta_{\tilde{A}} \), for channel memory \( M = 1, 2, \ldots, 10 \). It is evident that so-called direct methods that determine \( \Gamma \) as the largest root of the characteristic equation of a matrix will become inefficient as \( M \) increases since these methods do not take advantage of the fact that \( A \) and

<table>
<thead>
<tr>
<th>( M )</th>
<th>( P )</th>
<th>( N_A )</th>
<th>( \eta_A )</th>
<th>( N_{\tilde{A}} )</th>
<th>( \eta_{\tilde{A}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>27</td>
<td>0.333</td>
<td>13</td>
<td>0.520</td>
</tr>
<tr>
<td>3</td>
<td>13</td>
<td>81</td>
<td>0.111</td>
<td>40</td>
<td>0.204</td>
</tr>
<tr>
<td>4</td>
<td>40</td>
<td>243</td>
<td>0.0370</td>
<td>121</td>
<td>0.0720</td>
</tr>
<tr>
<td>5</td>
<td>121</td>
<td>729</td>
<td>0.0123</td>
<td>364</td>
<td>0.0245</td>
</tr>
<tr>
<td>6</td>
<td>364</td>
<td>2187</td>
<td>0.00412</td>
<td>1093</td>
<td>0.00820</td>
</tr>
<tr>
<td>7</td>
<td>1093</td>
<td>6561</td>
<td>0.00137</td>
<td>3280</td>
<td>0.00274</td>
</tr>
<tr>
<td>8</td>
<td>3280</td>
<td>19683</td>
<td>0.000457</td>
<td>9841</td>
<td>0.000914</td>
</tr>
<tr>
<td>9</td>
<td>9841</td>
<td>59049</td>
<td>0.000152</td>
<td>29524</td>
<td>0.000305</td>
</tr>
<tr>
<td>10</td>
<td>29524</td>
<td>177147</td>
<td>0.0000508</td>
<td>88573</td>
<td>0.000102</td>
</tr>
</tbody>
</table>
\( \tilde{A} \) are very sparse matrices for large \( M \). In addition, storage requirements become significant for large \( M \). Iterative methods avoid these problems since they require only storage of the nonzero matrix elements. The fact that only the dominant eigenvalue is of interest also favors the iterative approach. Iterative methods based on the classical power method determine \( \Gamma \) directly and to any desired degree of accuracy [ZUR(58), Chapter VI].

For nonhermitian matrices with the properties of \( \tilde{A} \), \( \Gamma \) may be obtained as

\[
\Gamma = \lim_{i \to \infty} \rho_i ,
\]

(6.33a)

where \( \rho_i \) is the generalized Rayleigh quotient [LAN(69), p. 111], [ZUR(58), p. 238], i.e.,

\[
\rho_i = \frac{\langle x_i, \tilde{A} y_i \rangle}{\langle x_i, y_i \rangle} , \quad 0 \leq i < \infty ,
\]

(6.33b)

where the column vectors

\[
\begin{align*}
  x_i &= \tilde{A}^T x_{i-1} , \quad x_0 \neq 0 , \\
  y_i &= \tilde{A} y_{i-1} , \quad y_0 \neq 0 .
\end{align*}
\]

(6.33c)

In (6.33), \( x_i \) and \( y_i \) converge towards the left and right eigenvectors, respectively, associated with \( \Gamma \), provided that the initial vectors, \( x_0 \) and \( y_0 \), belong to the respective spaces spanned by the sets of left and right eigenvectors of \( \tilde{A} \). The iterative method described by (6.33) is based on the stationary properties of the (generalized) Rayleigh quotient [ZUR(58), Section 15.2] and [FAD-FAD(73), pp. 375-376]. Let \( \lambda_2 \) be the eigenvalue of \( \tilde{A} \) which has the second-largest magnitude, i.e., \( \Gamma > |\lambda_2| \). It can then be shown that

\[
\rho_i = \Gamma + O\left(\left|\lambda_2 / \Gamma \right|^{2i}\right) , \quad i >> 1 ,
\]

(6.34)
indicating that the rate of convergence of $\rho_l$ depends primarily on the ratio $|\lambda_2|/\Gamma$ and that the error may be made as small as desired by increasing the number of iterations. In practice, (6.53) has to be modified to avoid loss of accuracy and to prevent underflow or overflow due to excessive decrease or increase of the iterated vector components. This may be achieved, for example, by normalizing the iterated vectors so that their largest element has magnitude one. The algorithm used to generate the numerical results is given below.

**Algorithm 6.1:**

Given are:

- $\tilde{A}$, $(P + 1) \times (P + 1)$ matrix;
- $x_0 > 0$, $(P + 1)$-element column vector;
- $y_0 > 0$, $(P + 1)$-element column vector;
- $N_{\text{max}}$, maximum number of iterations;
- $0 < \varepsilon < 1$, required accuracy.

Iteration steps:

1) $i = 0$, $\rho_{-1} = -1$.
2) $\mu_i = \max_k |x_k^i|$, $\bar{x}_i = x_i/\mu_i$.
   $\nu_i = \max_k |y_k^i|$, $\bar{y}_i = y_i/\nu_i$.
3) $y_{i+1} = \tilde{A}\bar{y}_i$.
4) $\rho_i = \frac{\langle \bar{x}_i, y_{i+1} \rangle}{\langle \bar{x}_i, \bar{y}_i \rangle}$.
5) If $|\rho_i - \rho_{i-1}| < \varepsilon$, then go to step 10.
6) $x_{i+1} = \tilde{A}^T\bar{x}_i$.
7) $i = i + 1$. 
8) If \( i \leq N_{\text{max}} \), then go to step 2.

9) Stop procedure (accuracy not reached).

10) \( \Gamma \leftarrow \rho_i \).

**Example 6.3:** The convergence properties of Algorithm 6.1 are first explored for a channel with unit-sample response of unit energy and memory \( M = 2 \), \( h^3 = \sqrt{2/3} (1, 1/2, -1/2) \). Figures 6.6 and 6.7 show \( |\rho_i - \rho_{i-1}| \) as a function of iteration step \( i \), using different methods to initialize \( x_0 \) and \( y_0 \) for each value of \( E_s/N_o \). A representative range of values for \( E_s/N_o \) was chosen to cover the range of \( \Gamma \) indicated in (6.52). As expected, the number of iterations required to reach a prescribed accuracy, e.g., \( \varepsilon = 10^{-10} \), depends primarily on the choice of \( x_0 \) and \( y_0 \) and on the value of \( E_s/N_o \). The latter determines the eigenvalue ratios \( |\lambda_n|/\Gamma, 2 \leq n \leq P + 1 \), where \( \Gamma = \lambda_1 > |\lambda_2| \geq ... \geq |\lambda_{P+1}| \) (\( P = 4 \) in this example). Table 6.2 lists \( \Gamma, |\lambda_2|/\Gamma, \) and \( |\lambda_3|/\Gamma \), for each value of \( E_s/N_o \) used in Figures 6.6 and 6.7. The rate of convergence is influenced by \( \lambda_3 \) in the case where \( |\lambda_2| \) becomes large and \( |\lambda_3| \approx |\lambda_2| \). The convergence properties observed in this example hold similarly in the general case. Accordingly, computation of \( \Gamma \) for channels with large memory, \( M \), over a wide range of values for \( E_s/N_o \), is most efficient when using the method of Figure 6.7b. In practice, the prescribed accuracy, \( \varepsilon \), may be reduced to the range \( 10^{-7} \leq \varepsilon \leq 10^{-3} \). Figure 6.8 shows the performance of Algorithm 6.1 for the nontrivial case where \( M = 8 \), i.e., \( P = 3280 \) (see also Table 6.1). In the case where \( \Gamma \) is evaluated repeatedly for large \( M \), e.g., when searching for a unit-sample response having particular properties with respect to \( R_o^b(E_s) \), characteristics of the type shown in Figure 6.8 are helpful in the design of computationally efficient procedures.
Fig. 6.6a Convergence of Algorithm 6.1 (Example 6.3): Initialization with $x_0 = y_0 = [1, 1, 1, 1, 1]^T$ for each value of $E_s/N_o$ ($\varepsilon = 10^{-10}$).

Fig. 6.6b Convergence of Algorithm 6.1 (Example 6.3): Initialization with $x_0 = y_0 = [1, 0, 0, 0, 0]^T$ for each value of $E_s/N_o$ ($\varepsilon = 10^{-10}$).
Fig. 6.7a Convergence of Algorithm 6.1 (Example 6.3): Initialization with $x_0 = y_0 = [1, 1, 1, 1, 1]^T$ for $E_s/N_o = -25$ dB; otherwise, $x_0$ and $y_0$ are the final iterated vectors $x_i$ and $y_i$, respectively, as obtained for the nearest lower value of $E_s/N_o$ ($\varepsilon = 10^{-10}$).

Fig. 6.7b Convergence of Algorithm 6.1 (Example 6.3): Initialization with $x_0 = y_0 = [1, 0, 0, 0, 0]^T$ for $E_s/N_o = 8$ dB; otherwise, $x_0$ and $y_0$ are the final iterated vectors $x_i$ and $y_i$, respectively, as obtained for the nearest higher value of $E_s/N_o$ ($\varepsilon = 10^{-10}$).
Fig. 6.8. Performance of Algorithm 6.1 (Example 6.3):
(a) Required iterations for $10^{-7} \leq \varepsilon \leq 10^{-3}$ as a function of $E_s/N_o$ for a channel with memory $M = 8$. The method of Figure 6.7b was used with modified initial vectors, $x_0 = y_0 = [1, \Delta, \Delta, \ldots, \Delta]^T$, where $\Delta = 10\varepsilon$.
(b) $R^b_\epsilon(E_s)$ for the above channel determined with $\varepsilon = 10^{-7}$. The unit-sample response is given as $h = (0.14, 0.26, 0.36, 0.43, 0.45, 0.43, 0.36, 0.26, 0.14)$. 
Table 6.2: (Example 6.3) $\Gamma, |\lambda_2|/\Gamma$, and $|\lambda_3|/\Gamma$, for a channel with unit-sample response $h^3 = \sqrt{2/3} (1, 1/2, -1/2)$. The prescribed accuracy, for the computation of $\Gamma$ with Algorithm 6.1, is specified with $\varepsilon = 10^{-10}$.

| $(E_b/N_0)_\text{dB}$ | $\Gamma$         | $|\lambda_2|/\Gamma$ | $|\lambda_3|/\Gamma$ |
|------------------------|------------------|-----------------------|-----------------------|
| -25                    | 0.9984220500     | 0.0327                | 0.0322                |
| -5                     | 0.8689428373     | 0.3242                | 0.2753                |
| 0                      | 0.7010054820     | 0.5039                | 0.3840                |
| 2                      | 0.6202484465     | 0.5527                | 0.4078                |
| 8                      | 0.5009798714     | 0.5014                | 0.5012                |

Note: The eigenvalues, $\lambda_2$ and $\lambda_3$, were determined with a standard (direct) method.

B. $R_{a}(E_b)$ for Partial-Response Channels and Minimum-Distance Channels

The symmetric cutoff rate, $R_{a}(E_b)$, is first determined for the class of partial-response (PR) channels investigated extensively by Kabal and Pasupathy [KAB-PAS(75), Table I]. These results are then compared to those obtained for the so-called minimum-distance channels (MDC’s) introduced earlier by Magee and Proakis [MAG-PRO(73)].

Partial-Response Channels

Table 6.3 identifies and characterizes the PR channels in terms of their system polynomial, $F(D)$, ISI memory, $M$, and unit-sample response, $h^{M+1}$ (normalized to unit energy). The PR channels listed in Table 6.3 form the four groups \{A, B, C\}, \{D, E\}, \{F, G\}, and \{H, I\}, reflecting the equivalence (duality) properties between certain channels [KOB(71)], [KAB-PAS(75)]. Assuming that the received noise components are additive and uncorrelated, the basic decoder structure (maximum-likelihood sequence decoder) and its error rate performance remain unchanged for certain transformations of the system polynomial $F(D)$. For some given system, $F(D)$, the transformations $F(-D)$
Table 6.3: Identification and characterization of partial-response (PR) channels in terms of system polynomial $F(D)$, ISI memory $M$, and unit-sample response, $h^{M+1}$, where $\|h^{M+1}\|^2 = 1$ (see also [KAB-PAS(75), Tables I and II]).

<table>
<thead>
<tr>
<th>Channel</th>
<th>$F(D)$</th>
<th>$M$</th>
<th>$h^{M+1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$1 + D$</td>
<td>1</td>
<td>$(1, 1)/\sqrt{2}$</td>
</tr>
<tr>
<td>B</td>
<td>$1 - D$</td>
<td>1</td>
<td>$(1, -1)/\sqrt{2}$</td>
</tr>
<tr>
<td>C</td>
<td>$1 - D^2$</td>
<td>2 (1)</td>
<td>$(1, 0, -1)/\sqrt{2}$</td>
</tr>
<tr>
<td>D</td>
<td>$1 + 2D + D^2$</td>
<td>2</td>
<td>$(1, 2, 1)/\sqrt{6}$</td>
</tr>
<tr>
<td>E</td>
<td>$1 - 2D^2 + D^4$</td>
<td>4 (2)</td>
<td>$(1, 0, -2, 0, 1)/\sqrt{6}$</td>
</tr>
<tr>
<td>F</td>
<td>$2 + D - D^2$</td>
<td>2</td>
<td>$(2, 1, -1)/\sqrt{6}$</td>
</tr>
<tr>
<td>G</td>
<td>$2 - D^2 - D^4$</td>
<td>4 (2)</td>
<td>$(2, 0, -1, 0, -1)/\sqrt{6}$</td>
</tr>
<tr>
<td>H</td>
<td>$1 + D - D^2 - D^3$</td>
<td>3</td>
<td>$(1, 1, -1, -1)/\sqrt{4}$</td>
</tr>
<tr>
<td>I</td>
<td>$1 - D - D^2 + D^3$</td>
<td>3</td>
<td>$(1, -1, -1, 1)/\sqrt{4}$</td>
</tr>
</tbody>
</table>

and $F(D^m)$, $m = 1, 2, \ldots$, yield equivalent systems with respect to performance and basic decoder structure. The latter transformation represents a system that is an $m$-fold interleaved form of the system $F(D)$. In Table 6.3, for example, channel C may be considered as two interleaved channels of type B which operate independently at half the symbol rate of channel C. In general, $F(\pm D^m)$, $m = 1, 2, \ldots$, has a dual relationship with $F(D)$. Examples in Table 6.3 are channels D and F which yield the duals E and G, respectively, after application of the transformation $F(-D^2)$. Conversely, the memory of some channel, $M$, may be reduced, if the greatest common divisor (gcd) of the exponents of the $D$-operator in $F(D)$ is greater than unity. In Table 6.3, such a memory reduction is possible for channels C, E, and G, respectively, where in
each case the gcd of the exponents of the $D$-operator in $F(D)$ equals two. With respect to computing $R_0^b(E_s)$, this memory reduction is advantageous since it reduces the number of states in the machine of Figure 6.1. As a result, the size of the state transition matrix, $A$, and thus of $\tilde{A}$, can be reduced (see also Table 6.1).

Figure 6.9 shows the symmetric cutoff rate, $R_0^b(E_s)$, for the PR channels of Table 6.3, as a function of the signal-to-noise ratio, $E_s/N_0$. The equivalence described above between channels of the same group holds also with respect to $R_0^b(E_s)$. This result is not unexpected; it is explained by the fact that the Euclidean norms (distances) of the product (output) sequences $F(D) \epsilon(D)$, and $F(-D) \epsilon(-D)$, are identical, i.e., $\|F(D) \epsilon(D)\|^2 = \|F(-D) \epsilon(-D)\|^2$, where $\epsilon(D)$ represents some error sequence in $D$-operator notation [KOB(71)]. This implies that the ensemble of squared distance norms appearing in (6.5b), $\|d_{mn}^N\|^2$, $N = 1, 2, \ldots$, is invariant between channels $F(D)$ and $F(-D)$.

Figure 6.9 indicates that the degradation in $R_0^b(E_s)$, with respect to the memoryless channel (no ISI), is minimal for channels $\{F, G\}$, while channels $\{D, E\}$ exhibit the worst degradation. Interestingly, the channels in these two groups have all (equivalent) ISI memory $M = 2$. This observation leads naturally to the problem of defining and characterizing those channels which minimize $R_0^b(E_s)$, given that $M$, $E_s/N_0$, and $\|h^{M+1}\|^2$, are fixed. This problem will be addressed subsequently.

Comparison of Minimum-Distance Channels and Partial Response Channels

As a preliminary step towards the identification of the channels which minimize $R_0^b(E_s)$, the MDC's originally introduced by Magee and Proakis [MAG-PRO(73)] are considered first. These channels were later also addressed in [AND-FOS(75)] and [CAN-BUT(76-1)].

\[1\] [MAG-PRO(73)] and [CAN-BUT(76-1)] use $l$ for the channel length and channel duration, respectively, to characterize the time-dispersion of the channel, while
Fig. 6.9. Symmetric cutoff rate, $R_o^b(E_s)$, as a function of signal-to-noise ratio, $E_s/N_o$, for the partial-response (PR) channels defined in Table 6.3.
Figure 6.10 shows $R_p^b(E_s)$ for the MDC's having memory $M = 1, 2,$ and $3,$ as a function of $E_s/N_0$. Clearly, for some fixed value of $E_s/N_0$, $R_p^b(E_s)$ degrades with increasing $M$. Comparing these results with those obtained in Figure 6.9 confirms that MDCI is identical to the PR channel A of Table 6.3, and thus equivalent to PR channels B and C. As indicated in Figure 6.10, the MDC's are not unique for any fixed $M$. For example, the unit-sample response $h^2 = (1, a)/\sqrt{1 + a^2}, -\infty < a < \infty,$ is a MDC [MAG-PRO(73), Table 1]. In what follows this channel will be called MDC1a; it has $d_{\min}^2 = 4,$ independent of $a$. Let $b_h^{M+1}$ denote a unit-sample response in backward order, i.e., given the channel $h^{M+1} = (h_0, h_1, \ldots, h_M)$, $b_h^{M+1} = (b_0, b_1, \ldots, b_M) = (h_M, h_{M-1}, \ldots, h_0)$. One can easily verify that $h^{M+1}$ and $b_h^{M+1}$ represent a pair of equivalent channels with respect to $d_{\min}^2$ and $R_p^b(E_s)$ (e.g., their autocorrelation sequences are identical). Thus, without loss of generality, the range of $a$ specifying the MDC1a may be restricted to $-1 \leq a \leq 1$. Figure 6.11 indicates that $R_p^b(E_s)$ for the MDC1a is minimized when $|a| = 1$ independent of $E_s/N_0$ (formal proof of this result is given below). For large and small values of $E_s/N_0$, $R_p^b(E_s)$ is not sensitive to changes in $a$. Figure 6.11 seems to suggest that the channel responses which minimize $R_p^b(E_s)$ are independent of the value of $E_s/N_0$. However, it will be shown that this conclusion does not generalize; thus, the MDC1a represents a special case among the class of channels minimizing $R_p^b(E_s)$.

[AND-FOS(75)] calls the channel memory $v$, i.e., our $M = L - 1 = v$.

It is worth mentioning that the unit-sample responses listed in [MAG-PRO(73), Table 1] for $L \geq 7$, are not true MDC's, i.e., the error sequence $e(D) = 1 - D$ does not minimize $d_{\min}^2$ for channels with $M \geq 6$. As pointed out in [CAN-BUT(76-1)], Anderson and Foschini [AND-FOS(75)] provided the counterexample for the case $M = 6$. They found that the error sequence $e(D) = 1 - D - D^2 + D^3 + D^4 - D^5$ yields the true $d_{\min}^2 = 0.524 (= 4 \times 0.131)$. The (nonunique) MDC in this case has the unit-sample response $h^7 = (0.176, 0.316, 0.476, 0.532, 0.476, 0.316, 0.176)$. 
Fig. 6.10. Symmetric cutoff rate, $R^b_0(E_s)$, as a function of signal-to-noise ratio, $E_s/N_o$, for minimum-distance channels (MDC's) having ISI memory $M = 1$ (MDC1), $M = 2$ (MDC2), and $M = 3$ (MDC3), respectively. The unit-sample responses of unit energy are for the MDC1: $h^2 = (0.707, 0.707)$, MDC2: $h^3 = (0.500, 0.707, 0.500)$, and MDC3: $h^4 = (0.372, 0.602, 0.602, 0.372)$. 
Fig. 6.11. Symmetric cutoff rate, $R_{o}^b(E_s)$, for different values of $E_s/N_o$, for unit-energy channel with ISI memory $M = 1$, having unit-sample response $h^2 = (1, \alpha)/\sqrt{(1 + \alpha^2)}$, $-1 \leq \alpha \leq 1$, as a function of $|\alpha|$ (MDC1\alpha).
Table 6.4 ranks the PR channels listed in Table 6.3 and the corresponding MDC's (Figure 6.10) on the basis of decreasing $R_0^b(E_s)$, and it relates this rank to the Euclidean distance properties of the channels. In addition to $d^2_{\text{min}}$, Table 6.4 lists also the five smallest squared Euclidean distances found for each channel, i.e., $d^2_{\text{min}} = d_1^2 < d_2^2 < d_3^2 < d_4^2 < d_5^2$. The latter were determined using a procedure that is based on a stack algorithm which identifies and ranks error sequences according to their induced Euclidean distance [OEL(86)]. It is interesting to note from Table 6.4 that the first four channels (groups) have the same $d^2_{\text{min}} = 4$. In these cases, $d^2_{\text{min}}$ gives no clue about the rank with respect to $R_0^b(E_s)$. However, by considering the next higher ( squared) distances, a correspondence with the $R_0^b(E_s)$ properties becomes apparent. The reason for the small degradation in $R_0^b(E_s)$ of channels $\{F, G\}$ is found in the fact that, for these channels, only single errors produce $d^2_{\text{min}} = 4$, while for channels $\{A, B, C\}$, MDC1, and $\{H, I\}$, error sequences of length greater than one may also produce $d^2_{\text{min}} = 4$. The large degradation in $R_0^b(E_s)$ of channels $\{D, E\}$ is clearly a result of $d^2_{\text{min}} = 2.667$, which is not far from the lower-bound value, 2.343, obtained for the corresponding MDC with $M=2$ (MDC2).

Furthermore, the well performing channels $\{F, G\}$ (Table 6.4) belong to the class of channels with system polynomial of the form

$$F(D) = k + D - D^2 + ... - (-D)^k, \quad k = 1, 2, ... . \quad (6.55)$$

Particularly for large $k$, this class of channels is known to have minimal performance loss in terms of signal-to-noise ratio degradation due to ISI [KAB-PAS(75)]. As a final observation, it is noted that the ranking obtained for the PR systems listed in Table 6.3 with respect to decreasing $R_0^b(E_s)$, is identical with the ranking on the basis of increasing signal-to-noise ratio.
degradation as reported for the same systems in [KAB-PAS(75), Table V (Model 2, no precoding)].

Table 6.4: Ranking of the partial-response channels (Table 6.3) and minimum-distance channels (Figure 6.10) on the basis of decreasing symmetric cutoff rate in bits/symbol, $R_o^b(E_s)$, evaluated for $E_s/N_0 \equiv 0$ dB. The rank is independent of the signal-to-noise ratio, $E_s/N_0$. The five smallest squared Euclidean distances (for i.i.d., equally-likely binary inputs), $d_i^2$, $i = 1, 2, \ldots, 5$, $d_{\text{min}}^2 = d_1^2 < d_2^2 < d_3^2 < d_4^2 < d_5^2$, are also listed.

<table>
<thead>
<tr>
<th>Channel</th>
<th>$R_o^b(E_s)$</th>
<th>$d_{\text{min}}^2$</th>
<th>$d_2^2$</th>
<th>$d_3^2$</th>
<th>$d_4^2$</th>
<th>$d_5^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>no ISI</td>
<td>0.5481</td>
<td>4 1)</td>
<td>8</td>
<td>12</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>{F, G}</td>
<td>0.5125</td>
<td>4 1)</td>
<td>5.333</td>
<td>6.667</td>
<td>8</td>
<td>9.333</td>
</tr>
<tr>
<td>{A, B, C}, MDC1</td>
<td>0.4863</td>
<td>4 2)</td>
<td>8</td>
<td>12</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>{H, I}</td>
<td>0.4595</td>
<td>4 2)</td>
<td>6</td>
<td>8</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>{D, E}</td>
<td>0.4415</td>
<td>2.667</td>
<td>4</td>
<td>5.333</td>
<td>6.667</td>
<td>8</td>
</tr>
<tr>
<td>MDC2</td>
<td>0.4266</td>
<td>2.343</td>
<td>2.686</td>
<td>3.029</td>
<td>3.373</td>
<td>3.716</td>
</tr>
<tr>
<td>MDC3</td>
<td>0.3825</td>
<td>1.528</td>
<td>1.689</td>
<td>1.950</td>
<td>2.111</td>
<td>2.273</td>
</tr>
</tbody>
</table>

1) Only single errors produce $d_{\text{min}}^2$.
2) Error sequences of length greater than one may produce $d_{\text{min}}^2$. 
C. Channels Minimizing $R_o^b(E_s)$

The results obtained in the previous section have led to consider the problem of identifying those channels which minimize the symmetric cutoff rate, $R_o^b(E_s)$. These worst-case channels are defined and characterized in this section in terms of their unit-sample response of unit energy, given that the channel memory, $M$, and the signal-to-noise ratio, $E_s/N_o$, are specified.

Problem Definition and Example

Let $\mathcal{H}_M$ be the class of channels with unit-sample responses of unit energy, $h^{m+1} = (h_0, h_1, ..., h_m)$, $h_0 h_m \neq 0$, where $0 \leq m \leq M$, i.e., $\mathcal{H}_M$ defines the vector space

$$\mathcal{H}_M = \left\{ h^{m+1} : \|h^{m+1}\|^2 = 1, h_0 h_m \neq 0, 0 \leq m \leq M \right\}, \quad (6.56)$$

such that $\mathcal{H}_{M-1} \in \mathcal{H}_M$. Denote the minimum value of $R_o^b(E_s)$ over $\mathcal{H}_M$ as $\tilde{R}_o^b(E_s)$ and let $\tilde{h}^{m+1} \in \mathcal{H}_M$ be a channel response which achieves $\tilde{R}_o^b(E_s)$. Formally, the optimization problem to be solved becomes

$$\tilde{R}_o^b(E_s) = \min_{h^{m+1} \in \mathcal{H}_M} R_o^b(E_s), \quad (6.57)$$

or equivalently, by making use of (6.20),

$$\bar{\Gamma}_M = \max_{h^{m+1} \in \mathcal{H}_M} \Gamma, \quad (6.58)$$

where $\Gamma$ is the dominant eigenvalue of the reduced state transition matrix, $\tilde{A}$, corresponding to $h^{m+1} \in \mathcal{H}_M$ and some fixed value of $E_s/N_o$ (Theorem 6.2). The fact that $\Gamma$ is not described by a simple differentiable function for all $h^{m+1} \in \mathcal{H}_M$ makes a direct analytical approach to finding $\tilde{h}^{m+1}$ rather difficult. For the case $M = 1$, an analytical solution for $\tilde{h}^{m+1}$, and thus $\bar{\Gamma}_1$, ...
can be given; for the general case, however, a numerical optimization procedure is required.

**Example 6.1 (continued):** Consider the case $M = 1$, where $\tilde{A}$ is given by (6.50) and (6.37b). Note that $h^2(\alpha) = (h_0, h_1) = (1, \alpha)\sqrt{1 + \alpha^2}$, $-\infty < \alpha < \infty$, belongs to $\mathcal{H}_1$, so that $h^2(\alpha = 0) \equiv h^0 \in \mathcal{H}_0$ represents the memoryless channel. From $\det(\tilde{A} - y I) = 0$, where $I$ denotes the identity matrix, one obtains the characteristic equation

$$\gamma^2 - \gamma p(\alpha) - q(\alpha) = 0 , \quad (6.59a)$$

where

$$p(\alpha) = a_0 + a_2 + a_4$$

$$= \frac{1}{2} \left[ 1 + e^{-E_s/N_0} \cosh\left(\frac{2\alpha}{1 + \alpha^2} (E_s/N_0)\right) \right] , \quad (6.59b)$$

and

$$q(\alpha) = 2a_1a_3 - (a_0 + a_2)a_4$$

$$= \frac{1}{4} e^{-E_s/N_0} \left[ 1 - \cosh\left(\frac{2\alpha}{1 + \alpha^2} (E_s/N_0)\right) \right]$$

$$= \frac{1}{2} \left[ 1 + e^{-E_s/N_0} \right] - p(\alpha)$$

$$= \frac{p(0) - p(\alpha)}{2} . \quad (6.59c)$$

Note that $p(\alpha) \geq p(0) \geq 1/2$, so that $q(\alpha) \leq 0$, as can be seen from (6.59b) and (6.59c). The solution of (6.59a) is the dominant eigenvalue
where the expression under the square-root is always positive since 
\[ p^2(\alpha) + 4q(\alpha) = p^2(\alpha) - 2p(\alpha) + 2p(0) \geq p^2(\alpha) - 2p(\alpha) + 1 = [p(\alpha) - 1]^2 \geq 0. \]

From \( \frac{\partial y(\alpha)}{\partial \alpha} = 0 \), the condition for the extreme points of \( y(\alpha) \) is obtained as

\[
\frac{\partial p(\alpha)}{\partial \alpha} \left[ \sqrt{p^2(\alpha) + 4q(\alpha)} + p(\alpha) - 1 \right] = 0. \tag{6.61}
\]

One can verify that the expression in the square brackets of (6.61) can only be zero if \( p(\alpha) = 1 \), a condition which is only satisfied in the limit as \( E_s/N_o \to 0 \) as can be seen from (6.59b). Therefore, assuming that \( E_s/N_o > 0 \), the extrema of \( y(\alpha) \) are determined by the condition \( \frac{\partial p(\alpha)}{\partial \alpha} = 0 \). From (6.59b), the condition

\[
\frac{1 - \alpha^2}{1 + \alpha^2} \sinh\left( \frac{2\alpha}{1 + \alpha^2} (E_s/N_o) \right) = 0 \tag{6.62}
\]

is obtained, where \( 0 < E_s/N_o < \infty \) is assumed. Clearly, (6.62) is satisfied for

i) \( \alpha = 0 \), ii) \( \alpha = \pm \infty \), and iii) \( \alpha = \pm 1 \).

Cases i) and ii) are equivalent, i.e.,

\[ y(0) = y(\pm \infty) = p(0) = (1 + e^{-E_s/N_o})/2, \]

and they represent the memoryless channel \((M = 0)\). One can verify that, for \( \alpha = \pm 1 \), the second derivative of \( y(\alpha) \) with respect to \( \alpha \), \( \frac{\partial^2 y(\alpha)}{\partial \alpha^2} \), is indeed negative for \( 0 < E_s/N_o < \infty \); similarly, for \( \alpha = 0 \), it is positive. Therefore, \( y(\pm 1) = \max y(\alpha) \), i.e., \( y(\alpha) \geq y(0) \), where equality holds only in one of the two equivalent cases, \( \alpha = 0 \) and \( \alpha = \pm \infty \). Thus, for \( M = 1 \) and \( 0 < E_s/N_o < \infty \), \( R^b_0(E_s) \) is minimized when the unit-sample response of the channel is given as

\[ \bar{h}^2_1 = (1, 1)/\sqrt{2}, \]

or \( \bar{h}^2_2 = (1, -1)/\sqrt{2} \), i.e., the minimizing response, \( \bar{h}^{m+1} \), is
not unique and it has memory \( m = 1 = M \). Furthermore, with \( \alpha = \pm 1 \), (6.60) yields the maximized eigenvalue as

\[
\bar{\gamma}_1 = \frac{3 + e^{-2E_s/N_0}}{8} \left[ 1 + \sqrt{1 - 8 \left( \frac{1 - e^{-E_s/N_0}}{3 + e^{-2E_s/N_0}} \right)^2} \right], \tag{6.63}
\]

so that \( \bar{R}_o^b(E_s) = -\log \bar{\gamma}_1 \) (see also Figure 6.11).

For \( M = 1 \), the system polynomial minimizing \( R_o^b(E_s) \) was found to be of the form \( F(D) = 1 \pm D \); thus, any channel with system polynomial \( F(D) = 1 \pm D^k, \ k = 1, 2, \ldots \), is a so-called minimum-\( R_o^b(E_s) \) channel. Figure 6.12 illustrates the above analysis by showing the dominant eigenvalue, \( \gamma(\alpha) \), in the range \(-10 \leq \alpha \leq 10\), for different values of \( E_s/N_0 \). Note that \( \gamma(\alpha) = \gamma(1/\alpha) \), as can be seen from (6.59b), (6.59c) and (6.60). It is thus sufficient to consider \( \alpha \) in the range where \( |\alpha| \leq 1 \), as shown earlier for the MDC1\( \alpha \). Furthermore, Figure 6.12 shows that \( \gamma(\alpha) \) is not a convex function over this range of \( \alpha \).

The above example has indicated that \( \Gamma \) in (6.58), and hence \( R_o^b(E_s) \) in (6.57), are in general not convex over the class of channel responses in \( \mathcal{H}_M \). It has also been shown that the minimum-\( R_o^b(E_s) \) channels can be found, at least in principle, by solving the characteristic equation \( \det(\tilde{A} - \gamma I) = 0 \) for its largest root. However, such a direct analytical approach becomes intractable as \( M \) increases. Since any \( h^{m+1} \in \mathcal{H}_M \) determines \( \Gamma \) in a rather indirect way, the gradient of \( \Gamma \) with respect to \( h^{m+1} \) is generally not directly available so that nonlinear optimization techniques [DI(X(72))] must be used to determine \( \bar{R}_o^b(E_s) \).

In applications of optimization techniques that use numerical estimates of the gradients, difficulties often arise when numerical round-off errors affect these estimates and interact with the chosen convergence criterion. To avoid these potential difficulties, it was decided to adopt a so-called direct search technique (DST) [DI(X(72), Chapter 5)]. This technique requires only evaluation of the
Fig. 6.12. Dominant eigenvalue, $\gamma(\alpha) = \gamma(1/\alpha)$, for unit-energy channel with memory $M = 1$, having unit-sample response $h^2 = (1, \alpha)/\sqrt{1 + \alpha^2}$, for different values of $E_s/N_o$ (Example 6.1). $\gamma(\alpha)$ is maximized, independent of $E_s/N_o$, for $|\alpha| = 1$, i.e., $\Gamma_1 = \gamma(\pm 1)$, and $R_o^b(E_s) = -\log_2 \gamma(\pm 1)$ (see also Figure 6.11).
function to be optimized and does not estimate its gradient. The DST used here
is based on an efficient and very robust algorithm originally proposed by
Rosenbrock [ROS(60)] and later also described in [FLE(69), pp. 191-193] and
[DIX(72), Section 5.2]. For the present problem, the method was modified in
the way the constraint function is included. The algorithm is described below.

Direct Search Algorithm to Determine $\overline{y}_M$

The problem being considered is that of finding the maximum of the
objective function

$$\Gamma = f(z^{M+1})$$  \hfill (6.64a)

subject to the constraint

$$\|z^{M+1}\|^2 = 1,$$  \hfill (6.64b)

where

$$z^{M+1} = (z_0, z_1, \ldots, z_M)$$  \hfill (6.64c)

$$= (h_0, h_1, \ldots, h_m, 0, 0, \ldots, 0), \quad 0 \leq m \leq M,$$

e.i., $h^{m+1} = (h_0, h_1, \ldots, h_m) \in \mathcal{H}_M$. Constraint (6.64b) means that the
end-point of $z^{M+1}$ must lie on the $M$-dimensional surface of a unit sphere
centered at the origin of the $(M+1)$-dimensional variable space. Thus, (6.64b)
may be written in the equivalent form

$$z_j = \pm \left[1 - \sum_{i=0}^{M^*} z_i^2\right]^{1/2}, \quad j \in \{0, 1, \ldots, M\},$$  \hfill (6.65)

where $\Sigma^*$ indicates that $i \neq j$, and the sign of $z_j$ is arbitrary, since
$f(z^{M+1}) = f(-z^{M+1})$ in the present case. Thus, any of the $z_j, 0 \leq j \leq M$,
may be expressed in terms of the remaining $M$ components of $z^{M+1}$ and can be eliminated from the argument of the objective function in (6.64a). Defining the new vector variable

$$x^M = (x_0, x_1, \ldots, x_{M-1})$$

(6.66a)

$$= (z_0, z_1, \ldots, z_{j-1}, z_{j+1}, \ldots, z_M),$$

the original problem can now be treated as an essentially unconstrained problem in $M$ dimensions with the objective to maximize the modified function

$$\Gamma = f(x^M).$$

(6.66b)

Obviously, (6.65) implies that we must have $\|x^M\|^2 \leq 1$, a condition which is not difficult to handle, e.g., the norm of $x^M$ can be normalized to unity whenever $\|x^M\| > 1$ is encountered during the search process.

Rosenbrock's unconstrained direct search algorithm [ROS(60)], as described in Appendix 6.A, can now be applied to the modified optimization problem formulated above.

Example 6.2 (continued): The direct search algorithm, described above and in Appendix 6.A, was applied to the example $M = 2$ where the channel's unit-sample response has the form $h^3 = (h_0, h_1, h_2)$. The vector variable $x^2 = (x_0, x_1) \equiv (h_1, h_2)$ was used, together with $h_0 = \sqrt{1 - h_1^2 - h_2^2}$ to satisfy $\|h^3\|^2 = 1$. The performance of this algorithm for $E_s/N_0 = 0$ dB, and the solutions obtained for $\overline{R}^{\overline{m} + 1}$ and $R_o^b(E_s)$, are illustrated in Figure 6.13. The initial vector variable was chosen to be $x_0^2 = (0, 0)$, i.e., the initial unit-sample response, $h^3 = (1, 0, 0)$, corresponds to the memoryless channel. The search
Fig. 6.13. Performance of direct search algorithm (Example 6.2):
(a) Convergence of energy normalized unit-sample response as a function of stage iterations for $E_s/N_0 = 0$ dB. In Algorithm 6.2 (Appendix 6.A), the vector variable $x^2 = (x_0, x_1) = (h_1, h_2)$, and the parameters $\tilde{\alpha} = 2$, $\tilde{\beta} = 1/2$, and $\Delta = 1/10$. The final channel response, to within 6 significant digits, is obtained after 11 stage iterations as $\tilde{h}^3 = (0.509278, 0.693737, 0.509278)$.
(b) Corresponding convergence of the symmetric cutoff rate towards its minimum value, $\bar{R}_o^b(E_s) = 0.426458$ bits/symbol.
parameters used in Algorithm 6.2 (Appendix 6.A) were $\tilde{\alpha} = 2$, $\tilde{\beta} = 1/2$, and $\Delta = 1/10$. The objective function, $\Gamma = f(x^2)$, was computed iteratively using Algorithm 6.1 with accuracy $\varepsilon = 10^{-12}$. To be sure, the asymptotic results shown in Figure 6.13 were confirmed for different initial vectors, $x_0^2$; it can be concluded, that $R_0^b(E_s)$ has a unique minimum value. Furthermore, the minimizing unit-sample response, $h^{M+1}$, has the longest possible memory $\bar{m} = 2 = M$. The fact that the minimum-$R_0^b(E_s)$ channel has a symmetric unit-sample response, i.e., $h_0 = h_2$, is significant, and it is intuitively pleasing in view of the properties of the MDC's considered earlier. The (asymptotic) solution shown in Figure 6.13(a) is not unique; the complete set of channel responses yielding $R_0^b(E_s)$, for $E_s/N_0 \equiv 0$ dB, is found to be $h_1^3 = (0.509278, 0.693737, 0.509278)$, $h_2^3 = (-0.5092785, 0.693737, -0.509278)$, $h_3^3 = -h_1^3$, and $h_4^3 = -h_2^3$. Figure 6.14(a) demonstrates that the channel coefficients minimizing $R_0^b(E_s)$ are not constant, but they vary as a function of $E_s/N_0$ around the corresponding channel coefficients of the MDC2, $h_0 = h_2 = 1/2$, and $h_1 = 1/\sqrt{2}$. While the largest difference between the corresponding coefficients of the MDC and the minimum-$R_0^b(E_s)$ channel, $|h_i - h_i|$, $0 \leq i \leq 2$, is numerically significant, the resulting difference between the respective cutoff rates, $R_0^b(E_s)$ and $R_0^b(E_s)$, is not significant (for all practical purposes), as can be seen in Figure 6.14(c). 1

Summary of Results

Results similar to those obtained in the previous example for $M = 2$ apply for $M > 2$; the generalized results are summarized below. Reflecting the nature of the problem considered here, they were obtained by a combination of extensive numerical evaluation and functional analysis.

1 For clarity, the unit-sample response of the MDC, and the corresponding symmetric cutoff rate, are called $h^{M+1}$, and $R_0^b(E_s)$, respectively.
Fig. 6.14. Minimum-$R_o^b(E_s)$ channel for $M = 2$ (Example 6.2):
(a) Energy normalized unit-sample response coefficients, $\bar{h}_i$, $i = 0, 1, 2$, as a function of $E_s/N_o$.
(b) Resulting minimized cutoff rate, $\bar{R}_o^b(E_s)$, in bits/symbol.
(c) Difference between the symmetric cutoff rate of the MDC2, $\hat{R}_o^b(E_s)$, and the minimized cutoff rate, $\bar{R}_o^b(E_s)$, in bits/symbol.
Conjecture 6.1: Let

\[ X_M^s = \{ X_M : h_i = h_{M-i}, \ 0 \leq i \leq M, \ h_0h_M \neq 0 \} \quad (6.67a) \]

and

\[ X_M^a = \{ X_M : h_i = -h_{M-i}, \ 0 \leq i \leq M, \ h_0h_M \neq 0 \} \quad (6.67b) \]

be subsets of \( X_M \) defined in (6.56), where the superscripts \( s \) and \( a \) denote, respectively, symmetric and antisymmetric (also skew symmetric). Then, the channel unit-sample responses minimizing \( R_o^b(E_o) \) in (6.57), \( \hat{h}^{M+1} \), belong to the subsets \( X_M^s \) and \( X_M^a \), i.e., \( \hat{h}^{M+1} \in \{ X_M^s \cup X_M^a \} \), and \( M = M \). For \( M > 1 \), the coefficients of the channel's unit-sample response minimizing \( R_o^b(E_o) \), \( \hat{h}_i, i = 0, 1, \ldots, M \), vary significantly with \( E_s/N_o \) in the vicinity of the corresponding coefficients of the MDC response with the same memory, \( \hat{h}_i, i = 0, 1, \ldots, M \). However, \( \bar{R}_o^b(E_o) \) is not significantly different (for all practical purposes) from \( R_o^b(E_o) \), the symmetric cutoff rate of this MDC. Moreover, as \( E_s/N_o \to \infty \), \( \hat{h}^{M+1} \to \hat{h}^{M+1} \), and for \( E_s/N_o \to 0 \), the actual numerical value of \( \bar{h}^{M+1} \) becomes unimportant since \( R_o^b(E_o) \to 0.7213E_s/N_o \) (bits/symbol), for any unit-energy channel. Thus, for all practical purposes,

\[ \bar{R}_o^b(E_o) \approx \hat{R}_o^b(E_o), \quad (6.68) \]

i.e., the MDC unit-sample response, \( \hat{h}^{M+1} \), yielding \( \hat{R}_o^b(E_o) \), provides a practical approximation for the minimum-\( R_o^b(E_o) \) channel unit-sample response, \( \bar{h}^{M+1} \). Also, \( \bar{R}_o^b(E_o) \) represents a very tight upper bound on \( R_o^b(E_o) \), independent of \( E_s/N_o \).

---

1 In general, the \( \hat{h}_i \) do not necessarily attain the same values as \( \hat{h}_i \) for some finite, intermediate value of \( E_s/N_o \), as might be deduced from Figure 6.14(a) in Example 6.2. The counterexample can be given for \( M = 6 \), where the \( \hat{h}_i \) attain the values of the \( \hat{h}_i \) only asymptotically as \( E_s/N_o \to \infty \).
Supporting arguments for the validity of Conjecture 6.1 may be constructed as follows. Let

\[
\gamma_{N+M} = \frac{1}{K^2} \sum_{m=0}^{K-1} \sum_{n=0}^{K-1} \exp \left\{ - \left( E_z / N_o \right) \|d_{mn}^{N+M} \|^2 / 4 \right\} \tag{6.69a}
\]

where \( K = 2^N \) and

\[
\|d_{mn}^{N+M} \|^2 = 4 \sum_{k=0}^{N+M-1} \left[ \sum_{i=P}^{Q} h_i e_{k-i}^{mn} \right]^2, \tag{6.69b}
\]

with \( P = \max(0, k - N + 1) \), \( Q = \min(M, k) \), \( e_{i}^{mn} \in \{-1, 0, 1\} \), \( 0 \leq i < N \), and \( 0 \leq m, n < K \). Since \( M \) is finite, it is easily seen that \( R_o^b(E_\delta) \) as defined in (6.1) is also obtained as

\[
R_o^b(E_\delta) = \lim_{N \to \infty} - \frac{1}{N} \log \gamma_{N+M}. \tag{6.70}
\]

Define the \((M + 1) \times (M + 1)\) matrix \( B_N \) corresponding to some error sequence \( e^N = (e_0, e_1, \ldots, e_{N-1}) \) as

\[
B_N = E_N^T E_N = [b_{ik}], \quad 0 \leq i, k \leq M, \tag{6.71a}
\]

where \( E_N \) is the \((N + M) \times (M + 1)\) matrix.
and $T$ denotes transpose. Then, (6.69b) can be written as

$$\mathbf{E}_N = \begin{bmatrix} e_0 & 0 & 0 & \ldots & 0 \\ e_1 & e_0 & 0 \\ \vdots & e_1 & \ddots \\ \vdots & \vdots & \ddots & \ddots \\ 0 & e_{N-1} & 0 & \ldots & 0 & e_{N-1} \end{bmatrix}$$ (6.71b)

Consider now the case where $E_s/N_0 \to \infty$. Clearly, among the $K(K-1)$ exponential terms in (6.69a) which are not unity ($m \neq n$), those where $\|d_{mn}\|$ assumes the smallest (nonzero) value will dominate the double sum. Thus, the minimum value of $\|d_{mn}\|^2$ is of interest, where the minimum is taken over all channel responses in $\mathcal{H}_M$ and over the sets $\mathbf{E}_N$, $N = 1, 2, \ldots$, of all nonzero error sequences, $e^N$. Formally,

$$\Delta_{min}^2 = \min_{h^{M+1} \in \mathcal{H}_M} \min_{e^N \in \mathbf{E}_N} 4 \langle h^{M+1}, B_N h^{M+1} \rangle$$ (6.73)
is being sought. Without loss of generality, \( e_0 = 1 \) may be assumed in (6.71b). It is well known that, for any given error sequence, \( e^N = (e_0, e_1, \ldots, e_{N-1}) \), the unit-sample response of unit energy which minimizes the inner product (squared Euclidean distance) in (6.73), equals the eigenvector of \( B_N \) associated with the smallest eigenvalue, \( \lambda_{\text{min}} \); furthermore, \( \lambda_{\text{min}} \) represents the value of this minimum inner product [MAG-PRO(73)], [MAK(81)]. The eigenvectors of \( B_N \) are either symmetric or antisymmetric (skew symmetric) because \( B_N \) is a persymmetric (also called symmetric centrosymmetric or doubly symmetric) matrix, i.e., 
\[
\hat{b}_{ik} = \hat{b}_{ki} = b_{M-i,M-k}, \quad 0 \leq i, k \leq M
\]
[CAN-BUT(76-1)], [CAN-BUT(76-2)]. The above properties hold for all \( e^N \in E_N \) and in particular for the error sequences which provide the solutions for (6.73). Thus, \( \hat{d}_{\text{min}}^2 \) is obtained for some minimizing unit-sample response \( \hat{h}^{M+1} \in \{H_M^s \cup H_M^a \} \), with \( H_M^s \) and \( H_M^a \) defined in (6.67). The response \( \hat{h}^{M+1} \) characterizes the MDC with channel memory \( M \) (a different method, described in [AND-FOS(75)], leads to the same result). Since \( \hat{d}_{\text{min}}^2 \) will dominate the double sum in (6.69a) for \( E_s/N_o >> 1 \), it is expected that \( \hat{h}^{M+1} \rightarrow \hat{h}^{M+1} \) and thus \( \hat{R}_o^b(E_s) \rightarrow \hat{R}_o^b(E_s) \), as \( E_s/N_o \rightarrow \infty \). This proves the validity of (6.68) in the limit as \( E_s/N_o \rightarrow \infty \) (see also Figure 6.14). The proof for the case where \( E_s/N_o \rightarrow 0 \) is provided by Corollary 6.2. For intermediate values of \( E_s/N_o \), it is observed that for \( e^N \in E_N, \forall N \), each exponential in (6.69a) is individually maximized for some \( h^{M+1} \in \{H_M^s \cup H_M^a \} \). From this observation, it seems plausible to have \( \hat{h}^{M+1} \in \{H_M^s \cup H_M^a \} \) as the maximizing unit-sample response in (6.58). Thus, based on numerical results and the arguments given above in support of Conjecture 6.1, strong evidence has been established that the MDC's are not only important in their own right but that they are also closely related to the minimum-\( R_o^b(E_s) \) channels.

Using the results outlined in Conjecture 6.1 and applying the direct search algorithm described above, we determined \( \hat{R}_o^b(E_s) \) for the channels with memory
$M = 1, 2, \ldots, 8$. Figure 6.15 shows, for different values of $E_s/N_0$, how $\hat{R}_o^b(E_s)$ is reduced as $M$ increases. As predicted by Corollary 6.2, $\hat{R}_o^b(E_s)$ is not sensitive to increasing $M$ for small values of $E_s/N_0$. For intermediate values of $E_s/N_0$, however, $\hat{R}_o^b(E_s)$ is rather sensitive to the length of the channel memory.

The results in Figure 6.15 establish worst-case limits (lower bounds) for the symmetric cutoff rate, $R_o^b(E_s)$, of any unit-energy channel with fixed memory, $M$, and given signal-to-noise ratio, $E_s/N_0$. In particular, these results represent very tight lower bounds on $\hat{R}_o^b(E_s)$ of the corresponding MDC's, as shown in Example 6.2 and verified numerically for $M = 1, 2, \ldots, 6$. 

Figure 6.16 presents the results of Figure 6.15 from a different perspective. Figure 6.16(a) indicates that, for $R_o^b(E_s)$ fixed at 0.6 bits/symbol, the noise margin is reduced by about 1 dB for each unit increment (one normalized symbol interval) in the channel memory. In other words, for any channel with ISI memory in the range $1 \leq M \leq 8$, operating at a signal-to-noise ratio, $E_s/N_0$, such that $R_o^b(E_s) \approx 0.6$ bits/symbol, the expected worst-case loss in noise margin, with respect to the memoryless channel, amounts to about $M$ dB.

Figure 6.16(b) shows that the relative worst-case reduction of $R_o^b(E_s)$, due to channel memory, is largest at $E_s/N_0 \equiv 1$ dB for $M = 1$, and $E_s/N_0 \equiv 3$ dB for $M = 8$. Thus, the worst-case effect of channel memory, in the range $1 \leq M \leq 8$, occurs roughly when the channel is operating at a signal-to-noise ratio in the range $1 \leq E_s/N_0 \leq 3$ dB, where the symmetric cutoff rate falls in the range $1/3 \leq R_o^b(E_s) \leq 2/3$ bits/symbol. This also happens to be the range of rates that is often of practical interest when coding is considered (see, e.g., [WOL-UNG(86)])

---

1 For $M > 6$, the MDC's are not known.
Fig. 6.15. Minimum (worst-case) symmetric cutoff rate, $\bar{R}_o^b(E_s)$, in bits/symbol, as a function of the channel memory, $M = 0, 1, \ldots, 8$, for different values of $E_s/N_o$. 
Fig. 6.16. (a) Minimum (worst-case) symmetric cutoff rate, $R_o^b(E_s)$, in bits/symbol, as a function of $E_s/N_o$, for channel memory $M = 1, 2, \ldots, 8$.

(b) Resulting relative rate loss, $R_o^b(E_s)/R_o^b(E_s)_{\text{no ISI}}$, representing a measure for expected worst-case reduction of symmetric cutoff rate, due to ISI memory of length $M = 1, 2, \ldots, 8$.
APPENDIX 6.A

Rosenbrock's Unconstrained Direct Search Algorithm

An efficient and very robust direct search algorithm, originally introduced by Rosenbrock [ROS(60)], and later also described in [FLE(69)], and [DIX(72), Section 5.2], is introduced in this appendix. The algorithm solves the basic *unconstrained* optimization problem, where \( f(x^n) \) is a given objective function to be *maximized* over the \( n \) components of the vector variable \( x^n = (x_0, x_1, \ldots, x_{n-1}) \). For notational convenience, the dimensionality index, \( n \), will be dropped in the following vector notation.

**Algorithm 6.2:** Let \( r_j, 1 \leq j \leq n \), be a set of orthonormal direction vectors, and denote the associated *step-lengths* to be taken along them by \( \rho_j, 1 \leq j \leq n \). The initial set of directions is chosen as the set of unit vectors parallel to the coordinate axes and the initial step-lengths are either arbitrary or they are chosen in relation to the anticipated variable changes. The best *current* function value is denoted by \( f_0 \) and the corresponding vector variable is called \( x_0 \), i.e., \( f_0 = f(x_0) \). A *success* is defined to be a new step, taken in any of the above directions, that results in a function value *larger or equal* to \( f_0 \). Similarly, a *failure* is a new step, taken in any of the above directions, that results in a function value *smaller* than \( f_0 \).

An *iteration* consists of taking a single step in each direction in turn until, for each direction, a success has been followed by a failure. After a success, the step-length associated with the successful direction is multiplied by the factor \( \tilde{\alpha} > 1 \) before a new step is taken in the same direction; after a failure, the step-length is similarly multiplied by the factor \( \tilde{\beta}, 0 < \tilde{\beta} < 1 \). Furthermore, during each iteration and for each direction \( r_j, 1 \leq j \leq n \), the algebraic sum, \( \omega_j, 1 \leq j \leq n \), of all the successful step-lengths, \( \rho_j \), is recorded. Note that an iteration must necessarily complete, because the step-lengths become smaller and...
smaller after repeated failures so that the function value will not change at some point and a success will be recorded.

After each iteration defined above, a new set of orthonormal direction vectors is computed from the auxiliary vectors

\[ a_j = \sum_{k=j}^{n} \omega_k r_k, \quad 1 \leq j \leq n \quad (6.A.1) \]

by applying the Gram-Schmidt procedure [LUE(69), Section 3.5] to the set of vectors generated by (6.A.1). Thus, the new set of direction vectors is obtained from

\[ \tilde{r}_1 = a_1 / \|a_1\|, \quad (6.A.2a) \]

and

\[ \tilde{r}_j = b_j / \|b_j\|, \quad 2 \leq j \leq n, \quad (6.A.2b) \]

where

\[ b_j = a_j - \sum_{k=1}^{j-1} \langle a_j, \tilde{r}_k \rangle \tilde{r}_k, \quad (6.A.2c) \]

by letting \( r_j \leftarrow \tilde{r}_j, \quad 1 \leq j \leq n \). Geometrically, the algorithm tends to rotate the search directions, \( r_j, \quad 1 \leq j \leq n \), until \( r_1 \) becomes parallel to the direction of fastest advance, \( r_2 \) approaches the best direction normal to \( r_1 \), and so on (see also [DIX(72), Figure 5.2] for an illustration of the performance of this algorithm on a narrow ridge). An iteration combined with the computation of new search directions, according to (6.A.2), is called a stage iteration.

Figure 6.A.1 shows the basic flow chart of the direct search algorithm described above. The convergence criterion used to terminate the search depends usually on the function to be maximized (e.g., the step-lengths,
\( \rho_j, 1 \leq j \leq n, \) may be used to construct a criterion for the termination of the algorithm).
Fig. 6.A.1: Flow chart of the basic direct search algorithm (Algorithm 6.2).
CHAPTER 7

New Applications and Properties of the Symmetric Cutoff Rate

We consider the discrete-time Gaussian channel with intersymbol interference (ISI) and i.i.d. binary inputs of equal probability as described in Chapter 4 (Figure 4.1). For this channel model, the symmetric cutoff rate, \( R_o^b(E_s) \), was defined and evaluated in Chapter 6. This chapter presents some new applications and properties of \( R_o^b(E_s) \) that have not been noted previously. Section 7.1 introduces a simple approximation of the minimum squared Euclidean distance, \( d_{\text{min}}^2 \), for uncoded binary signaling over channels with time-dispersive memory. A method for approximating the unit-sample response of so-called minimum-distance channels (MDC's) is described in Section 7.2. Finally, Section 7.3 provides a simple, but tight, lower bound on \( R_o^b(E_s) \).

7.1 Approximation of \( d_{\text{min}}^2 \) by Means of \( R_o^b(E_s) \)

The minimum distance, \( d_{\text{min}} \), is the basic performance parameter for (uncoded) systems with maximum-likelihood sequence estimation (MLSE), where the received signals suffer from ISI and are corrupted by additive white Gaussian noise (AWGN) [FOR(72)], [UNG(74)], [FOS(75)]. The fundamental importance of \( d_{\text{min}} \) stems from the fact that upper and lower bounds on the probability of symbol error are of the form [FOR(72)]

\[
c Q \left( \frac{1}{2} \sqrt{\frac{E_s}{N_0}} d_{\text{min}}^2 \right)
\]

where \( c \) is some constant which is independent of the one-sided noise power spectral density, \( N_0 \), where \( E_s \) is the average symbol energy and where
\[ Q(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} e^{-y^2/2} dy \]

is the error function. The parameter \( d_{\text{min}}^2 \) can be viewed as the square of the smallest possible Euclidean distance between noiseless received signals which are caused by pairs of transmitted symbol sequences differing in a finite number of symbols. Alternatively, \( d_{\text{min}}^2 \) can be defined in terms of a minimization problem over all nonzero error sequences (see also Sections 6.1 and 6.4.3). In this case, \( d_{\text{min}}^2 \) is called the minimum (Euclidean) weight of any (nonzero) error sequence [MAG-PRO(73)]. For binary signaling over the memoryless channel (no ISI) with unit-energy response, \( d_{\text{min}}^2 = 4 \) as can be seen from (6.5b) or (6.69b).

In many situations, it is rather difficult to determine or bound \( d_{\text{min}}^2 \), owing mainly to the lack of suitable "closed-form" expressions for \( d_{\text{min}}^2 \) for MLSE systems of interest. The usual methods of solution are to use a Monte-Carlo computer search technique [OEL(86)] or to apply a combination of functional-analysis and computer-search [AND-FOS(75)]. For channels with moderately long ISI memory (e.g., \( M > 5 \)), such procedures can be very costly in terms of computer run-time. Moreover, there exists often some uncertainty as to whether the true value for \( d_{\text{min}}^2 \) was found when the search is terminated. In any case, an efficient method to determine or approximate \( d_{\text{min}}^2 \) is of great practical interest.

In this section, we consider systems where i.i.d. (equally likely) binary symbols are transmitted over the discrete-time AWGN channel with finite ISI memory, \( M \), having unit-sample response \( h^{M+1} = (h_0, h_1, \ldots, h_M) \), \( \|h^{M+1}\|^2 = 1 \), and where the receiver performs MLSE. For such (uncoded) MLSE systems, we introduce a simple expression based on the symmetric cutoff rate of the channel, \( R_o^b(E_s) \), which yields \( d_{\text{min}}^2 \) asymptotically as the signal-to-noise ratio, \( E_s/N_o \), increases. Successful application of this new method relies on the ability
to compute $R_o^b(E_s)$ efficiently with a high degree of accuracy as it approaches unity (1 bit/symbol).

\textit{A. Derivation of Main Result}

We have

$$R_o^b(E_s) = \lim_{N \to \infty} - \frac{1}{N} \log_2 \gamma_{N+M}$$

(7.1a)

where

$$\gamma_{N+M} = E \left[ \exp \left\{ - (E_s / N_0) \|d_m^{N+M} \|^2 / 4 \right\} \right], \quad 0 \leq m, n < K, \quad (7.1b)$$

$K = 2^N$, and $\|d_m^{N+M} \|^2$ is given by (6.69b) with the distance vectors, $d_m^{N+M}$, being equally likely with probability $K^{-2}$. As the signal-to-noise ratio increases (i.e., as $E_s / N_0 \to \infty$), only those terms in the expectation of (7.1b) will contribute significantly where the corresponding error sequence in (6.69b) is either the zero sequence (i.e., $\|d_m^{N+M} \|^2 = 0$) or one of minimum Euclidean weight (i.e., $\|d_m^{N+M} \|^2 = d_{min}^2$). Recall that the symmetric cutoff rate, $R_o^b(E_s)$, was obtained from the \textit{random-coding bound} for the average block error probability. The latter is expressed as the expectation of the Bhattacharyya bound for codes with two codewords of length $N$ [MAS(83), Chapter 5]. From the point of view of random coding, the behaviour of $R_o^b(E_s)$ in the case where $E_s / N_0 \to \infty$, is equivalent to considering a reduced ensemble of codes which consists only of those codes with two codewords where the codewords are either identical (i.e., $\|d_m^{N+M} \|^2 = 0$) or least different (i.e., $\|d_m^{N+M} \|^2 = d_{min}^2$). In other words, the poorest codes in the ensemble of two randomly chosen codewords dominate the average block error probability as $E_s / N_0 \to \infty$, even though these extreme codes are quite improbable. Moreover, these extreme codes exhibit the same
Bhattacharyya bound as for all codes with two codewords so that there is no need for averaging over the ensemble of codes. \(^1\)

Thus, one may choose any pair among the poorest codes to represent the ensemble of extreme codes, and consider an equivalent memoryless channel where the codeword length \(N = 1\) . For this case, using (7.1) and (6.20), one obtains

\[
R_o^b(E_s) = - \log_2 \Gamma \\
\approx - \log_2 \left[ \frac{1 + e^{- (E_s/N_0) d_{\min}^2 / 4}}{2} \right], \quad E_s / N_0 \gg 1,
\]

where the last expression in (7.2) has the same form as the cutoff rate of the memoryless channel,

\[
R_o^b(E_s) \bigg|_{M=0} = - \log_2 \left[ \frac{1 + e^{- E_s / N_0}}{2} \right].
\]

Requiring that

\[
R_o^b(E_s) = R_o^b(E_s^*) \bigg|_{M=0} \bigg|_{h_0=1},
\]

one obtains from (7.2) and (7.4)

\[
d_{\min}^2 \approx 4 \left( E_s^* / N_0 \right) / (E_s / N_0), \quad E_s / N_0 \gg 1,
\]

---

\(^1\) Equivalently, we are saying that expurgation of all the codes which are better than any of the two types of poorest codes (those yielding \( \|d_{mn}^{N+M} \|^2 = 0 \) or \( \|d_{mn}^{N+M} \|^2 = d_{\min}^2 \)) does not change the bound of the average block error probability, and thus \( R_o^b(E_s) \), in the limit as \( E_s / N_0 \to \infty \). In a way, we are doing the opposite of what leads to an improved random-coding bound at low rates by expurgating the poor codewords in an ensemble of codes [GA1.(68), Section 5.7 ].
where

\[ E_s^* / N_o = - \ln \left( 2^{1 - R_0^b(E_s)} - 1 \right). \]  

(7.5b)

Defining the quantity

\[ D^2 = \frac{4 (E_s^* / N_o)}{(E_s / N_o)} \]

\[ = - 4 \ln \left( 2^{1 - R_0^b(E_s)} - 1 \right) \]

\[ = \frac{- 4 \ln (2 \Gamma - 1)}{(E_s / N_o)} \]

(7.6a)

one obtains as the final result

\[ d_{\text{min}}^2 = \lim_{E_s / N_o \to \infty} D^2. \]  

(7.6b)

The quantity \( D^2 \) is a function of \( R_0^b(E_s) \), or equivalently, of the dominant eigenvalue, \( \Gamma \), of the reduced state transition matrix, \( \tilde{A} \) (see Theorem 6.2), and of the signal-to-noise ratio, \( E_s / N_o \), for which \( R_0^b(E_s) \), or equivalently \( \Gamma \), is evaluated. The simple expression for \( D^2 \) yields \( d_{\text{min}}^2 \) asymptotically as \( E_s / N_o \to \infty \) while, for finite but large values of \( E_s / N_o \), \( D^2 \) approximates \( d_{\text{min}}^2 \). The significance of this result stems from the fact that \( R_0^b(E_s) \) and \( \Gamma \) can be evaluated efficiently and with high accuracy for large values of \( E_s / N_o \) so that the quantity \( D^2 \) should provide a good approximation for \( d_{\text{min}}^2 \). The following applications demonstrate the practical usefulness of this (to our knowledge) new approach of estimating \( d_{\text{min}}^2 \) for uncoded binary signaling on channels with ISI memory.
B. Numerical Results

Figure 7.1 demonstrates that the quantity $1 - R_0^b(E_s)$ in (7.6a) can be evaluated accurately for large values of $E_s/N_0$. The similar appearance of these curves compared with usual probability-of-error performance curves is not unexpected in view of the foregoing discussion. As a reminder, probability of error as a modulation system- or channel design criterion is only useful for systems without coding or those where the demodulator delivers only hard decisions [MAS(74)]. Here, we are interested in relating $R_0^b(E_s)$ to the probability of symbol error of uncoded MLSE systems via the squared minimum Euclidean distance, $d_{\text{min}}^2$.

For a fixed value of $1 - R_0^b(E_s)$, let

$$\Delta_{\text{SNR}} = (E_s/N_0)_d B - (E_{s}^*/N_0)_d B$$

(7.7a)

where $E_{s}^*/N_0$ is the signal-to-noise ratio for the memoryless channel (C1 in Figure 7.1) and $E_s/N_0$ corresponds to the ISI channel of interest. From (7.6), one then obtains

$$d_{\text{min}}^2 \approx D^2 = 4 \times 10^{-\Delta_{\text{SNR}}/10}, \quad E_s/N_0 \gg 1.$$  

(7.7b)

The numerical results obtained from Figure 7.1 for $1 - R_0^b(E_s) = 10^{-20}$ are given in Table 7.1. For the channels C2 and C3, the approximation for $d_{\text{min}}^2$ is exact, and for the channels C4 and C5, $D^2$ over-estimates $d_{\text{min}}^2$ by about 1% and 4%, respectively.

Figure 7.2 shows the quantity $D^2$ as a function of $E_s/N_0$ for the channels of Figure 7.1. Whether $D^2$ is an upper or a lower bound on $d_{\text{min}}^2$ depends generally on the channel itself and the signal-to-noise ratio, $E_s/N_0$. However, it is significant that, for large values of $E_s/N_0$, $D^2$ yields $d_{\text{min}}^2$ exactly (channels C2 and C3), or approximates it closely (channels C4 and C5). In general, the
Fig. 7.1. $1 - R_{\theta}(E_s)$ as a function of $E_s/N_o$ for channels C1, C2, C3, C4 and C5 with unit-energy response characterized as follows:

C1: memoryless channel (dashed line);

C2: $h^3 = (2, 1, -1)/\sqrt{6}$;

C3: $h^3 = (1, 2, 1)/\sqrt{6}$;

C4: MDC3, $h^4 = (0.372, 0.602, 0.602, 0.372)$;

C5: MDC6, $h^7 = (0.176, 0.316, 0.476, 0.532, 0.476, 0.316, 0.176)$. 
Table 7.1: $\Delta_{SNR}$, $D^2$ and $d_{\text{min}}^2$ for the channels of Figure 7.1. The numerical values of $\Delta_{SNR}$ were determined for $1 - R_0^b(E_s) = 10^{-20}$.

<table>
<thead>
<tr>
<th>Channel</th>
<th>$\Delta_{SNR}$ (dB)</th>
<th>$D^2$</th>
<th>$d_{\text{min}}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>C3</td>
<td>1.76</td>
<td>2.667</td>
<td>2.667</td>
</tr>
<tr>
<td>C4</td>
<td>4.15</td>
<td>1.54</td>
<td>1.53</td>
</tr>
<tr>
<td>C5</td>
<td>8.60</td>
<td>0.55</td>
<td>0.53</td>
</tr>
</tbody>
</table>

Accuracy is limited by the precision with which $R_0^b(E_s)$ and $\Gamma$ can be computed. Note that

$$\lim_{E_s/N_0 \to 0} D^2 = 4$$

for all channels satisfying $\|h^{M+1}\|^2 = 1$ as follows from the fact that $R_0^b(E_s)$ becomes more and more independent of the channel memory, $M$, as $E_s/N_0$ decreases (Corollary 6.2).

In conclusion, this section has shown a further use of the (symmetric) cutoff rate parameter, $R_0^b(E_s)$. This parameter, which is of great significance when modulation and coding is considered [MAS(74)], can also be used to estimate the error rate performance of uncoded MLSE systems based on $d_{\text{min}}^2$ [FOR(72)]. There is then no need to determine the minimum weight error sequences which give rise to $d_{\text{min}}^2$. Alternatively, the proposed method provides initial estimates for $d_{\text{min}}^2$ that can help to reduce the search time required when using more refined procedures to determine $d_{\text{min}}^2$, e.g., as described in [AND-FOS(75)]. The new method has been applied to the case where the input symbols are binary; it should be possible to use it also in the case where multi-level input symbols are transmitted.
Fig. 7.2. $D^2$ as a function of $E_s/N_0$ for channels C1, C2, C3, C4 and C5 as specified in Figure 7.1. $D^2$ approaches $d^2_{\text{min}}$ (dashed or dotted lines) asymptotically as $E_s/N_0$ increases.
7.2 Approximating MDC's by Means of $\hat{R}_o^h(E_s)$

It was pointed out in Chapter 6 that the MDC's and the channels minimizing $R_o^h(E_s)$ (the so-called $\hat{R}_o^h(E_s)$-channels) are closely related. The MDC's have attracted wide attention because they establish worst-case limits (upper bounds) on the error-rate performance of uncoded binary systems that use maximum-likelihood sequence estimation (MLSE) [FOR(72)], [UNG(74)]. MDC's are constrained in the energy and the duration (memory) of their unit-sample response [MAG-PRO(73)], [AND-FOS(75)], [CAN-BUT(76-2)], [PRO(83), Section 6.7.1] (related references are [FRE(74)] and [FOS(75)]). The problem one has to solve to determine the MDC's is expressed by (6.73).

For $1 < M < 5$, the minimum of $d_{min}$, i.e., $d_{min}^2$, and the corresponding channel response, $h_{M+1}$, are known [CAN-BUT(76-2)]. For $M = 6$, a solution for $\hat{h}_{M+1}$ and $\hat{d}_{min}^2$ was reported in [AND-FOS(75)]. For longer channel memory, $M > 6$, the MDC's are not known, in part because of the formidable computational efforts which must be expended to solve (6.73) for large values of $M$ and $N$ (in practice, $N$ must be chosen sufficiently large to ensure that $d_{min}$ does not change as $N \rightarrow \infty$). The results obtained in Section 6.4.3 by studying channels that minimize $R_o^h(E_s)$ suggest that an approximation to the minimizing channel response, $\hat{h}_{M+1}$, may be obtained by determining $\hat{h}_{M+1}$ for large values of $E_s/N_0$ using the direct search algorithm described in Appendix 6.A. Thus, for $E_s/N_0 \gg 1$, $\hat{h}_{M+1} \approx \bar{h}_{M+1}$; the resulting value for the minimum distance may then be determined by some known method such as that described in [AND-FOS(75)], or it may be determined approximately using the method proposed in Section 7.1.

Figure 7.3 illustrates this approach for the case $M = 3$. The solution of (6.73) is known to be $\hat{d}_{min}^2 = 1.53$, and is produced by the channel unit-sample response $\hat{h}_4 = (0.372, 0.602, 0.602, 0.372)$ with the error sequence $e(D) = 1 - D$, or $\hat{h}_2 = (0.372, -0.602, 0.602, -0.372)$ with $e(D) = 1 + D$ ($\hat{h}_3 = -\hat{h}_4$).
Fig. 7.3. Sensitivity of $d_{\text{min}}^2$ and $R_o^b(E_s)$ for symmetric channel with memory $M = 3$, i.e., $h^4 = (h_0, h_1, h_2, h_3)$, where $h_0 = h_3$, and $h_1 = h_2 = \sqrt{0.5 - h_0^2}$:

(a) $d_{\text{min}}^2$ for the error sequence $e(D) = 1 - D$ as a function of $h_0 = h_3$;

(b) $R_o^b(E_s)$ for $E_s/N_o = 4$ dB as a function of $h_0 = h_3$;

(c) $R_o^b(E_s)$ for $E_s/N_o = 0$ dB as a function of $h_0 = h_3$. 
and \( \hat{h}_4 = -\hat{h}_2 \), are trivial additional solutions). In this example, the minima of \( d_{\text{min}}^2 \) and \( R_o^b(E_s) \) coincide at \( E_s/N_o = 4 \) dB, when plotted as a function of \( h_0 = h_3 \) (although in general, such common minima of \( d_{\text{min}}^2 \) and \( R_o^b(E_s) \) do not exist for finite \( E_s/N_o \)). Thus, for \( E_s/N_o = 4 \) dB, one obtains exact results, viz., \( \hat{h}^4 = \overline{h}^4 \) and \( \hat{d}_{\text{min}}^2 = \overline{d}_{\text{min}}^2 \). However, the common minima of \( d_{\text{min}}^2 \) and \( R_o^b(E_s) \), if they exist, are generally unknown, so that \( \overline{h}^{M+1} \) and \( \overline{d}_{\text{min}}^2 \), obtained for some value of \( E_s/N_o \), can only approximate \( \hat{h}^{M+1} \) and \( \hat{d}_{\text{min}}^2 \), respectively. For \( E_s/N_o = 0 \) dB, Figure 7.3 indicates that the approximation \( \hat{d}_{\text{min}}^2 \approx \overline{d}_{\text{min}}^2 = 1.55 \) is close to the actual value, \( \hat{d}_{\text{min}}^2 = 1.53 \).

The above method performs even better if special care is taken with respect to numerical accuracy when programming the procedure to compute \( R_o^b(E_s) \). In this case, one can make use of the fact that \( \overline{h}^{M+1} \rightarrow \hat{h}^{M+1} \) as \( E_s/N_o \rightarrow \infty \) to obtain very close approximations for \( \hat{h}^{M+1} \), and thus \( \hat{d}_{\text{min}}^2 \). Application of this idea to the example of Figure 7.3 yields for \( E_s/N_o = 19 \) dB, \( \hat{h}_1^4 \approx \overline{h}_1^4 = (0.370, 0.603, 0.603, 0.370) \) and \( \hat{d}_{\text{min}}^2 \approx \overline{d}_{\text{min}}^2 = 1.53 \). In this case, where \( \hat{h}_1^4 = (0.372, 0.602, 0.602, 0.372) \) and \( \hat{d}_{\text{min}}^2 = 1.53 \) [CAN-BUT(76-2)], the approximations, \( \overline{h}_1^4 \) and \( \overline{d}_{\text{min}}^2 \), respectively, are close to exact.

Finally, the case \( M = 6 \) is considered to verify the method. With \( E_s/N_o = 20 \) dB, one obtains \( \hat{h}^7 = (0.178, 0.324, 0.471, 0.531, 0.471, 0.324, 0.178) \), and \( \hat{d}_{\text{min}}^2 = 0.528 \) for the error sequence \( \varepsilon(D) = 1 - D - D^2 + D^3 + D^4 - D^5 \). These results are in good agreement with the previously published values for the MDC with \( M = 6 \), \( \hat{h}^7 = (0.176, 0.316, 0.476, 0.532, 0.476, 0.316, 0.176) \), and \( \hat{d}_{\text{min}}^2 = 0.524 \ (= 4 \times 0.131) \), respectively [AND-FOS(75)]. They demonstrate the practical merits of the proposed method which should be very useful when \( \hat{h}^{M+1} \) and \( \hat{d}_{\text{min}}^2 \) are of interest for MDC’s with long ISI memory (\( M > 6 \)).

---

1 The actual value to three significant digits is \( \hat{d}_{\text{min}}^2 = 0.522 \).
7.3 $R_o^G(E_s)$ as a Lower Bound on $R_o^b(E_s)$

In this section, the symmetric cutoff rate for independent, equally likely binary inputs, $R_o^b(E_s)$, and the cutoff rate for i.i.d. Gaussian inputs, $R_o^G(E_s)$, are related to each other.

A. Some Basic Observations

The simple upper bound on $R_o^b(E_s)$ in (6.28) coincides with the one obtained for $R_o^G(E_s)$ in Chapter 5. In view of Corollary 4.3 (Chapter 4), which asserts that $I^b(E_s) \leq I^G(E_s)$ for all values of $E_s/N_o$ and in particular as $E_s/N_o \to 0$, one might expect that the similar inequality $R_o^b(E_s) \leq R_o^G(E_s)$ should hold. However, it shall be demonstrated that this latter relation does not hold true for values of $E_s/N_o$ which are smaller than some finite, positive number, $\delta$. In fact, it will be shown that $R_o^G(E_s)$ provides a lower bound on $R_o^b(E_s)$ whenever $0 \leq E_s/N_o \leq \delta$. $R_o^b(E_s)$ and $R_o^G(E_s)$ are both convex functions over $0 \leq E_s/N_o < \infty$ and both are identical zero at $E_s/N_o = 0$. Since two convex functions can intersect in at most two points, there exists possibly a second value for $E_s/N_o$, different from zero, where $R_o^b(E_s) = R_o^G(E_s)$.

Consider the case where the channel has no ISI (i.e., $M = 0$, $h_0 = 1$) so that $R_o^b(E_s)$ is given as in (6.22b), and $R_o^G(E_s) = \log_2(1 + E_s/N_o)$ bits/symbol. Let $x = E_s/N_o$ and define the difference function

$$g(x) = R_o^b(x) - R_o^G(x)$$

$$= - \log_2 \left[ \frac{(1 + e^{-x})(1 + x)^{1/2}}{2} \right], \quad x \geq 0. \quad \text{(7.9)}$$

It follows, as indicated in Figure 7.4, that $g(0) = g(\delta) = 0$ and $g(x) > 0$, $0 < x < \delta$, where $\delta = 2.311$. Furthermore, $g(x)$ has a unique maximum at $x = 1.257$. In other words, for the memoryless channel, $R_o^b(E_s) \geq R_o^G(E_s)$ whenever $0 \leq E_s/N_o \leq 2.311$, and $R_o^b(E_s)$ may exceed $R_o^G(E_s)$ by up to 0.0516
Fig. 7.4. The difference function, $g(x)$, defined in Equation (7.9).
bits/symbol, or by almost 9%. The point \( g(x = \Delta = 3) = -0.0701 \), as marked in Figure 7.4, corresponds to \( E_s / N_o = 3 \), where \( R_o^G(E_s) \) has reached unity, the asymptotic upper limit of \( R_o^b(E_s) \) in bits/symbol. Since \( \delta \leq \Delta \), \( R_o^G(E_s) \) serves as a lower bound on \( R_o^b(E_s) \) over the entire range of rates that is of interest for coded binary signaling; Figure 7.5 emphasizes this observation.

An intuitive reason for \( R_o^b(E_s) > R_o^G(E_s) \) to be possible (for \( E_s / N_o \) smaller than some finite, positive number) may be seen in the fact that neither the Gaussian nor the binary input probability measure achieves the true (optimized) cutoff rate for the class of inputs satisfying an average-energy constraint. ¹ This is in contrast to the result of Corollary 4.3 when applied to the memoryless channel; in this case, the Gaussian probability density achieves capacity, i.e., for \( M = 0 \), \( I^G(E_s) = C(E_s) \).

**Conjecture 7.1:** For every discrete-time Gaussian channel with unit-sample response of finite energy, \( h^{M+1} \), \( \|h^{M+1}\|^2 < \infty \), there exists a positive number, \( \delta \), such that

\[
R_o^b(E_s) \geq R_o^G(E_s), \quad 0 \leq E_s / N_o \leq \delta. 
\] (7.10)

Furthermore, \( \delta \leq \Delta \), where \( \Delta \) is the value of \( E_s / N_o \) where \( R_o^G(E_s) = 1 \), the asymptotic upper limit of \( R_o^b(E_s) \) in bits/symbol. Equality holds in (7.10) for \( E_s / N_o = 0 \) and \( E_s / N_o = \delta \), and the lower bound provided by \( R_o^G(E_s) \) becomes asymptotically tight as \( E_s / N_o \to 0 \).

**B. Numerical Results**

In the following, numerical evidence is given to support Conjecture 7.1 which states that

¹ The optimizing input probability measure is not known (see also Chapter 5).
Fig. 7.5. \( R^b_o(E) \) and \( R^G_o(E) \) for the memoryless channel \((M = 0, h_0 = 1)\) as a function of \( E_s/N_0 \).
\[ R_o^G(E_s) = \frac{1}{2\pi} \int_0^\pi \log_2 \left[ 1 + \left( \frac{E_s}{N_0} \right) |H(\lambda)|^2 \right] d\lambda , \]  
(7.11a)

where

\[ H(\lambda) = \sum_{m=0}^{M} h_m e^{-jm\lambda}, \quad j = \sqrt{-1}, \]  
(7.11b)

provides a lower bound on \( R_o^b(E_s) \) as given in (6.20) of Theorem 6.1. More specifically, \( R_o^G(E_s) \leq R_o^b(E_s) \) when the signal-to-noise ratio, \( E_s/N_0 \), is in the range \( 0 \leq E_s/N_0 \leq \delta \leq \Delta \), with \( \delta \) such that \( R_o^G(\delta N_0) = R_o^b(\delta N_0) \), and \( \Delta \) such that \( R_o^G(\Delta N_0) = R_o^b(E_s \rightarrow \infty) = 1 \) bit/symbol.

Figure 7.6 shows \( R_o^G(E_s) \) and \( R_o^b(E_s) \) for the memoryless channel \( (M = 0) \) and the class of minimum-distance channels (MDC's) with \( M = 1, 2, \ldots, 6 \) [CAN-BUT(76-1)], [AND-FOS(75)] (see also Chapter 6). Clearly, \( R_o^G(E_s) \) provides a rather tight lower bound on \( R_o^b(E_s) \) for all \( M \), and in particular for large \( M \), say \( M \geq 4 \). This implies that good approximations (tight lower bounds) may be obtained for \( R_o^b(E_s) \) in the signal-to-noise ratio range \( 0 \leq E_s/N_0 \leq \delta \), by evaluating the simple expression for \( R_o^G(E_s) \) in (7.11), rather than using the more complicated procedure required to compute \( R_o^b(E_s) \).

In all cases in Figure 7.6, the error in the noise margin is at most 0.5 dB \( (M = 0) \); as \( M \) increases, it is considerably less. Table 7.2 lists the channel responses used to generate the curves in Figure 7.6 and gives the corresponding values for \( \delta \) and \( \Delta \) as defined above. Generally, \( \delta \) is unknown for arbitrary channels; however, Table 7.2 indicates that \( \delta \) can be closely approximated by solving \( R_o^G(\delta N_0) = 0.9 \) bits/symbol. This confirms, at least for the MDC's, that \( R_o^G(E_s) \) provides a (tight) lower bound on \( R_o^b(E_s) \) over the entire range of rates which is of practical interest for coded binary signaling.

The question, whether Conjecture 7.1 generalizes to systems where multi-level symbols are transmitted cannot be answered without further investigation.
Fig. 7.6. $R^G_o(E_s)$ (dashed lines) providing a lower bound for $R^b_o(E_s)$ (solid lines):
(a) Memoryless channel ($M = 0$) and minimum-distance channels (MDC's) with memory $M = 2, 4, 6$;
(b) MDC's with memory $M = 1, 3, 5$ (rate units are bits/symbol).
Table 7.2: $R_o^G(E_s)$ providing a lower bound for $R_o^b(E_s)$ for a signal-to-noise ratio in the range $0 \leq E_s/N_o \leq \delta < \Delta$, where $\delta$ and $\Delta$ are, respectively, such that $R_o^G(\delta N_o) = R_o^b(\delta N_o)$, and $R_o^G(\Delta N_o) = R_o^b(E_s \to \infty) = 1$ bit/symbol (see also Figure 7.6).

<table>
<thead>
<tr>
<th>$M$</th>
<th>$h_i$, $i = 0, 1, \ldots, M$</th>
<th>$R_o^G(\delta N_o) = R_o^b(\delta N_o)$ (bits/symbol)</th>
<th>$\delta$ (dB)</th>
<th>$\Delta$ (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>0.8635</td>
<td>3.637</td>
<td>4.771</td>
</tr>
<tr>
<td>1</td>
<td>0.707, 0.707</td>
<td>0.9076</td>
<td>5.168</td>
<td>6.021</td>
</tr>
<tr>
<td>2</td>
<td>0.500, 0.707, 0.500</td>
<td>0.9083</td>
<td>6.645</td>
<td>7.606</td>
</tr>
<tr>
<td>3</td>
<td>0.372, 0.602, 0.602, 0.372</td>
<td>0.9004</td>
<td>8.010</td>
<td>9.173</td>
</tr>
<tr>
<td>4</td>
<td>0.289, 0.500, 0.577, 0.500, 0.289</td>
<td>0.8958</td>
<td>9.343</td>
<td>10.668</td>
</tr>
<tr>
<td>5</td>
<td>0.232, 0.418, 0.521, 0.418, 0.232</td>
<td>0.8946</td>
<td>10.648</td>
<td>12.078</td>
</tr>
<tr>
<td>6</td>
<td>0.176, 0.316, 0.476, 0.532, 0.476, 0.316, 0.176</td>
<td>0.9008</td>
<td>11.995</td>
<td>13.472</td>
</tr>
</tbody>
</table>

1) Coefficients are given to three significant digits (see [CAN-BUT(76-1)] for exact expressions for $M = 1, 2, \ldots, 5$)
However, in the case of the memoryless channel with *equispaced* input symbols, where no symbols can be greater than $\sqrt{E_s}$, such a generalization does *not* hold. This follows directly from the results obtained by Wozencraft and Jacobs [WOZ-JAC(65), Figures 5.17 and 5.18] for the cutoff rate of the memoryless channel with equispaced input symbols; except for binary input symbols, $R^G_\rho(E_s) = \log_2 \sqrt{1 + E_s / N_\rho}$ (bits/symbol) always exceeds the cutoff rate for the multi-level systems, independent of whether or not the symbol probabilities are optimized.
CHAPTER 8

Summary and Conclusions

The present study has been inspired by a recent interest among digital communication and coding specialists in the search for and the application of effective codes for noisy digital communication channels which exhibit ISI memory. The exact origin of the ISI memory which characterizes the overall transmission system is not important in this context, what matters for the coding system is the equivalent discrete-time channel which is created by the combination of the modulator-transmitter filter, the physical (waveform) channel with the additive noise source, and the receiver filter. Assuming that the channel is linear and time-invariant, such a channel model is given in the form of a discrete-time transversal filter with a finite number of constant coefficients and with i.i.d. Gaussian noise samples added to the filter output. This basic channel model, which specifies the conditional probability for the output given some specific input, has been used for modelling channels in both data transmission and data storage applications.

The discrete-time channel model that is relevant to the coding system of a digital communication scheme which exhibits ISI memory has been analyzed from an information-theoretical viewpoint. The main objective was to determine how the two parameters, channel capacity and cutoff rate, or lower bounds thereon, are influenced by the combined effects of the ISI memory and the signal-to-noise ratio of the channel. A variety of rate parameters have been defined and determined for different channel models where each model consists of 1) the basic channel model and 2) the constraints on channel usage (input constraints). Two types of inputs were considered, namely, the class of input symbols which are constrained in their average energy (per symbol or per block) and the class of independent, antipodal, binary symbols of equal probability.
further objective was the development of computationally efficient algorithms for the numerical evaluation of the analytical results. The numerical results were generated under the commonly-used assumption that the finite unit-sample responses of the transversal filters which model the ISI part of the channels are normalized to have unit energy. Furthermore, all information-rate quantities have been derived and computed for channels with amplitude-continuous outputs, i.e., the results in this thesis represent upper bounds for the corresponding channels with finite output quantization. For ease of reference, the principal parameters which have been considered in this thesis are listed in Table 8.1. 

A. Summary of Main Results

A new approach was introduced in Chapter 3 to obtain the capacity formula of the DTGC. The capacity of the DTGC, $C(E_s)$, was derived from $\tilde{C}_N(E_s)$, the capacity of the newly defined $N$-circular Gaussian channel (NCGC). $\tilde{C}_N(E_s)$ was readily found by using the discrete Fourier transform (DFT). This simple DFT approach allowed us further to prove that the asymptotic capacity of the more commonly used $N$-block DTGC, $\hat{C}(E_s)$, is indeed the capacity of the DTGC. It was shown (Corollary 3.1) that the asymptotic capacity of the NCFC, $\tilde{C}(E_s)$, is identical to both $\hat{C}(E_s)$ and $C(E_s)$. The solution for the information rate of the DTGC-G, where the inputs are chosen to be i.i.d. Gaussian random variables with mean zero and variance $E_s$, $I^G(E_s)$, was derived by the DFT method. Apart from being a lower bound on $C(E_s)$, $I^G(E_s)$ establishes an upper bound on the information rate that can be achieved on the basic DTGC with independent average-energy constrained inputs. It was also shown (Corollary 3.6) that $I^G(E_s)$ provides a tight lower asymptote for $C(E_s)$ in the limit of very large signal-to-noise ratio $(E_s/N_0 \to \infty)$.

A better intuitive and quantitative understanding of both the DTGC and the $N$-block DTGC has been gained by first studying the NCFC in the discrete-
Table 8.1: The principal parameters considered in this thesis.

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Defined by Eq.</th>
<th>Given by Eq.</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C(F_S) )</td>
<td>(3.6)</td>
<td>(3.19)</td>
<td>Capacity of the DTGC</td>
</tr>
<tr>
<td>( \tilde{C}_N(F_S) )</td>
<td>(3.14)</td>
<td>(3.16)</td>
<td>Capacity of the NCGC</td>
</tr>
<tr>
<td>( \hat{C}(E_S) )</td>
<td>(3.9)</td>
<td>(3.18)</td>
<td>Asymptotic capacity of the N-block DTGC</td>
</tr>
<tr>
<td>( \tilde{C}(E_S) )</td>
<td>(3.15)</td>
<td>(3.18)</td>
<td>Asymptotic capacity of the NCGC</td>
</tr>
<tr>
<td>( I^G(E_S) )</td>
<td>(3.36), (3.39)</td>
<td>(3.47)</td>
<td>Information rate of the DTGC-G (i.i.d. Gaussian input symbols)</td>
</tr>
<tr>
<td>( I^b(E_S) )</td>
<td>(4.4)</td>
<td>*</td>
<td>Symmetric information rate of the basic DTGC with i.i.d., equally-likely binary input symbols</td>
</tr>
<tr>
<td>( C^b(E_S) )</td>
<td>(4.13)</td>
<td>*</td>
<td>Capacity of the basic DTGC with binary input symbols</td>
</tr>
<tr>
<td>( I_N^b(E_S) )</td>
<td>(4.4b), (4.5)</td>
<td>(4.10)</td>
<td>( N )-block average mutual information of the basic DTGC with i.i.d., equally-likely binary input symbols</td>
</tr>
<tr>
<td>( \hat{I}_N^b(E_S) )</td>
<td>(4.27a)</td>
<td>(4.32)</td>
<td>Upper bound on ( I_N^b(E_S) )</td>
</tr>
<tr>
<td>( \overline{I}_N^b(E_S) )</td>
<td>(4.52)</td>
<td>*</td>
<td>Practical approximation of ( I^b(E_S) )</td>
</tr>
<tr>
<td>( R_o(E_S) )</td>
<td>(5.1)</td>
<td>*</td>
<td>Cutoff rate of the DTGC</td>
</tr>
<tr>
<td>( R_o^G(E_S) )</td>
<td>(5.29)</td>
<td>(5.26)</td>
<td>Lower bound on ( R_o(E_S) )</td>
</tr>
<tr>
<td>( R_o^C(E_S) )</td>
<td>(5.31)</td>
<td>(5.32)</td>
<td>Cutoff rate of the DTGC-G (i.i.d. Gaussian input symbols)</td>
</tr>
<tr>
<td>( R_o^{*}(E_S) )</td>
<td>(5.39a)</td>
<td>(5.39)</td>
<td>Cutoff rate of the DTGC*</td>
</tr>
<tr>
<td>( R_o^C(E_S) )</td>
<td>(5.52)</td>
<td>(5.53)</td>
<td>Cutoff rate of the basic DTGC with input symbols that achieve ( C(E_S) )</td>
</tr>
<tr>
<td>( R_o^b(E_S) )</td>
<td>(6.1)</td>
<td>(6.20)</td>
<td>Symmetric cutoff rate of the basic DTGC with i.i.d., equally-likely binary input symbols</td>
</tr>
<tr>
<td>( \overline{R}_o^b(E_S) )</td>
<td>(6.57)</td>
<td>*</td>
<td>Minimum value of ( R_o^b(E_S) ) over all unit-sample responses of unit energy and with some fixed, maximum ISI memory length (( M ))</td>
</tr>
</tbody>
</table>
frequency transform domain. A link between the three channel models is given through the results of Theorem 3.3 and Lemma 3.4. The DFT method is not directly applicable to the case where the ISI memory length of the channel, $M$, is unbounded ($M = \infty$). However, it was shown in Section 3.6 that the capacity result (Theorem 3.2) is also valid for channels with infinite memory as long as the unit-sample response of the transversal filter is absolutely summable. On the other hand, it should be possible to extend the DFT method to the more general case of the discrete-time Gaussian vector channel with finite ISI memory [TSY(65)], [TOM-BER(73)], [BRA-WYN(74)].

Numerical results have shown that $\tilde{C}_N(E_d)$ approximates $C(E_d)$ as closely as desired if the block-length, $N$, is chosen large enough (Fig. 3.2). It was also found (Corollary 3.6) and illustrated (Fig. 3.3) that $C(E_d)$ is not strictly upper bounded by the capacity of the memoryless channel ($M = 0$) when the signal-to-noise ratio is small. This result is counterintuitive and is an artifact of the chosen channel response normalization, i.e., all channels have the same unit-sample response energy.

In Chapter 4 the symmetric information rate of the basic DTGC with i.i.d., equally-likely binary input symbols, $^I(E_d)$, was dealt with. $^I(E_d)$ serves as a lower bound on the asymptotic capacity of the basic DTGC with binary input symbols, $C(E_d)$. An expression for the $N$-block average mutual information, $^I_N(E_d)$, was derived from which the known capacity of the corresponding memoryless channel was obtained as a special case (Corollary 4.1). $^I_N(E_d)$ was shown to depend on all possible $N$-dimensional Euclidean distance vectors which occur between the noiseless channel output signal sequences.

It was argued that in general $^I_N(E_d)$, and thus also $^I(E_d)$, are not effectively computable. Instead, an efficient Monte-Carlo integration technique was formulated (see also Appendix 4.C) and used to obtain statistical estimates of $^I_N(E_d)$ and of an upper bound thereon, $\hat{^I}_N(E_d)$. It was then proposed to consider the unweighted average of these estimates, $\overline{^I}_N(E_d)$, as a "practical"
approximation to $I^h(E_3)$, provided that $N >> M$; the experimental results (Fig. 4.6) support the usefulness of this approximation. However, for channels with long ISI memory ($M > 6$), the required computer run-time may still be excessive in practice. The numerical results indicate that, in the range where $I^h(E_3) \leq 1/2 \text{ bit per symbol}$, $I^G(E_3)$ should provide a tight upper bound on $I^h(E_3)$ for the channels used in Figure 4.6. Corollary 4.3 suggests that this observation may be generalized to include all channels.

In Chapter 5 the cutoff rate of the DTGC, $R_o(E_3)$, has been considered. Theorem 5.1 asserts that $R_o(E_3)$ may be obtained, at least in principle, from the cutoff rate of the NCGC in the limit of increasing block-length, $N$. Unlike in the case of capacity, Gaussian input symbols do not achieve $R_o(E_3)$. Starting with the basic channel model of the NCGC, a lower bound on $R_o(E_3)$, denoted by $R_o^{G^*}(E_3)$, was defined and then derived (Corollary 5.3) by adopting the DFT method introduced in Chapter 3. It was shown that $R_o^{G^*}(E_3) = C(E_3/2)$, $0 \leq E_3/N_0 < \infty$, i.e., the curve for $R_o^{G^*}(E_3)$ plotted as a function of $E_3/N_0$ is simply the curve for $C(E_3)$ shifted by 3 dB to the right; otherwise, $R_o^{G^*}(E_3)$ and $C(E_3)$ exhibit the same characteristics. The cutoff rate of the DTGC-G, where the inputs are chosen to be i.i.d. Gaussian random variables with mean zero and variance $E_3$, denoted by $R_o^G(E_3)$, was obtained next as a lower bound on $R_o^{G^*}(E_3)$. Not unexpectedly, it was found that $R_o^G(E_3) = I^G(E_3/2)$, $0 \leq E_3/N_0 < \infty$.

The zero-rate intercept of the straight-line portion of Gallager's random-coding exponent was defined as the cutoff rate of the DTGC*, $R_o^*(E_3)$, where the * indicates that the basic DTGC is used such that each input sequence is constrained in allowed average energy per symbol. The previously known result for $R_o^*(E_3)$ of the memoryless channel was obtained as a special case of the result for the general DTGC*. It was also shown that $C(E_3)/2 \leq R_o^{G^*}(E_3) \leq R_o^*(E_3) \leq C(E_3)$, $0 \leq E_3/N_0 < \infty$, (Corollary 5.6).
Numerical results were presented which illustrate the analytical results in terms of their spectral and asymptotic properties. Specifically, it was found that the cutoff rate of the basic DTGC with inputs that achieve $C(E_s)$, denoted by $R^C_o(E_s)$, is not significantly smaller than $R^G_o(E_s)$. Numerical results were presented which support the conjecture that, at any signal-to-noise ratio, $R_o(E_s)$ is smaller than $C(E_s)$ by at most $1/2$ bit per symbol.

Chapter 6 provides a detailed study on the symmetric cutoff rate of the basic DTGC with i.i.d., equally-likely binary input symbols, $R^b_o(E_s)$. A computationally efficient procedure was developed to determine $R^b_o(E_s)$ as the negative logarithm of the dominant eigenvalue of a certain (reduced) state transition matrix whose size depends on the length of the channel's ISI memory. Examples were given to illustrate the major steps in the development of the procedure, such as the formalized matrix reduction method (Theorem 6.2).

$R^b_o(E_s)$ was evaluated for partial-response channels (PRC's) and for minimum-distance channels (MDC's). The ranking of the PRC's and the MDC's with respect to decreasing values of $R^b_o(E_s)$ was interpreted in terms of the corresponding Euclidean distances that are induced by the five most dominant error sequences (Table 6.4). Among channels which exhibit the same minimum squared Euclidean distance, $d_{\min}^2$, it was shown that the multiplicity of the error sequences which produce $d_{\min}^2$ and the magnitudes of the next larger distances determine an individual channel's rank.

The numerical procedure by which $R^b_o(E_s)$ is computed was enhanced with an efficient direct search algorithm. This combined procedure was then used to determine those unit-sample responses which minimize $R^b_o(E_s)$, given that both the signal-to-noise ratio, $E_s/N_o$, and the maximum ISI memory length, $M$, are fixed. The symmetric cutoff rate of these minimum-$R^b_o(E_s)$ channels, denoted by $\bar{R}^b_o(E_s)$, was shown to differ very little from $R^b_o(E_s)$ of the MDC with the same ISI memory. For $M > 1$, it was found, however, that the unit-sample responses of the minimum-$R^b_o(E_s)$ channels vary with the signal-to-noise ratio.
and that they exhibit the same symmetry properties as the MDC's (Conjecture 6.1). For $M = 1$ the system polynomial that minimizes $R_{0}^{b}(E_{s})$ was found analytically to be of the form $F(D) = 1 \pm D$; thus, any channel with a system polynomial $F(D) = 1 \pm D^{k}$, $k = 1, 2, \ldots$, is both a MDC and a minimum-$R_{0}^{b}(E_{s})$ channel, independent of the signal-to-noise ratio.

Finally, worst-case limits were established for the $R_{0}^{b}(E_{s})$ of any unit-energy channel with ISI memory in the range $1 \leq M \leq 8$, and with a signal-to-noise ratio in the range $-15 \text{ dB} \leq (E_{s}/N_{o})_{\text{dB}} \leq 20 \text{ dB}$ (Fig. 6.15 and Fig. 6.16). As a rule, it was found that the worst-case loss in noise margin with respect to the memoryless channel amounts to about $M$ dB for any channel with ISI memory $M \leq 8$ and with $E_{s}/N_{o}$ such that $R_{0}^{b}(E_{s}) \approx 0.6$ bits per symbol.

In Chapter 7 some new applications and properties of $R_{0}^{b}(E_{s})$ were presented that have not been noted previously. The ability to compute $R_{0}^{b}(E_{s})$ efficiently and with a high degree of accuracy as it approaches 1 bit per symbol was exploited and combined with the availability of an efficient direct-search algorithm to identify minimum-$R_{0}^{b}(E_{s})$ channels (see also Chapter 6).

Using random-coding arguments, a simple expression in terms of $R_{0}^{b}(E_{s})$ and $E_{s}/N_{o}$ was derived that yields a computationally efficient approximation to the minimum squared Euclidean distance, $d_{\text{min}}^{2}$, the parameter which determines performance bounds for maximum-likelihood sequence estimation (MLSE) of uncoded (binary) sequences which are transmitted over Gaussian channels with ISI. This novel approach does not require identification of a specific error-event which produces $d_{\text{min}}^{2}$. Examples were given to support the practical merits of this technique (Table 7.1). It should be possible to extend it to other types of modulation schemes (input constraints) for which the symmetric cutoff rate can be defined and computed (e.g., multi-amplitude signaling).

A method was described which can be used to approximate the unit-sample responses of MDC's by means of computing $R_{0}^{b}(E_{s})$. The method makes use of the fact that, in the limit as $E_{s}/N_{o} \to \infty$, the unit-sample responses of the
newly defined minimum-$R^b_o(E_s)$ channels converge towards the unit-sample responses of MDC's with the same ISI memory. There is again no need to know any specific error-event which produces the smallest minimum distance.

In view of the inequality $I^b(E_s) \leq I^G(E_s)$, $0 \leq E_s/N_o < \infty$, obtained in Chapter 4 (Corollary 4.3), one might intuitively expect that the similar inequality $R^b_o(E_s) \leq R^G_o(E_s)$ should hold. It was found, however, that this latter relation does not hold for values of $E_s/N_o$ which are smaller than some positive number, $\delta$. In fact, for a number of channels, it has been shown (Fig. 7.6 and Table 7.2) that $R^G_o(E_s)$ provides a fairly tight lower bound on $R^b_o(E_s)$ whenever $0 \leq E_s/N_o < \delta$. In terms of $E_s/N_o$ within the given range, $R^b_o(E_s)$ and $R^G_o(E_s)$ are typically within 0.5 dB. This result is of practical interest since $R^G_o(E_s)$ is much simpler to compute than $R^b_o(E_s)$. It holds when $R^G_o(E_s) \leq 0.9$ bits per symbol. This means that $R^G_o(E_s)$ provides a tight lower bound on $R^b_o(E_s)$ for all rates that are of practical interest for coded binary signaling. This relation between $R^G_o(E_s)$ and $R^b_o(E_s)$ does not appear to have been noticed previously. An intuitive understanding of this result may be obtained by considering the memoryless channel. In this case, it is known that $I^b(E_s) = C^b(E_s) \leq I^G(E_s) = C(E_s)$. On the other hand, it is not known what type of input distribution achieves the true cutoff rate of the memoryless channel with average-energy constrained inputs. Thus, there is actually no reason to presume that $R^b_o(E_s)$ should be strictly upper bounded by $R^G_o(E_s)$.

B. Conclusions

The basic DTGC with ISI memory and with various constraints on its usage has been analyzed from an information-theoretical viewpoint. Some new as well as previously known results have been derived and/or generalized by introducing alternative methods of analysis and numerical evaluation. A number of relations between the various parameters studied have been proved and some others conjectured on the basis of strong evidence; some of them do not seem to have
been noticed previously. For example, an interesting asymptotic relation was found between the squared Euclidean distance, $d_{\min}^2$, which is the important parameter for uncoded binary signaling over ISI channels, and the corresponding symmetric cutoff rate, $R_o^b(E_s)$, which is relevant for coded signaling over the same channel. Concrete numerical data and examples which illustrate the properties and characteristics of the studied parameters were generated; while such an approach is perhaps not characteristic in information-theoretical work, it has led to some interesting relations and interpretations which were not obvious from the analytical expressions.

The results of this work permit one to make some quantitative and qualitative statements about the effects of the channel's ISI memory length, $M$, on the various rate parameters (Table 8.1), as well as about the necessary (additional) complexity of coding systems for such channels. A consolidated view of the numerical results is presented by Figure 8.1. It shows the typical dependence on $M$ of the parameters $C(E_s)$, $I^G(E_s)$, $I^b(E_s)$, $R_o^s(E_s)$, $R_o^G(E_s)$, $R_o^G(E_s)$, and $R_o^b(E_s)$, at low and high values of the signal-to-noise ratio, $E_s/N_o$. The channel responses $C_1 ... C_4$ have been defined in Figure 3.1; $C_1$ represents the memoryless channel ($M = 0$), $C_2$ ($M = 1$) and $C_3$ ($M = 2$) represent MDC's and $C_4$ closely approximates a MDC for $M = 6$ (see also Figure 2.3 in Chapter 2 for the frequency-domain characteristics of these channels). For fixed values of $E_s/N_o$, all rate parameters decrease with increasing $M$ except at low values of $E_s/N_o$ where $C(E_s)$, $R_o^s(E_s)$ and $R_o^G(E_s)$ increase with $M$. The latter three parameters have in common that they are achieved with inputs which do not make use of the full channel bandwidth, while all other parameters assume flat (white) input power spectral densities. For this reason it is possible that $R_o^s(E_s) > I^G(E_s)$ and that $R_o^G(E_s) > I^G(E_s)$ for certain values of $E_s/N_o$. In Figure 8.1, for example, these inequalities hold for chan-

---

1 The capacity-achieving spectral distribution of the allotted input energy is determined by the water-filling principle [GAI.(68), Section 8.3] (see also Figure 5.4).
Fig. 8.1: Summary of the numerical results: (a) Dependence of $C(E_s)$, $I^G(E_s)$, and $I^h(E_s)$, on the ISI memory length of the channel, $M$, at low and high values of the signal-to-noise ratio, $E_s/N_o$. (b) Dependence of $R_o^*(E_s)$, $R_o^{G^*}(E_s)$, $R_o^{G}(E_s)$, and $R_o^{h}(E_s)$, on $M$ at low and high values of $E_s/N_o$. The channel responses C1 ... C4 are defined in Figure 3.1.
nels C3 and C4 when \((E_s/N_o)_{dB} = -8\) dB, and for channel C4 alone when \((E_s/N_o)_{dB} = 6\) dB.

In summary, the characteristics of \(C(E_s)\), \(R_o^*(E_s)\) and \(R_o^{G*}(E_s)\) indicate that the encoder in a coded digital communication system which makes use of channels with ISI memory should introduce redundancy in the modulating symbol sequence in such a way that the transmitted energy is concentrated at those frequencies where the channel exhibits the least attenuation. This principle becomes apparently more important as the signal-to-noise ratio decreases and/or the ISI memory length increases. A number of authors have proposed code design procedures which are based on this water-filling interpretation of the capacity result \([\text{RUI-CIO}(87)], [\text{GIL-THO}(76)]\). On the other hand, if a certain reliability is to be maintained, the results obtained for the worst-case symmetric cutoff rate, \(R_o^b(E_s)\), indicate that binary coded systems must provide up to about 1 dB additional coding gain when the ISI memory length is increased from \(M\) to \(M + 1\). This implies that the required complexity of the coding system is an increasing function of \(M\).

For example, consider maximum-likelihood demodulation-decoding of the ensemble of time-varying binary convolutional codes of rate \(R\) and constraint length \(N_t\) (measured in terms of encoded symbols). It has been shown by Viterbi and Omura \([\text{VIT-OMU}(79), \text{Theorem 5.8.1}]\) that the bit error probability is bounded as

\[
P_b \leq \frac{\gamma R e^{-N_t R_o^b(E_s)}}{\left[1 - \gamma R e^{-N_t R_o^b(E_s)}\right]^2} \approx \gamma R e^{-N_t R_o^b(E_s)}, \quad R < R_o^b(E_s), \quad (8.1)
\]

where \(\gamma R\) is a factor that depends on \(R\) but not on \(N_t\). Assuming that \(P_b\) is to be maintained when the ISI memory length of the channel increases from \(M = 0\) to \(M = 6\) while \(E_s/N_o\) is kept constant at 3 dB, it follows from Figure 6.15 and (8.1) that \(N_t\) must be about doubled since, in the worst case,
$R_0^b(E_s)$ could be halved. Alternatively, if $N_t$ is kept constant for both channels, then $E_s/N_o$ would have to be increased by about 7 dB for the channel with $M = 6$ in order to achieve the same $P_b$. In this comparison, we have tacitly ignored the fact that the required receiver complexity increases rapidly with $M$ and $N_t$. Note also from Figure 6.15 and (8.1) that, for some given $N_t$, the effect of the channel memory on $P_b$ vanishes as $E_s/N_o$ gets small. This observation is interesting since Corollary 6.2 indicates that $R_0^b(E_s)/E_s \lesssim 0.721/N_o$ whenever $E_s/N_o \ll 1$, i.e., the signal energy is used most efficiently when $E_s/N_o$ is small. The benefits of (low-rate) coding are thus optimal on the very noisy channel, independent of its ISI memory structure. In practice, however, a lower limit on $E_s/N_o$ is imposed since reliable symbol synchronization in the receiver requires roughly that $E_s/N_o \geq 1$ (\(0\) dB) [HAG(80)], [WIN-LUE(69)].

Finally, note that with equally-likely, antipodal binary inputs, the channel filter transforms each binary input digit into a cluster of signal points, one point (not necessarily unique) for each channel state. Thus, one may view the ensemble of different outputs of the channel filter as a set of signals centered about the origin with average energy $E_s$. Such signal sets were shown to be optimal with respect to energy efficiency for coded transmission on the AWGN channel at small values of $E_s/N_o$ because they yield the cutoff rate $R_0 \approx 0.721(E_s/N_o)$ bits per symbol [MAS(76)]. This result is in agreement with our present interpretation of the properties of $R_0^b(E_s)$ and $\overline{R}_0^b(E_s)$, respectively, at small values of $E_s/N_o$. 
REFERENCES


[OEL(86)] S. Ölzec, private communication and unpublished notes, March 1986.


CURRICULUM VITAE

Walter Hirt, born on December 3, 1947, in Windisch, AG, Switzerland.

1954 -1959 Primary School, Windisch, AG.

1959 -1963 Secondary School, Brugg, AG.

Diploma awarded in April 1967.

1968 -1971 Student, Höhere Technische Lehranstalt, Brugg-Windisch, AG.
El.-Ing. HTL diploma (Hasler Prize) awarded in October 1971.


1974 -1975 Teaching Assistant and part-time student, Dept. of Electrical
Engineering, University of Toronto, Toronto, Ontario, Canada.

1975 -1977 Undergraduate studies (advanced standing), Dept. of Electrical
Engineering, University of Toronto, Toronto, Ontario, Canada.
B.A.Sc. (Honours) degree awarded in June 1977.

1977 -1979 Graduate studies and part-time employment, Dept. of Electrical
Engineering, University of Toronto, Toronto, Ontario, Canada.
M.A.Sc. degree awarded in November 1979.

1979 -1980 Research Engineer, Wave Sciences Laboratory, Dept. of Electrical
Engineering, University of Toronto, Toronto, Ontario, Canada.

since 1980 Research Staff Member, IBM Research Division, Zurich Research
Laboratory, Rüschlikon, ZH.

1982 -1984 Candidate, Institute for Signal and Information Processing, Swiss
Federal Institute of Technology (ETH), Zurich.
Admission Examination passed in February 1984.

1984 -1988 Doctoral student, Institute for Signal and Information Processing,
Swiss Federal Institute of Technology (ETH), Zurich.