Diss. ETH No. 10314

Communication Theory and Coding for Channels with Intersymbol Interference

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZÜRICH

for the degree of
Doctor of Technical Sciences

presented by
FREDY D. NEESER
dipl. El.-Ing. ETH
born September 27, 1960
citizen of Schlossrued, Aargau

accepted on the recommendation of
Prof. Dr. J. L. Massey, referee
Prof. Dr. J. Hagenauer, co-referee

1993

To Andrea and Livia

Acknowledgments

Working as a research assistant at the Signal and Information Processing Laboratory of ETH Zürich has been a rewarding and enjoyable experience.

I wish to express my sincere gratitude to Professor James L. Massey for his advice, constructive critique and support given to me in the course of this research. I gratefully remember many stimulating discussions during which I learned to attack problems by trying a different point of view.

I am also indebted to Professor Joachim Hagenauer for acting as co-referee.

I would like to thank Professor G. S. Moschytz for giving me the opportunity to pursue an interesting industry project during my first two years at the Laboratory.

Jürg Ruprecht and Peter Kartaschoff of the Swiss PTT deserve special thanks for continuously supporting my work.

During my time at the Laboratory, I have appreciated very much the kind interest and assistance of Jürg Ganz, Richard Gut, Markus Hufschmid, Gerhard Krämer, Urs Loher, Andi Löliger, Thomas Mittelholzer, Marcel Rupf, and many other colleagues. I am particularly grateful to Gerhard Krämer for his review of, and useful comments on, Chapter 4. It is also a pleasure to remember the great time I spent with my friends of the running team, catching some fresh air for the next mental challenge, and the many restful days on Rigi Scheidegg with our friends Ruth and Marcel Ledergerber.

I owe very special thanks to my wife Andrea for the love, encouragement, and understanding I received from her during all the highs and lows I have experienced in this research and to my parents for their loving care and for supporting my studies.

Leer - Vide - Empty

Abstract

This thesis is primarily concerned with the analysis of discrete-time communication channels with intersymbol interference (ISI) and additive white Gaussian noise (AWGN).

Complex random variables and processes are useful for describing the baseband equivalent of bandpass communication channels with memory and - as shown in this thesis - for deriving the capacity of ISI channels with a complex unit-sample response and complex AWGN. The complex noise process in the equivalent baseband channel usually exhibits certain symmetries in the autocovariance and crosscovariance function of its real and imaginary part. We show that in order to achieve capacity the information-carrying process must obey the same covariance symmetries.

It is shown that the 'covariance' of complex random variables and processes, when defined consistently with the corresponding notion for real random variables, is specified by the conventional (complex) covariance and a quantity called the *pseudo-covariance.* A characterization of uncorrelatedness and wide-sense stationarity in terms of covariance and pseudo-covariance is given. The above covariance symmetries correspond to a vanishing pseudo-covariance. Complex random variables and processes with a zero pseudo-covariance are called *proper.* It is shown that properness is preserved under affine transformations and that the complex-multivariate Gaussian density assumes a natural form only for proper random variables. The maximum-entropy theorem is generalized to the complex-multivariate case; the differential entropy of a complex random vector with a fixed correlation matrix is shown to be maximum if and only if the random vector is proper, Gaussian and zero-mean. The notion of *circular stationarity* is introduced. For proper complex random variables, a discrete Fourier transform correspondence is derived that relates circular stationarity in the time domain to uncorrelatedness

in the frequency domain.

As an application of the theory, the capacity of an ISI channel with complex inputs, a finite complex unit-sample response, and proper complex AWGN is determined. This derivation is considerably simpler than an earlier derivation by Hirt and Massey for the *real* ISI channel with AWGN, whose capacity is obtained as a by-product of the results for the complex channel.

Motivated by the fact that the capacity of an ISI channel with AWGN does not depend on the phase response of the channel filter, we introduce a transfer-function equivalence class, whose members are equal up to an allpass factor. It is shown that the mutual information between a finite-length input block and the relevant channel outputs is the same for all filters in such an equivalence class, regardless of the input probability distribution. This result allows a simpler proof of a lower bound on the information rate for independent, identically distributed inputs due to Shamai, Ozarow, and Wyner.

The state transitions of a finite-state, time-invariant trellis encoder can be described by a directed graph with exactly K branches leaving every node, called a *K-ary state-transition diagram* (STD). The problem of finding all isomorphism classes of K-ary STD's with certain topological constraints is investigated. K-ary STD's are constructed recursively from so-called partial K-ary STD's. Necessary and sufficient conditions are derived for when a partial K-ary STD is extendable to a (complete) strongly connected K-ary STD. We consider K-ary STD's with *maximum detour memory*, i.e., with the property that the shortest detour in the associated trellis has maximum length. Such STD's are shown to be strongly connected and to have the same number of branches ending at every state. All non-isomorphic binary STD's with maximum detour memory and $N = 2^M$ states, $M \leq 4$, are determined. These binary STD's can be used, e.g., for the construction of matched spectral-null codes for partial-response channels.

The performance of trellis-coded data transmission over channels with finite ISI and AWGN is analyzed by investigating the trellis encoder that results from the cascade of the channel encoder (the outer encoder) with the subsequent channel filter (the inner encoder). Such a composite trellis encoder usually has a non-uniform distance spectrum. A well-known upper bound on bit error probability for convolutional encoders and maximum-likelihood decoding is generalized to trellis encoders with a non-uniform distance spectrum. The generalized upper bound involves an *average distance spectrum*, which can be obtained by using a modified

Viterbi algorithm. As an application, some bipolar trellis encoders for the dicode channel are compared. It is shown that, for trellis encoders with a non-uniform distance spectrum, the average number of bit errors over all detours at free distance can be much smaller than one and is therefore a more important parameter than in the case of a uniform distance spectrum.

Keywords. Intersymbol interference, proper complex random variables, capacity of ISI channel, information rate for i.i.d. inputs, equivalent ISI channels, strongly connected state-transition diagram, non-isomorphic state-transition diagrams, detour memory, trellis-coded ISI channel, steady-state encoder, average distance spectrum.

Leer - Vide - Empty

Übersicht

Die vorliegende Dissertation befasst sich vorwiegend mit der Analyse von zeitdiskreten Kommunikationskanälen mit Intersymbol-Interferenz (ISI) und additivem, weissem Gauss'schem Rauschen (AWGN).

Komplexe Zufallsgrössen und -prozesse eignen sich zur äquivalenten Basisband-Darstellung eines Bandpass-Kanals mit Gedächtnis und - wie in der Dissertation gezeigt wird - zur Herleitung der Kapazität von ISI-Kanälen mit einer komplexen Impulsantwort und komplexem AWGN. Real- und Imaginärteil des komplexen Rauschens in einem äquivalenten Basisband-Kanal besitzen im allgemeinen gewisse Symmetrien in der Autokovarianz- und Kreuzkovarianzfunktion. Wir zeigen, dass die Kapazität nur erreicht werden kann, wenn der informationstragende Zufallsprozess dieselben Kovarianz-Symmetrien aufweist.

Wird die 'Kovarianz' von komplexen Zufallsgrössen und -prozessen widerspruchsfrei zum entsprechenden Begriff für reelle Zufallsgrössen definiert, so ist sie durch die konventionelle (komplexe) Kovarianz und eine zweite Grösse bestimmt, die wir als *Pseudo-Kovarianz* bezeichen. Unkorreliertheit und schwache Stationarität werden mittels Kovarianz und Pseudo-Kovarianz ausgedrückt. Die obigen Kovarianz-Symmetrien entsprechen einer verschwindenden Pseudo-Kovarianz. Komplexe Zufallsgrössen und -prozesse mit solchen Symmetrien werden als *eigentlich* bezeichnet. Es wird gezeigt, dass eigentliche komplexe Zufallsgrössen durch affine Transformationen wiederum in eigentliche komplexe Zufallsgrössen übergehen und dass die komplexe, multivariate Gausssche Wahrscheinlichkeitsdichte nur für eigentliche Zufallsgrössen eine natürliche Form annimmt. Der Satz über die maximale Entropie wird verallgemeinert auf den komplexen, multivariaten Fall; die differentielle Entropie eines komplexen Zufallsvektors mit einer gegebenen Korrelationsmatrix ist maximal genau dann, wenn der Zufallsvektor eigentlich, Gaussisch und mittelwertfrei ist. Der Begriff der *zirkulären Stationa-*

rität wird eingeführt. Für eigentliche komplexe Zufallsgrössen und die diskrete Fourier-Transformation wird eine Korrespondenz zwischen zirkulärer Stationarität im Zeitbereich und Unkorreliertheit im Frequenzbereich hergeleitet.

Die Theorie wird angewendet zur Bestimmung der Kapazität eines ISI-Kanals mit komplexen Eingängen, einer komplexen Impulsantwort und eigentlichem AWGN. Dabei ergibt sich eine wesentliche Vereinfachung gegenüber einer früheren Herleitung von Hirt und Massey für den *reellen* ISI-Kanal, dessen Kapazität als Nebenprodukt der Resultate für den komplexen Kanal erhalten wird.

Die Tatsache, dass die Kapazität eines ISI-Kanals mit AWGN nicht vom Phasengang des Kanalfilters abhängt, legt die Einführung einer Äquivalenzklasse von Übertragungsfunktionen nahe, deren Mitglieder sich nur durch einen Allpass-Faktor unterscheiden. Es wird nachgewiesen, dass die gegenseitige Information zwischen einem Eingangsblock endlicher Länge und den relevanten Kanalausgängen für alle Filter in einer solchen Äquivalenzklasse gleich ist, und zwar unabhängig von der Eingangs-Wahrscheinlichkeitsverteilung. Dieses Resultat erlaubt einen einfacheren Beweis einer unteren Schranke von Shamai, Ozarov und Wyner für die Informationsrate bei statistisch unabhängigen, gleichverteilten Eingängen.

Die Zustandsübergänge eines zeitinvarianten Trellis-Encoders können durch ein K-wertiges Zustandsdiagramm (STD[1]) dargestellt werden, d.h. durch einen gerichteten Graphen mit genau K von jedem Knoten ausgehenden Zweigen. Wir untersuchen das Problem der Bestimmung aller Isomorphie-Klassen von K-wertigen STD's, die bestimmten topologischen Anforderungen genügen. K-wertige STD's werden durch Erweiterung gewisser unvollständiger K-wertiger STD's in rekursiver Weise konstruiert. Die notwendigen und hinreichenden Bedingungen werden hergeleitet, unter denen ein solches Teil-STD zu einem vollständigen, stark zusammenhängenden K-wertigen STD erweitert werden kann. Wir untersuchen K-wertige STD's mit *maximalem Umweg-Gedächtnis*, d.h. mit der Eigenschaft, dass der kürzeste Umweg im zugehörigen Trellis die grösstmögliche Länge hat. Es wird gezeigt, dass solche STD's stark zusammenhängend sind und dass in jedem Zustand gleich viele Zweige enden. Alle nicht-isomorphen binären STD's mit maximalem Umweg-Gedächtnis und $N = 2^M$ Zuständen, $M \leq 4$, werden bestimmt. Diese binären STD's lassen sich zur Konstruktion

[1] Von englisch 'state-transition diagram'.

von Trellis-Codes mit kanalangepassten spektralen Nullen für Partial-Response-Kanäle verwenden.

Zur Untersuchung trellis-codierter Datenübertragung auf Kanälen mit endlicher ISI und AWGN untersuchen wir denjenigen Trellis-Encoder, der aus der Kaskade des Kanal-Encoders (äusserer Encoder) mit dem Kanalfilter (innerer Encoder) entsteht. Im allgemeinen hat ein derartiger Encoder ein ungleichförmiges Distanzspektrum. Eine bekannte obere Schranke für die Bitfehler-Wahrscheinlichkeit bei Verwendung eines Faltungs-Encoders und Maximum-Likelihood Decodierung wird verallgemeinert auf Trellis-Encoder mit ungleichförmigem Distanzspektrum. In der verallgemeinerten oberen Schranke tritt ein *gemitteltes Distanzspektrum* auf, welches mit Hilfe eines modifizierten Viterbi-Algorithmus bestimmt werden kann. Als Anwendung werden einige bipolare Trellis-Encoder für den Dicode-Kanal miteinander verglichen. Anhand von Beispielen wird gezeigt, dass bei ungleichförmigem Distanzspektrum die mittlere Anzahl der Bitfehler über alle Umwege in freier Distanz deutlich kleiner als eins sein kann, so dass diesem Parameter die grössere Bedeutung zukommt als im Falle eines gleichförmigen Distanzspektrums.

Stichwörter. Intersymbol-Interferenz, eigentliche komplexe Zufallsgrössen, Kapazität des ISI-Kanals, Informationsrate bei statistisch unabhängigen und gleichverteilten Eingängen, äquivalente ISI-Kanäle, stark zusammenhängendes Zustandsdiagramm, nicht-isomorphe Zustandsdiagramme, Umweg-Gedächtnis, Trellis-codierter ISI-Kanal, Steady-State Encoder, gemitteltes Distanzspektrum.

Leer - Vide - Empty

Contents

Chapter 1

Introduction

This dissertation is primarily concerned with the phenomenon of intersymbol interference (ISI) on communication channels with additive white Gaussian noise (AWGN). In the discrete-time channel created by the modulator, the waveform channel, and the demodulator, ISI results from processing the sequence of modulation symbols by a linear filter.

Perhaps one of the most intriguing aspects of ISI is the fact that it can be avoided - at least in theory. For instance, Nyquist showed that for a channel having an ideal lowpass filter characteristic with cutoff frequency W, transmission of pulse amplitude modulation (PAM) signals is possible without generating ISI provided that the symbol rate does not exceed $2W$ [1], [2, p. 337]. More interestingly, ISI can also be avoided on channels whose transfer function is not constant over its passband, although not when PAM signals are used. Shannon [3, p. 169] pointed out long ago that a continuous-time channel with a linear filter and additive colored Gaussian noise can be divided into a large number of parallel narrowband channels with statistically independent noise processes, where each of the parallel channels has a small passband with nearly constant amplitude response and almost flat noise power spectral density within that passband. In other words, the channel can be divided into a bank of independent AWGN channels. (Shannon concluded from this observation that the capacity of a linear channel with colored Gaussian noise and an average power constraint is the sum of the capacities of the parallel AWGN channels when the transmitter power is optimally distributed among the parallel channels.)

This might be a good place to reflect on how a discrete-time channel can be created from a waveform channel and whether ISI can be

avoided in this discrete-time channel. In the now classical communication theory as formulated by Shannon [3], the waveform transmitted over a continuous-time channel is represented as

$$x_i(t) = \sum_{j=1}^{N} x_{ij}\phi_j(t), \quad 1 \leq i \leq m,$$

where $\{\phi_j(t) : 1 \leq j \leq N\}$ is a set of orthonormal functions that are usually chosen to be convenient for modulation and demodulation, cf. [4, pp. 223]. Each possible waveform $x_i(t)$ can thus be represented by a point $\underline{x}_i = [x_{i1}, x_{i2}, \ldots, x_{iN}]^T$ in an N-dimensional 'signal space'. Assuming a 'straight-wire' channel with AWGN, the modulator, the channel, and the demodulator combine to form an equivalent[1] memoryless vector channel

$$\underline{y} = \underline{x} + \underline{w},$$

where $\underline{y}$ is the received vector, $\underline{x}$ is a point chosen from the signal set $\{\underline{x}_i : 1 \leq i \leq m\}$, and $\underline{w}$ is a vector of independent Gaussian random variables with zero mean and variance $N_0/2$. Interestingly, this vector-channel model can also be used with a nontrivial linear channel filter [5] - the transmitted waveform $x_i(t)$ is then obtained by convolving the modulator waveform with the impulse response of the channel. However, such a filtering operation complicates the computation of the orthonormal functions $\phi_j(t)$ needed to represent the transmitted waveforms $x_i(t)$ and, typically, requires a signal space with large dimension N. In many applications, the dependence of the $\phi_j(t)$ on the channel impulse response is highly undesirable as this impulse response may be unknown or time-varying.

In PAM systems, a controlled amount of ISI is sometimes introduced to shape the transmitted power spectrum or to reduce the sensitivity to timing errors. Such design issues have led to the proposal of partial-response systems [6] in which ISI is deliberately introduced.

In this dissertation, we will assume that one is either not able or not willing to create an ISI-free discrete-time channel. Assuming PAM on a linearly distorted AWGN channel, the combination of the modulator, the channel, and the demodulator can be reduced (without loss of information) to a time-invariant discrete-time channel called the *ISI*

[1] Here, 'equivalent' means that no information is lost by replacing the continuous channel output with the vector $\underline{y}$.

channel with AWGN [or the *discrete-time Gaussian channel* (DTGC)] and described by

$$y_i = \sum_{m=0}^{\infty} h_m x_{i-m} + w_i, \quad -\infty < i < \infty,$$

where $\{w_i\}$ is white Gaussian noise [7], [8]. Here, the ISI is just the contribution of the past modulation symbols $x_{i-1}, x_{i-2}, \ldots$ to the current output y_i.

A formally equivalent channel is obtained in the case of quadrature amplitude modulation (QAM), where the modulation symbols x_i can be complex and the noise $\{w_i\}$ is a complex white Gaussian process whose real part and imaginary part are independent real white Gaussian processes with the same sample variance. We shall have further occasion to see the usefulness of such complex random processes.

The capacity of the DTGC with an average symbol-energy constraint was recently computed by Hirt and Massey [9]. By introducing a hypothetical circular channel model, they were able to avoid the formidable asymptotics of Toeplitz matrices found in other derivations. Following Shannon's original derivation [3, p. 169] and a later treatment by Gallager [10, p. 385] of the capacity of continuous-time channels with a linear filter, additive Gaussian noise and an average power constraint, Hirt and Massey converted the circular channel to a bank of independent, memoryless, frequency-domain channels by means of a *real* discrete Fourier transform. To simplify their derivation further, we will introduce a *complex* circular channel model that allows the use of the more familiar *discrete Fourier transform* (DFT).

Trellis-coded data transmission over ISI channels with AWGN is of much current interest in various applications, e.g., mobile radio systems, high-rate digital subscriber lines (HDSL's) [11], voiceband modems [12], and magnetic recording [13]. Assuming that the tasks of channel coding and modulation have been appropriately separated, we can show the communication system of interest as in Figure 1.1. It is obvious that a finite-state trellis encoder followed by a discrete-time channel filter with a finite unit-sample response[2] itself forms a finite-state trellis encoder. We will show how to upper-bound the bit error probability for maximum-likelihood decoding of this latter trellis code by using an

[2]We will use the more precise term 'unit-sample response' rather than the corruption 'impulse response' for discrete-time filters, except for the standard terminology of a 'finite-impulse-response' (FIR) or 'infinite-impulse-response' (IIR) filter [14].

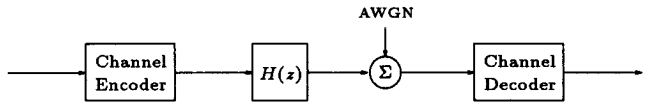


Figure 1.1: Coded communication over an ISI channel with a channel filter $H(z) = \sum_{m=0}^{\infty} h_m z^{-m}$ and AWGN

average distance spectrum that can be evaluated efficiently by means of a *modified Viterbi algorithm.*

Trellis codes designed for the 'straight-wire' channel with AWGN are often used in the presence of ISI because of the unpredictable nature of ISI on some channels. The design of trellis codes for a large free Euclidean distance at the *output* of a subsequent partial-response channel is a relatively new idea [15], [13], [16]. In particular, matched spectral-null (MSN) codes for partial-response channels have received considerable attention recently because of an interesting lower bound on the free squared Euclidean distance at the channel output for MSN codes [13, Prop. 6], [16, Thm. 6]. MSN trellis encoders can be constructed from 'canonical graphs' with the desired spectral nulls [17], [18], [13], [16] by using *state-splitting algorithms* [19], [20]. However, currently available state-splitting algorithms are unable to attain a large free Euclidean distance. Moreover, how to construct minimal MSN encoders is an open problem and easy-to-use criteria for avoiding catastrophicity are not available.

We propose, as an alternative to state-splitting, an exhaustive search for (n, k) trellis codes with a first-order spectral null at zero frequency and large free Euclidean distance. Our approach assigns code n-tuples to the branches of 2^k-ary state-transition diagrams. (A *K-ary state-transition diagram* (STD) is a directed graph with exactly K branches leaving every node.) It turns out to be desirable to use *strongly connected* 2^k-ary STD's, i.e., 2^k-ary STD's in which there is a path from any node to any other node. Our approach raises the following questions:

(i) How can we construct all strongly connected 2^k-ary STD's with a given number of nodes?

(ii) How can we eliminate 'isomorphic' 2^k-ary STD's, i.e., 2^k-ary STD's that differ only in the names of their nodes and

branches?

(iii) Which strongly connected 2^k-ary STD's are well-suited for constructing trellis encoders?

These questions will be pursued in a self-contained chapter on state-transition diagrams that may be of interest for other applications of trellis codes.

The outline of this dissertation is as follows. Chapter 2 provides a rounded treatment of certain complex random variables and processes that we will call *proper* and is motivated by the simplifications that arise from using such random variables in the analysis of ISI channels[3]. In Section 2.1, second-order statistical properties, such as uncorrelatedness and wide-sense stationarity of complex random variables and processes, are characterized in terms of the conventional covariance and an unconventional quantity that we call the *pseudo-covariance.* In Section 2.2, we introduce the class of proper complex random variables and processes, which is characterized by a vanishing pseudo-covariance. We show in Section 2.2.1 that several results from the theory of real random variables can be generalized in a natural way to *proper* complex random variables. For instance, the famous maximum-entropy theorem [10, Thm. 7.4.1], [23, Thm. 9.6.5] is generalized to complex random vectors. It is shown in Section 2.2.2 that when a bandpass communication channel with wide-sense stationary noise is represented by an equivalent complex baseband channel [4], then the additive noise process in this baseband channel is proper complex. In Section 2.3, a DFT correspondence is presented that relates circular stationarity in the time domain to uncorrelatedness in the frequency domain for sequences of proper complex random variables.

Chapter 3 is devoted to the computation of information rates for real or complex ISI channels with AWGN and makes extensive use of the results derived in Chapter 2. In Section 3.1, Hirt and Massey's derivation of the capacity of the real DTGC [9] is simplified by first generalizing to the complex DTGC. It is shown that the DFT permits the conversion of a complex circular channel with proper complex Gaussian noise to a bank of independent, memoryless, frequency-domain channels. To aid in selecting the information rate for coded data transmission over a real [or complex] ISI channel with AWGN, one often wishes to have a

[3]Chapter 2 together with Section 3.1 will appear in the *IEEE Transactions on Information Theory* [22].

good lower bound on the channel capacity C_S for a given 1-dimensional real [or complex] *signal set*. According to [24, Eq. (4.13a), pp. 4-7], C_S is lower-bounded by $I_{\text{i.i.d.}}$, the per-symbol mutual information with the same signal set under the additional constraint that the modulation symbols are independent and identically distributed. However, computing $I_{\text{i.i.d.}}$ requires the numerical evaluation of an N-dimensional integral, where N approaches infinity. In general, such an integral can only be approximated, e.g., by Monte Carlo methods [24, pp. 4-17]. As an alternative to computing $I_{\text{i.i.d.}}$, Shamai, Ozarow, and Wyner recently derived a lower bound on $I_{\text{i.i.d.}}$ [25] that can be interpreted as the mutual information for a *memoryless* AWGN channel with the given signal set and a *degraded* signal-to-noise ratio. In Section 3.2, we provide a simpler proof of this lower bound that is based on the information-theoretic equivalence of certain allpass-transformed ISI channels.

The self-contained Chapter 4 deals with K-ary STD's. In Section 4.1, the framework for the investigation of K-ary STD's is developed. The problem of constructing strongly connected K-ary STD's from so-called partial K-ary STD's is addressed in Section 4.2, where necessary and sufficient conditions are derived for when a partial K-ary STD can be extended to a (complete) strongly connected K-ary STD with N nodes. In Section 4.3, an algorithm is presented for the systematic construction of all non-isomorphic K-ary STD's with N nodes and given topological constraints. A particularly useful constraint for a K-ary STD is the *detour memory*, defined in Section 4.1 as the smallest nonnegative integer M such that the STD contains a pair of parallel paths of length $M+1$. In Section 4.4, we investigate K-ary STD's with *maximum detour memory*, i.e., K-ary STD's with $N = K^M$ nodes, where M is the detour memory. We prove that such STD's are strongly connected and have the same number of branches ending at every node. Tables containing all non-isomorphic *binary* STD's with maximum detour memory for $N = 1, 2, 4, 8$, and 16 nodes are provided.

In Chapter 5, we study trellis-coded data transmission over ISI channels. In Section 5.1, trellis encoders and trellis codes are characterized by means of labeled directed graphs and an upper bound on the free distance of a trellis code is given. In Section 5.2, we analyze the trellis encoder created by cascading an outer trellis encoder with a finite-impulse-response (FIR) channel filter. More specifically, we show how to obtain the so-called *steady-state encoder* from the *composite trellis encoder* whose state contains both the state of the outer trellis encoder and the state of the channel filter. In Section 5.3, we generalize the well-

known upper bound on bit error probability for convolutional encoders and maximum-likelihood decoding [26, Sec. 4.4, pp. 242], [27, Sec. 6.E] to trellis encoders with a *non-uniform distance spectrum*, i.e., with a distance spectrum that depends on the reference path. The generalized upper bound involves an *average distance spectrum*, which can be evaluated by means of a *modified Viterbi algorithm* as described in Section 5.4. An analysis of bipolar trellis encoders for the dicode channel in Section 5.5 demonstrates the importance of the average number of bit errors over all detours at free distance in the case of a non-uniform distance spectrum.

In Chapter 6, the results obtained in this thesis are summarized.

Leer - Vide - Empty

Chapter 2

Properness in Complex Probability Theory

The purpose of this chapter is to provide a rounded treatment of certain complex random variables and processes, which we will call *proper*, and to show their usefulness in statistical communication theory. It will be shown, for instance, that the probability density function of a complex Gaussian random vector assumes the anticipated 'natural' form only for proper random vectors.

Complex signals were introduced by electrical engineers to incorporate phase information in a convenient way. An important application of complex signals is the analysis of linear bandpass communication channels. Such channels can be represented in baseband by an equivalent two-dimensional channel with two quadrature inputs and two quadrature outputs [4], [2]. In general, for a passband channel with memory, the quadrature components interfere so that the two-dimensional equivalent baseband channel does not reduce to a pair of independent quadrature channels as in the memoryless case. To simplify notation, most communication engineers describe the equivalent baseband channel in terms of complex signals and complex impulse responses. Formulations of linear systems for complex-valued signals are also increasingly employed in adaptive signal processing, see e.g., [28]. Somewhat paradoxically, one finds in the literature very few treatments of complex random variables and processes. In fact, many investigators resort to the two-dimensional real representation of systems with complex signals whenever a probabilistic treatment is needed. Notable exceptions are Doob [29], who gives considerable attention to complex Gaussian

random processes, and Wooding [30], who first derived the probability density function for a complex Gaussian random vector whose real and imaginary part have certain symmetries in their autocovariance and crosscovariance matrix - symmetries, which are equivalent to properness in our terminology.

The organization of this chapter is as follows. In Section 2.1, we characterize second-order statistical properties such as uncorrelatedness and wide-sense stationarity of complex random variables and processes. We show that to specify the four covariances arising between the real and imaginary parts of two complex random variables X and Y, one needs both the conventional covariance $c_{XY} \triangleq \mathrm{E}[(X - m_X)(Y - m_Y)^*]$ and the unconventional quantity $\tilde{c}_{XY} \triangleq \mathrm{E}[(X - m_X)(Y - m_Y)]$, which we will call the *pseudo-covariance*. Complex random variables and processes with a vanishing pseudo-covariance will be called *proper*. In Section 2.2, we justify the terminology 'proper' by demonstrating several natural results for the class of proper complex random variables and processes that do not hold in general. For instance, the probability density function and the entropy of a proper complex Gaussian random vector are specified solely by the vector of means and the matrix of (conventional) covariances. For bandpass communication channels with real wide-sense stationary noise, it is shown that the complex noise at the demodulator output is proper. In Section 2.3, we prove a general discrete Fourier transform correspondence between *circular stationarity* in the time domain and uncorrelatedness in the frequency domain for sequences of proper complex random variables. An application of this correspondence will be given in Section 3.1, where an earlier derivation of capacity for discrete-time Gaussian channels with memory [9] is considerably simplified by first generalizing to complex channels.

2.1 Preliminaries

2.1.1 Complex Random Variables

A complex random variable X is defined as a random variable of the form

$$X = X_c + j\,X_s \quad , \quad j = \sqrt{-1},$$

where the real and imaginary parts, X_c and X_s, are real random variables [29, p. 7]. The subscripts 'c' and 's', borrowed from [4] and [2], suggest the cosine and sine components of an equivalent baseband sig-

nal. The expectation of a real random variable is naturally generalized to the complex case [31, p. 472] as $\mathrm{E}[X] \triangleq \mathrm{E}[X_c] + j\,\mathrm{E}[X_s]$. The statistical properties of $X = X_c + j\,X_s$ are determined by the joint probability density function (p.d.f.) $p_{X_c X_s}(x_c, x_s)$ of X_c and X_s, provided of course that the p.d.f. exists. For convenience, we introduce the notation $p_X(x_c + j\,x_s) \triangleq p_{X_c X_s}(x_c, x_s)$.

Let F be a complex-valued function whose domain includes the range $X(\Omega)$ of the complex random variable X, where Ω is the sample space. The expectation of $F(X)$ can be expressed in terms of two expectations of real functions in the real random variables X_c and X_s as

$$\mathrm{E}[F(X)] \triangleq \mathrm{E}[\mathrm{Re}\{F(X_c + j\,X_s)\}] + j\,\mathrm{E}[\mathrm{Im}\{F(X_c + j\,X_s)\}].$$

Equivalently,

$$\mathrm{E}[F(X)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(x_c + j\,x_s)\,p_X(x_c + j\,x_s)\,dx_c\,dx_s.$$

To specify the 'covariance' of two complex random vectors $\underline{X} = \underline{X}_c + j\,\underline{X}_s$ and $\underline{Y} = \underline{Y}_c + j\,\underline{Y}_s$, the four covariance matrices

$$\mathrm{Cov}[\underline{X}_c, \underline{Y}_c];\quad \mathrm{Cov}[\underline{X}_s, \underline{Y}_s];\quad \mathrm{Cov}[\underline{X}_s, \underline{Y}_c];\quad \mathrm{Cov}[\underline{X}_c, \underline{Y}_s] \tag{2.1}$$

are needed, where the covariance matrix of two real random vectors $\underline{U}$ and $\underline{V}$ is defined as

$$\mathrm{Cov}[\underline{U}, \underline{V}] \triangleq \mathrm{E}\left[(\underline{U} - \mathrm{E}[\underline{U}])\,(\underline{V} - \mathrm{E}[\underline{V}])^T\right]. \tag{2.2}$$

The *covariance matrix*

$$\mathbf{\Lambda}_{\underline{X}\underline{Y}} \triangleq \mathrm{E}\left[(\underline{X} - m_{\underline{X}})\,(\underline{Y} - m_{\underline{Y}})^*\right], \tag{2.3}$$

where $m_{\underline{X}} \triangleq \mathrm{E}[\underline{X}]$, $m_{\underline{Y}} \triangleq \mathrm{E}[\underline{Y}]$ and '$*$' denotes conjugate-transpose, is widely used in the literature. We define also the *pseudo-covariance matrix*

$$\tilde{\mathbf{\Lambda}}_{\underline{X}\underline{Y}} \triangleq \mathrm{E}\left[(\underline{X} - m_{\underline{X}})\,(\underline{Y} - m_{\underline{Y}})^T\right], \tag{2.4}$$

which will play a key role in what follows. To simplify the notation for (pseudo-)autocovariance matrices, we will write $\mathbf{\Lambda}_{\underline{X}}$ (or $\tilde{\mathbf{\Lambda}}_{\underline{X}}$) instead of $\mathbf{\Lambda}_{\underline{X}\underline{X}}$ (or $\tilde{\mathbf{\Lambda}}_{\underline{X}\underline{X}}$). The 'covariance' of two complex random vectors can

be specified alternatively by the complex covariance and the pseudo-covariance since it follows from (2.2) - (2.4) that

$$\begin{aligned}
\mathrm{Cov}\,[\underline{X}_c, \underline{Y}_c] &= \frac{1}{2}\,\mathrm{Re}\left\{\mathbf{\Lambda}_{\underline{X}\underline{Y}} + \tilde{\mathbf{\Lambda}}_{\underline{X}\underline{Y}}\right\} \\
\mathrm{Cov}\,[\underline{X}_s, \underline{Y}_c] &= \frac{1}{2}\,\mathrm{Im}\left\{\mathbf{\Lambda}_{\underline{X}\underline{Y}} + \tilde{\mathbf{\Lambda}}_{\underline{X}\underline{Y}}\right\} \\
\mathrm{Cov}\,[\underline{X}_s, \underline{Y}_s] &= \frac{1}{2}\,\mathrm{Re}\left\{\mathbf{\Lambda}_{\underline{X}\underline{Y}} - \tilde{\mathbf{\Lambda}}_{\underline{X}\underline{Y}}\right\} \\
\mathrm{Cov}\,[\underline{X}_c, \underline{Y}_s] &= \frac{1}{2}\,\mathrm{Im}\left\{\tilde{\mathbf{\Lambda}}_{\underline{X}\underline{Y}} - \mathbf{\Lambda}_{\underline{X}\underline{Y}}\right\}.
\end{aligned} \tag{2.5}$$

To define uncorrelatedness consistently with the corresponding notion for real random vectors, the complex random vectors $\underline{X}$ and $\underline{Y}$ are called *uncorrelated* if all four covariances in (2.1) vanish. From (2.5) we now obtain the following simple result.

Lemma 2.1: The complex random vectors $\underline{X}$ and $\underline{Y}$ are uncorrelated if and only if $\mathbf{\Lambda}_{\underline{X}\underline{Y}} = \mathbf{0}$ and $\tilde{\mathbf{\Lambda}}_{\underline{X}\underline{Y}} = \mathbf{0}$, i.e., if and only if both the covariance matrix and the pseudo-covariance matrix vanish. ———

2.1.2 Complex Random Processes

A continuous-time (or discrete-time) complex random process is defined as a random process of the form $X(t) \triangleq X_c(t) + j\,X_s(t)$ (or $X[k] \triangleq X_c[k] + j\,X_s[k]$), where $X_c(t)$ and $X_s(t)$ (or $X_c[k]$ and $X_s[k]$) are a pair of real continuous-time (or discrete-time) random processes. By definition, a complex random process is wide-sense stationary (w.s.s.) if its real and imaginary parts are jointly w.s.s.. The following result [32, p. 121] characterizes wide-sense stationarity in terms of the mean $m_X(t) \triangleq \mathrm{E}\,[X(t)]$, the *autocorrelation function*

$$r_X(\tau, t) \triangleq \mathrm{E}\,[X(t+\tau)\,X^*(t)]\,,$$

and the *pseudo-autocorrelation function*

$$\tilde{r}_X(\tau, t) \triangleq \mathrm{E}\,[X(t+\tau)\,X(t)]$$

of the complex random process $X(.)$.

Lemma 2.2: A complex random process $X(.)$ is w.s.s. if and only if $m_X(t)$, $r_X(\tau, t)$ and $\tilde{r}_X(\tau, t)$ are independent of t. ———

The corresponding result for discrete-time processes is obvious.

2.2 Proper Complex Random Variables and Processes

2.2.1 Proper Complex Random Variables

Definition 2.1: A complex random vector $\underline{Z} = \underline{Z}_c + j\,\underline{Z}_s$ will be called *proper* if its pseudo-covariance $\tilde{\mathbf{\Lambda}}_{\underline{Z}}$ vanishes. The complex random vectors $\underline{Z}_1$ and $\underline{Z}_2$ will be called *jointly proper* if the composite random vector having $\underline{Z}_1$ and $\underline{Z}_2$ as subvectors is proper. ———

Note that any subvector of a proper random vector is also proper. However, two individually proper random vectors are not necessarily also jointly proper. Defining $\mathbf{\Lambda}_{cc} \triangleq \mathrm{Cov}\,[\underline{Z}_c, \underline{Z}_c]$, $\mathbf{\Lambda}_{ss} \triangleq \mathrm{Cov}\,[\underline{Z}_s, \underline{Z}_s]$, $\mathbf{\Lambda}_{sc} \triangleq \mathrm{Cov}\,[\underline{Z}_s, \underline{Z}_c]$, $\mathbf{\Lambda}_{cs} \triangleq \mathrm{Cov}\,[\underline{Z}_c, \underline{Z}_s]$ and using the fact that $\mathbf{\Lambda}_{cs} = \mathbf{\Lambda}_{sc}^T$, the covariance and the pseudo-covariance of a complex random vector $\underline{Z}$ can be written as

$$\mathbf{\Lambda}_{\underline{Z}} = \mathbf{\Lambda}_{cc} + \mathbf{\Lambda}_{ss} + j\,(\mathbf{\Lambda}_{sc} - \mathbf{\Lambda}_{sc}^T) \tag{2.6}$$

and

$$\tilde{\mathbf{\Lambda}}_{\underline{Z}} = \mathbf{\Lambda}_{cc} - \mathbf{\Lambda}_{ss} + j\,(\mathbf{\Lambda}_{sc} + \mathbf{\Lambda}_{sc}^T), \tag{2.7}$$

respectively. Thus, the vanishing of $\tilde{\mathbf{\Lambda}}_{\underline{Z}}$ is equivalent to the conditions that

$$\mathbf{\Lambda}_{cc} = \mathbf{\Lambda}_{ss} \quad \text{and} \quad \mathbf{\Lambda}_{sc} = -\mathbf{\Lambda}_{sc}^T, \tag{2.8}$$

i.e., $\tilde{\mathbf{\Lambda}}_{\underline{Z}}$ vanishes if and only if $\underline{Z}_c$ and $\underline{Z}_s$ have identical autocovariance matrices and a skew-symmetric crosscovariance matrix. We conclude that a proper complex random vector $\underline{Z}$ has the covariance matrix

$$\mathbf{\Lambda}_{\underline{Z}} = 2\,(\mathbf{\Lambda}_{cc} + j\,\mathbf{\Lambda}_{sc}). \tag{2.9}$$

Note that the skew-symmetry of $\mathbf{\Lambda}_{sc}$ implies that $\mathbf{\Lambda}_{sc}$ has a zero main diagonal, which means that the real and imaginary part of each component Z_k of $\underline{Z}$ are uncorrelated. The vanishing of $\tilde{\mathbf{\Lambda}}_{\underline{Z}}$ does not, however, imply that the real part of Z_k and the imaginary part of Z_l are uncorrelated for $k \neq l$. It should be pointed out that a *real* random vector is a proper complex random vector if and only if it is constant (with probability 1), since $\mathbf{\Lambda}_{ss} = \mathbf{0}$ and (2.8) imply $\mathbf{\Lambda}_{cc} = \mathbf{0}$.

The appropriateness of the term 'proper' in connection with complex random vectors is supported by the following lemma dealing with closure

under affine transformations as well as by a number of other results to follow.

Lemma 2.3: Let $\underline{Z}$ be a proper complex n-dimensional random vector, i.e., $\tilde{\mathbf{\Lambda}}_{\underline{Z}} = \mathbf{0}$. Then any random vector obtained from $\underline{Z}$ by a linear or affine transformation, i.e., any random vector $\underline{Y}$ of the form $\underline{Y} = \mathbf{A}\underline{Z} + \underline{b}$, where $\mathbf{A} \in \mathbb{C}^{m \times n}$ and $\underline{b} \in \mathbb{C}^m$ are constant, is also proper.

Proof: Since $m_{\underline{Y}} = \mathbf{A}\, m_{\underline{Z}} + \underline{b}$ and $\underline{Y} - m_{\underline{Y}} = \mathbf{A}\,(\underline{Z} - m_{\underline{Z}})$, we have

$$\tilde{\mathbf{\Lambda}}_{\underline{Y}} = \mathrm{E}\left[(\underline{Y} - m_{\underline{Y}})(\underline{Y} - m_{\underline{Y}})^T\right] = \mathbf{A}\,\tilde{\mathbf{\Lambda}}_{\underline{Z}}\,\mathbf{A}^T = \mathbf{0}. \qquad \square$$

Note that $\underline{Y}$ and $\underline{Z}$ as in Lemma 2.3 are automatically jointly proper, since the vector having $\underline{Y}$ and $\underline{Z}$ as subvectors is obtained by the affine transformation

$$\begin{bmatrix} \underline{Y} \\ \underline{Z} \end{bmatrix} = \hat{\mathbf{A}}\,\underline{Z} + \begin{bmatrix} \underline{b} \\ \underline{0}_n \end{bmatrix} \quad , \quad \hat{\mathbf{A}} \triangleq \begin{bmatrix} \mathbf{A} \\ \mathbf{I}_n \end{bmatrix} \quad .$$

Lemma 2.4: Let $\underline{Z}_1$ and $\underline{Z}_2$ be two independent complex random vectors and let $\underline{Z}_2$ be proper. Then the linear combination $\underline{Y} = a_1\underline{Z}_1 + a_2\underline{Z}_2$, where a_1 and a_2 are complex numbers and $a_1 \neq 0$, is proper if and only if $\underline{Z}_1$ is also proper.

Proof: The independence of $\underline{Z}_1$ and $\underline{Z}_2$ and the properness of $\underline{Z}_2$ imply

$$\tilde{\mathbf{\Lambda}}_{\underline{Y}} = a_1^2\,\tilde{\mathbf{\Lambda}}_{\underline{Z}_1} + a_2^2\,\tilde{\mathbf{\Lambda}}_{\underline{Z}_2} = a_1^2\,\tilde{\mathbf{\Lambda}}_{\underline{Z}_1}.$$

Thus, $\tilde{\mathbf{\Lambda}}_{\underline{Y}}$ vanishes if and only if $\tilde{\mathbf{\Lambda}}_{\underline{Z}_1}$ vanishes. $\square$

Lemma 2.1 immediately implies the following result.

Lemma 2.5: Two jointly proper, complex random vectors $\underline{Z}_1$ and $\underline{Z}_2$ are uncorrelated if and only if their covariance matrix $\mathbf{\Lambda}_{\underline{Z}_1\underline{Z}_2}$ vanishes.

A *complex Gaussian random vector* $\underline{Z}$ is defined as a vector with jointly Gaussian real and imaginary parts. Following Feller [31, p. 86], we consider Gaussian distributions to include degenerate distributions

concentrated on a lower-dimensional manifold. In such degenerate cases, the $2n \times 2n$-covariance matrix

$$\Phi \triangleq \mathrm{Cov}\left[\begin{bmatrix} \underline{Z}_c \\ \underline{Z}_s \end{bmatrix}, \begin{bmatrix} \underline{Z}_c \\ \underline{Z}_s \end{bmatrix}\right] = \begin{bmatrix} \Lambda_{cc} & \Lambda_{cs} \\ \Lambda_{sc} & \Lambda_{ss} \end{bmatrix}, \tag{2.10}$$

is singular and the probability density function (p.d.f.) does not exist unless one admits generalized functions.

Note that two jointly proper Gaussian random vectors $\underline{Z}_1$ and $\underline{Z}_2$ are independent if and only if $\Lambda_{\underline{Z}_1 \underline{Z}_2} = \mathbf{0}$, which follows from Lemma 2.5 and the fact that uncorrelatedness and independence are equivalent for Gaussian random variables.

Wooding [30] was apparently the first to derive the p.d.f. of a complex Gaussian random vector satisfying the conditions (2.8), i.e., of a *proper* complex Gaussian random vector. Goodman [33] gave an alternative derivation based on the observation that the multiplication of certain orthogonal 2×2-matrices is isomorphic to the multiplication of related complex numbers. The complex-multivariate Gaussian p.d.f. is also found in [34], [35], and [36]. The results can be stated as follows.

Theorem 2.1: Let $\underline{Z}$ be a proper complex n-dimensional Gaussian random vector with mean $\underline{m}$ and nonsingular covariance matrix $\Lambda \triangleq \mathrm{E}[(\underline{Z} - \underline{m})(\underline{Z} - \underline{m})^*]$. Then the p.d.f. of $\underline{Z}$ is given by

$$p_{\underline{Z}}(\underline{z}) \triangleq p_{\underline{Z}_c \underline{Z}_s}(\underline{z}_c, \underline{z}_s) = \frac{1}{\pi^n \det(\Lambda)} \, \mathrm{e}^{-(\underline{z}-\underline{m})^* \Lambda^{-1} (\underline{z}-\underline{m})}. \tag{2.11}$$

Conversely, let the p.d.f. of a complex random vector $\underline{Z}$ be given by (2.11), where Λ is Hermitian and positive definite. Then $\underline{Z}$ is proper complex and Gaussian with covariance matrix Λ and mean $\underline{m}$. Moreover, for a proper complex $\underline{Z}$,

$$\begin{aligned} \Lambda &= 2(\Lambda_{cc} + j\,\Lambda_{sc}) \\ \Lambda^{-1} &= \tfrac{1}{2}\,\Delta^{-1}\left(\mathbf{I} - j\,\Lambda_{sc}\,\Lambda_{cc}^{-1}\right) \\ \Delta &\triangleq \Lambda_{cc} + \Lambda_{sc}\,\Lambda_{cc}^{-1}\,\Lambda_{sc} \\ \det(\Lambda) &= 2^n \sqrt{\det(\Lambda_{cc})\,\det(\Delta)}. \end{aligned} \tag{2.12}$$

Note that the p.d.f. (2.11) is completely specified by the vector of means and the conventional covariance matrix. The fact that the function (2.11) integrates to one over $\underline{z}_c$ and $\underline{z}_s$ for any positive-definite Hermitian matrix Λ was proved by Bellman without connection to p.d.f.'s

[37, Chap. 6, § 10]. This property of (2.11) can be used also to prove that $\det(\mathbf{\Lambda})$ is convex–$\cap$ over the positive-definite Hermitian matrices $\mathbf{\Lambda}$ [37, Chap. 8, § 5]. The matrix $\mathbf{\Delta}$ defined in (2.12) is known as the *Schur complement* [38, p. 46] of $\mathbf{\Lambda}_{cc}$ in the matrix $\mathbf{\Phi}$ of (2.10) . A proof of Theorem 2.1 is included in Appendix 2.A.

To specify that a random variable X is Gaussian [or proper complex Gaussian] with mean m and variance σ^2, we will sometimes write $X \sim \mathcal{N}(m, \sigma^2)$ [or $X \sim \mathcal{N}_p(m, \sigma^2)$]. Analogous notation will be used for Gaussian [or proper complex Gaussian] random vectors.

Example 2.1: Let $\underline{Z} \sim \mathcal{N}_p(\underline{0}, N_0\,\mathbf{I})$ so that the components of $\underline{Z}$ are independent and have equal variance N_0. Then the p.d.f. of $\underline{Z}$ is given by

$$p_{\underline{Z}}(\underline{z}) = \frac{1}{(\pi N_0)^n}\, e^{-\|\underline{z}\|^2/N_0}.$$

As an application of Theorem 2.1, we generalize the maximum-entropy theorem [10, Thm. 7.4.1], [23, Thm. 9.6.5] to the complex multivariate case. The result will be used in Section 3.1 to compute the capacity of a channel with proper complex Gaussian noise. For a complex Gaussian random vector $\underline{Z} = \underline{Z}_c + j\,\underline{Z}_s$, the *differential entropy* is appropriately defined as the joint differential entropy of its real and imaginary part, i.e., $h(\underline{Z}) \triangleq h(\underline{Z}_c\,\underline{Z}_s)$.

Theorem 2.2: Let $\underline{Z}$ be a complex, continuous, n-dimensional random vector with nonsingular correlation matrix $\mathbf{R}_{\underline{Z}} \triangleq \mathrm{E}\,[\underline{Z}\,\underline{Z}^*]$. Then

$$h(\underline{Z}) \leq \log\left[(\pi e)^n \det\left(\mathbf{R}_{\underline{Z}}\right)\right]$$

with equality if and only if $\underline{Z}$ is *proper* and Gaussian with zero mean.

Note that no real random vector maximizes entropy for a given correlation matrix $\mathbf{R}_{\underline{Z}}$ when complex random vectors are allowed. The proof of the analog to Theorem 2.2 for real random variables [10, p. 336], [23, p. 234] can be easily generalized to a proof of Theorem 2.2 by using the Gaussian density (2.11). For a scalar complex random variable, we give a different proof of Theorem 2.2 that better illustrates the role of properness.

Proof of Theorem 2.2 (scalar case): Let $Z = Z_c + j\, Z_s$ be a scalar complex random variable with the constraint $\mathrm{E}\left[\,|Z|^2\,\right] = S$. According to the maximum-entropy theorem for real random vectors [23, p. 234], the real random vector $\underline{W} \triangleq [Z_c,\, Z_s]^T$, for which $\mathbf{R}_{\underline{W}} \triangleq \mathrm{E}\left[\underline{W}\,\underline{W}^T\right]$ is nonsingular, satisfies

$$h(Z) = h(\underline{W}) \leq \tfrac{1}{2}\, \log\left[(2\pi e)^2\, \det(\mathbf{R}_{\underline{W}})\right]$$

with equality if and only if $\underline{W}$ is zero-mean Gaussian. By hypothesis, $\mathrm{E}\left[\,|Z|^2\,\right] = \mathrm{E}\left[\,Z_c^2\,\right] + \mathrm{E}\left[\,Z_s^2\,\right] = S$ and thus

$$\det(\mathbf{R}_{\underline{W}}) = E\left[Z_c^2\right] E\left[Z_s^2\right] - (\mathrm{E}\left[\,Z_c\, Z_s\,\right])^2 \leq E\left[Z_c^2\right] E\left[Z_s^2\right] \leq S^2/4,$$

where equality holds at both places if and only if $\mathrm{E}\left[\,Z_c\, Z_s\,\right] = 0$ and $E\left[Z_c^2\right] = E\left[Z_s^2\right]$. Therefore, $h(Z) = h(Z_c\, Z_s) = h(\underline{W}) \leq \log[\pi e S]$ with equality if and only if Z is proper and Gaussian with zero mean. □

Note that $h(\underline{Z}) = h(\underline{Z}_c\, \underline{Z}_s) = h(\underline{Z}_c) + h(\underline{Z}_s \,|\, \underline{Z}_c)$ for a complex random vector $\underline{Z} = \underline{Z}_c + j\, \underline{Z}_s$ and, when $\underline{Z}$ is Gaussian,

$$h(\underline{Z}_c) = \tfrac{1}{2}\, \log\left[(2\,\pi\, e)^n\, \det(\mathbf{\Lambda}_{cc})\right], \tag{2.13}$$

where $\mathbf{\Lambda}_{cc}$ is as in Theorem 2.1. It follows from (2.13), the proof of Theorem 2.1 (see (2.A.9) in Appendix 2.A), and from Theorem 2.2 that, for a proper complex Gaussian random vector $\underline{Z}$,

$$h(\underline{Z}_s \,|\, \underline{Z}_c) = \tfrac{1}{2}\, \log\left[(2\,\pi\, e)^n\, \det(\mathbf{\Delta})\right].$$

Example 2.2: Let $\underline{Z} = \underline{Z}_c + j\, \underline{Z}_s \sim \mathcal{N}_p(\underline{0}, \mathbf{\Lambda}_{\underline{Z}})$, where $\mathrm{Im}\left\{\mathbf{\Lambda}_{\underline{Z}}\right\} = \mathbf{0}$. Then (2.9) implies $\mathbf{\Lambda}_{sc} = \mathbf{0}$ so that the 'quadrature vectors' $\underline{Z}_c$ and $\underline{Z}_s$ are independent and $\mathbf{\Lambda}_{\underline{Z}} = 2\, \mathbf{\Lambda}_{cc}$. It follows from Theorem 2.2 that

$$\begin{aligned} h(\underline{Z}) &= \log\left[(\pi e)^n\, \det(2\, \mathbf{\Lambda}_{cc})\right] = \log\left[(2\pi e)^n\, \det(\mathbf{\Lambda}_{cc})\right] \\ &= 2\, h(\underline{Z}_c) = 2\, h(\underline{Z}_s). \end{aligned}$$

For a complex Gaussian random vector $\underline{Z}$ with zero mean and covariance matrix $\mathbf{\Lambda} = \mathbf{\Lambda}_c + j\, \mathbf{\Lambda}_s$, Theorem 2.2 implies the following nontrivial result in matrix theory, kindly suggested to us by Roger Cheng [private communication]:

Corollary 2.1: For any symmetric matrix $\mathbf{\Lambda}_c \in \mathbb{R}^{n\times n}$ and skew-symmetric matrix $\mathbf{\Lambda}_s \in \mathbb{R}^{n\times n}$ such that $\mathbf{\Lambda}_c + j\,\mathbf{\Lambda}_s$ is positive definite,

$$\max_{\substack{\begin{bmatrix}\mathbf{\Lambda}_{cc} & \mathbf{\Lambda}_{sc}^T \\ \mathbf{\Lambda}_{sc} & \mathbf{\Lambda}_{ss}\end{bmatrix} \geq 0 \\ \mathbf{\Lambda}_{cc}+\mathbf{\Lambda}_{ss}=\mathbf{\Lambda}_c \\ \mathbf{\Lambda}_{sc}-\mathbf{\Lambda}_{sc}^T=\mathbf{\Lambda}_s \\ \mathbf{\Lambda}_{cc}^T=\mathbf{\Lambda}_{cc},\, \mathbf{\Lambda}_{ss}^T=\mathbf{\Lambda}_{ss}}} \det\begin{bmatrix}\mathbf{\Lambda}_{cc} & \mathbf{\Lambda}_{sc}^T \\ \mathbf{\Lambda}_{sc} & \mathbf{\Lambda}_{ss}\end{bmatrix} = \frac{1}{2^{2n}} \det(\mathbf{\Lambda}_c) \det(\mathbf{\Lambda}_c + \mathbf{\Lambda}_s \mathbf{\Lambda}_c^{-1} \mathbf{\Lambda}_s),$$

where the maximum is achieved if and only if $\mathbf{\Lambda}_{cc} = \mathbf{\Lambda}_{ss} = \frac{1}{2}\mathbf{\Lambda}_c$ and $\mathbf{\Lambda}_{sc} = \frac{1}{2}\mathbf{\Lambda}_s$. ———

As one might expect, the differential entropy of real and complex random variables is affected differently by *scaling*. For any matrix $\mathbf{A} = \mathbf{A}_c + j\,\mathbf{A}_s \in \mathbb{C}^{n\times n}$, we can represent $\underline{Y} = \mathbf{A}\,\underline{X}$ by $\begin{bmatrix}\underline{Y}_c \\ \underline{Y}_s\end{bmatrix} = \mathbf{B}\begin{bmatrix}\underline{X}_c \\ \underline{X}_s\end{bmatrix}$, where $\mathbf{B} \triangleq \begin{bmatrix}\mathbf{A}_c & -\mathbf{A}_s \\ \mathbf{A}_s & \mathbf{A}_c\end{bmatrix}$. The scaling property for real random vectors [23, p. 234] and the fact that $\det(\mathbf{B}) = |\det(\mathbf{A})|^2$ [33, p. 156] now imply

$$h(\mathbf{A}\underline{X}) = h(\underline{X}_c\,\underline{X}_s) + \log|\det(\mathbf{B})| = h(\underline{X}) + 2\,\log|\det(\mathbf{A})|. \quad (2.14)$$

For a complex, non-degenerate, scalar random variable X, (2.14) yields

$$e^{h(aX)} = |a|^2\, e^{h(X)}, \quad a \in \mathbb{C},$$

which is plausible since the *entropy power* of a random variable can be interpreted as the effective size of its support set and the support set of X is an area in the complex plane.

2.2.2 Proper Complex Random Processes

The *covariance function* of a complex random process is defined as

$$c_Z(\tau,t) \triangleq \mathrm{E}\left[(Z(t+\tau) - m_Z(t+\tau))\,(Z(t) - m_Z(t))^*\right] \quad (2.15)$$

for continuous-time processes and as

$$c_Z[k,n] \triangleq \mathrm{E}\left[(Z[n+k] - m_Z[n+k])\,(Z[n] - m_Z[n])^*\right] \quad (2.16)$$

for discrete-time processes, where obvious notation has been used for the means. Analogously, we will define the *pseudo-covariance function* of a complex random process as

$$\tilde{c}_Z(\tau,t) \triangleq \mathrm{E}\left[(Z(t+\tau) - m_Z(t+\tau))\,(Z(t) - m_Z(t))\right] \quad (2.17)$$

for continuous-time processes and as

$$\tilde{c}_Z[k,n] \triangleq \mathrm{E}\left[\left(Z[n+k]-m_Z[n+k]\right)\left(Z[n]-m_Z[n]\right)\right] \tag{2.18}$$

for discrete-time processes.

Definition 2.2: A complex random process will be called *proper* if its pseudo-covariance function vanishes identically. ———

Using similar arguments as in Section 2.2.1 one can show that any linear or affine transformation of a proper complex random process is proper and that a linear combination of independent proper complex random processes is also proper. Moreover, any vector of samples taken from a proper complex random process is also proper.

Proper complex random processes arise in equivalent baseband representations of bandpass communication systems, as we show next. Consider the (real) additive noise channel together with the receiver front-end shown in Figure 2.1. The real process $X_0(.)$ is assumed to be

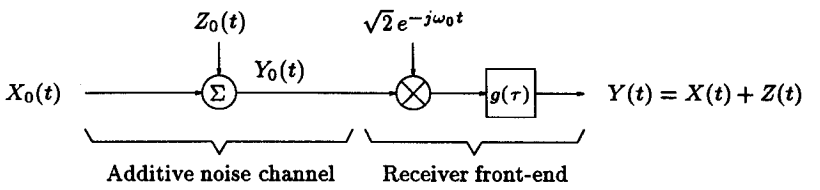


Figure 2.1: Additive noise channel and receiver front-end

w.s.s. and bandlimited to frequencies ω such that $||\omega| - \omega_0| \leq 2\pi W$, where $\omega_0 > 2\pi W$, and the real noise process $Z_0(.)$ is w.s.s. with zero mean and power spectral density $S_{Z_0}(\omega)$. The channel output $Y_0(.)$, another real process, is converted to baseband by a complex demodulator and an ideal lowpass filter $g(\tau)$ with frequency response

$$G(\omega) = \begin{cases} 1 & \text{if } |\omega| \leq 2\pi W \\ 0 & \text{otherwise.} \end{cases} \tag{2.19}$$

Let the complex random processes $X(t)$ and $Z(t)$ denote the response of the receiver-front-end, which is a time-varying linear system, to the

real random processes $X_0(t)$ and $Z_0(t)$, respectively. Wozencraft and Jacobs have shown that $Y(t) = X(t) + Z(t)$ provides sufficient statistics for an optimum receiver [4, pp. 496]. Of particular interest here are the properties of the demodulated noise $Z(.)$, which were proved in [4, pp. 498] and can be summarized in our terminology as follows.

Theorem 2.3: Let the real w.s.s. process $Z_0(.)$ with zero mean and power spectral density $S_{Z_0}(\omega)$ be the input to a complex demodulator with angular frequency ω_0 followed by an ideal lowpass filter (2.19). Then, if $\omega_0 > 2\pi W$, the complex random process

$$Z(t) \triangleq \sqrt{2} \int_{-\infty}^{\infty} Z_0(u)\, e^{-j\omega_0 u} g(t-u)\, du$$

at the lowpass filter output is w.s.s., *proper*, zero-mean, and has the autocorrelation function

$$r_Z(\tau) \triangleq \mathrm{E}[\, Z(t+\tau)\, Z^*(t)\,] = \frac{1}{\pi} \int_{-2\pi W}^{2\pi W} S_{Z_0}(\omega + \omega_0)\, e^{j\omega\tau}\, d\omega. \quad (2.20)$$

In particular, if $Z_0(.)$ is white noise with power spectral density $S_{Z_0}(\omega) = N_0/2$, then

$$r_Z(\tau) = N_0\, \frac{\sin 2\pi W \tau}{\pi \tau}. \quad (2.21)$$

Since $Z(.)$ is w.s.s., proper and zero-mean, $c_Z(\tau) = r_Z(\tau)$ and the pseudo-covariance vanishes, i.e.,

$$\tilde{c}_Z(\tau) = \tilde{r}_Z(\tau) \equiv 0. \quad (2.22)$$

Property (2.22) is equivalent to the symmetry relations [1]

$$r_{Z_c Z_c}(\tau) = r_{Z_s Z_s}(\tau) \quad \text{and} \quad r_{Z_s Z_c}(\tau) = -r_{Z_s Z_c}(-\tau), \quad (2.23)$$

i.e., the real and imaginary part of $Z(.)$ have the same autocorrelation function and an odd crosscorrelation function. Equivalent symmetry relations were found by Dugundji and Zakai for a real process $X(.)$

[1] We define $r_{UV}(\tau, t) \triangleq \mathrm{E}[U(t+\tau)V(t)]$ for any real processes $U(.)$ and $V(.)$ and write $r_{UV}(\tau)$ when $U(.)$ and $V(.)$ are jointly w.s.s..

and its Hilbert transform $\hat{X}(.)$ [39], [40], [41][2]. The process $Z(.) \triangleq X(.)+j\hat{X}(.)$ was called the 'pre-envelope' of $X(.)$ or an 'analytic signal' and satisfies (2.22). However, the requirement that the imaginary part is the Hilbert transform of the real part is more stringent than the symmetry relations (2.23) and the concept of the pre-envelope, unlike properness, is not appropriate for single random variables.

It should be mentioned that baseband complex random processes with nonzero mean are usually not of interest, since a 'complex envelope' $X(.)$ with nonzero mean corresponds to a non-stationary bandpass process. To see this, let $X_0(t) = \mathrm{Re}\{X(t)\sqrt{2}e^{j\omega_0 t}\}$ and note that $\mathrm{E}[X_0(t)] = \mathrm{Re}\{\mathrm{E}[X(t)]\sqrt{2}e^{j\omega_0 t}\} \neq$ constant if $\mathrm{E}[X(t)] \neq 0$.

We are particularly interested in the class of proper complex *Gaussian* random processes. Doob has given conditions that in our terminology are the necessary and sufficient conditions for the existence of such processes [29, Thm. 3.1]. Theorem 2.3 shows that demodulated Gaussian noise belongs to this class. The *proper complex AWGN channel* will be defined as a channel of the form $Y(.) = X(.) + Z(.)$, where $X(.)$ and $Z(.)$ are independent complex processes and $Z(.)$ is proper complex AWGN with power spectral density N_0. The proper complex *white* noise idealization is supported by the following consideration. If we choose a large bandwidth W and a carrier frequency $\omega_0 > 2\pi W$ in Theorem 2.3, then the correlation function $r_Z(\tau)$ closely approximates $N_0\,\delta(\tau)$.

2.3 Circular Stationarity

In this section, upper-case and lower-case letters denote frequency-domain and time-domain variables, respectively. For convenience, a length-N sequence $x[0]$, $x[1]$, ... , $x[N-1]$ will be written as $x[0, N-1]$. All indices in square brackets are understood in this section to be taken modulo the integer N.

Definition 2.3: A sequence of complex random variables $z[0, N-1]$ will be called *circularly wide-sense stationary (c.w.s.s.)*, if $\mathrm{E}[z[n]] = m_z$ is independent of n and if

$$\mathrm{E}[z[n]\,z^*[i]] = r_z[n-i] \quad \text{and} \quad \mathrm{E}[z[n]\,z[i]] = \tilde{r}_z[n-i] \tag{2.24}$$

[2]I. Bar-David is gratefully acknowledged for providing the references [39], [40], and [41].

holds for $0 \leq i, n < N$, i.e., if the correlation of two samples depends only on their time difference modulo N. We will call $r_z[0, N-1]$ and $\tilde{r}_z[0, N-1]$ the *circular correlation sequence* and *circular pseudo-correlation sequence*, respectively, of the c.w.s.s. sequence $z[0, N-1]$. Analogously, a sequence of real random variables $x[0, N-1]$ will be called c.w.s.s., if $\mathrm{E}[x[n]] = m_x$ is independent of n and if

$$\mathrm{E}[x[n]\, x[i]] = r_x[n-i], \quad 0 \leq i, n < N. \qquad \text{———}$$

A proper complex, non-trivial c.w.s.s. sequence $z[0, N-1]$ can be generated as the circular convolution of a proper complex white noise sequence $w[0, N-1]$ with some complex weighting sequence $h[0, N-1]$, i.e., $z[n] = \sum_{k=0}^{N-1} h[n-k]\, w[k]$, where $m_w = 0$ and $r_w[i] = \sigma^2\, \delta[i]$ with

$$\delta[i] = \begin{cases} 1 & \text{for } i = 0 \\ 0 & \text{otherwise.} \end{cases}$$

A simple calculation shows that $z[0, N-1]$ is c.w.s.s. with $m_z = 0$ and circular correlation function

$$r_z[i] = \mathrm{E}[z[n+i]\, z^*[n]] = \sigma^2 \sum_{k=0}^{N-1} h[k+i]\, h^*[k].$$

We now show that circular stationarity of a proper complex time-domain sequence corresponds to uncorrelatedness of the components of its *discrete Fourier transform* (DFT). This fact will be used in Section 3.1 to find the capacity of an ISI channel with proper complex AWGN.

Recall that the DFT of a complex sequence $z[0, N-1]$ is the sequence $Z[0, N-1]$ given by

$$Z[k] \triangleq \sum_{n=0}^{N-1} z[n]\, \Omega_N^{-kn}, \quad 0 \leq k < N, \tag{2.25}$$

where $\Omega_N \triangleq e^{j2\pi/N}$ is a primitive N-th root of unity. The time-domain sequence $z[0, N-1]$ can be recovered from the frequency-domain sequence $Z[0, N-1]$ by the *inverse discrete Fourier transform* (IDFT)

$$z[n] \triangleq \frac{1}{N} \sum_{k=0}^{N-1} Z[k]\, \Omega_N^{kn}, \quad 0 \leq n < N. \tag{2.26}$$

Note also that

$$\sum_{n=0}^{N-1} \Omega_N^{nk} = N\,\delta[k]. \tag{2.27}$$

If $z[0,N-1]$ is a sequence of complex random variables, then so also is $Z[0,N-1]$. Clearly, $z[0,N-1]$ is zero-mean if and only if $Z[0,N-1]$ is zero-mean. Moreover, by Lemma 2.3 and the invertibility of the DFT, $z[0,N-1]$ is proper if and only if $Z[0,N-1]$ is proper.

Theorem 2.4: Let $z[0,N-1]$ and its DFT $Z[0,N-1]$ be proper complex sequences with zero mean. Then the *time-domain sequence* $z[0,N-1]$ is *c.w.s.s.* if and only if the *frequency-domain sequence* $Z[0,N-1]$ is *uncorrelated*, i.e., if and only if

$$\mathrm{E}\,[\,Z[k]\,Z^*[l]\,] = N\,R_z[k]\,\delta[k-l], \tag{2.28}$$

where $R_z[0,N-1]$ is the DFT of the circular correlation sequence $r_z[0,N-1]$, i.e.,

$$R_z[k] = \sum_{n=0}^{N-1} r_z[n]\,\Omega_N^{-kn}\,; \qquad r_z[n] = \frac{1}{N}\sum_{k=0}^{N-1} R_z[k]\,\Omega_N^{kn}. \tag{2.29}$$

Proof: Suppose that the proper complex random sequence $z[0,N-1]$ is c.w.s.s. with circular correlation sequence $r_z[0,N-1]$ and let $R_z[0,N-1]$ be the DFT of $r_z[0,N-1]$. Then

$$\begin{aligned}
\mathrm{E}\,[\,Z[k]\,Z^*[l]\,] &= \sum_{i=0}^{N-1} \Omega_N^{li} \sum_{n=0}^{N-1} \mathrm{E}\,[\,z[n]\,z^*[i]\,]\;\Omega_N^{-kn} \\
&= \sum_{i=0}^{N-1} \Omega_N^{li} \sum_{n=0}^{N-1} r_z[n-i]\,\Omega_N^{-kn} \\
&= R_z[k] \sum_{i=0}^{N-1} \Omega_N^{(l-k)i} = N\,R_z[k]\,\delta[k-l],
\end{aligned}$$

where the second, third and last equality follow from (2.24), the shifting property of the DFT [14, p. 92] and from (2.27), respectively. It now follows from the properness of $Z[0,N-1]$ and from the fact that

$Z[0,N-1]$ has zero mean that the components of $Z[0,N-1]$ are uncorrelated. Conversely, suppose that $Z[0,N-1]$ satisfies (2.28) and let $r_z[0,N-1]$ be the IDFT of $R_z[0,N-1]$. Then

$$\begin{aligned} \mathrm{E}\left[z[n]\,z^*[i]\right] &= \frac{1}{N^2}\sum_{k=0}^{N-1}\sum_{l=0}^{N-1}\mathrm{E}\left[Z[k]\,Z^*[l]\right]\Omega_N^{kn-li} \\ &= \frac{1}{N}\sum_{k=0}^{N-1}R_z[k]\sum_{l=0}^{N-1}\delta[k-l]\,\Omega_N^{kn-li} \\ &= \frac{1}{N}\sum_{k=0}^{N-1}R_z[k]\,\Omega_N^{k(n-i)} = r_z[n-i]. \end{aligned}$$

The circular pseudo-correlation sequence of $z[0,N-1]$ vanishes because $z[0,N-1]$ is proper with zero mean. It follows that $z[0,N-1]$ is c.w.s.s.. □

Note that circular stationarity of $z[0,N-1]$ generally does not also imply circular stationarity of $Z[0,N-1]$ since in general $\mathrm{E}\left[\,|Z[k]|^2\,\right]$ depends on k. Since uncorrelatedness and independence are equivalent for Gaussian random variables, we immediately have the following result.

Corollary 2.2: Let $z[0,N-1]$ and its DFT $Z[0,N-1]$ be proper complex *Gaussian* sequences with zero mean. Then the *time-domain sequence* $z[0,N-1]$ is *c.w.s.s.* if and only if the components of the *frequency-domain sequence* $Z[0,N-1]$ are *independent.* ———

It was recently shown by Hirt and Massey that the *real* DFT of a sequence of real, i.i.d.[3], zero-mean Gaussian random variables is another sequence of real, i.i.d., zero-mean Gaussian random variables [9, Lemmas 1, 2]. Similarly, it was shown that the inverse *real* DFT of a (non-stationary) frequency-domain sequence with real, independent, zero-mean Gaussian components is a sequence of real, correlated, zero-mean Gaussian random variables [9, Lemma 3]. Note that the Lemmas 1-3 of [9] are special cases of the analog to Corollary 2.2 for real random variables. Thus, to establish a correspondence between circular stationarity and uncorrelatedness, the *real* DFT is needed in the case of real random variables, while the *ordinary* DFT is adequate for proper complex random variables.

[3] independent, identically distributed

Note that Theorem 2.4 is easily generalized to K-channel systems, $K \geq 2$, if the length-N sequences $z[0, N-1]$ and $Z[0, N-1]$ are replaced by length-N sequences $\underline{z}[0, N-1]$ and $\underline{Z}[0, N-1]$ of K-dimensional vectors, the correlation sequences $r_z[0, N-1]$ and $R_z[0, N-1]$ are replaced by $K \times K$-matrix sequences $\mathbf{r}_z[0, N-1]$ and $\mathbf{R}_z[0, N-1]$, and the DFT of a vector (or matrix) sequence is defined to be the vector (or matrix) of DFT's. This generalization is useful in the study of multi-user channels with finite memory.

Appendix 2.A Proof of Theorem 2.1

To prove Theorem 2.1, the following result on quadratic forms is needed:

Lemma 2.A.1: Let $\mathbf{M}_{cc}$, $\mathbf{M}_{ss}$, $\mathbf{M}_{sc}$ and $\mathbf{M}_{cs}$ be real $n \times n$-matrices, where $\mathbf{M}_{cc}$ and $\mathbf{M}_{ss}$ are symmetric and $\mathbf{M}_{cs}^T = \mathbf{M}_{sc}$. Define the Hermitian $n \times n$-matrix

$$\mathbf{M} = \mathbf{M}_c + j\,\mathbf{M}_s \triangleq \mathbf{M}_{cc} + \mathbf{M}_{ss} + j\,(\mathbf{M}_{sc} - \mathbf{M}_{sc}^T)$$

and the symmetric $2n \times 2n$-matrix

$$\mathbf{\Psi} \triangleq 2 \begin{bmatrix} \mathbf{M}_{cc} & \mathbf{M}_{cs} \\ \mathbf{M}_{sc} & \mathbf{M}_{ss} \end{bmatrix}. \tag{2.A.1}$$

Then the quadratic forms

$$\mathcal{E} \triangleq \underline{z}^*\,\mathbf{M}\,\underline{z}$$

and

$$\mathcal{E}' \triangleq [\underline{z}_c^T\ \underline{z}_s^T]\ \mathbf{\Psi} \begin{bmatrix} \underline{z}_c \\ \underline{z}_s \end{bmatrix}, \tag{2.A.2}$$

are equal for all $\underline{z} \triangleq \underline{z}_c + j\,\underline{z}_s$ if and only if

$$\mathbf{M}_{cc} = \mathbf{M}_{ss} \quad \text{and} \quad \mathbf{M}_{sc} = -\mathbf{M}_{sc}^T. \tag{2.A.3}$$

Moreover, under the conditions (2.A.3) $\mathbf{M}$ is positive (semi)definite if and only if $\mathbf{\Psi}$ is positive (semi)definite.

Proof: Since $\mathcal{E}$ is a Hermitian form, it is real for all $\underline{z}$. Hence

$$\begin{aligned} \mathcal{E} = \mathrm{Re}\{\underline{z}^*\,\mathbf{M}\,\underline{z}\} &= \underline{z}_c^T\,\mathbf{M}_c\,\underline{z}_c + \underline{z}_s^T\,\mathbf{M}_c\,\underline{z}_s + \underline{z}_s^T\,\mathbf{M}_s\,\underline{z}_c - \underline{z}_c^T\,\mathbf{M}_s\,\underline{z}_s \\ &= [\underline{z}_c^T\ \underline{z}_s^T] \begin{bmatrix} \mathbf{M}_c & -\mathbf{M}_s \\ \mathbf{M}_s & \mathbf{M}_c \end{bmatrix} \begin{bmatrix} \underline{z}_c \\ \underline{z}_s \end{bmatrix} \\ &= [\underline{z}_c^T\ \underline{z}_s^T] \begin{bmatrix} \mathbf{M}_{cc} + \mathbf{M}_{ss} & -\mathbf{M}_{sc} + \mathbf{M}_{sc}^T \\ \mathbf{M}_{sc} - \mathbf{M}_{sc}^T & \mathbf{M}_{cc} + \mathbf{M}_{ss} \end{bmatrix} \begin{bmatrix} \underline{z}_c \\ \underline{z}_s \end{bmatrix}, \end{aligned} \tag{2.A.4}$$

by definition of $\mathbf{M}$. Comparing (2.A.4) to (2.A.2) shows that (2.A.3) gives the necessary and sufficient conditions for the two quadratic forms to be identical. But $\mathcal{E} \equiv \mathcal{E}'$ shows that $\mathbf{M}$ is positive (semi)definite if and only if this is also true for $\mathbf{\Psi}$. □

Proof of Theorem 2.1: We first prove the direct part for $\underline{m} = \underline{0}$. Recalling that any covariance matrix $\mathbf{\Lambda}$ is positive semidefinite, we see that $\det(\mathbf{\Lambda}) \neq 0$ implies that $\mathbf{\Lambda}$ is in fact positive definite. Thus, $\mathbf{\Phi}$ defined by (2.10) is positive definite by Lemma 2.A.1. Since $\underline{Z}_c$ and $\underline{Z}_s$ are jointly Gaussian,

$$p_{\underline{Z}_c \underline{Z}_s}(\underline{z}_c, \underline{z}_s) = \frac{1}{(2\pi)^n \sqrt{\det \mathbf{\Phi}}} \exp\left\{ -\tfrac{1}{2} \left[\underline{z}_c^T \; \underline{z}_s^T \right] \mathbf{\Phi}^{-1} \begin{bmatrix} \underline{z}_c \\ \underline{z}_s \end{bmatrix} \right\}. \tag{2.A.5}$$

We now show that the exponents of (2.11) and (2.A.5) are equal. Using a standard result for inverting block matrices [42, p. 656] and the properness of $\underline{Z}$, which implies $\mathbf{\Lambda}_{cc} = \mathbf{\Lambda}_{ss}$ and $\mathbf{\Lambda}_{cs} = \mathbf{\Lambda}_{sc}^T = -\mathbf{\Lambda}_{sc}$, we obtain

$$\mathbf{\Phi}^{-1} = \begin{bmatrix} \mathbf{\Delta}^{-1} & \mathbf{\Lambda}_{cc}^{-1} \mathbf{\Lambda}_{sc} \mathbf{\Delta}^{-1} \\ -\mathbf{\Delta}^{-1} \mathbf{\Lambda}_{sc} \mathbf{\Lambda}_{cc}^{-1} & \mathbf{\Delta}^{-1} \end{bmatrix}, \tag{2.A.6}$$

where $\mathbf{\Delta}$, defined by (2.12), is symmetric. Note that $\mathbf{\Phi}^{-1}$ is nonsingular since $\mathbf{\Phi}$ is nonsingular, which implies that $\mathbf{\Lambda}_{cc}^{-1}$ and $\mathbf{\Delta}^{-1}$ exist. Moreover, $\mathbf{\Phi}^{-1}$ is symmetric, since the inverse of a symmetric matrix is symmetric. Next, we show that the upper-right block of $\mathbf{\Phi}^{-1}$ is skew-symmetric. Observing that

$$\mathbf{\Delta}\, \mathbf{\Lambda}_{cc}^{-1}\, \mathbf{\Lambda}_{sc} = \mathbf{\Lambda}_{sc} + \mathbf{\Lambda}_{sc}\, \mathbf{\Lambda}_{cc}^{-1}\, \mathbf{\Lambda}_{sc}\, \mathbf{\Lambda}_{cc}^{-1}\, \mathbf{\Lambda}_{sc} = \mathbf{\Lambda}_{sc}\, \mathbf{\Lambda}_{cc}^{-1}\, \mathbf{\Delta},$$

we obtain

$$\mathbf{\Lambda}_{cc}^{-1}\, \mathbf{\Lambda}_{sc}\, \mathbf{\Delta}^{-1} = \mathbf{\Delta}^{-1}\, \mathbf{\Lambda}_{sc}\, \mathbf{\Lambda}_{cc}^{-1} = (\mathbf{\Lambda}_{cc}^{-1}\, \mathbf{\Lambda}_{sc}^T\, \mathbf{\Delta}^{-1})^T = -(\mathbf{\Lambda}_{cc}^{-1}\, \mathbf{\Lambda}_{sc}\, \mathbf{\Delta}^{-1})^T,$$

i.e., the upper-right block, and thus also the lower-left block, is skew-symmetric. Thus, $\mathbf{\Phi}^{-1}$ has the same properties as $\mathbf{\Psi}$ in (2.A.1), namely symmetry, equal diagonal blocks, and skew-symmetry of off-diagonal blocks. Therefore Lemma 2.A.1 applies for $\mathbf{\Psi} \triangleq \frac{1}{2}\, \mathbf{\Phi}^{-1}$ and $\mathbf{M} \triangleq \frac{1}{2}\, \mathbf{\Delta}^{-1} \left(\mathbf{I} - j\, \mathbf{\Lambda}_{sc}\, \mathbf{\Lambda}_{cc}^{-1} \right)$. By the properness of $\underline{Z}$, $\mathbf{\Lambda}$ is given as in (2.12). Multiplying out $\mathbf{M}\, \mathbf{\Lambda}$ yields the identity matrix. Therefore $\mathbf{M} = \mathbf{\Lambda}^{-1}$ and the exponents of (2.11) and (2.A.5) are equal. It remains to show that $2^n \sqrt{\det(\mathbf{\Phi})} = \det(\mathbf{\Lambda})$. Note that the determinant of a Hermitian

matrix is always real. Using a well-known result for the determinant of block matrices (cf. [42, p. 650] or [38, p. 46]) and the fact that $\mathbf{\Lambda}_{cs} = -\mathbf{\Lambda}_{sc}$, we obtain

$$\det(\mathbf{\Phi}) = \det(\mathbf{\Lambda}_{cc}) \det(\mathbf{\Delta}). \tag{2.A.7}$$

We now determine $\det(\mathbf{\Lambda})$. Note that $\mathbf{\Lambda}^T = 2\,(\mathbf{\Lambda}_{cc} - j\,\mathbf{\Lambda}_{sc}) = 2\,(\mathbf{I} - j\,\mathbf{\Lambda}_{sc}\,\mathbf{\Lambda}_{cc}^{-1})\,\mathbf{\Lambda}_{cc}$. Therefore $\mathbf{\Lambda}^{-1} = \frac{1}{4}\,\mathbf{\Delta}^{-1}\,\mathbf{\Lambda}^T\,\mathbf{\Lambda}_{cc}^{-1}$. But

$$1 = \det\left(\mathbf{\Lambda}\,\mathbf{\Lambda}^{-1}\right) = \det\left(\tfrac{1}{4}\,\mathbf{\Lambda}\,\mathbf{\Delta}^{-1}\,\mathbf{\Lambda}^T\,\mathbf{\Lambda}_{cc}^{-1}\right) = \frac{[\det(\mathbf{\Lambda})]^2}{2^{2n}\,\det(\mathbf{\Delta})\,\det(\mathbf{\Lambda}_{cc})}. \tag{2.A.8}$$

Combining (2.A.8) and (2.A.7) yields

$$\det(\mathbf{\Lambda}) = 2^n\,\sqrt{\det(\mathbf{\Lambda}_{cc})\,\det(\mathbf{\Delta})} = 2^n\,\sqrt{\det(\mathbf{\Phi})}. \tag{2.A.9}$$

Now let $\underline{Z}$ have nonzero mean $\underline{m}$. Then $\underline{Z} - \underline{m}$ is zero-mean Gaussian and has the p.d.f. (2.11).

We now turn to the converse part. Since $\mathbf{\Lambda}$ is positive definite, so also is $\mathbf{M} = \mathbf{\Lambda}^{-1}$. According to Lemma 2.A.1, there exists a unique symmetric, positive-definite matrix $\mathbf{\Psi}$ such that in the case $\underline{m} = \underline{0}$ one has $\mathcal{E} = \mathcal{E}'$ for all $\underline{z} \triangleq \underline{z}_c + j\,\underline{z}_s$. In the words of Feller [31, p. 84], $\mathbf{\Psi}$ induces a normal density in $2n$ dimensions. Thus, the composite vector $[\underline{Z}_c^T\,,\,\underline{Z}_s^T]^T$ is Gaussian with mean $[\underline{m}_c^T\,,\,\underline{m}_s^T]^T$ and covariance matrix $\mathbf{\Phi} = \frac{1}{2}\,\mathbf{\Psi}^{-1}$. By Lemma 2.A.1, $\mathbf{\Psi}$ has equal diagonal blocks and skew-symmetric off-diagonal blocks, and by the argument in the direct part of the proof, the matrix $\mathbf{\Phi}$ enjoys the same properties. This implies the properness of $\underline{Z}$ and the claim follows. □

Leer - Vide - Empty

Chapter 3

Capacity and Information Rates of Channels with Intersymbol Interference and White Gaussian Noise

3.1 Capacity of the Intersymbol-Interference Channel with AWGN - A Simplified Derivation

Hirt and Massey [9] recently computed the capacity C of the intersymbol-interference (ISI) channel with AWGN or, to use their terminology, of the *discrete-time Gaussian channel* (DTGC), assuming finite ISI and an average symbol-energy constraint. Their derivation was based on a hypothetical channel model, the *N-circular Gaussian channel* (NCGC). Using a real version of the DFT, they showed the equivalence of the NCGC to a set of N parallel, decoupled memoryless channels. The per-symbol capacity C_N of the NCGC was then obtained using the 'water-filling theorem' [10, Thm 7.5.1]. Moreover, they proved that the DTGC and the NCGC are asymptotically equivalent in the sense that

$$C = \lim_{N \to \infty} C_N. \tag{3.1}$$

Verdú [43] independently used a circular convolution approach to determine the capacity region of the symbol-asynchronous Gaussian multiple-

access channel.

As an application of proper complex random variables, Hirt and Massey's derivation is simplified in this section and their results are generalized to channels with proper complex AWGN and a complex unit-sample response. A similar approach can be used to simplify the computation of capacity of Gaussian multiple-access channels with memory [44] as well as their complex generalizations. The notation is as in Section 2.3. We first consider a *real* DTGC whose channel filter has a real unit-sample response $(h_0, h_1, \ldots, h_M)$ and assume further that $h_0 \neq 0$ and $h_M \neq 0$. Consider now that one has available two instances of this DTGC, viz.,

$$y_{cn} = \sum_{m=0}^{M} h_m \, x_{cn-m} + w_{cn}, \quad -\infty < n < \infty, \tag{3.2}$$

and

$$y_{sn} = \sum_{m=0}^{M} h_m \, x_{sn-m} + w_{sn}, \quad -\infty < n < \infty, \tag{3.3}$$

where $\{w_{cn}\}$ and $\{w_{sn}\}$ are independent zero-mean white Gaussian noise (WGN) sequences, each sample of which has variance $N_0/2$, and where the real inputs are subject to the *symbol-energy constraints*

$$\mathrm{E}\left[x_{cn}^2\right] \leq E_s/2 \quad \text{and} \quad \mathrm{E}\left[x_{sn}^2\right] \leq E_s/2, \quad -\infty < n < \infty. \tag{3.4}$$

This pair of real DTGC's can be represented by the (one-dimensional) complex (or two-dimensional real) channel

$$y_n = \sum_{m=0}^{M} h_m \, x_{n-m} + w_n, \quad -\infty < n < \infty, \tag{3.5}$$

where $x_n \triangleq x_{cn} + j\, x_{sn}$, $w_n \triangleq w_{cn} + j\, w_{sn}$, and $y_n \triangleq y_{cn} + j\, y_{sn}$. Since $\{w_{cn}\}$ and $\{w_{sn}\}$ have the same autocorrelation function, a vanishing crosscorrelation function and zero means, it follows that $\{w_n\}$ is a *proper complex* WGN sequence. Moreover, $\mathrm{E}[w_n] = 0$ and $\mathrm{E}\left[|w_n|^2\right] = N_0$, for all n. The constraints (3.4) are now replaced by the weaker condition

$$\mathrm{E}\left[|x_n|^2\right] \leq E_s, \quad -\infty < n < \infty.$$

Clearly, capacity can be achieved on the channel (3.5) by independent sequences $\{x_{cn}\}$ and $\{x_{sn}\}$, since a real unit-sample response produces no 'crosstalk' between the real and imaginary component channel and because the real and imaginary noise sequence are independent. If the capacity-achieving input distribution also satisfies (3.4), then $C^{2D} = 2\,C^{1D}$, where C^{1D} and C^{2D} are the capacities of the one-dimensional real channel (3.2) [or (3.3)] and the two-dimensional real (or complex) channel (3.5), respectively. We obtain additional generality by allowing the unit-sample response in (3.5) to be complex. The resulting channel (3.5) will be called the *complex* DTGC. Following [9], we define the *complex* NCGC[1] by

$$y[n] = \sum_{i=0}^{N-1} h[i]\, x[n-i] + w[n], \quad 0 \le n < N, \tag{3.6}$$

where $N > M$, where the sequence $h[0, N-1]$ is obtained by padding $h_0, h_1, \ldots, h_M$ with zeros in the manner

$$h[i] \triangleq \begin{cases} h_i, & \text{if } 0 \le i \le M \\ 0, & \text{if } M < i < N \end{cases},$$

and where $w[0, N-1]$ is proper complex c.w.s.s. Gaussian noise with zero mean and circular correlation sequence

$$r_w[i] = \mathrm{E}\,[\,w[n+i]\,w^*[n]\,] = N_0\,\delta[i], \quad 0 \le i < N.$$

For brevity, $w[0, N-1]$ will be called a *proper complex WGN sequence.* Moreover, the input data $x[0, N-1]$ are subject to the *symbol-energy constraint*

$$\mathrm{E}\,[\,|x[n]|^2\,] \le E_s, \quad 0 \le n < N. \tag{3.7}$$

It can be easily shown that the complex DTGC and the complex NCGC are asymptotically equivalent in the sense of (3.1) by essentially the same argument as given in [9].

Theorem 3.1: The per-symbol capacity of the complex NCGC (3.6) under the symbol-energy constraint (3.7) is given by

$$C_N^{2D} = \frac{1}{N} \sum_{k=0}^{N-1} \log\left[\max\left(\beta\,|H[k]|^2/N_0\,,\ 1\right)\right], \tag{3.8}$$

[1] As in Section 2.3, a length-N sequence $x[0], x[1], \ldots, x[N-1]$ is written in this section as $x[0, N-1]$ and all indices in square brackets are understood to be taken modulo N.

where $H[0, N-1]$ is the DFT of $h[0, N-1]$ and where the parameter β is determined from the condition

$$\sum_{k=0}^{N-1} \epsilon[k] = N E_s \tag{3.9}$$

in which the *spectral energy distribution* $\epsilon[0, N-1]$ depends on β through

$$\epsilon[k] = \max\left(\beta - N_0/|H[k]|^2\,,\, 0\right), \quad 0 \le k < N. \tag{3.10}$$

Moreover, capacity is achieved if and only if the input sequence $x[0, N-1]$ is proper, Gaussian, and c.w.s.s. with zero mean and circular correlation sequence equal to the IDFT of $\epsilon[0, N-1]$, i.e.,

$$r_x[i] \triangleq \mathrm{E}\left[x[n+i]\, x^*[n]\right] = \frac{1}{N} \sum_{k=0}^{N-1} \epsilon[k]\, \Omega_N^{ik}, \quad 0 \le i < N. \tag{3.11}$$

Notice that the spectral energy distribution $\epsilon[0, N-1]$ determined by (3.9) and (3.10) has the water-filling interpretation given in [3, p. 169] and [10, p. 389].

Proof: It was proved in [9] that the capacity of the (real) NCGC equals the supremum of the (average) mutual information between the input and output sequence over all p.d.f.'s satisfying a weaker block-energy constraint. Analogously, we will show for the complex NCGC that

$$C_N^{2D} = I_N^{2D}, \tag{3.12}$$

where

$$I_N^{2D} \triangleq \sup_{p_N} \frac{1}{N} I(x[0, N-1]\,;\, y[0, N-1]),$$

and where the supremum is over all p.d.f.'s p_N for $x[0, N-1]$ satisfying the *block-energy constraint*

$$\sum_{n=0}^{N-1} \mathrm{E}\left[|x[n]|^2\right] \le N E_s. \tag{3.13}$$

Because (3.7) implies (3.13), it suffices to show that the maximizing p_N under the constraint (3.13) also satisfies the stronger constraint (3.7).

Taking the DFT of (3.6) yields a set of N parallel, *memoryless Gaussian channels* (MGC's) described by

$$Y[k] = H[k]\,X[k] + W[k], \qquad 0 \le k < N, \tag{3.14}$$

where $H[0,N-1]$ is the DFT of $h[0,N-1]$ and where the $W[k]$ are i.i.d.[2] proper complex Gaussian random variables with zero mean and variance[3] NN_0 by Corollary 2.2 and (2.28). Using Parseval's relation [14], the constraint (3.13) becomes

$$\sum_{k=0}^{N-1} \mathrm{E}\left[\,|X[k]|^2\,\right] \le N^2 E_s \tag{3.15}$$

in the frequency domain. By (3.13) and the invertibility of the DFT,

$$I_N^{2D} = \sup_{q_N} \frac{1}{N}\, I(X[0,N-1]\,;\, Y[0,N-1]), \tag{3.16}$$

where the supremum is over all p.d.f.'s q_N for $X[0,N-1]$ satisfying (3.15). By a standard information-theoretic argument,

$$I(X[0,N-1]\,;\, Y[0,N-1]) \le \sum_{k=0}^{N-1} I(X[k]\,;\, Y[k]) \tag{3.17}$$

with equality if and only if the *outputs* $Y[k]$ are independent [10, p. 321]. By using (3.17), (3.16) can be written as a supremum over the allowed spectral energy distributions $\epsilon[0,N-1]$, viz.,

$$I_N^{2D} = \sup_{\substack{\epsilon[0,N-1]:\\ \sum_{i=0}^{N-1}\epsilon[i] \le NE_s}} \frac{1}{N} \sum_{k=0}^{N-1} C[k], \tag{3.18}$$

where

$$C[k] \triangleq \sup_{\substack{q[k]:\\ \mathrm{E}\left[\,|X[k]|^2\,\right] \le N\epsilon[k]}} I(X[k]\,;\, Y[k]) \tag{3.19}$$

is the capacity of the k-th MGC (3.14). The average energy at the output of this MGC is bounded by

$$\mathrm{E}\left[\,|Y[k]|^2\,\right] \le N\left(|H[k]|^2\epsilon[k] + N_0\right) = S[k]. \tag{3.20}$$

[2] independent, identically distributed

[3] In accordance with (2.3), the *variance* of a scalar complex random variable is defined as $\mathrm{Var}[X] \triangleq \mathrm{E}\left[\,|X - m_X|^2\,\right]$.

When $H[k] = 0$, equality holds trivially in (3.20); when $H[k] \neq 0$, equality holds in (3.20) if and only if $\mathrm{E}\left[\,|X[k]|^2\,\right] = N\epsilon[k]$.

It follows from the maximum-entropy theorem (Theorem 2.2) that

$$I(X[k]\,;\,Y[k]) = h(Y[k]) - h(Y[k]\,|\,X[k]) \leq \log\left[S[k]/(N\,N_0)\right] \quad (3.21)$$

with equality if and only if $Y[k]$ is proper and Gaussian with zero mean and variance $S[k]$. When $H[k] = 0$, equality holds trivially in (3.21). Now consider the case when $H[k] \neq 0$. According to Cramèr's Theorem [31], which states that the sum of two independent random variables is Gaussian if and only if each of the two random variables is itself Gaussian, and by Lemma 2.4, equality holds in (3.21) for $H[k] \neq 0$ if and only if $X[k]$ is proper and Gaussian with zero mean and variance $N\epsilon[k]$. Therefore,

$$C[k] = \log\left[1 + |H[k]|^2\,\epsilon[k]/N_0\right]. \quad (3.22)$$

To complete the maximization in (3.18), it remains to choose $\epsilon[0, N-1]$ so as to maximize the sum $\sum_{k=0}^{N-1} C[k]$ under the equality constraint $\sum_{k=0}^{N-1} \epsilon[k] = NE_s$. The solution to this maximization problem can be adopted from [10] without change and yields (3.8)-(3.10). Since capacity is achieved only for proper Gaussian inputs $X[k]$ and since the noise samples $W[k]$ are independent, the necessary and sufficient condition for equality in (3.17) is equivalent to the independence of the *inputs* $X[k]$. Thus, capacity is achieved if and only if the inputs $X[k]$ are independent, proper, and Gaussian with zero mean and variance $N\,\epsilon[k]$. Invoking Theorem 2.4 once more shows that capacity is achieved if and only if the input sequence $x[0, N-1]$ is proper, Gaussian, and c.w.s.s. with zero mean and circular correlation sequence (3.11). It follows from $r_x[0] = \mathrm{E}\left[\,|x[n]|^2\,\right] = E_s$, $0 \leq n < N$, that the maximizing p.d.f. for the block-energy constraint (3.13) also satisfies the symbol-energy constraint (3.7), which confirms (3.12). □

We now return to the special case of the real NCGC

$$v[n] = \sum_{i=0}^{N-1} h[i]\,u[n-i] + z[n], \quad 0 \leq n < N, \quad (3.23)$$

treated in [9], where the unit-sample response and all random variables are real. The noise sequence $z[0, N-1]$ is assumed to be white Gaussian with zero mean and energy

$$\mathrm{E}\left[\,(z[n])^2\,\right] = N_0/2, \quad 0 \leq n < N,$$

and the inputs are subject to the *symbol-energy constraint*

$$\mathrm{E}\left[(u[n])^2\right] \le E_s/2, \quad 0 \le n < N. \tag{3.24}$$

Corollary 3.1: The per-symbol capacity of the real NCGC (3.23) under the symbol-energy constraint (3.24) is given by

$$C_N^{1D} = C_N^{2D}/2, \tag{3.25}$$

where C_N^{2D} is given in Theorem 3.1. Moreover, capacity is achieved if and only if the input sequence $u[0, N-1]$ is Gaussian and c.w.s.s. with zero mean and circular correlation sequence equal to

$$r_u[i] \triangleq \mathrm{E}\left[u[n+i]\,u[n]\right] = \frac{1}{N}\sum_{k=0}^{N-1}\frac{\epsilon[k]}{2}\cos(2\pi ik/N), \quad 0 \le i < N. \tag{3.26}$$

Note that our β is related to the parameter θ used in [9] by $\beta = \theta N_0$. Note also that our $\epsilon[k]$ is defined to be twice the "$\epsilon[k]$" defined in [9].

Proof of Corollary 3.1: Theorem 3.1 applies when the sequence $h[0, N-1]$ is real, in which case no crosstalk is produced from the real (imaginary) part of the channel input to the imaginary (real) part of the channel output. Since capacity is achieved by zero-mean proper complex inputs and since $r_x[0] = E_s$, we obtain $\mathrm{E}\left[x_c^2[n]\right] = \mathrm{E}\left[x_s^2[n]\right] = E_s/2$, $0 \le n < N$. Further, since $h[0, N-1]$ is real, $H[0]$ is also real and $H[k] = H^*[N-k]$, $1 \le k < N$ [14, p. 110]. Therefore, (3.10) yields $\epsilon[k] = \epsilon[N-k]$, $1 \le k < N$, and (3.11) gives

$$\begin{aligned} r_x[i] &= \frac{1}{N}\left[\epsilon[0] + \sum_{k=1}^{N-1}\epsilon[k]\,\Omega_N^{ik}\right] \\ &= \frac{1}{N}\left[\epsilon[0] + \frac{1}{2}\left(\sum_{k=1}^{N-1}\epsilon[k]\,\Omega_N^{ik} + \sum_{k=1}^{N-1}\epsilon[N-k]\,\Omega_N^{ik}\right)\right] \\ &= \frac{1}{N}\sum_{k=0}^{N-1}\epsilon[k]\cos(2\pi ik/N), \quad 0 \le i < N. \end{aligned}$$

Since the capacity-achieving inputs are proper, Gaussian and zero-mean and since the circular correlation sequence is real, the sequences $x_c[0, N-1]$ and $x_s[0, N-1]$ are independent[4] and $r_u[i] = r_{x_c}[i] =$

[4] cf. Example 2.2.

$r_{x_s}[i] = r_x[i]/2, \quad 0 \leq i < N$. We have thus shown that the complex NCGC (3.6) under the constraint (3.7) and used with the capacity-achieving input distribution reduces to a pair of independent real NCGC's (3.23) on each of which the constraint (3.24) is satisfied with equality. Hence, $2\,C_N^{1D}$ is lower-bounded by C_N^{2D}.

Conversely, two instances of the real NCGC (3.23) under constraint (3.24) can be represented by a complex NCGC under constraint (3.7) so that $2\,C_N^{1D}$ is upper-bounded by, and therefore equal to, C_N^{2D}. □

3.2 On a Lower Bound for the Information Rate of Intersymbol-Interference Channels with i.i.d. Inputs

Shamai, Ozarow, and Wyner [25] recently obtained a lower bound for the information rate that can be achieved with i.i.d.[5] inputs on intersymbol-interference (ISI) channels with AWGN. In this section, we present a new derivation of their result, avoiding the inversion of the channel transfer function and the subsequent innovation argument as well as the use of asymptotic properties of Toeplitz matrices that were employed in the original proof. The new derivation is based on the information-theoretic equivalence of certain allpass-transformed ISI channels. We are interested in both the real and the complex version of the ISI channel with AWGN, both of which can be described formally in the same way by

$$Y_n = \sum_{m=0}^{\infty} h_m X_{n-m} + W_n, \quad -\infty < n < \infty. \tag{3.27}$$

The unit-sample response $\{h_m\}$ is assumed to be causal and delayless[6]. Its z-transform $\sum_{m=-\infty}^{\infty} h_m z^{-m}$ will be called the *channel filter*. The *real* channel is characterized by a real input sequence $\{X_n\}$, a real unit-sample response $(h_0, h_1, \ldots)$, and a real AWGN process $\{W_n\}$, whereas the *complex* channel has complex inputs, a complex unit-sample response, and *proper* complex AWGN. For the complex channel, the inputs may still be real as, e.g., with binary antipodal signaling. For both versions of the channel, we assume a finite-energy unit-sample response and a zero-mean noise process with sample variance σ_W^2. The noise sample variance σ_W^2 is given by $N_0/2$ and N_0 for the real channel and the complex channel, respectively.

As in Section 2.2, we will write $X \sim \mathcal{N}(m, \sigma^2)$ [or $X \sim \mathcal{N}_p(m, \sigma^2)$] to specify a Gaussian [or proper complex Gaussian] random variable with mean m and variance σ^2. Analogous notation will be used for random vectors. In this section, we will use the terminology 'probability function' to stand for a probability *mass* function $P_X(.)$ in the case of a discrete random variable X and for a probability *density* function $p_X(.)$ in the case of a continuous random variable X.

[5] independent, identically distributed

[6] A sequence $\{h_m\}$ is called *causal* if $h_m = 0$ for $m < 0$. Moreover, a causal sequence is called *delayless* if $h_0 \neq 0$.

For channels with finite memory M and unbounded memory, respectively, we define the *information rate for i.i.d. inputs* governed by a probability function $P_X(.)$ or $p_X(.)$ as

$$I_{\text{i.i.d.}} \triangleq \lim_{N\to\infty} \frac{1}{N} I(X_0X_1\ldots X_{N-1}; Y_0Y_1\ldots Y_{N+M-1}) \quad (3.28)$$

and

$$I_{\text{i.i.d.}} \triangleq \lim_{N\to\infty} \lim_{\mu\to\infty} \frac{1}{N} I(X_0X_1\ldots X_{N-1}; Y_0Y_1\ldots Y_{N+\mu-1}), \quad (3.29)$$

where the inputs $X_0, X_1, \ldots, X_{N-1}$ are assumed to be preceded and followed by an all-zero sequence[7]. The lower bound of Shamai, Ozarow, and Wyner [25, Thm. 1], generalized to include the complex version of the channel, can be stated as follows.

Theorem 3.2: Let the real [or complex] ISI channel with AWGN (3.27) have a channel filter $H(z) = N(z)/D(z)$ whose numerator and denominator, respectively, are given as $N(z) = h_0 \prod_{n=1}^{n_z}(1-z_nz^{-1})$ and $D(z) = \prod_{n=1}^{n_p}(1-p_nz^{-1})$, where $h_0 \neq 0$, $0 < |z_n| < \infty$ for $1 \leq n \leq n_z$, $0 < |p_n| < 1$ for $1 \leq n \leq n_p$, and where the region of convergence of $H(z)$ is $\{z : |z| > \max_{1\leq n\leq n_p} |p_n|\}$. Then the information rate (3.28) (when $n_p = 0$) or (3.29) (when $n_p > 0$) for i.i.d. inputs with a probability function $P_X(.)$ or $p_X(.)$ satisfies

$$I_{\text{i.i.d.}} \geq I_L \triangleq I(X\,;\,\rho X + W), \quad (3.30)$$

where X is a real [or complex] random variable with probability function $P_X(.)$ or $p_X(.)$, where $W \sim \mathcal{N}(0,\sigma_W^2)$ [or $W \sim \mathcal{N}_p(0,\sigma_W^2)$ for the complex channel], and where the *degradation factor* ρ is given by

$$\begin{aligned}\rho = |h_0| \cdot \prod_{m:\,|z_m|>1} |z_m| &= \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|H(e^{j\Omega})\right| d\Omega\right\} \\ &= \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|N(e^{j\Omega})\right| d\Omega\right\}.\end{aligned} \quad (3.31)$$

[7] It was shown by Gallager [10] that, in the limit as $N \to \infty$, such a constraint has no influence on the information rate.

Remarks on Theorem 3.2:

- The assumptions on the poles p_n and on the region of convergence of $H(z)$ imply that $H(z)$ is stable [14] and that it is the z-transform of a causal sequence $\{h_m\}$. Moreover, $\lim_{z\to\infty} H(z) = h_0 \neq 0$ implies that $\{h_m\}$ is delayless.
- It seems somewhat remarkable that knowledge either of $|h_0|$ and the magnitudes of the zeros of $H(z)$ outside the unit circle or of the amplitude response $|N(e^{j\Omega})|$ is sufficient for the computation of the degradation factor ρ.
- The first expression for the degradation factor ρ in (3.31), which was not given in [25, Thm. 1], considerably simplifies its computation. From the second and third expression in (3.31), it follows that ρ is invariant to an allpass transformation of the transfer function $H(z)$ and is independent of $D(z)$. Thus, the first expression for ρ in (3.31) implies that ρ equals $|h_{\min 0}|$, the magnitude of the leading coefficient of the *minimum-phase version* [14] $H_{\min}(z)$ of $H(z)$, which is obtained from $H(z)$ by replacing every zero z_m outside the unit circle with a zero $1/z_m^*$. Therefore, $\rho = |h_0|$ if $H(z)$ *is* minimum-phase.
- As illustrated in Figure 3.1, the lower bound I_L in Theorem 3.2 can be interpreted as the mutual information

$$I(X_n\,;\,Y_n') = I(X_n\,;\,h_{\min 0}\,X_n + W_n),$$

where Y_n' is the output of a memoryless AWGN channel with input X_n and gain equal to the leading coefficient $h_{\min 0}$ of the minimum-phase version $H_{\min}(z)$ of $H(z)$. This memoryless channel can be thought of as created by the combination of an ISI channel with channel filter $H_{\min}(z) = h_{\min 0} + z^{-1}\,G(z)$ and additive noise $W_n \sim \mathcal{N}(0,\sigma_W^2)$ [or $W_n \sim \mathcal{N}_p(0,\sigma_W^2)$ for the complex channel] with a subsequent 'correct-decision-feedback receiver' in which the ISI is canceled by subtracting the output of a filter $G(z)$ from Y_n when this filter is fed with the true past data X_{n-1} that are assumed to be provided by a 'magic genie'. Figure 3.1 also indicates that the memoryless channel from X_n to Y_n' can be approximated by a realizable decision-feedback 'equalizer' (upper switch position) in which a hard decision $\hat{X}_{n-1}$ is substituted for X_{n-1} (cf. [25, p. 1529]). ———

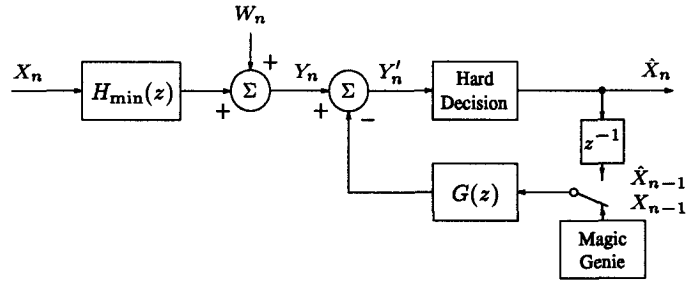


Figure 3.1: Interpretation of the lower bound on $I_{\text{i.i.d.}}$ in Theorem 3.2

Example 3.1: Consider an ISI channel with AWGN whose normalized (to energy 1) unit-sample response is

$$\begin{aligned} \underline{h} &= (1, -1, 3, -1, 1)/\sqrt{13} \\ &= (0.2774, -0.2774, 0.8321, -0.2774, 0.2774). \end{aligned}$$

The corresponding transfer function $H(z)$ is not minimum-phase (it is in fact linear-phase [14], as follows from the symmetry of $\underline{h}$ about its middle coefficient) and has the zeros $z_{1,2} = 0.3516 \pm j\ 1.4985$ (outside the unit circle) and $z_{3,4} = 0.1484 \pm j\ 0.6325$ (inside the unit circle), where $z_3 = 1/z_1^*$ and $z_4 = 1/z_2^*$. From the first expression for the degradation factor in (3.31), one gets $\rho = |z_1|\,|z_2|\,/\sqrt{13} = 0.6571$. The same result can be obtained by first converting $H(z)$ to its minimum-phase version

$$H_{\min}(z) = \left(|z_1|\,|z_2|\,/\sqrt{13}\right)\left[1 - (z_3 + z_4)\,z^{-1} + z_3 z_4\,z^{-2}\right]^2,$$

which corresponds to

$$\underline{h}_{\min} = (0.6571, -0.3900, 0.6126, -0.1646, 0.1171)$$

so that $\rho = |h_{\min 0}|$, as required. ———

In Section 3.2.1, we prove that, for the ISI channel with AWGN, the mutual information between a finite-length input block and the relevant channel outputs is the same for all transfer functions in a certain equivalence class, regardless of the input distribution. An alternative proof of this result for channels with finite memory is provided in Appendix 3.A.

The results of Section 3.2.1 will be used in Section 3.2.2 for the proof of Theorem 3.2. Appendix 3.B contains some useful properties of allpass filters and related matrices, while Appendix 3.C contains Jensen's integral formula that will be used for proving Theorem 3.2.

3.2.1 Allpass Filters and Equivalent Intersymbol-Interference Channels

The amplitude response $|H(e^{j\Omega})|$ of a channel filter $H(z)$ with n_z zeros is preserved upon replacing any zero z_n of $H(z)$ by $1/z_n^*$, i.e., when a zero is reflected at the unit circle. If $H(z)$ has n_1 zeros on the unit circle, any subset of the remaining zeros may be reflected at the unit circle, which yields $2^{n_z-n_1}$ transfer functions with the same amplitude response. For a nonzero constant c_0, for a possibly empty set of zeros

$$\{\zeta_n : 0 < |\zeta_n| < \infty,\ 1 \le n \le n_z\},$$

for an $n_p \ge 0$, and for a denominator $D(z) = 1+d_1 z^{-1}+\ldots+d_{n_p} z^{-n_p}$ all of whose zeros are inside the unit circle, we define the transfer-function *equivalence class*

$$\mathcal{H} = \left\{ c_0 \frac{A(z)}{D(z)} \prod_{n=1}^{n_z} (1-\zeta_n z^{-1}) :\ A(z) = \alpha \prod_{\substack{n:\, 1\le n\le n_z,\\ |\zeta_n|\neq 1}} \left(\frac{\zeta_n^* - z^{-1}}{1-\zeta_n z^{-1}} \right)^{i_n}, \right.$$

$$\left. \text{where } |\alpha| = 1 \quad \text{and} \quad i_n \in \{0,1\} \right\}. \tag{3.32}$$

Here and hereafter, the convergence regions of the transfer functions in $\mathcal{H}$ are taken as $\{z : |z| > R_{\max}\}$, where $R_{\max}$ is the largest magnitude of a zero of $D(z)$. Note that $A(z)$ in (3.32) is an allpass filter (cf. Appendix 3.B) associated with the subset of zeros not on the unit circle and that the indicators i_n specify whether or not a zero ζ_n is reflected at the unit circle. Transfer functions in the same equivalence class have the same amplitude response. For $n_p = 0$ and for $M \triangleq n_z$, a transfer function $H(z)$ in $\mathcal{H}$ corresponds to a finite-length unit-sample response $(h_0, h_1, \ldots, h_M)$ with $h_0 \neq 0$ and $h_M \neq 0$ and is called an *M-th order FIR filter* [14]. In this case, $\mathcal{H}$ will be called an *M-th order FIR-filter equivalence class.*

The following two theorems for channels with finite and infinite memory, respectively, show that ISI channels with channel filters in the same equivalence class and with AWGN processes having the same sample variance are equivalent with respect to mutual information. It should be remembered, however, that two channels equivalent with respect to mutual information might not be equally convenient for coding and decoding!

Theorem 3.3: For the class of real [or complex] ISI channels (3.27) with a channel filter $H(z)$ in an M-th order FIR-filter equivalence class $\mathcal{H}$ and with AWGN, for any $N \geq 1$ and for *any* (fixed) *probability function* for the real [or complex] input sequence $(X_0, X_1, \ldots, X_{N-1})$ (which is assumed to be preceded and followed by all-zeros),

$$I(X_0\, X_1 \ldots X_{N-1}\, ;\, Y_0\, Y_1 \ldots Y_{N+M-1})$$

is the same for all $H(z)$ in $\mathcal{H}$. ———

For ISI channels with an M-th order FIR channel filter and AWGN, i.e., for ISI channels with finite memory M, note that the outputs Y_0, Y_1, ..., Y_{N+M-1} are the only relevant observations. All other outputs are not affected by the inputs and do not depend on the noise components of $Y_0, Y_1, \ldots, Y_{N+M-1}$.

Given the fact that the capacity of an ISI channel with AWGN is preserved under an allpass transformation of the channel output, one might have expected that the per-symbol mutual information is preserved in the limit as N approaches infinity for (asymptotically) stationary Gaussian inputs. However, it is somewhat surprising that mutual information is preserved for *any* input probability function and - although the allpass filters involved may have unbounded memory - for *finite-length sequences.*

Proof of Theorem 3.3: Consider the transfer functions $H(z)$ and $\tilde{H}(z) = A(z)H(z)$ in $\mathcal{H}$, where $A(z) = \sum_{k=-\infty}^{\infty} a_k\, z^{-k}$ is an allpass filter as specified in (3.32). Let $\{Y_n\}$ be the output sequence of the channel with channel filter $H(z)$ for some finite-length input $(X_0, X_1, \ldots, X_{N-1})$ and let $\tilde{Y}_n = \sum_{k=-\infty}^{\infty} a_k\, Y_{n-k}$. Then, $\{\tilde{Y}_n\}$ can be interpreted as the output sequence of a channel with channel filter $\tilde{H}(z)$ for the same input sequence. Starting from the channel with channel

filter $H(z)$ gives

$$\begin{aligned} & I(X_0 X_1 \ldots X_{N-1}\,;\, Y_0 Y_1 \ldots Y_{N+M-1}) \\ &= \lim_{\mu\to\infty} I(X_0 X_1 \ldots X_{N-1}\,;\, Y_{-\mu} Y_{-\mu+1} \ldots Y_{N+M+\mu-1}) \\ &= \lim_{\mu\to\infty} I(X_0 X_1 \ldots X_{N-1}\,;\, \tilde{Y}_{-\mu} \tilde{Y}_{-\mu+1} \ldots \tilde{Y}_{N+M+\mu-1}) \\ &= I(X_0 X_1 \ldots X_{N-1}\,;\, \tilde{Y}_0 \tilde{Y}_1 \ldots \tilde{Y}_{N+M-1}). \end{aligned} \tag{3.33}$$

For any $\mu > 0$, note that the observations $Y_{-\mu} \ldots Y_{-1}$ and $Y_{N+M} \ldots Y_{N+M+\mu-1}$ are irrelevant since $H(z)$ corresponds to a causal, finite-length sequence $(h_0, h_1 \ldots h_M)$ and because the additive noise is white Gaussian. Similarly, for any $\mu > 0$, the observations $\tilde{Y}_{-\mu} \ldots \tilde{Y}_{-1}$ and $\tilde{Y}_{N+M} \ldots \tilde{Y}_{N+M+\mu-1}$ are irrelevant since $\tilde{H}(z)$ corresponds to a causal, finite-length sequence $(\tilde{h}_0, \tilde{h}_1 \ldots \tilde{h}_M)$ sequence and because the noise component of $\{\tilde{Y}_n\}$ is white Gaussian as well. The first and the last equality in (3.33) are due to the fact that adding or omitting irrelevant observations does not change mutual information and the second equality in (3.33) follows from the invertibility of the allpass filtering. □

For the reader suspicious of the addition or omission of an infinite number of (irrelevant) observations in the above proof, an alternative proof that avoids such operations is provided in Appendix 3.A.

Theorem 3.4: For the class of real [or complex] ISI channels (3.27) with a channel filter $H(z)$ in an equivalence class $\mathcal{H}$ given by (3.32) and with AWGN, for any $N \geq 1$, and for *any* (fixed) ***probability function*** for the real [or complex] input sequence $(X_0, X_1, \ldots, X_{N-1})$ (which is assumed to be preceded and followed by all-zeros),

$$\lim_{\mu\to\infty} I(X_0 X_1 \ldots X_{N-1}\,;\, Y_0 Y_1 \ldots Y_{N+\mu-1})$$

is the same for all $H(z)$ in $\mathcal{H}$. ———

It should be pointed out that the statement in Theorem 3.4 does *not* in general hold for any finite μ.

Proof of Theorem 3.4: Consider the transfer functions $H(z)$ and $\tilde{H}(z) = A(z)H(z)$ in $\mathcal{H}$, where $A(z) = \sum_{k=-\infty}^{\infty} a_k\, z^{-k}$ is an all-pass filter as specified in (3.32). Let $\{Y_n\}$ be the output sequence

of the channel with channel filter $H(z)$ for some finite-length input $(X_0, X_1, \ldots, X_{N-1})$ and let $\tilde{Y}_n = \sum_{k=-\infty}^{\infty} a_k Y_{n-k}$. Then, $\{\tilde{Y}_n\}$ can be interpreted as the output sequence of a channel with channel filter $\tilde{H}(z)$ for the same input sequence. Starting from the channel with channel filter $H(z)$ gives

$$\begin{aligned}
&\lim_{\mu\to\infty} I(X_0\, X_1 \ldots X_{N-1}\,;\, Y_0\, Y_1 \ldots Y_{N+\mu-1}) \\
&= \lim_{\mu\to\infty} I(X_0\, X_1 \ldots X_{N-1}\,;\, Y_{-\mu}\, Y_{-\mu+1} \ldots Y_{N+\mu-1}) \\
&= \lim_{\mu\to\infty} I(X_0\, X_1 \ldots X_{N-1}\,;\, \tilde{Y}_{-\mu}\, \tilde{Y}_{-\mu+1} \ldots \tilde{Y}_{N+\mu-1}) \\
&= \lim_{\mu\to\infty} I(X_0\, X_1 \ldots X_{N-1}\,;\, \tilde{Y}_0\, \tilde{Y}_1 \ldots \tilde{Y}_{N+\mu-1}). \qquad (3.34)
\end{aligned}$$

For any $\mu > 0$, note that the observations $Y_{-\mu} \ldots Y_{-1}$ are irrelevant since $H(z)$ corresponds to a causal sequence and because the additive noise is white Gaussian. Similarly, for any $\mu > 0$, the observations $\tilde{Y}_{-\mu} \ldots \tilde{Y}_{-1}$ are irrelevant since $\tilde{H}(z)$ corresponds to a causal sequence and because the noise component of $\{\tilde{Y}_n\}$ is white Gaussian as well. The first and the last equality in (3.34) are due to the fact that adding or omitting irrelevant observations does not change mutual information and the second equality in (3.34) follows from the invertibility of the allpass filtering. □

3.2.2 Proof of the Lower Bound

Theorem 3.2 will be proved by first deriving a simple and straightforward lower bound, which is then maximized over all transfer functions in the associated equivalence class.

Lemma 3.1: For the real [or complex] ISI channel (3.27) with a channel filter $H(z)$ of the form given in Theorem 3.2 and with AWGN,

$$I_{\text{i.i.d.}} \geq I(X\,;\, h_0\, X + W), \qquad (3.35)$$

where $h_0 = \lim_{z\to\infty} H(z)$, where X is a real [or complex] random variable with probability function $P_X(.)$ or $p_X(.)$ that specifies the i.i.d. input sequence, and where $W \sim \mathcal{N}(0, \sigma_W^2)$ [or $W \sim \mathcal{N}_p(0, \sigma_W^2)$ for the complex channel].

Notice that $h_0 = \lim_{z\to\infty} H(z)$ is the leading coefficient of the Laurent-series expansion of $H(z)$. Notice also that $I(X\,;\, h_0\, X + W) =$

$I(X\,;\,|h_0|X+W)$ since W has an even [or rotationally symmetric] p.d.f. for zero-mean real [or proper complex] Gaussian noise.

Proof: For a finite-length input sequence $(X_0, X_1, \ldots, X_{N-1})$ and any $\mu \geq 0$, we have

$$\begin{aligned} &I(X_0 \ldots X_{N-1}\,;\,Y_0 \ldots Y_{N+\mu-1}) \\ &\quad = h(Y_0 \ldots Y_{N+\mu-1}) - h(Y_0 \ldots Y_{N+\mu-1}\,|\,X_0 \ldots X_{N-1}) \\ &\quad = h(Y_0 \ldots Y_{N+\mu-1}) - h(W_0 \ldots W_{N+\mu-1}), \end{aligned} \tag{3.36}$$

where $h(\,.\,)$ denotes differential entropy. By the chain rule,

$$h(Y_0 \ldots Y_{N+\mu-1}) = \sum_{n=0}^{N+\mu-1} h(Y_n\,|\,Y_0 \ldots Y_{n-1}).$$

Each term in this expansion can be lower-bounded as

$$\begin{aligned} h(Y_n\,|\,Y_0 \ldots Y_{n-1}) &\geq h(Y_n\,|\,Y_0 \ldots Y_{n-1}\,,\,X_0 \ldots X_{n-1}) \\ &= h\left(\textstyle\sum_{m=0}^{n} h_m\, X_{n-m} + W_n \,\middle|\, X_0 \ldots X_{n-1}\,,\,W_0 \ldots W_{n-1}\right) \\ &= h\left(h_0\, X_n + W_n\,|\,X_0 \ldots X_{n-1}\,,\,W_0 \ldots W_{n-1}\right) \\ &= h\left(h_0\, X_n + W_n\right). \end{aligned} \tag{3.37}$$

The inequality in (3.37) follows from the fact that further conditioning cannot increase entropy. The first equality follows from (3.27) and the fact that (i) $(X_0 \ldots X_{n-1})$ and $(Y_0 \ldots Y_{n-1})$ uniquely determine $(W_0 \ldots W_{n-1})$ and (ii) $(X_0 \ldots X_{n-1})$ and $(W_0 \ldots W_{n-1})$ uniquely determine $(Y_0 \ldots Y_{n-1})$. The second equality holds since the conditional entropy of a random variable is not changed upon adding a deterministic function of conditioning random variables, and the last equality holds since both $(X_0 \ldots X_n)$ and $(W_0 \ldots W_n)$ are sequences of i.i.d. random variables. Since $\{W_n\}$ is WGN, the entropy of the noise sequence is

$$h(W_0 \ldots W_{N+\mu-1}) = \sum_{n=0}^{N+\mu-1} h(W_n).$$

Substituting these expressions into (3.36) and using the fact that $X_n = 0$ for $n \geq N$ yields

$$\begin{aligned} &\frac{1}{N}\, I(X_0 \ldots X_{N-1}\,;\,Y_0 \ldots Y_{N+\mu-1}) \geq \\ &\qquad \frac{1}{N} \sum_{n=0}^{N+\mu-1} h(h_0\, X_n + W_n) - h(W_n) = I(X\,;\,h_0\, X + W), \end{aligned} \tag{3.38}$$

where we have introduced the generic random variables X and W with distributions as described in the lemma. Let n_z and n_p be the number of zeros and poles of $H(z)$, respectively. For channels with finite memory $M \triangleq n_z$, i.e., when $n_p = 0$, the proof is completed by setting $\mu = M$ and then taking the limit of (3.38) as $N \to \infty$. For channels with unbounded memory, i.e., when $n_p > 0$, we first take the limit as $\mu \to \infty$ and then let $N \to \infty$. □

Lemma 3.2: Let $\mathcal{H}$ be an equivalence class as in (3.32) for transfer functions with $n_z \geq 0$ zeros and denominator $D(z) = 1 + d_1 z^{-1} + \ldots + d_{n_p} z^{-n_p}$. Then $h_0 = \lim_{z \to \infty} H(z)$ for a member $H(z) = N(z)/D(z)$ of $\mathcal{H}$ with zeros $z_1, z_2, \ldots, z_{n_z}$ satisfies

$$|h_0| \leq \rho, \tag{3.39}$$

where

$$\begin{aligned} \rho &= |h_0| \cdot \prod_{n:\,|z_n|>1} |z_n| \\ &= \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|H(e^{j\Omega})\right| d\Omega\right\} \\ &= \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|N(e^{j\Omega})\right| d\Omega\right\}, \end{aligned}$$

and equality is achieved in (3.39) if and only if $|z_n| \leq 1$, $1 \leq n \leq n_z$.

Note that the condition for equality in (3.39) is also the condition for $H(z)$ to be a minimum-phase filter [14]. Hence $\rho = |h_{\min 0}|$, where $h_{\min 0}$ is the leading coefficient of the Laurent-series expansion of a minimum-phase version $H_{\min}(z)$ of $H(z)$.

Proof: We express $H(z)$ as

$$H(z) = A(z)\, H_{\min}(z), \tag{3.40}$$

where $H_{\min}(z)$ is a minimum-phase version of $H(z)$, i.e., the zeros $c_1, c_2, \ldots, c_{n_z}$ of $H_{\min}(z)$ all lie inside or on the unit circle, and $A(z)$ is an allpass filter of the form

$$A(z) = \prod_{n=1}^{n_z} \left(\frac{c_n^* - z^{-1}}{1 - c_n z^{-1}}\right)^{i_n},$$

where the indicator i_n is 0 when $|c_n| = 1$ and $i_n \in \{0,1\}$ otherwise. Therefore, $H(z)$ has a zero at $1/c_n^*$ for all n such that $i_n = 1$ and at c_n for all other n, i.e., the zeros z_n of $H(z)$ are given by

$$z_n = \begin{cases} c_n, & \text{if } i_n = 0 \\ 1/c_n^*, & \text{if } i_n = 1 \end{cases}, \qquad 1 \le n \le n_z.$$

Note that a zero z_n is outside the unit circle if and only if $i_n = 1$. Recall from Section 3.2.1 that all transfer functions in $\mathcal{H}$ have the same poles and that $R_{\max} < 1$, where $R_{\max}$ is the largest magnitude of a pole. The transfer functions in $\mathcal{H}$ are thus analytic in the region $\{z : |z| > R_{\max}\}$, which is their region of convergence.

For the following, it will be convenient to define $H^+(z) \triangleq H^*(1/z^*)$, whose region of convergence is given by $\mathcal{R} \triangleq \{z : |z| < 1/R_{\max}\}$. Note that $H^+(z)$ is the z-transform of the time-reversed and conjugated sequence $\{h_{-k}^*\}$, which is anticausal[8].

We will apply Jensen's integral formula (cf. Lemma 3.C.1 in Appendix 3.C) to $H^+(z)$, which is analytic in the region $\mathcal{R}$. Consider the singularities of $H^+(z)$. Let R_1, $R_1 < 1$, be the largest magnitude of a zero of $H^+(z)$ inside the unit circle. If $H^+(z)$ has no zeros inside the unit circle, we take $R_1 = 0$. Further, let R_2, $1 < R_2 \le 1/R_{\max}$, be the smallest magnitude of a singularity (either zero or pole) of $H^+(z)$ outside the unit circle. If $H^+(z)$ has no singularities outside the unit circle, we take $R_2 = \infty$. Clearly, $H^+(z)$ is analytic for $|z| < R_2$. We will apply Jensen's integral formula twice, first for $1 < R < R_2$ and then for $R_1 < R < 1$. This will allow us to write the degradation factor either as a left-sided or as a right-sided limit as R approaches one. Showing that both limits exist and are equal will prove the convergence of the integrals in (3.31), even if these integrals are improper because of zeros on the unit circle.

For $1 < R < R_2$, the sum in (3.C.1) must be taken over all zeros of $F(z) \triangleq H^+(z)$ *inside and on* the unit circle. Thus,

$$\log|H^+(0)| = \log|h_0| = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|H(R^{-1}e^{j\Omega})\right| d\Omega + \sum_{\substack{n:\, i_n=1 \\ \text{or } |c_n|=1}} \log(|c_n|/R)$$

[8] A sequence $\{g_k\}$ is called *anticausal* if $g_k = 0$ for $k > 0$.

or, equivalently,

$$|h_0| = \rho(R) \cdot \prod_{\substack{n:\, i_n=1 \\ \text{or } |c_n|=1}} |c_n| \,/\, R,$$

where

$$\rho(R) \triangleq \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|H(R^{-1}e^{j\Omega})\right| \, d\Omega\right\}. \tag{3.41}$$

Notice that the integral in (3.41) becomes improper as $R \to 1$ when there are zeros on the unit circle. Nevertheless, the right-sided limit of $\rho(R)$ exists and is given by

$$\lim_{R\to 1^+} \rho(R) = |h_0| \;/ \prod_{n:\, i_n=1} |c_n|. \tag{3.42}$$

When Jensen's integral formula is applied to $H^+(z)$ for $R_1 < R < 1$, the sum in (3.C.1) is over all zeros of $H^+(z)$ *inside* the unit circle, so that

$$|h_0| = \rho(R) \cdot \prod_{n:\, i_n=1} |c_n| \,/\, R. \tag{3.43}$$

Hence, the left-sided limit $\lim_{R\to 1^-} \rho(R)$ exists as well and is equal to the right-sided limit (3.42). Thus, $\rho \triangleq \rho(1)$ is well-defined and given by

$$\begin{aligned}\rho &= |h_0| \;/ \prod_{n:\, i_n=1} |c_n| = |h_0| \cdot \prod_{n:\, |z_n|>1} |z_n| \\ &= \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|H(e^{j\Omega})\right| \, d\Omega\right\} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|N(e^{j\Omega})\right| \, d\Omega\right\},\end{aligned}$$

where the last equality follows from the facts that the poles of $H(z) = N(z)/D(z)$ do not appear in the expressions on the first line and that the denominator $D(z)$ has a leading coefficient 1. Noting that $|c_n| \le 1$, we obtain the upper bound

$$|h_0| = \rho \prod_{n:\, i_n=1} |c_n| \;\le\; \rho. \tag{3.44}$$

Thus, $|h_0|$ is largest when (3.44) holds with equality, i.e., when $i_n = 0$ for all n or, equivalently, when $|z_n| \le 1$ for $1 \le n \le n_z$. □

We now turn to the proof of Theorem 3.2.

Proof: Let $\mathcal{H}$ be the transfer-function equivalence class of which $H(z)$ is a member. We know from Theorem 3.3 (or Theorem 3.4) and Lemma 3.1 that, for any $\tilde{H}(z)$ in $\mathcal{H}$,

$$I_{\text{i.i.d.}}(H(z)) = I_{\text{i.i.d.}}(\tilde{H}(z)) \geq I(X\,;\,\tilde{h}_0\,X + W), \tag{3.45}$$

where $\tilde{h}_0$ is the leading coefficient of the Laurent-series expansion of $\tilde{H}(z)$. We are thus free to maximize the lower bound by choosing the $\tilde{H}(z)$ with the largest $|\tilde{h}_0|$. According to Lemma 3.2, the magnitude of the leading coefficient is maximized when all zeros of $\tilde{H}(z)$ are inside or on the unit circle, i.e., when $\tilde{H}(z)$ equals a minimum-phase version $H_{\min}(z)$ of $H(z)$. The proof is completed by noting that in Lemma 3.2 all transfer functions in $\mathcal{H}$ have the same upper bound ρ on their $|\tilde{h}_0|$. □

Appendix 3.A Alternative Proof of Theorem 3.3

In the following proof, some allpass-filter properties from Appendix 3.B will be used.

Proof of Theorem 3.3: We begin with the proof for complex channels. Assuming an FIR channel filter $H(z) = \sum_{m=0}^{M} h_m z^{-m}$ with associated $(N+M) \times N$ Toeplitz matrix

$$\mathbf{H} \triangleq \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & & \vdots \\ \vdots & h_1 & \ddots & 0 \\ h_M & & & h_0 \\ 0 & h_M & & h_1 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & h_M \end{bmatrix}$$

and defining $\underline{X} \triangleq [X_0, X_1, \ldots, X_{N-1}]^T$, $\underline{W} \triangleq [W_0, W_1, \ldots, W_{N-1}]^T$, and $\underline{Y} \triangleq [Y_0, Y_1, \ldots, Y_{N+M-1}]^T$, we may write the relevant channel outputs as

$$\underline{Y} = \mathbf{H}\,\underline{X} + \underline{W} \tag{3.A.1}$$

where $\underline{W} \sim \mathcal{N}_p(\underline{0}, \sigma_W^2\,\mathbf{I})$.

We now take any allpass filter $A(z)$ of the form given in (3.32) and of order $K \leq M$ so that $\tilde{H}(z) \triangleq A(z)H(z)$ is another M-th order FIR filter in the equivalence class $\mathcal{H}$. Note that K zeros are reflected at the unit circle by multiplying $H(z)$ with $A(z)$. Correspondingly, we transform the observation vector $\underline{Y}$ by premultiplying it with the $(N+M) \times (N+M)$ allpass matrix $\mathbf{A}$ associated with $A(z)$ (cf. (3.B.4)). Notice that the noise component $\mathbf{A}\,\underline{W}$ of $\mathbf{A}\,\underline{Y} = \mathbf{A}\,\mathbf{H}\,\underline{X} + \mathbf{A}\,\underline{W}$ has the correlation matrix[1] $\sigma_W^2\,\mathbf{A}\mathbf{A}^*$, which is not equal to $\sigma_W^2\,\mathbf{I}$ in general. Nevertheless, an observation vector with a *white* noise component can be obtained from $\mathbf{A}\underline{Y}$ by adding an independent, proper complex Gaussian random vector $\underline{V} \sim \mathcal{N}_p(\underline{0}, \mathbf{\Phi}_{\underline{V}})$, where $\mathbf{\Phi}_{\underline{V}} \triangleq \sigma_W^2\,(\mathbf{I} - \mathbf{A}\mathbf{A}^*)$, to give

$$\underline{\tilde{Y}} \triangleq \mathbf{A}\,\underline{Y} + \underline{V}. \tag{3.A.2}$$

[1] We use $\mathbf{A}^*$ to denote the conjugate-transpose of a matrix $\mathbf{A}$.

As required for a covariance matrix, $\Phi_{\underline{V}}$ is nonnegative definite, since the eigenvalues of $\mathbf{A}\mathbf{A}^*$ do not exceed one by Lemma 3.B.2[2]. We will show that the addition of $\underline{V}$ results in no information loss. Substituting (3.A.1) into (3.A.2) yields

$$\underline{\tilde{Y}} = \tilde{\mathbf{H}}\,\underline{X} + \underline{\tilde{W}}, \tag{3.A.3}$$

where $\tilde{\mathbf{H}} \triangleq \mathbf{A}\,\mathbf{H}$ and $\underline{\tilde{W}} \triangleq \mathbf{A}\,\underline{W} + \underline{V}$. Since $\tilde{\mathbf{H}}$ corresponds to the transfer function $\tilde{H}(z)$, it must be a Toeplitz matrix with the same structure as $\mathbf{H}$. Moreover, it follows from

$$\Phi_{\underline{\tilde{W}}} = \sigma_W^2\,\mathbf{A}\mathbf{A}^* + \sigma_W^2\,(\mathbf{I} - \mathbf{A}\mathbf{A}^*) = \sigma_W^2\,\mathbf{I}$$

and from Lemma 2.4 that $\underline{W}$ is a segment of proper complex *white* Gaussian noise. Thus, $\underline{\tilde{Y}}$ is the observation vector obtained at the output of an ISI channel with channel filter $\tilde{H}(z)$ and with the same noise statistics as the original channel.

Now observe from (3.A.1) and (3.A.2) that the sequence $(\underline{X}, \underline{Y}, \underline{\tilde{Y}})$ is a Markov chain. Hence, the data-processing inequality [23, p. 32] yields

$$I(\underline{X}\,;\underline{Y}) \;\geq\; I(\underline{X}\,;\underline{\tilde{Y}}). \tag{3.A.4}$$

But had we started with the channel filter $\tilde{H}(z)$ and its associated Toeplitz matrix $\tilde{\mathbf{H}}$, the transformation of the observation vector by the allpass matrix $\mathbf{A}^*$ (which is shown in Appendix 3.B to be associated with $1/A(z)$) and the subsequent addition of an independent, proper complex Gaussian random vector with covariance matrix $\sigma_W^2\,(\mathbf{I} - \mathbf{A}^*\mathbf{A})$ would have led to the ISI channel with channel filter $H(z)$ and AWGN, which proves that (3.A.4) also holds in the reverse direction. Hence (3.A.4) must hold with equality.

To prove the theorem for *real* channels where all filter coefficients are real[3] and, in all transfer functions, zeros and poles appear in complex conjugate pairs, we simply replace the proper complex distributions by real distributions, i.e., $\underline{V} \sim \mathcal{N}(\underline{0}, \Phi_{\underline{V}})$ and $\underline{\tilde{W}} \sim \mathcal{N}(\underline{0}, \Phi_{\underline{\tilde{W}}})$. □

[2]Lemma 3.B.1 shows further that $\Phi_{\underline{V}}$ has at most K nonzero eigenvalues.

[3]Notice that Lemma 3.B.2 used above applies to real matrices $\mathbf{A}$ for which $\mathbf{A}^*$ can be replaced by $\mathbf{A}^T$.

Appendix 3.B Allpass-Filter Properties

For $K \geq 1$, a *K-th order allpass filter* $A(z) = \sum_{i=-\infty}^{\infty} a_i z^{-i}$ is given by

$$A(z) = \alpha \prod_{k=1}^{K} \frac{z - z_k}{z z_k^* - 1}, \quad |\alpha| = 1, \ 0 < |z_k| < \infty, \ |z_k| \neq 1, \tag{3.B.1}$$

where, for each zero z_k, there is a corresponding pole at $1/z_k^*$ that is obtained by reflecting this zero at the unit circle. An allpass filter of order zero is simply $A(z) = \alpha$. It is sometimes convenient to use the equivalent form

$$A(z) = \alpha' \prod_{k=1}^{K} \frac{z_k'^* z - 1}{z - z_k'}, \quad |\alpha'| = 1, \ 0 < |z_k'| < \infty, \ |z_k'| \neq 1, \tag{3.B.2}$$

where the *poles* are at z_k'.

For the following, it will be convenient to define $A^+(z) \triangleq A^*(1/z^*)$, which is the z-transform of the time-reversed and conjugated sequence $\{a_{-i}^*\}$. It is easy to check that

$$A^+(z) = \alpha^* \prod_{k=1}^{K} \frac{z^{-1} - z_k^*}{z_k z^{-1} - 1} = \frac{1}{\alpha} \prod_{k=1}^{K} \frac{z z_k^* - 1}{z - z_k} = 1/A(z). \tag{3.B.3}$$

This implies that $A(z)A^+(z) = 1$ and hence that $|A(e^{j\Omega})| = 1$, $0 \leq \Omega < 2\pi$, which is the reason for the name 'allpass filter'.

If all poles of $A(z)$ are inside the unit circle and $1/z_1^*$ is the pole closest to the unit circle, we take $\{z : |z| > |1/z_1|\}$ to be the region of convergence (ROC) of $A(z)$ so that the sequence $\{a_i\}$ is causal. Moreover, $\{a_i\}$ is delayless since $\lim_{z\to\infty} A(z) = a_0 \neq 0$. If all poles are outside the unit circle and $1/z_1^*$ is the pole closest to the unit circle, we take $\{z : |z| < |1/z_1|\}$ to be the ROC of $A(z)$ so that the sequence $\{a_i\}$ is anticausal[1]. Moreover, $\lim_{z\to 0} A(z) = a_0 \neq 0$. Finally, if there are poles inside and outside the unit circle, we let the ROC of $A(z)$ be the largest ring centered at the origin and containing the unit circle but no poles. In this case, the sequence $\{a_i\}$ is two-sided and is neither causal nor anticausal.

[1] A sequence $\{a_i\}$ is called *anticausal* if $a_i = 0$ for $i > 0$.

In the following, we investigate, for all $N \geq 1$, some properties of the $N \times N$ Toeplitz matrix

$$\mathbf{A} \triangleq \begin{bmatrix} a_0 & a_{-1} & & \cdots & a_{-N+1} \\ a_1 & a_0 & a_{-1} & & \\ & a_1 & a_0 & & \vdots \\ \vdots & & & & \\ & & & & a_{-1} \\ a_{N-1} & & \cdots & a_1 & a_0 \end{bmatrix} \tag{3.B.4}$$

associated with $A(z)$. Notice that the $N \times N$ Toeplitz matrix associated with $1/A(z) = A^+(z)$ is given by $\mathbf{A}^*$.

Lemma 3.B.1: Let $A(z) = \sum_{i=-\infty}^{\infty} a_i z^{-i}$ be a K-th order allpass filter with *poles* at $z_1, z_2, \ldots, z_K$ and define the K-th order FIR filter

$$H(z) \triangleq h_0 \prod_{k=1}^{K} (1 - z_k z^{-1}) = \sum_{k=0}^{K} h_k z^{-k}.$$

For $N \geq K+1$, let $\mathbf{A}$ be the $N \times N$ Toeplitz matrix associated with $A(z)$ as defined in (3.B.4). Then $\mathbf{A}^*\mathbf{A}$ has a unit eigenvalue of multiplicity at least $N - K$ and the corresponding eigenvectors can be selected as

$$\underline{v}_n = [\underline{0}_{n-1}^T : \underline{h}^T : \underline{0}_{N-K-n}^T]^T, \quad 1 \leq n \leq N - K,$$

where $\underline{h} \triangleq [h_0, h_1, \ldots, h_K]^T$ and $\underline{0}_n$ denotes an all-zero vector of length n.

Proof: The transfer function $\tilde{H}(z) \triangleq A(z)H(z)$ is a K-th order FIR filter since the poles of $A(z)$ and the zeros of $H(z)$ cancel out. Explicitly, $\tilde{H}(z) = \tilde{h}_0 \prod_{k=1}^{K}(1 - 1/(z z_k^*))$. The coefficients of $\tilde{H}(z)$ are denoted by $\underline{\tilde{h}} \triangleq [\tilde{h}_0, \tilde{h}_1, \ldots, \tilde{h}_K]^T$. Expressing the convolution of $\{a_i\}$ and $\{h_i\}$ in matrix form gives

$$\mathbf{A} \begin{bmatrix} \underline{0}_{n-1} \\ \underline{h} \\ \underline{0}_{N-K-n} \end{bmatrix} = \begin{bmatrix} \underline{0}_{n-1} \\ \underline{\tilde{h}} \\ \underline{0}_{N-K-n} \end{bmatrix},$$

where $1 \leq n \leq N-K$. Using (3.B.3) gives $H(z) = A^{-1}(z)\tilde{H}(z) = A^{+}(z)\tilde{H}(z)$. Therefore,

$$\mathbf{A}^* \begin{bmatrix} \underline{0}_{n-1} \\ \tilde{\underline{h}} \\ \underline{0}_{N-K-n} \end{bmatrix} = \mathbf{A}^*\mathbf{A} \begin{bmatrix} \underline{0}_{n-1} \\ \underline{h} \\ \underline{0}_{N-K-n} \end{bmatrix} = \begin{bmatrix} \underline{0}_{n-1} \\ \underline{h} \\ \underline{0}_{N-K-n} \end{bmatrix},$$

where $1 \leq n \leq N-K$. □

Notice that Lemma 3.B.1 holds also for $A^{+}(z)$. It follows that, for $N \geq K+1$, the matrix $\mathbf{A}\mathbf{A}^*$ has a unit eigenvalue of multiplicity at least $N-K$ and the corresponding eigenvectors can be chosen as $\underline{v}_n = [\underline{0}_{n-1}^T : \tilde{\underline{h}}^T : \underline{0}_{N-K-n}^T]^T$.

Lemma 3.B.2: Let $A(z) = \sum_{i=-\infty}^{\infty} a_i z^{-i}$ be a K-th order allpass filter and let $\mathbf{A}$ be the $N \times N$ Toeplitz matrix associated with $A(z)$, as defined in (3.B.4). Then the eigenvalues λ_n of the matrix $\mathbf{A}\mathbf{A}^*$ satisfy $0 \leq \lambda_n \leq 1$, $1 \leq n \leq N$, and are all nonzero if the sequence $\{a_i\}$ is causal or anticausal. Moreover, for $N \geq K+1$, at least $N-K$ eigenvalues equal one.

Proof: Consider the experiment in Figure 3.B.1, where a white, proper complex random process $\{W_n\}$ with zero mean and unit sample variance is input to the upper allpass filter $A(z)$ for $0 \leq n < N$ and to the lower identical allpass filter at all other times[2] and where the outputs of these allpass filters are added to form the process $\{\tilde{W}_n\}$, where $\tilde{W}_n \triangleq Z_n + V_n$ for all n. It follows that $\{\tilde{W}_n\}$ is another white, proper complex random process with unit sample variance. Since the two all-

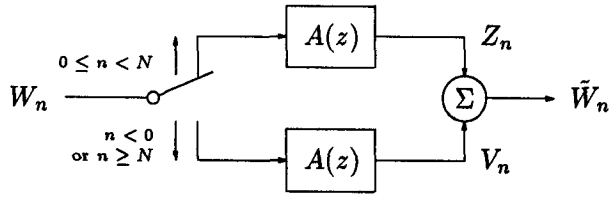


Figure 3.B.1: Allpass experiment

pass filters are fed by uncorrelated subsequences of $\{W_n\}$, the vectors

[2]While a filter is not connected to W_n, zeros are assumed to be the input.

$\underline{Z} = [Z_0, Z_1, \ldots, Z_{N-1}]^T \triangleq \mathbf{A}\underline{W}$ (where $\underline{W} = [W_0, W_1, \ldots, W_{N-1}]^T$) and $\underline{V} = [V_0, V_1, \ldots, V_{N-1}]^T$ are uncorrelated as well. Thus, the covariance matrices (cf. (2.3)) of the vectors $\underline{\tilde{W}} = [\tilde{W}_0, \tilde{W}_1, \ldots, \tilde{W}_{N-1}]^T$, $\underline{Z}$ and $\underline{V}$ are related by $\mathbf{\Phi}_{\underline{\tilde{W}}} = \mathbf{I} = \mathbf{\Phi}_{\underline{Z}} + \mathbf{\Phi}_{\underline{V}}$. From this and since $\mathbf{\Phi}_{\underline{Z}}$ and $\mathbf{\Phi}_{\underline{V}}$ are nonnegative-definite Hermitian matrices, the eigenvalues λ_n of $\mathbf{\Phi}_{\underline{Z}} = \mathbf{A}\mathbf{A}^*$ satisfy $0 \leq \lambda_n \leq 1$. When the sequence $\{a_i\}$ is causal [anticausal], the matrix $\mathbf{A}$ is lower-triangular [upper-triangular] with $a_0 \neq 0$ on the main diagonal and therefore nonsingular, which implies $\lambda_n > 0$, $1 \leq n \leq N$. Moreover, for $N \geq K + 1$ the matrix $\mathbf{A}\mathbf{A}^*$ has at least $N - K$ unit eigenvalues by Lemma 3.B.1. □

Appendix 3.C Jensen's Integral Formula

For the proof of Theorem 3.2, we need the following result [45, p. 424]:

Lemma 3.C.1 (Jensen's integral formula): Let $F(z)$ be a function of a complex variable that is analytic in a region containing the disk $\mathcal{D}_R \triangleq \{z : |z| \leq R\}$ and that has finitely many (not necessarily distinct) zeros $c_1, c_2, \ldots, c_n$ in the interior of $\mathcal{D}_R$. Then, if $F(0) \neq 0$,

$$\log|F(0)| = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log\left|F(Re^{j\Omega})\right| d\Omega + \sum_{i=1}^{n} \log\left(|c_i|/R\right). \quad (3.C.1)$$

Leer - Vide - Empty

Chapter 4

Construction of K-ary State-Transition Diagrams for Trellis Encoders

In this chapter, we investigate directed graphs with exactly K branches emanating from every node, referred to as K-ary state-transition diagrams (STD's), for the construction of trellis encoders. (A trellis encoder will be defined as a labeled K-ary STD in Chapter 5.) Our original motivation for the investigation of K-ary STD's was their usefulness in finding good matched spectral-null codes [13], [16] by exhaustive search. However, K-ary STD's are also potentially useful for searching trellis codes with other properties.

An algorithm will be presented for finding all K-ary STD's that satisfy certain topological constraints. Various constraints can be imposed in order to keep the number of K-ary STD's manageable and to identify the K-ary STD's best suited for coding. For practical reasons, one usually desires a *controllable* (n, k) trellis encoder, i.e., a trellis encoder, which can be driven to any state from any given initial state by some information sequence. This encoder property is equivalent to the requirement that the underlying 2^k-ary STD is *strongly connected* [46, p. 132], [47, p. 3].

Certain directed graphs differ only in the names of their nodes and branches and are therefore considered equivalent or *isomorphic* [46, p. 6], [47, p. 154]. We are thus interested in the *isomorphism classes* of K-ary STD's. A straightforward approach for constructing

all non-isomorphic[1] K-ary STD's with N nodes and given topological constraints would be to generate a list of such K-ary STD's sequentially. This could be done by generating and testing all combinatorially possible K-ary STD's with N nodes. Every candidate that is neither isomorphic to an entry already in the list nor in conflict with a constraint would be appended to the list; every other candidate would be rejected. However, such an approach is unfeasible already for $N \gtrsim 8$ and $K = 2$ since the number of candidates is too large and because the number of isomorphisms that must be checked becomes prohibitive.

To alleviate these problems, we show that K-ary STD's can be constructed in a recursive way by successively extending nodes of so-called *partial K-ary STD's*. The topological constraints on the (complete) K-ary STD's also constrain the partial K-ary STD's. Depending on the constraints, this will allow us to reject a large fraction of the combinatorially possible partial K-ary STD's and thus to keep the number of processed candidates manageable.

The outline of this chapter is as follows. In Section 4.1, after some terminology for directed graphs, different ways of representing K-ary STD's are considered and the notion of an isomorphism is examined in terms of these representations. In Section 4.1, we also define the *detour memory* of a K-ary STD. Large detour memory will turn out to be a useful criterion for identifying K-ary STD's from which good trellis codes can be constructed. (In Chapter 5, we will present a simple upper bound on the free distance of a trellis code that involves only the detour memory of a K-ary STD and the maximum distance between two elements of the coding alphabet.) In Section 4.2, necessary and sufficient conditions are derived for when a partial K-ary STD can be extended to some complete, strongly connected K-ary STD with N nodes. Moreover, it is shown that the parameters needed to test these conditions can be computed recursively. In Section 4.3, an algorithm is presented for the systematic construction of all non-isomorphic K-ary STD's with N nodes and given topological constraints. This algorithm generates a sequence of N ordered lists, where the n-th list, $1 \leq n \leq N$, contains the partial K-ary STD's with n extended nodes. In Section 4.4, K-ary STD's with *maximum detour memory* are investigated and tabulated for $K = 2$ and $N = 1$, 2, 4, 8, and 16 nodes. Appendix 4.A contains some properties of the n-th power of a directed graph that are used in Sections 4.1 and 5.1 for the investigation of paths in K-ary STD's.

[1] When we speak of "all non-isomorphic K-ary STD's", we mean one representative for every isomorphism class.

4.1 Graph Preliminaries

A directed graph can be defined as follows [46, p. 125], [47, p. 3].

Definition 4.1: A *directed graph* or *digraph* G is a four-tuple $(\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$ where $\mathcal{S}$ (the set of *nodes*) and $\mathcal{B}$ (the set of *branches*) are disjoint sets and where σ and ϵ are mappings from $\mathcal{B}$ to $\mathcal{S}$. ———

Unless stated otherwise, we assume in this thesis that both $\mathcal{S}$ and $\mathcal{B}$ are finite and that $\mathcal{S}$ is non-empty[2]. For a branch $b \in \mathcal{B}$, $\sigma(b)$ and $\epsilon(b)$ denote the *start-node* and *end-node*, respectively. A branch b with $\sigma(b) = \epsilon(b)$, i.e., with the same start-node and end-node, is called a *self-loop*. Two branches b and b' with $\sigma(b) = \sigma(b')$ and $\epsilon(b) = \epsilon(b')$, i.e., with a common start-node and a common end-node, are said to be *parallel*. For any $s \in \mathcal{S}$, let

$$\mathcal{B}_{\text{out}}(s) \triangleq \{b : b \in \mathcal{B}, \sigma(b) = s\} \quad \text{and} \quad \mathcal{B}_{\text{in}}(s) \triangleq \{b : b \in \mathcal{B}, \epsilon(b) = s\},$$

i.e., $\mathcal{B}_{\text{out}}(s)$ and $\mathcal{B}_{\text{in}}(s)$ denote the subset of branches emanating from and ending at node s, respectively. The *out-degree* $d_{\text{out}}(s)$ and *in-degree* $d_{\text{in}}(s)$ are the numbers of branches starting from and ending at node s, respectively, i.e., $d_{\text{out}}(s) = |\mathcal{B}_{\text{out}}(s)|$ and $d_{\text{in}}(s) = |\mathcal{B}_{\text{in}}(s)|$. A digraph is said to have *uniform out-degree* (*uniform in-degree*) d if $d_{\text{out}}(s) = d$ ($d_{\text{in}}(s) = d$) for every $s \in \mathcal{S}$. The nodes in $\{\epsilon(b) : b \in \mathcal{B}_{\text{out}}(s)\}$ and $\{\sigma(b) : b \in \mathcal{B}_{\text{in}}(s)\}$ are called the *successors* and *predecessors* of s, respectively. A node that is neither a start-node nor an end-node for any branch is said to be *isolated*.

A *path* in a digraph is a non-empty sequence of branches such that, for any two subsequent branches b_i and b_{i+1} in this sequence, the end-node of b_i is the start-node of b_{i+1}. Paths can be finite, semi-infinite, or infinite. For a finite or right-sided semi-infinite path γ, the start-node of the first branch is called the *start-node of the path* and is denoted by $\sigma(\gamma)$. For a finite or left-sided semi-infinite path γ, the end-node of the last branch is called the *end-node of the path* and is denoted by $\epsilon(\gamma)$. Two finite paths are said to be *parallel* if they have a common start-node and a common end-node. A *cycle* or *cyclic path* is a finite path whose start-node and end-node coincide. A digraph without cycles is called *acyclic*. A digraph with node set $\mathcal{S}$ is said to be *strongly connected* [46, p. 132], [47, p. 3] if there is a finite path from every node $s \in \mathcal{S}$ to

[2] In Section 4.2, it will be convenient to recognize an 'empty digraph' [46, p. 2] that has both an empty node set and an empty branch set.

every node $s' \in \mathcal{S}$. A graph consisting of an isolated node is, by way of convention, considered to be strongly connected[3]. The *period* P of a strongly connected digraph with a non-empty branch set is defined as the greatest common divisor of the lengths of all cycles. A digraph with period $P>1$ (P=1) is said to be *periodic* (*aperiodic*).

In this chapter, we are interested in a special kind of directed graph that will be used for the construction of trellis encoders in Chapter 5.

Definition 4.2: A *K-ary state-transition diagram* (STD), where $K > 0$, is a digraph with uniform out-degree K.

The name 'state-transition diagram' anticipates the use of such digraphs for describing the state transitions of certain finite-state machines [48]. In Chapter 5, a trellis encoder will be defined as a 'labeled' K-ary STD that is obtained from an 'unlabeled' K-ary STD as in Definition 4.2 by assigning 'inputs' and 'outputs' to the branches.

A K-ary STD may also have a uniform in-degree, which then must also be K.

Example 4.1: Figure 4.1 shows a strongly connected, aperiodic binary STD with uniform in-degree 2. Note that the state transitions of a

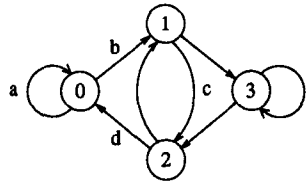


Figure 4.1: A binary state-transition diagram

feedforward shift register with two serial 1-bit stages can be described by such a binary STD. In Figure 4.1, $\mathcal{B}_{\text{out}}(0) = \{a, b\}$ and $\mathcal{B}_{\text{in}}(0) = \{a, d\}$. The paths (a, a, a) and (b, c, d) are parallel.

Certain K-ary STD's will be considered essentially the same or 'isomorphic' [46, p. 6], [47, p. 154] in the following sense.

[3]This will be relevant for the definition of the component-reduced digraph in Section 4.2.

Definition 4.3: Two directed graphs $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$ and $G' = (\mathcal{S}', \mathcal{B}', \sigma', \epsilon')$ are called *isomorphic* if there are bijective maps $\mu : \mathcal{S} \to \mathcal{S}'$ and $\beta : \mathcal{B} \to \mathcal{B}'$ such that, for every $b \in \mathcal{B}$,

$$\mu(\sigma(b)) = \sigma'(\beta(b)) \quad \text{and} \quad \mu(\epsilon(b)) = \epsilon'(\beta(b)). \tag{4.1}$$

The condition (4.1) means that, for every $b \in \mathcal{B}$, the start-node and end-node of b is mapped by μ to the start-node and end-node of $b' \triangleq \beta(b)$, respectively, as shown in Figure 4.2. We write $G \cong G'$ to

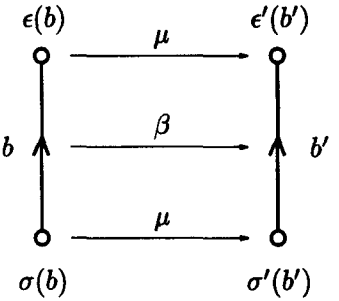


Figure 4.2: Illustration of condition (4.1) with $b' \triangleq \beta(b)$

indicate that G and G' are isomorphic. The pair of maps (μ, β) in Definition 4.3 is an *isomorphism* of G onto G' [46, p. 6]. An isomorphism (μ, β) from G onto itself is called an *automorphism.* The relation of isomorphism between digraphs is easily verified to be reflexive, symmetrical, and transitive, i.e., it satisfies the conditions for an equivalence relation [46, p. 6]. It therefore partitions the class of digraphs into disjoint subclasses called *isomorphism classes.*

Every branch in a digraph without parallel branches is uniquely identified by its start-node and end-node. Such a digraph is therefore isomorphic to a digraph with the same set of nodes, say $\mathcal{S}$, and a set of branches $\mathcal{B}$ that can be taken to be a subset of $\mathcal{S} \times \mathcal{S}$. The maps σ (for the start-node) and ϵ (for the end-node) are then simply the projections of $\mathcal{B}$ onto its first and second component, respectively, and need not be specified explicitly. We will write simply $G = (\mathcal{S}, \mathcal{B})$ when $\mathcal{B}$ is a subset of $\mathcal{S} \times \mathcal{S}$.

The *adjacency matrix* of a digraph $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$ with node set $\mathcal{S} = \{0, 1, \ldots, N-1\}$ is the $N \times N$ matrix[4] $\mathbf{A} = [a_{ij}]$, where a_{ij} is the number of branches from node i to node j, $0 \leq i, j < N$. We do not constrain the entries of $\mathbf{A}$ to be zero or one, as this would disallow parallel branches. In the remainder of this section, the node set is always taken to be $\mathcal{S} = \{0, 1, \ldots, N-1\}$.

If $\mathbf{A}$ is the adjacency matrix of a digraph G then it is obvious that a digraph G' isomorphic to G can be constructed from $\mathbf{A}$. Thus, $\mathbf{A}$ represents the isomorphism class of which G is a member. The representation of a digraph (and its isomorphism class) by its adjacency matrix is the bridge between graph theory and the theory of nonnegative matrices. For instance, a digraph is strongly connected if and only if its adjacency matrix is *irreducible* [38, Thm. 15.1, p. 529]. (An $N \times N$ matrix $\mathbf{M}$, where $N \geq 2$, is said to be *reducible* if there is an $N \times N$ permutation matrix $\mathbf{P}$ such that

$$\mathbf{P}^T \mathbf{M} \mathbf{P} = \begin{bmatrix} \tilde{\mathbf{M}}_{11} & \tilde{\mathbf{M}}_{12} \\ \mathbf{0} & \tilde{\mathbf{M}}_{22} \end{bmatrix},$$

where every submatrix has at least one row and column. If no such $\mathbf{P}$ exists, then $\mathbf{M}$ is called *irreducible* [38, p. 374]. A 1×1 matrix is, by way of convention, considered to be irreducible.) Further, the period of a strongly connected digraph G with a non-empty branch set equals the so-called *index of imprimitivity* of its adjacency matrix $\mathbf{A}$ [38, p. 544], which is defined as the number of eigenvalues of $\mathbf{A}$ with largest magnitude. A nonnegative, irreducible matrix with index of imprimitivity one is called *primitive* [38, p. 544]. Hence, the adjacency matrix $\mathbf{A}$ of a strongly connected digraph G with a non-empty branch set is primitive if and only if G is aperiodic. This relationship will be useful in checking the strong connectivity of the n-th power of a digraph (cf. Appendix 4.A).

The adjacency matrix $\mathbf{A}$ of a K-ary STD with N nodes is sparse when $K \ll N$, since every row of $\mathbf{A}$ is nonzero in at most K positions. We now present a different, more compact description of a K-ary STD $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$ (with or without parallel branches), which is better suited for manipulation on a computer. For the i-th node, where $0 \leq i < N$, the branches emanating from node i can be numbered in any order from 0 to $K-1$. In particular, they can be numbered as $b_j(i)$, $0 \leq j < K$, where the end-nodes of $b_0(i), b_1(i), \ldots, b_{K-1}(i)$ are in non-increasing order, i.e.,

$$\epsilon(b_0(i)) \geq \epsilon(b_1(i)) \geq \ldots \geq \epsilon(b_{K-1}(i)). \tag{4.2}$$

[4]Matrix rows and columns are indexed starting from zero in this chapter.

We arrange the NK numbers $f_{ij} \triangleq \epsilon(b_j(i))$, $0 \leq i < N$, $0 \leq j < K$, as an $N \times K$ matrix $\mathbf{F} = [f_{ij}]$, called the *next-node matrix* of G. The i-th row of $\mathbf{F}$, which is denoted by $\underline{f}_i$, is an ordered K-tuple whose entries are a combination of K not necessarily distinct nodes. From (4.2), it follows that the coefficients f_{ij} satisfy

$$f_{i0} \geq f_{i1} \geq \cdots \geq f_{i,K-1}. \tag{4.3}$$

If G has no parallel branches then

$$f_{i0} > f_{i1} > \cdots > f_{i,K-1} \tag{4.4}$$

since, for any node i, the branches $b_0(i), b_1(i), \ldots, b_{K-1}(i)$ have distinct end-nodes.

Example 4.2: The binary STD from Example 4.1 has the next-node matrix

$$\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 3 & 2 \\ 1 & 0 \\ 3 & 2 \end{bmatrix}.$$

Let $\mathbf{F}$ denote the next-node matrix of a K-ary STD G. Then it is obvious that a K-ary STD G' isomorphic to G can be constructed from $\mathbf{F}$. Thus, $\mathbf{F}$ represents the isomorphism class of which G is a member. We will soon give a justification for calling certain next-node matrices 'isomorphic'.

Definition 4.4: Two $N \times K$ next-node matrices $\mathbf{F}$ and $\mathbf{F}'$ are called *isomorphic* if there is a bijective map $\mu : \{0, 1, \ldots, N-1\} \rightarrow \{0, 1, \ldots, N-1\}$ such that

$$\mathbf{F}' = \rho\left(\mathbf{P}_\mu \, \mu(\mathbf{F})\right), \tag{4.5}$$

where $\mu(\mathbf{F})$ denotes the result of applying μ to each entry of $\mathbf{F}$, where $\mathbf{P}_\mu$ is the $N \times N$ permutation matrix such that $[0, 1, \ldots, N-1]^T = \mathbf{P}_\mu \, [\mu(0), \mu(1), \ldots, \mu(N-1)]^T$, and where $\rho(\mathbf{M})$ denotes the operation of reordering each row $\underline{m}_i$ of a matrix $\mathbf{M}$ so that the elements of $\underline{m}_i$ are in non-increasing order.

The multiplication by $\mathbf{P}_\mu$ in (4.5) permutes the rows of $\mu(\mathbf{F})$ in such a way that row n becomes row $\mu(n)$ for $0 \leq n < N$. Note that

$$\mathbf{P}_\mu \, \mu(\mathbf{F}) = \mu(\mathbf{P}_\mu \, \mathbf{F}).$$

Example 4.3: We claim that the next-node matrices

$$\mathbf{F} = \begin{bmatrix} 1 & 0 \\ 3 & 2 \\ 1 & 0 \\ 3 & 2 \end{bmatrix} \quad \text{and} \quad \mathbf{F}' = \begin{bmatrix} 2 & 0 \\ 2 & 0 \\ 3 & 1 \\ 3 & 1 \end{bmatrix}$$

are isomorphic. To see this, let μ be the bijective map defined by $\mu(0) = 3$, $\mu(1) = 1$, $\mu(2) = 2$, and $\mu(3) = 0$. Then

$$\mu(\mathbf{F}) = \begin{bmatrix} 1 & 3 \\ 0 & 2 \\ 1 & 3 \\ 0 & 2 \end{bmatrix}, \quad \mathbf{P}_\mu\,\mu(\mathbf{F}) = \begin{bmatrix} 0 & 2 \\ 0 & 2 \\ 1 & 3 \\ 1 & 3 \end{bmatrix},$$

and $\mathbf{F}' = \rho(\mathbf{P}_\mu\,\mu(\mathbf{F}))$. The reader may want to check that the bijection μ defined by $\mu(0) = 0$, $\mu(1) = 2$, $\mu(2) = 1$, and $\mu(3) = 3$ yields the same $\mathbf{F}'$. ———

We write $\mathbf{F} \cong \mathbf{F}'$ to indicate that $\mathbf{F}$ and $\mathbf{F}'$ are isomorphic. It should be obvious that the relation of isomorphism between $N \times K$ next-node matrices is reflexive, symmetrical, and transitive, as required for an equivalence relation. The bijective map μ in Definition 4.4 induces an *isomorphism* η_μ of $\mathbf{F}$ onto $\mathbf{F}'$, i.e., $\mathbf{F}' = \eta_\mu(\mathbf{F})$. When $\mathbf{F} = \eta_\mu(\mathbf{F})$, η_μ is called an *automorphism* of $\mathbf{F}$.

Example 4.4: The bijection μ defined by $\mu(0) = 1$ and $\mu(1) = 0$ induces an automorphism of the next-node matrix $\mathbf{F} = \left[\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right]$. ———

The following result justifies Definition 4.4.

Proposition 4.1: Let $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$ and $G' = (\mathcal{S}, \mathcal{B}', \sigma', \epsilon')$ denote two K-ary STD's with the same node set $\mathcal{S} = \{0, 1, \ldots, N-1\}$ and let $\mathbf{F}$ and $\mathbf{F}'$ be their next-node matrices. Then G and G' are isomorphic if and only if $\mathbf{F}$ and $\mathbf{F}'$ are isomorphic. ———

Proof: Suppose that $G \cong G'$ and let (μ, β) be an isomorphism of G onto G'. Since the K successors[5] of node n in G are the elements of the n-th row of $\mathbf{F}$, viz., $\underline{f}_n = [f_{n0}, f_{n1}, \ldots, f_{n,K-1}]$, the

[5]In this proof, the 'successors' of a node n are the K not necessarily distinct end-nodes of the branches leaving node n.

successors of node $\mu(n)$ in G' are the elements of the row vector $\underline{v}_{\mu(n)} \triangleq [\mu(f_{n0}), \mu(f_{n1}), \ldots, \mu(f_{n,K-1})]$. Hence, by definition of $\mathbf{F}'$, the elements of the $\mu(n)$-th row of $\mathbf{F}'$ are the elements of $\underline{v}_{\mu(n)}$, arranged in non-increasing order. This holds for every n, $0 \leq n < N$, so that $\mathbf{F} \cong \mathbf{F}'$.

Conversely, suppose that $\mathbf{F} \cong \mathbf{F}'$ and let μ be a bijective map that induces an isomorphism of $\mathbf{F}$ onto $\mathbf{F}'$. The K branches emanating from node n in G can be numbered in such a way as $b_k(n)$, $0 \leq k < K$, that

$$\epsilon(b_k(n)) = f_{nk}, \quad 0 \leq k < K. \tag{4.6}$$

Similarly, the K branches emanating from node $\mu(n)$ in G' can be numbered in such a way as $b'_k(\mu(n))$, $0 \leq k < K$, that

$$\epsilon'(b'_k(\mu(n))) = \mu(f_{nk}), \quad 0 \leq k < K. \tag{4.7}$$

Define the bijection

$$\begin{array}{rccc} \beta: & \mathcal{B} & \rightarrow & \mathcal{B}' \\ & b_k(n) & \mapsto & b'_k(\mu(n)) \,. \end{array}$$

Applying μ to (4.6) and comparing to (4.7) shows that, for every $b \in \mathcal{B}$,

$$\mu(\epsilon(b)) = \epsilon'(\beta(b)). \tag{4.8}$$

Moreover, $\sigma(b_k(n)) = n$ and $\sigma'(b'_k(\mu(n))) = \mu(n)$ by definition of $b_k(n)$ and $b'_k(\mu(n))$ so that, for every $b \in \mathcal{B}$,

$$\mu(\sigma(b)) = \sigma'(\beta(b)). \tag{4.9}$$

But (4.8) and (4.9) are precisely the conditions for $G \cong G'$. □

We now derive crude upper bounds on the number of all non-isomorphic K-ary STD's with N nodes, which equals the number of all non-isomorphic $N \times K$ next-node matrices by virtue of Proposition 4.1. Let $\underline{f}_i$ denote the i-th row of an $N \times K$ next-node matrix and observe that the number of choices for $\underline{f}_i$ satisfying (4.3) is exactly the number of combinations of K elements from a set of N elements when repetitions are allowed and is given by $\binom{N+K-1}{K}$. Similarly, the number of choices for $\underline{f}_i$ satisfying (4.4) is the number of combinations without repetitions of K elements from a set of N elements and is given by $\binom{N}{K}$. Let $\gamma_1(N,K)$ and $\gamma_2(N,K)$ denote the number of all non-isomorphic

K-ary STD's with N nodes when parallel branches are allowed and disallowed, respectively. It follows that

$$\gamma_1(N,K) \leq \binom{N+K-1}{K}^N \quad \text{and} \quad \gamma_2(N,K) \leq \binom{N}{K}^N .$$

As an example, one finds $\gamma_1(8,2) \lesssim 2.8 \cdot 10^{12}$ and $\gamma_2(8,2) \lesssim 3.8 \cdot 10^{11}$. Unfortunately, these (loose) upper bounds cannot be tightened by dividing through the number $N!$ of isomorphisms between $N \times K$ next-node matrices, since some next-node matrices have a nontrivial automorphism group (see Example 4.4). The related problem of counting strongly connected finite automata with the same number of state transitions from every state was considered by Robinson [49, Table 2]. However, Robinson's results do not apply to our problem, since a finite automaton is defined via a state-transition function in which the assignment of the inputs to the K transitions leaving a state is relevant.

A useful topological constraint for a K-ary STD G is the *detour memory*, defined as the smallest nonnegative integer M such that G has a pair of parallel paths of length $M+1$. Equivalently, the detour memory of a K-ary STD G can be defined as the smallest nonnegative integer M such that an element of $\mathbf{A}^{M+1}$ exceeds one, where $\mathbf{A}$ is the adjacency matrix of G. A K-ary STD with $M = 0$ has parallel branches[6].

Lemma 4.1: The detour memory M of a K-ary STD G ($K \geq 2$) with N nodes, where $N \geq 1$, satisfies

$$M \leq \lfloor \log_K N \rfloor, \tag{4.10}$$

where $\log_K(\,.\,)$ denotes the logarithm to the base K.

Proof: Lemma 4.1 holds trivially if G has parallel branches, i.e., when $M = 0$. The STD G has detour memory $M \geq 1$ only if its M-th power G^M (cf. Appendix 4.A) is free of parallel branches. Observing that K^M branches are emanating from every node of G^M, we conclude that G^M can be free of parallel branches only if $K^M \leq N$ or $M \leq \lfloor \log_K N \rfloor$. □

A K-ary STD ($K \geq 2$) with detour memory M, where $M \geq 0$, and $N = K^M$ nodes achieves equality in (4.10) and is therefore said to have

[6]Such STD's exist for $K \geq 2$.

maximum detour memory. The above proof implies that such an STD is automatically strongly connected since, for any given start-node, there is exactly one path of length M to every node of the STD.

In Chapter 5, we will be particularly interested in 2^k-ary STD's, where k is a positive integer. By Lemma 4.1, the detour memory of a 2^k-ary STD ($k \geq 1$) with N nodes, where $N \geq 1$, satisfies

$$M \leq \lfloor \log_K N \rfloor = \lfloor \log_2 N / \log_2 K \rfloor = \lfloor \tfrac{1}{k} \log_2 N \rfloor. \tag{4.11}$$

Thus, a 2^k-ary STD ($k \geq 1$) with detour memory M, where $M \geq 0$, and $N = 2^{kM}$ nodes has *maximum* detour memory.

It should be noted that certain finite-state machines [48] have a 2^k-ary STD with maximum detour memory. Consider a Moore-type finite-state machine with input alphabet $\{0,1\}^k$, whose state is given by the contents of a (not necessarily feedforward) shift register with m k-bit stages, and suppose that, for any contents of the shift register, there is a one-to-one mapping from the input of the finite-state machine to the input of the shift register. Starting from any of the $N = 2^{km}$ states, the 2^k inputs generate 2^k transitions to 2^k not necessarily distinct successors. Hence, the state transitions can be described by a 2^k-ary STD G, which we now show has detour memory $M = m$. By definition, the detour memory of G is the smallest nonnegative integer M such that there is a pair of parallel paths of length $M + 1$. One of these parallel paths can be considered as the 'correct path' and the other as the 'detour'. The correct path and the detour have N-ary state sequences $(s_0, s_1, \ldots, s_{M+1})$ and $(\hat{s}_0, \hat{s}_1, \ldots, \hat{s}_{M+1})$, respectively, where $s_0 = \hat{s}_0$, $s_{M+1} = \hat{s}_{M+1}$, and $s_i \neq \hat{s}_i$, $0 < i < M$. Since the states are equal at time zero, their difference at time one must correspond to a difference in the first stage of the shift register. Such a difference remains in the shift register for m time units. The proof that $M = m$ is completed by noting that, for any given start-state $s_0 = \hat{s}_0$, two different inputs for the finite-state machine result in two different next states s_1 and $\hat{s}_1$, which is a consequence of the one-to-one mapping from the input of the finite-state machine to the input of the shift register.

In Section 4.4, we will say more about K-ary STD's with maximum detour memory.

4.2 Recursive Construction of Strongly Connected K-ary State-Transition Diagrams from Partial K-ary State-Transition Diagrams

Strongly connected K-ary state-transition diagrams can be constructed in a recursive way by successively extending nodes of partial K-ary state-transition diagrams, defined as follows.

Definition 4.5: A *partial K-ary state-transition diagram* (STD), where $K > 0$, is a digraph $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$ such that, for every node s in $\mathcal{S}$, either $d_{\text{out}}(s) = K$, or $d_{\text{out}}(s) = 0$ and $d_{\text{in}}(s) > 0$, i.e., every node has either K outgoing branches or none, in which case it has at least one entering branch. ———

A partial K-ary STD has no isolated nodes. For clarity, a K-ary STD will sometimes be called *complete*. It would be useful to know whether a partial K-ary STD can be 'grown' to a complete, strongly connected K-ary STD with N nodes in the following sense.

Definition 4.6: A partial K-ary STD is called *strongly N-connectable* if it is a subdigraph[7] of a complete, strongly connected K-ary STD with N nodes. ———

Example 4.5: Figure 4.3 shows a partial binary STD G. Note that G

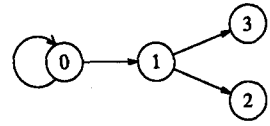


Figure 4.3: A partial binary state-transition diagram

is strongly 4-connectable since it is a subdigraph of the binary STD in Figure 4.1. ———

[7] A digraph $G = (\mathcal{S}', \mathcal{B}', \sigma', \epsilon')$ is called a *subdigraph* of a digraph $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$ if $\mathcal{S}' \subseteq \mathcal{S}$, $\mathcal{B}' \subseteq \mathcal{B}$, and each branch of G' has the same start-node and end-node in G' as in G [46, p. 125].

In Section 4.3, we will construct all non-isomorphic, strongly connected K-ary STD's with N nodes by extending partial K-ary STD's. If a partial K-ary STD is not strongly N-connectable, it will be discarded. We will thus avoid having to generate, test, and reject a large number of complete K-ary STD's that are not strongly connected.

In this section, conditions for a partial K-ary STD to be strongly N-connectable will be given in terms of its *component-reduced digraph.* In order to define the latter, we first introduce the notion of a maximal strongly connected component [47, p. 64]. Let $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$ be a digraph. For $s, s' \in \mathcal{S}$, we use the notation $s \rightleftarrows s'$ to indicate that there are finite paths from s to s' and back. By way of convention, $s \rightleftarrows s$. Clearly, '$\rightleftarrows$' is an equivalence relation. It partitions $\mathcal{S}$ into disjoint non-empty sets $\mathcal{S}_i$ (the equivalence classes), i.e., $\mathcal{S} = \bigcup_{i=1}^{m} \mathcal{S}_i$, where $1 \leq m \leq |\mathcal{S}|$. The *maximal strongly connected components* (or simply the *components*) of G are the subdigraphs $G_i = (\mathcal{S}_i, \mathcal{B}_i, \sigma, \epsilon)$ of G, $1 \leq i \leq m$, where

$$\mathcal{B}_i \triangleq \{b : \ b \in \mathcal{B},\ \sigma(b) \in \mathcal{S}_i \text{ and } \epsilon(b) \in \mathcal{S}_i\},$$

i.e., $\mathcal{B}_i$ is the subset of those branches in $\mathcal{B}$ whose start-node and end-node are both in $\mathcal{S}_i$. Note that a branch set $\mathcal{B}_i$ can be empty and that $\mathcal{B} \supseteq \bigcup_{i=1}^{m} \mathcal{B}_i$.

Definition 4.7: For any digraph $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$ whose maximal strongly connected components are denoted by $G_i = (\mathcal{S}_i, \mathcal{B}_i, \sigma, \epsilon)$, $1 \leq i \leq m$, the *component-reduced digraph* (CRD) of G is the digraph $G_c = (\mathcal{S}_c, \mathcal{B}_c)$, where $\mathcal{S}_c \triangleq \{\mathcal{S}_1, \mathcal{S}_2, \ldots, \mathcal{S}_m\}$ and $\mathcal{B}_c \triangleq \{(\mathcal{S}_i, \mathcal{S}_j) : \ i \neq j$, there exists a $b \in \mathcal{B}$ with $\sigma(b) \in \mathcal{S}_i$ and $\epsilon(b) \in \mathcal{S}_j\}$.

Hence, the CRD G_c of a digraph G is obtained by collapsing every component G_i of G to a node $\mathcal{S}_i$ of G_c, i.e., by letting the sets $\mathcal{S}_i$ become the nodes of G_c and by providing a branch from node $\mathcal{S}_i$ to node $\mathcal{S}_j$ for every pair $(\mathcal{S}_i, \mathcal{S}_j)$, $i \neq j$, for which G contains a branch from a node in $\mathcal{S}_i$ to a node in $\mathcal{S}_j$. By definition, the CRD has no parallel branches and no self-loops. Moreover, the CRD is *acyclic* since if there was a cycle that connects two different nodes $\mathcal{S}_j$ and $\mathcal{S}_k$ in G_c, all nodes of G contained in the sets $\mathcal{S}_j$ and $\mathcal{S}_k$ would be connected by a path in each direction, in contradiction with the definition of these sets.

Example 4.6: For the partial binary STD G in Figure 4.4, the node equivalence classes $\mathcal{S}_i$ induced by the relation '$\rightleftarrows$' are the sets $\{0\}$, $\{1,2\}$, $\{3\}$, $\{4\}$, $\{5\}$, and $\{6\}$. Notice that the CRD G_c of G (also shown

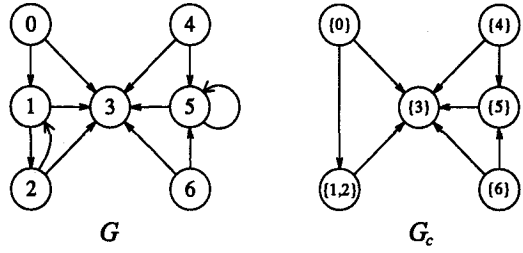


Figure 4.4: A partial binary STD G and its CRD G_c

in Figure 4.4) is acyclic.

Consider now the CRD $G_c = (\mathcal{S}_c, \mathcal{B}_c)$ of a partial K-ary STD $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$. The nodes of G_c can be classified according to their out-degrees and in-degrees. Node $\mathcal{S}_i \in \mathcal{S}_c$ is called a *source node* if $d_{\text{out}}(\mathcal{S}_i) > 0$ and $d_{\text{in}}(\mathcal{S}_i) = 0$, an *intermediate node* if $d_{\text{out}}(\mathcal{S}_i) > 0$ and $d_{\text{in}}(\mathcal{S}_i) > 0$, a *sink node* if $d_{\text{out}}(\mathcal{S}_i) = 0$ and $d_{\text{in}}(\mathcal{S}_i) > 0$, and an *isolated node* if $d_{\text{out}}(\mathcal{S}_i) = d_{\text{in}}(\mathcal{S}_i) = 0$. Every unextended node s in G constitutes a sink node $\mathcal{S}_i = \{s\}$ in G_c since $d_{\text{out}}(s) = 0$ implies $d_{\text{in}}(s) > 0$ according to Definition 4.5. Such a sink node will be referred to as a *resource node* because it represents an extendable node of G. Every other sink node or isolated node will be called a *dead node* because it represents a component of G that neither can be left on any path nor contains an extendable node. We now state the main result of this section, whose proof can be found in Appendix 4.B.

Theorem 4.1: Let G be a partial K-ary STD with $V \geq 1$ nodes and let G_c be its CRD with S source nodes, R resource nodes, and D dead nodes. Then G is strongly N-connectable for some $N \geq V$ if and only if

$$D = 0 \quad \text{and} \quad RK + (N - V)(K - 1) \geq S \tag{4.12}$$

or

$$D = 1, \quad S = 0, \quad \text{and} \quad V = N. \tag{4.13}$$

Example 4.6 (cont.): The partial binary STD G in Figure 4.4 has $V = 7$ nodes. Its CRD G_c has $S = 3$ source nodes ($\{0\}$, $\{4\}$, and $\{6\}$), $I = 2$ intermediate nodes ($\{1, 2\}$ and $\{5\}$), $R = 1$ resource node ($\{3\}$), and $D = 0$ dead nodes. Note that (4.12) is satisfied for $N = 8$, but not for $N = 7$, i.e., G is strongly 8-connectable, but not strongly 7-connectable. Indeed, G can be extended to a complete strongly connected binary STD with $N = 8$ nodes by appending a binary subtree with three leaves to node 3, such that the three leaves coincide with the nodes 0, 4, and 6. ———

In the remainder of this section, we show how to determine the parameters in Theorem 4.1 recursively. Let G_n denote a partial K-ary STD with n extended nodes, which are numbered from 0 to $n-1$, and let G_{c_n} be its CRD. Suppose that we extend node n of G_n by adding the branches $\mathcal{B}_{\text{out}}(n)$, which yields G_{n+1}. In order to apply Theorem 4.1 to G_{n+1}, we have to find its CRD $G_{c_{n+1}}$. One solution is to determine $G_{c_{n+1}}$ directly from G_{n+1} using the *Tarjan algorithm*[8] [50], [47], [51]. However, it is much more efficient to determine $G_{c_{n+1}}$ recursively from G_{c_n} using a simplified version of the Tarjan algorithm, as we now show.

Let G_n be a strongly N-connectable partial K-ary STD and let the unextended nodes of G_n be numbered contiguously with the extended nodes. This ensures that, for $1 \leq n < N$, node n has $d_{\text{in}}(n) > 0$ in G_n so that node $\{n\}$ is a resource node in G_{c_n}. Let G_{c_0} be the empty digraph, i.e., let $\mathcal{S}_{c_0} \triangleq \{\}$ and $\mathcal{B}_{c_0} \triangleq \{\}$. The following algorithm for updating the CRD consists of two steps. First, node $\{n\}$ of $G_{c_n} = (\mathcal{S}_{c_n}, \mathcal{B}_{c_n})$ is extended, which yields a digraph $E = (\mathcal{S}, \mathcal{B})$, and second, E is reduced to the desired CRD $G_{c_{n+1}}$.

Algorithm 4.1 (Updating the CRD): Let G_n, where $0 \leq n < N$, denote a partial K-ary STD with n extended nodes, numbered from 0 to $n-1$, and let $G_{c_n} = (\mathcal{S}_{c_n}, \mathcal{B}_{c_n})$ be its CRD. Further, let G_{n+1} be the partial K-ary STD that results from G_n by extending node n, i.e., by adding the branches $\mathcal{B}_{\text{out}}(n)$. Then the CRD $G_{c_{n+1}}$ can be obtained from G_{c_n} as follows.

Step 1: Determine an extended digraph $E = (\mathcal{S}, \mathcal{B})$: Initialize $\mathcal{S} \leftarrow \mathcal{S}_{c_n}$ and $\mathcal{B} \leftarrow \mathcal{B}_{c_n}$. If $n = 0$, add the node[9] $\{0\}$ to $\mathcal{S}$. Extend node

[8] The Tarjan algorithm is a depth-first search method for finding the maximal strongly connected components of a digraph.

[9] Recall that the nodes of a component-reduced digraph are disjoint sets.

$\{n\}$ of E as follows. For every branch $b \in \mathcal{B}_{\text{out}}(n)$ whose end-node $\epsilon(b)$ is not contained in any node of $\mathcal{S}$, add the node $\{\epsilon(b)\}$ to $\mathcal{S}$. Finally, for every branch $b \in \mathcal{B}_{\text{out}}(n)$ that is not a self-loop, add the branch $(\{n\}, \mathcal{S}_{\epsilon(b)})$ to $\mathcal{B}$, where $\mathcal{S}_{\epsilon(b)}$ is that node of $\mathcal{S}$, which contains $\epsilon(b)$.

Step 2: *Reduce E to $G_{c_{n+1}}$:* Starting from node $\{n\}$ of E, run the Tarjan algorithm to find the maximal strongly connected component containing $\{n\}$. Let $\mathcal{S}_n$ be the subset of those nodes of G_{n+1}, which are contained in the nodes of this component. Collapse the component just found to one node $\mathcal{S}_n$ by redirecting branches appropriately and by removing self-loops.

Remarks on Algorithm 4.1:

- The branches added to G_{c_n} in Step 1 may have completed one or more cycles through node $\{n\}$ in E. Since G_{c_n} is acyclic, there can be no other cycles in E. The node $\{n\}$ together with all other nodes on these cycles constitute a maximal strongly connected component in E, which can be found using the Tarjan algorithm[10].
- In Step 2, parallel branches are removed automatically since there can be at most one branch from one node of E to another.
- In Step 2, the collapsed node $\mathcal{S}_n$ becomes a dead node if it has no successor; otherwise, it becomes an intermediate node if it has a predecessor or a source node if it has none.
- The parameters S, R, and D needed in Theorem 4.1 can be updated as follows. Increase R by the number of new nodes added in Step 1. Decrease S and R according to the types of the nodes merged in Step 2. Increment S (or D) if the collapsed node $\mathcal{S}_n$ is a source node (or dead node). ---

[10]The Tarjan algorithm uses a stack for storing the nodes in the order they are visited first during a depth-first search. Each time a maximal strongly connected component is found, the corresponding nodes are popped from the stack. The fact that $\{n\}$ is the last node popped from the stack is used as the stopping criterion.

Example 4.7: Algorithm 4.1 is illustrated in Figure 4.5 for $K = 2$ and $N = 4$. For $n = 1$ and $n = 2$, the extended digraph E is already in

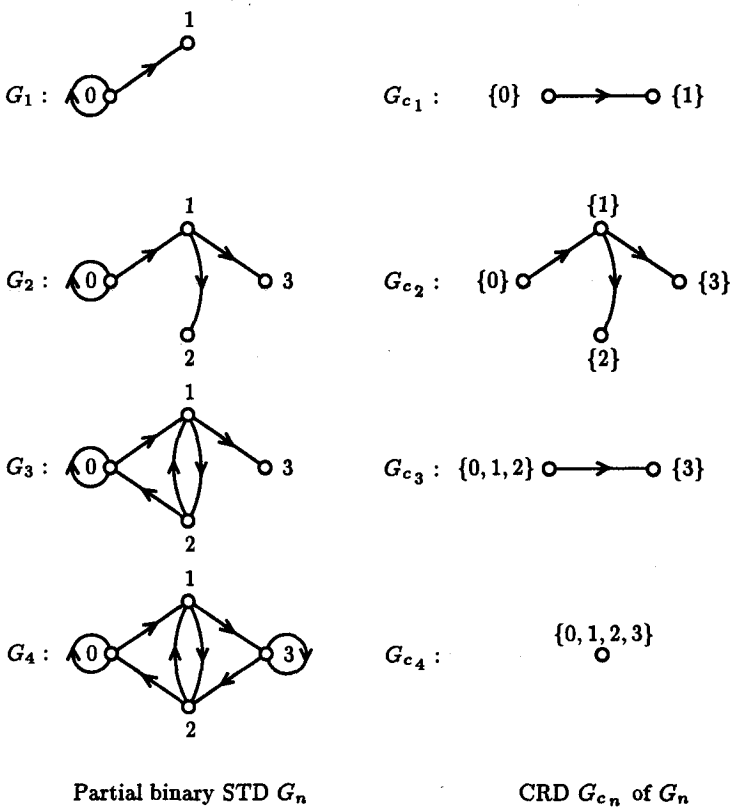


Figure 4.5: Illustration of Algorithm 4.1

component-reduced form and is therefore not modified by Step 2. For $n = N$, observe that the CRD G_{c_n} of G_n is a trivial digraph with one (dead) node and no branches. ———

4.3 Systematic Construction of All Non-Isomorphic K-ary State-Transition Diagrams with N Nodes and Given Topological Constraints

The concept of extending partial STD's[11] is developed further in this section to an algorithm for the systematic construction of all non-isomorphic STD's with N nodes and given topological constraints. We wish to generate a sequence of N ordered lists of partial STD's, where the order relation is yet to be defined and where the n-th list, $1 \leq n \leq N$, contains partial STD's with n extended nodes. The idea is to construct these lists sequentially as follows. Each entry in the list of partial STD's with $n-1$ extended nodes is extended in all combinatorially possible ways. An extended partial STD is appended to the list of partial STD's with n extended nodes when it satisfies the given constraints; otherwise, it is discarded. Clearly, the STD isomorphism classes that can be generated by extending a partial STD G can be generated also by extending a partial STD G' isomorphic to G. Hence, only non-isomorphic partial STD's are retained in the lists.

In this section, the node set of a partial STD will always be a subset of $\{0, 1, \ldots, N-1\}$ (where N is a fixed positive integer) and, for a partial STD with n extended nodes, $1 \leq n \leq N$, the subset of extended nodes will always be the set

$$\mathcal{E} = \{0, 1, \ldots, n-1\}. \tag{4.14}$$

Since *every* partial STD with n extended nodes is isomorphic to a partial STD whose extended nodes are given by (4.14), no isomorphism class will be excluded by imposing (4.14).

We now give a compact matrix description of a partial STD G whose extended nodes are given by the set $\mathcal{E}$ in (4.14). For the i-th extended node, where $0 \leq i < n$, the branches emanating from node i can be numbered in any order from 0 to $K-1$. In particular, they can be numbered as $b_j(i)$, $0 \leq j < K$, where the end-nodes of $b_0(i), b_1(i), \ldots, b_{K-1}(i)$ are in non-increasing order, i.e.,

$$\epsilon(b_0(i)) \geq \epsilon(b_1(i)) \geq \ldots \geq \epsilon(b_{K-1}(i)).$$

[11] In this section, 'STD' always refers to a K-ary STD.

We arrange the nK numbers $f_{ij} \triangleq \epsilon(b_j(i))$, $0 \le i < n$, $0 \le j < K$, as an $n \times K$ matrix $\mathbf{F} = [f_{ij}]$, called the *partial next-node matrix* of G. Since the node set of G is a subset of $\{0, 1, \ldots, N-1\}$, the elements of $\mathbf{F}$ are also in that subset. By definition, every row $\underline{f}_i$ of $\mathbf{F}$ satisfies (4.3) and, when G has no parallel branches, the stronger condition (4.4). Notice that the set of unextended nodes of G can be written as

$$\mathcal{U} = \{f_{ij} : 0 \le i < n, 0 \le j < K, f_{ij} \ge n\}. \tag{4.15}$$

We will soon give a justification for the following definition[12].

Definition 4.8: Two $n \times K$ partial next-node matrices $\mathbf{F}$ and $\mathbf{F}'$ (where $1 \le n \le N$) are called *isomorphic* if there is a bijective map $\mu : \{0, 1, \ldots, N-1\} \to \{0, 1, \ldots, N-1\}$, which satisfies $\mu(i) < n$ for all $i < n$, such that

$$\mathbf{F}' = \rho\left(\mathbf{P}_\mu \, \mu(\mathbf{F})\right), \tag{4.16}$$

where $\mu(\mathbf{F})$ denotes the result of applying μ to each entry of $\mathbf{F}$, where $\mathbf{P}_\mu$ is the $n \times n$ permutation matrix such that $[0, 1, \ldots, n-1]^T = \mathbf{P}_\mu \, [\mu(0), \mu(1), \ldots, \mu(n-1)]^T$, and where $\rho(\mathbf{M})$ denotes the operation of reordering each row $\underline{m}_i$ of a matrix $\mathbf{M}$ so that the elements of $\underline{m}_i$ are in non-increasing order.

We write $\mathbf{F} \cong \mathbf{F}'$ to indicate that $\mathbf{F}$ and $\mathbf{F}'$ are isomorphic. The bijective map μ in Definition 4.8 induces an *isomorphism* η_μ of $\mathbf{F}$ onto $\mathbf{F}'$, i.e., $\mathbf{F}' = \eta_\mu(\mathbf{F})$. When $\mathbf{F} = \eta_\mu(\mathbf{F})$, η_μ is called an *automorphism* of $\mathbf{F}$. The following slight generalization of Proposition 4.1 justifies Definition 4.8.

Proposition 4.2: Let $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$ and $G' = (\mathcal{S}', \mathcal{B}', \sigma', \epsilon')$ denote two partial K-ary STD's with n extended nodes (where $1 \le n \le N$) given by the set $\mathcal{E} = \{0, 1, \ldots, n-1\}$, a subset of both $\mathcal{S}$ and $\mathcal{S}'$. Then G and G' are isomorphic if and only if their $n \times K$ partial next-node matrices $\mathbf{F}$ and $\mathbf{F}'$ are isomorphic.

Example 4.8: Let $N = 3$. The partial next-node matrices $\mathbf{F} = \left[\begin{smallmatrix} 1 & 0 \\ 2 & 0 \end{smallmatrix}\right]$ and $\mathbf{F}' = \left[\begin{smallmatrix} 2 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ are isomorphic since $\mathbf{F}' = \eta_\mu(\mathbf{F})$, where the bijective map μ is defined by $\mu(0) = 1$, $\mu(1) = 0$, and $\mu(2) = 2$. Figure 4.6 shows two partial binary STD's G and G' with partial next-node matrix $\mathbf{F}$

[12] See also Definition 4.4.

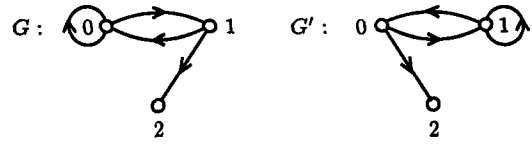


Figure 4.6: Isomorphic partial binary STD's

and $\mathbf{F}'$, respectively. As predicted by Proposition 4.2, G and G' are isomorphic.

The algorithm proposed in this section generates a sequence of ordered lists of partial next-node matrices. For enumerating and ordering partial next-node matrices, we define order relations as follows. For an $n \times K$ partial next-node matrix $\mathbf{F}$ whose rows are denoted by $\underline{f}_i$, let $\underline{f}_0, \underline{f}_1, \ldots, \underline{f}_{n-1}$ be the order of decreasing significance. Similarly, for the i-th row $\underline{f}_i = [f_{i0}, f_{i1}, \ldots, f_{i,K-1}]$ of $\mathbf{F}$, let $f_{i0}, f_{i1}, \ldots, f_{i,K-1}$ be the order of decreasing significance. Thus, the *order relations* are defined by

$$\begin{aligned} \mathbf{F}' > \mathbf{F} \iff & (\underline{f}'_0 > \underline{f}_0) \vee \\ (\underline{f}'_0 = \underline{f}_0) \wedge & [(\underline{f}'_1 > \underline{f}_1) \vee \ldots \\ & \vdots \\ (\underline{f}'_{n-2} = \underline{f}_{n-2}) \wedge & (\underline{f}'_{n-1} > \underline{f}_{n-1})] \ldots] \end{aligned} \tag{4.17}$$

and

$$\begin{aligned} \underline{f}'_i > \underline{f}_i \iff & (f'_{i0} > f_{i0}) \vee \\ (f'_{i0} = f_{i0}) \wedge & [(f'_{i1} > f_{i1}) \vee \ldots \\ & \vdots \\ (f'_{i,K-2} = f_{i,K-2}) \wedge & (f'_{i,K-1} > f_{i,K-1})] \ldots], \end{aligned} \tag{4.18}$$

where '$\vee$' and '$\wedge$' are the boolean 'or' and 'and' operators.

Let $\mathcal{L}_{n-1} = \{\mathbf{F}^{(1)}_{n-1}, \mathbf{F}^{(2)}_{n-1}, \ldots, \mathbf{F}^{(l_{n-1})}_{n-1}\}$ denote a list of $(n-1) \times K$ partial next-node matrices, ordered in such a way that $i < j$ implies $\mathbf{F}^{(i)}_{n-1} < \mathbf{F}^{(j)}_{n-1}$. A list $\mathcal{L}_n = \{\mathbf{F}^{(1)}_n, \mathbf{F}^{(2)}_n, \ldots, \mathbf{F}^{(l_n)}_n\}$ is

generated by 'extending' $\mathcal{L}_{n-1}$ as follows. We first let $\mathcal{L}_n$ be the empty list. Beginning with $i = 1$, we extend $\mathbf{F}_{n-1}^{(i)}$ to

$$\mathbf{F}_n = \begin{bmatrix} \mathbf{F}_{n-1}^{(i)} \\ \underline{f}_{n-1} \end{bmatrix}, \tag{4.19}$$

for every row vector $\underline{f}_{n-1}$ satisfying (4.3) or (4.4). The row vectors $\underline{f}_{n-1}$, and thus also the extended matrices $\mathbf{F}_n$, are generated with increasing order according to (4.18). Extended matrices that pass all topological tests and that are not isomorphic to any entry already in the list $\mathcal{L}_n$ are appended to this list in the order they are generated. The process of extending $\mathbf{F}_{n-1}^{(i)}$ is repeated for $i = 2, 3, \ldots, l_{n-1}$. Observe that the list $\mathcal{L}_n$ is ordered as well, i.e., $i < j$ implies $\mathbf{F}_n^{(i)} < \mathbf{F}_n^{(j)}$, since the last row is the least significant row.

It is desirable to check whether a partial next-node matrix represents a strongly N-connectable partial STD. To do such a check efficiently, we also maintain a list $\mathcal{L}_{c_n} = \{G_{c_n}^{(1)}, G_{c_n}^{(2)}, \ldots, G_{c_n}^{(l_n)}\}$, where $G_{c_n}^{(i)}$ is the component-reduced digraph (CRD)[13] corresponding to to $\mathbf{F}_n^{(i)}$. The list $\mathcal{L}_{c_n}$ is obtained by updating the entries of $\mathcal{L}_{c_{n-1}}$ according to Algorithm 4.1.

The complete algorithm starts from $\mathcal{L}_0 = \{\mathbf{F}_0\}$ and $\mathcal{L}_{c_0} = \{G_{c_0}\}$, where $\mathbf{F}_0$ is a dummy matrix with zero rows and G_{c_0} is the empty digraph, and extends $\mathcal{L}_{n-1}$ to $\mathcal{L}_n$ and $\mathcal{L}_{c_{n-1}}$ to $\mathcal{L}_{c_n}$ for $n = 1, 2, \ldots, N$.

In the remainder of this section, we show how to do the topological tests and how to recognize isomorphisms. We first summarize the topological tests. Let the $n \times K$ matrix (4.19) be the candidate to be examined and let G_{c_n} be its CRD. As a first test, we will require that, for some $u \geq 0$,

$$\mathcal{U} = \{n, n+1, \ldots, n+u-1\}, \tag{4.20}$$

where $\mathcal{U}$ is the set of unextended nodes given by (4.15). This test will greatly reduce the number of candidates to be processed further, without excluding any isomorphism class. Another test will reject candidates violating the detour-memory constraint. The detour memory of an STD, which is constructed from a *partial* STD with adjacency matrix $\mathbf{A}$, is easily seen to be upper-bounded by the smallest nonnegative integer m such that an element of $\mathbf{A}^{m+1}$ exceeds one. Note that

[13]Note that the CRD is uniquely determined by the partial next-node matrix.

the adjacency matrix is uniquely determined by the partial next-node matrix.

The candidate $\mathbf{F}_n$ is discarded in any of the following cases.

(i) The constraint (4.20) is violated.

(ii) A uniform in-degree is desired and, for some j, $0 \leq j < N$, the number of elements of $\mathbf{F}_n$ equal to j exceeds K.

(iii) A detour memory $M > 0$ is desired and the smallest nonnegative integer m such that an element of $\mathbf{A}^{m+1}$ exceeds one is less than M.

(iv) The partial STD represented by $\mathbf{F}_n$ is not strongly N-connectable.

The following algorithm searches the list $\mathcal{L}_n$ for a partial next-node matrix that is isomorphic to a given candidate $\mathbf{F}_n$.

Algorithm 4.2 (Search for an Isomorphic Partial Next-node Matrix): Let $\mathcal{L}_n = \{\mathbf{F}_n^{(1)}, \mathbf{F}_n^{(2)}, \ldots, \mathbf{F}_n^{(L)}\}$ be a list of non-isomorphic $n \times K$ partial next-node matrices ($1 \leq n \leq N$) with the property (4.20), and let the list be ordered in such a way that $i < j$ implies $\mathbf{F}_n^{(i)} < \mathbf{F}_n^{(j)}$. Let $\mathbf{F}_n = [f_{ij}]$ be a candidate $n \times K$ partial next-node matrix with the property (4.20) for some $u \geq 0$. Finally, let $\mathcal{M} = \{\mu_1, \mu_2, \ldots, \mu_J\}$ be a set of bijective maps $\mu_j : \{0, 1, \ldots, N-1\} \to \{0, 1, \ldots, N-1\}$. The list $\mathcal{L}_n$ is searched for an entry isomorphic to $\mathbf{F}_n$ as described by the

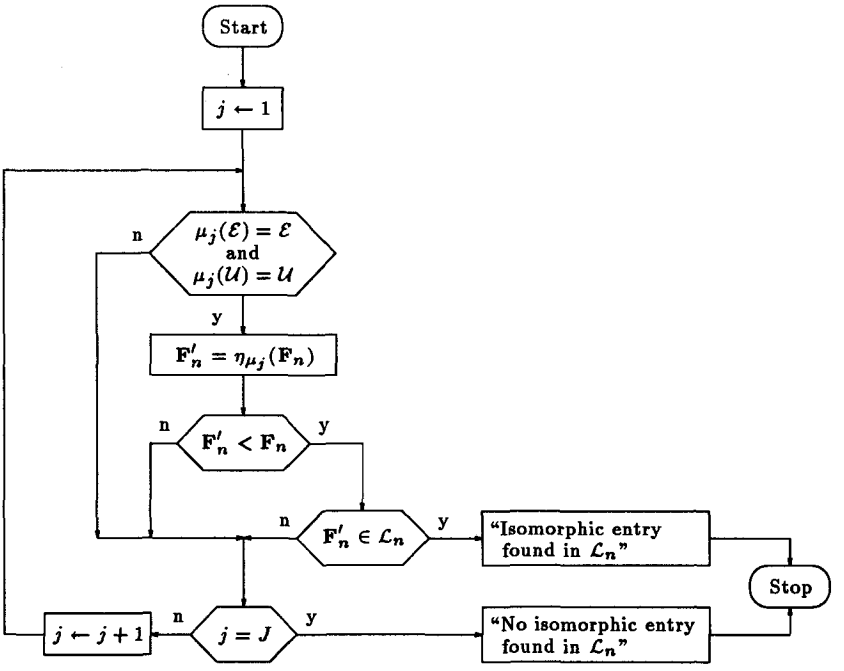


Figure 4.7: Searching the list $\mathcal{L}_n$ for an isomorphic partial next-node matrix

flowchart in Figure 4.7.

Remarks on Algorithm 4.2:

- Since the entries of $\mathcal{L}_n$ and the candidate $\mathbf{F}_n$ have the property (4.20), any bijective map μ_j that induces an isomorphism of $\mathbf{F}_n$ onto an entry of $\mathcal{L}_n$ satisfies

$$\mu_j(\mathcal{E}) = \mathcal{E} \quad \text{and} \quad \mu_j(\mathcal{U}) = \mathcal{U}, \tag{4.21}$$

  where $\mathcal{E} = \{0, 1, \ldots, n-1\}$ and $\mathcal{U} = \{n, n+1, \ldots, n+u-1\}$.

- When $\mathcal{M}$ is chosen as the set of all $n!u!$ bijective maps satisfying condition (4.21)[14], the result "No isomorphic entry found in $\mathcal{L}_n$" implies that no isomorphic entry *exists* in $\mathcal{L}_n$. For $N > 8$, it becomes unfeasible to check all $n!\,u!$ bijective maps for $1 \leq n \leq N$. Fortunately, a much smaller set of bijective maps exists, which still allows the detection of all isomorphisms. Such a solution will be presented in Section 4.4 for searching all non-isomorphic binary STD's with $N = 16$ nodes and maximum detour memory.

- Due to the fact that the list $\mathcal{L}_n$ is ordered, the test $\mathbf{F}'_n \in \mathcal{L}_n$ in Figure 4.7 can be implemented efficiently as a binary search [51]. Nevertheless, the computational costs for searching an isomorphic entry were found to be considerably higher than those for doing the topological tests.

[14] In that case, condition (4.21) need not be checked in Figure 4.7.

4.4 K-ary State-Transition Diagrams with Maximum Detour Memory

In this section, our interest is in all non-isomorphic binary ($K = 2$) STD's with $N = 2^M$ nodes, $M \geq 0$, and *maximum detour memory*, which is M according to (4.11). The results in Tables 4.1 to 4.4 were obtained as described in Section 4.3. For $N \leq 8$, the set $\mathcal{M}$ in Algorithm 4.2 was chosen as the set of all $n!\,u!$ bijective maps satisfying condition (4.21).

STD No.	Successors of node 0
1	0, 0

Table 4.1: The only binary STD with 1 node and detour memory 0.

STD No.	Successors of node 0	1
1	1, 0	1, 0

Table 4.2: The only binary STD with 2 nodes and detour memory 1.

STD No.	Successors of node 0	1	2	3
1	1,0	3,2	1,0	3,2

Table 4.3: The only binary STD with 4 nodes and detour memory 2.

STD No.	Successors of node 0	1	2	3
1...3	1,0	3,2	5,4	7,6

STD No.	Successors of node 4	5	6	7
1	1,0	3,2	5,4	7,6
2	1,0	6,3	5,4	7,2
3	1,0	6,5	7,4	3,2

Table 4.4: All non-isomorphic binary STD's with 8 nodes and detour memory 3.

We know from Section 4.1 that binary STD's with maximum detour memory are *automatically strongly connected.* (Nevertheless, rejecting partial binary STD's that are not strongly N-connectable was still useful for reducing the number of candidates with 1 to $N-1$ extended nodes.) Observe that all STD's in Tables 4.1 to 4.4 have exactly two self-loops and uniform in-degree, although they were not constrained to have these

properties. The proof of the author's conjecture that, for $N = 2^M$, every binary STD with maximum detour memory has exactly two self-loops, and the generalization to K-ary STD's, is gratefully acknowledged to Gerhard Krämer. His proof encouraged the author to verify a number of additional conjectures. The properties of K-ary STD's with maximum detour memory are summarized in

Theorem 4.2: Let G denote a K-ary STD with $N = K^M$ nodes and detour memory M, where $K \geq 2$ and $M > 0$. Then, from any given start-node, G contains exactly one path of length M and K parallel paths of length $M + 1$ to each of its nodes. Moreover, G has *exactly* K *self-loops* and *uniform in-degree* K.

Notice also the singular case of a K-ary STD ($K \geq 2$) with one node and detour memory zero. Such an STD has obviously K parallel paths of length one, K self-loops, and uniform in-degree K.

Before we turn to the proof of Theorem 4.2, a few immediate consequences should be mentioned. The existence of a path of length M between any two nodes implies that G is strongly connected. Moreover, the presence of self-loops implies that G is aperiodic. Note that reversing the branch directions of a K-ary STD with uniform in-degree and detour memory M yields also a K-ary STD with uniform in-degree and detour memory M. We thus have the following consequence of Theorem 4.2[15].

Corollary 4.1: Let G denote a K-ary STD with $N = K^M$ nodes and detour memory M, where $K \geq 2$ and $M \geq 0$. Then the digraph obtained by reversing all branch directions of G is also a K-ary STD with detour memory M.

Proof of Theorem 4.2 (Part 1): Choose any integer m, where $1 \leq m \leq M$. Let $\mathbf{A}$ denote the adjacency matrix of G. Clearly, the i-th row sum of $\mathbf{A}^m$ equals the number of paths of length m starting at node i, which must be K^m. On the other hand, no element of $\mathbf{A}^m$ exceeds one by definition of the detour memory. These facts imply that every row of $\mathbf{A}^m$ contains K^m ones and $N - K^m$ zeros. In particular, $\mathbf{A}^M$ is the all-ones matrix. Thus, for any given start-node, there is exactly one path of length M to every node of G. It is easy to check that

[15] Corollary 4.1 holds trivially for $M = 0$.

$\mathbf{A}^{M+1}$ equals K times the all-ones matrix. Thus, from any given start-node, there are exactly K parallel paths of length $M+1$ to every node of G. Note that $\mathbf{A}^M$ has rank 1 since $\mathbf{A}^M = \underline{u}\,\underline{u}^T$, where $\underline{u}$ is the all-ones vector. It is well-known that an $N \times N$ matrix of rank 1, viz., $\underline{v}_1\,\underline{v}_2^T$, has the characteristic polynomial $\lambda^{N-1}(\lambda - \underline{v}_2^T\underline{v}_1)$ [42, p. 651 and p. 658][16]. Hence

$$\det\left(\lambda\mathbf{I} - \mathbf{A}^M\right) = \lambda^{N-1}(\lambda - N) = \lambda^{N-1}(\lambda - K^M). \tag{4.22}$$

Observe that $\lambda_1 = K$ is an eigenvalue of $\mathbf{A}$ (with eigenvector $\underline{u}$). Moreover, (4.22) and [38, Thm. 9.4.6, p. 312] imply that the remaining eigenvalues of $\mathbf{A}$ are zero, i.e., $\lambda_n = 0$, $2 \leq n \leq N$. But

$$\operatorname{tr}\mathbf{A} = \sum_{n=1}^{N} \lambda_n = K$$

so that $\mathbf{A}$ must have K ones on its main diagonal and G must have K self-loops[17]. □

The uniform in-degree was first proved by the author using graph isomorphisms. His graph-theoretic argument is postponed to the following simple proof by Gerhard Krämer.

Proof of Theorem 4.2 (Part 2): Note that a K-ary STD G has uniform out-degree (in-degree) if and only if all row-sums (column-sums) of its adjacency matrix $\mathbf{A}$ equal K. Consider the identities

$$\mathbf{A}\mathbf{A}^M = \mathbf{A}^{M+1} = \mathbf{A}^M\mathbf{A}.$$

Since $\mathbf{A}^M$ is the all-ones matrix and since the row-sums of $\mathbf{A}$ are K, $\mathbf{A}^{M+1}$ is the all-K's matrix. But the second identity implies that the column-sums of $\mathbf{A}$ are K and thus that G has uniform in-degree K. □

The following lemma is required for the author's graph-theoretic proof of the uniform in-degree and will also lead to a shortcut in the search for isomorphic K-ary STD's.

Lemma 4.2: For $K \geq 2$ and $M > 0$, every K-ary STD with $N = K^M$ nodes and detour memory M is isomorphic to an STD G_N with node

[16] The converse is not true in general, i.e., an $N \times N$ matrix ($N \geq 2$) with characteristic polynomial $\lambda^{N-1}(\lambda - \alpha)$ may have a rank greater than one.

[17] Similarly, one can show that G^m, $m \geq 1$, has K^m self-loops.

set $\mathcal{S}_N = \{0, 1, \ldots, N-1\}$ such that G_N contains a partial K-ary STD $G_{N/K}$ as a subdigraph, where $G_{N/K}$ has the partial next-node matrix $\mathbf{F}_{N/K} = [f_{nk}]$ given by

$$f_{nk} = K(n+1) - k - 1, \quad 0 \leq n < N/K, \quad 0 \leq k < K. \tag{4.23}$$

Example 4.9: For $K = 2$, $M = 3$, and thus $N = 8$, equation (4.23) yields the partial next-node matrix[18] $\mathbf{F}_4 = [1, 0;\ 3, 2;\ 4, 3;\ 7, 6]$. The partial STD G_4 is shown in Figure 4.8.

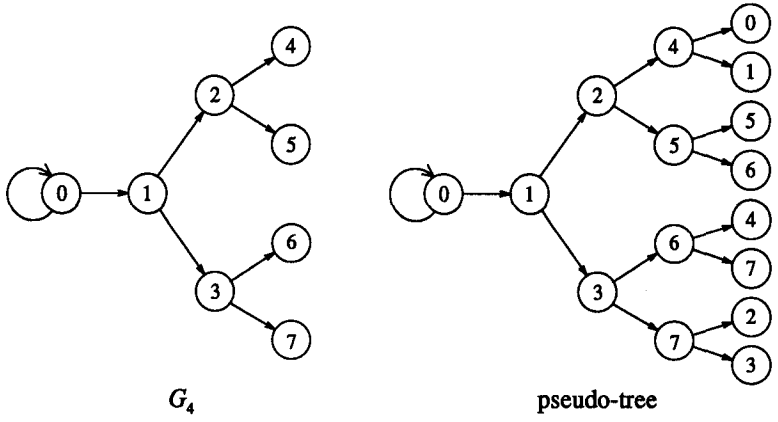


Figure 4.8: A partial binary STD G_4 and the pseudo-tree of depth 4 obtained by extending nodes 4 to 7 according to STD No. 3 in Table 4.4.

Observe that each of the $N = K^M$ nodes of $\mathcal{S}_N$ appears exactly once in $\mathbf{F}_{N/K}$ and that $G_{N/K}$ has a tree-like topology with one node at depth 0 and $(K-1)K^{m-1}$ nodes at depth m, $1 \leq m \leq M$.

Proof of Lemma 4.2: The proof of Theorem 4.2 (Part 1) showed that the K-ary STD's considered have K self-loops. Hence, we can choose a G_N having one self-loop at node 0. The other successors of node 0 must

[18] In this section, the rows of a matrix are sometimes written on one line and separated by semicolons.

be different nodes. We are free to choose $\mathbf{F}_1 = [K-1, K-2, \ldots, 0]$. (Notice that the index of $\mathbf{F}_n$ is the number of its rows, which is also the number of extended nodes of G_n.) If $M = 1$ and thus $N = K$, we are done. If $M \geq 2$ and thus $N \geq K^2$, we extend the nodes 1 to $K-1$. We may not choose any node already extended as a successor of node 1 to $K-1$, since there would be two parallel paths of length 2 from node 0 to such a successor and that would imply $M \leq 1$. Choosing the $(K-1)K$ nodes $K, K+1, \ldots, K^2-1$ as the successors, we obtain the K-row matrix

$$\mathbf{F}_K = [K-1, \ldots, 0;\ 2K-1, \ldots, K;\ \ldots\ ;\ K^2-1, \ldots, K^2-K].$$

If $M = 2$ and thus $N = K^2$, we are done. If $M \geq 3$ and thus $N \geq K^3$, we extend the nodes K to K^2-1. Again, we may not choose any node already extended as a successor, since this would imply $M \leq 2$. By continuing in this way, we finally obtain $\mathbf{F}_{N/K}$. □

We are now prepared for the graph-theoretic proof of the uniform in-degree.

Proof of Theorem 4.2 (Part 2): Lemma 4.2 shows that every K-ary STD G with $N = K^M$ nodes and detour memory M is isomorphic to a K-ary STD G_N with a tree-like partial K-ary STD $G_{N/K}$ as a subdigraph. Recall that G_N and $G_{N/K}$ have the same node set $\mathcal{S}_N$. Hence, G_N can be obtained by extending every node of $G_{N/K}$ at depth M. Observe that the nodes 1 to $K-1$ in $G_{N/K}$ are the root nodes of $K-1$ K-ary trees of depth $M-1$. Thus, the nodes yet to be extended are the nodes of these $K-1$ trees at tree depth $M-1$. For one of these trees, extending the K^{M-1} nodes at depth $M-1$ requires $N = K^M$ successors. But these successors must be distinct, for otherwise the detour memory would be smaller than M. Hence, the set of successors must be the node set $\mathcal{S}_N$. Repeating the extension for all $K-1$ trees yields a 'pseudo-tree' of depth $M+1$ as shown in Figure 4.8 for $K = 2$. Since every node appears exactly K times in this pseudo-tree, every node of G_N has in-degree K. Hence, G has uniform in-degree K. □

The fact that there are only 3 binary STD's with $N = 8$ nodes and maximum detour memory indicates that the latter constraint might be restrictive enough to keep the number of binary STD's small also for $N = 16$. However, even when the number of complete STD's is small, this need not be so for the partial STD's. In fact, the size l_n of the list $\mathcal{L}_n$ in Section 4.3 generally has a very large maximum for some $n < N$

when a detour memory greater than zero is required. This behavior can be explained as follows. In the third topological test in Section 4.3, a necessary (but not sufficient) condition is checked for when it is possible to construct a complete STD with detour memory at least M from a given partial STD. The maximum of the list size l_n for an $n < N$ is due to the fact that this condition is not sufficient. The algorithm, as described in Section 4.3, failed for $N = 16$ because of the excessive amount of temporary storage (> 32 Mb). We will soon return to the improvements that were necessary to solve the memory problem.

Another obstacle was the impossibility of checking all $n!\,u!$ bijective maps satisfying the condition (4.21). This problem was solved by choosing a much smaller set of bijective maps, which still allowed the detection of all isomorphisms. We will return to the choice of this set shortly.

Table 4.5 shows all non-isomorphic binary STD's with 16 nodes and detour memory 4. The reader is invited to check the properties predicted by Theorem 4.2. He should also notice the peculiar property of Tables 4.1 to 4.5 that reversing all branch directions of an STD in such a table yields an STD isomorphic to some STD in the same table.

STD No.	Successors of node							
	0	1	2	3	4	5	6	7
1 ... 32	1,0	3, 2	5, 4	7, 6	9, 8	11,10	13,12	15,14

STD No.	Successors of node							
	8	9	10	11	12	13	14	15
1	1,0	3, 2	5, 4	7, 6	9, 8	11,10	13,12	15,14
2	1,0	3, 2	5, 4	7, 6	9, 8	14,13	15,12	11,10
3	1,0	3, 2	5, 4	7, 6	10, 9	11, 8	13,12	15,14
4	1,0	3, 2	5, 4	7, 6	11,10	12, 8	13, 9	15,14
5	1,0	3, 2	5, 4	12, 6	15,14	13, 7	9, 8	11,10
6	1,0	3, 2	6, 4	7, 5	13,12	15,14	9, 8	11,10
7	1,0	3, 2	7, 6	12, 4	11,10	9, 8	13, 5	15,14
8	1,0	3, 2	7, 6	12, 5	9, 8	11,10	13, 4	15,14
9	1,0	3, 2	7, 6	13,12	9, 8	11,10	5, 4	15,14
10	1,0	3, 2	11,10	13,12	9, 8	15,14	5, 4	7, 6
11	1,0	3, 2	11,10	13,12	14, 8	15, 9	6, 5	7, 4
12	1,0	6, 2	5, 4	7, 3	13,12	15,14	9, 8	11,10
13	1,0	6, 2	5, 4	7, 3	13,12	15,14	10, 9	11, 8
14	1,0	6, 3	5, 4	7, 2	9, 8	11,10	13,12	15,14
15	1,0	6, 3	5, 4	7, 2	10, 9	11, 8	13,12	15,14
16	1,0	6, 3	7, 2	12, 4	11,10	9, 8	13, 5	15,14
17	1,0	6, 3	7, 2	12, 5	9, 8	11,10	13, 4	15,14
18	1,0	6, 3	12, 2	13, 7	11,10	9, 8	5, 4	15,14
19	1,0	6, 3	13,12	14, 7	9, 8	11,10	5, 4	15, 2
20	1,0	6, 5	12,11	14, 9	15, 8	13,10	7, 4	3, 2
21	1,0	7, 6	3, 2	5, 4	10, 8	14,13	15,12	11, 9
22	1,0	10, 2	7, 6	12, 5	9, 8	11, 3	13, 4	15,14
23	1,0	10, 3	5, 4	12, 6	15,14	13, 7	9, 8	11, 2
24	1,0	10, 3	12, 5	14, 7	9, 8	11, 2	13, 4	15, 6
25	1,0	10, 6	5, 4	7, 3	13,12	15,14	9, 8	11, 2
26	1,0	10, 9	12,11	15,14	13, 8	3, 2	5, 4	7, 6
27	1,0	10, 9	13,12	3, 2	11, 8	15,14	5, 4	7, 6
28	1,0	11,10	5, 4	7, 6	3, 2	14, 9	12, 8	15,13
29	1,0	12, 3	5, 4	6, 2	15,14	13, 7	9, 8	11,10
30	1,0	12, 3	7, 6	15,14	5, 4	13, 2	9, 8	11,10
31	2,1	3, 0	5, 4	7, 6	9, 8	14,13	15,12	11,10
32	3,2	8, 0	5, 4	12, 6	15,14	13, 7	9, 1	11,10

Table 4.5: All non-isomorphic binary STD's with 16 nodes and detour memory 4.

Let us now return to the aforementioned improvements of the algorithm in Section 4.3, the first of which is due to Theorem 4.2. The existence of self-loops in every binary STD with maximum detour memory and Lemma 4.2 allowed us to start the algorithm of Section 4.3 from $\mathcal{L}_{N/2} = \mathcal{L}_8 = \{\mathbf{F}_8\}$ and $\mathcal{L}_{c_{N/2}} = \mathcal{L}_{c_8} = \{G_{c_8}\}$, where $\mathbf{F}_8$ is given by (4.23) and G_{c_8} is given by G_8 when the self-loop at node 0 is removed. This allowed a considerable reduction of temporary storage. We were able to reduce the amount of temporary storage by another factor of about 1000 by changing the strategy after having computed the lists $\mathcal{L}_{10}$ and $\mathcal{L}_{c_{10}}$. Instead of extending all 774 entries of $\mathcal{L}_{10}$ ($\mathcal{L}_{c_{10}}$) simultaneously to obtain $\mathcal{L}_{11}$ ($\mathcal{L}_{c_{11}}$), we extended one entry at the time up to $N = 16$.

The second improvement of the algorithm in Section 4.3 resulted from choosing a small set of bijective maps $\mathcal{M}$ for Algorithm 4.2 that still allows the detection of all isomorphisms and is explained in the following. Consider two binary STD's G_N and G'_N with maximum detour memory and let $\mathcal{S}_N = \{0, 1, \ldots, N-1\}$ be the common node set of these STD's. Without loss of essential generality, we may assume that both G_N and G'_N have a self-loop at node 0 and a subdigraph $G_{N/2}$ as defined in Lemma 4.2. Let $s_0 \neq 0$ be the node of G_N with the second self-loop. Clearly, an isomorphism of G_N onto G'_N must map node 0 onto node 0 or node s_0 onto node 0.

We consider first those bijections, which map node 0 onto node 0. Since both G_N and G'_N have $G_{N/2}$ as a subdigraph, any such bijection must induce an automorphism (μ, β) of $G_{N/2}$. For $N = 8$ and $M = 3$, Figure 4.9 shows how to obtain such an automorphism by 'twisting' the graphical representation of $G_{N/2}$ at any node from depth 1 to depth $M-1$, i.e., by interchanging the two binary subtrees stemming out from that node. A twist can occur at any node n with $1 \leq n \leq 2^{M-1} - 1 = N/2 - 1$. Let us introduce binary numbers i_n, where $i_n = 1$ ($i_n = 0$) indicates a twist (no twist) at node n. The $2^{N/2-1}$ automorphisms of $G_{N/2}$ can be indexed by $i \triangleq \sum_{n=1}^{N/2-1} i_n \; 2^{N/2-1-n}$. We are interested only in the node-mapping components $\mu_i : \mathcal{S}_N \to \mathcal{S}_N$ of these automorphisms, which can be represented by the permutation vectors $\underline{\mu}_i = [\mu_{i0}, \mu_{i1}, \ldots, \mu_{i,N-1}]$, where $\mu_{in} \triangleq \mu_i(n)$. When the twisted representation of $G_{N/2}$ is drawn from left to right as shown by G_4 in Figure 4.9, the permutation vector $\underline{\mu}_i$ is obtained by reading off the nodes columnwise from depth 0 to depth M. Figure 4.9, for instance, shows the automorphism number $i = (i_1, i_2, i_3)_2 = 5$, which

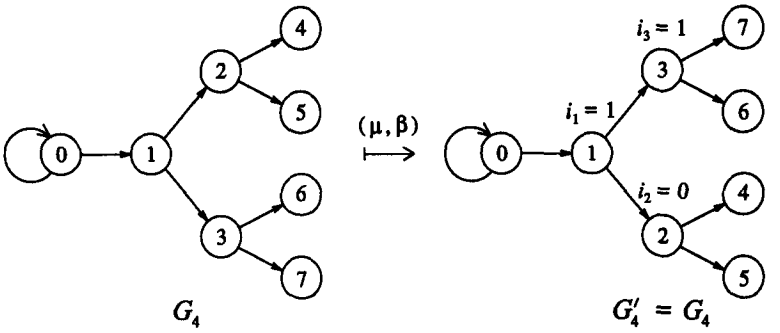


Figure 4.9: The partial binary STD G_4 and an automorphism (μ, β) of G_4

yields $\underline{\mu}_5 = [0, 1, 3, 2, 7, 6, 4, 5]$. For $N = 8$, the $2^{N/2-1} = 8$ permutation vectors $\underline{\mu}_0, \underline{\mu}_1, \ldots, \underline{\mu}_7$ are the rows of

$$\mathbf{M} = \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 0 & 1 & 2 & 3 & 4 & 5 & 7 & 6 \\ 0 & 1 & 2 & 3 & 5 & 4 & 6 & 7 \\ 0 & 1 & 2 & 3 & 5 & 4 & 7 & 6 \\ 0 & 1 & 3 & 2 & 6 & 7 & 4 & 5 \\ 0 & 1 & 3 & 2 & 7 & 6 & 4 & 5 \\ 0 & 1 & 3 & 2 & 6 & 7 & 5 & 4 \\ 0 & 1 & 3 & 2 & 7 & 6 & 5 & 4 \end{bmatrix}.$$

Note that the trivial bijection corresponding to $\underline{\mu}_0$ can be excluded from the set $\mathcal{M}$.

Now consider those bijections, which map node s_0 onto node 0. By the same argument as in the proof of Lemma 4.2, G_N has a tree-like subdigraph $H_{N/2}$ of depth M starting at node s_0 such that $\mathcal{S}(s_n) = \{s_{2n}, s_{2n+1}\}$ is the set of successors of node n, $0 \leq n < N/2$, and $\{s_0, s_1, \ldots, s_{N-1}\} = \mathcal{S}_N$. By the above 'twisting argument', there are $2^{N/2-1}$ isomorphisms of $H_{N/2}$ onto $G_{N/2}$. The first of these isomorphisms has the node-mapping component

$$\begin{array}{rccc} \nu: & \mathcal{S}_N & \rightarrow & \mathcal{S}_N \\ & s_n & \mapsto & n. \end{array}$$

The node-mapping components of the remaining isomorphisms are obtained by combining ν with the above bijections μ_i. Therefore, the

node-mapping components of the isomorphisms of $H_{N/2}$ onto $G_{N/2}$ are given by the composite bijections $\mu_i \circ \nu$, $0 \le i < 2^{N/2-1}$, defined by

$$\begin{array}{rccl} \mu_i \circ \nu: & \mathcal{S}_N & \to & \mathcal{S}_N \\ & n & \mapsto & \mu_i(\nu(n)) \,. \end{array}$$

Letting $\mu_{2^{N/2-1}+i} \triangleq \mu_i \circ \nu$ for $0 \le i < 2^{N/2-1}$ completes the definition of the bijective maps $\mathcal{M} = \{\mu_1, \mu_2, \ldots, \mu_J\}$ for Algorithm 4.2 with $J = 2^{N/2} - 1$.

For $N = 16$, only this choice of $\mathcal{M}$ rendered the search for isomorphisms feasible: For $n = N = 16$ and $u = 0$, the number of bijective maps in Algorithm 4.2 was reduced from $N! \approx 2 \cdot 10^{13}$ to $2^{N/2}-1 = 255$.

Appendix 4.A Properties of the n-th Power of a Digraph

The n-th power of a digraph is a useful concept for the investigation of paths in state-transition diagrams. Its meaning will be evident, once we have defined the product of two digraphs with the same node set.

Definition 4.A.1: The *product* of two digraphs $G_1 = (\mathcal{S}_1, \mathcal{B}_1, \sigma_1, \epsilon_1)$ and $G_2 = (\mathcal{S}_2, \mathcal{B}_2, \sigma_2, \epsilon_2)$ with $\mathcal{S}_1 = \mathcal{S}_2 = \mathcal{S}$ is defined as the graph $G = G_1 G_2 = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$, where

$$\mathcal{B} \triangleq \{(b_1, b_2) : \; b_1 \in \mathcal{B}_1, \; b_2 \in \mathcal{B}_2, \; \epsilon_1(b_1) = \sigma_2(b_2)\},$$

$$\begin{array}{rccl} \sigma: & \mathcal{B} & \to & \mathcal{S} \\ & (b_1, b_2) & \mapsto & \sigma_1(b_1) \end{array}$$

and

$$\begin{array}{rccl} \epsilon: & \mathcal{B} & \to & \mathcal{S} \\ & (b_1, b_2) & \mapsto & \epsilon_2(b_2). \end{array}$$

The set $\mathbb{G}$ of all digraphs with node set $\mathcal{S}$ is obviously closed under multiplication. Note also that the operation '$*$' of multiplying graphs is associative. It follows that the system $\langle \mathbb{G}, * \rangle$ is a semigroup. As a consequence of Definition 4.A.1, the *n-th power* of a digraph G, denoted

by G^n, is a digraph with the same set of nodes as G and branches that represent the paths of length n in G. It is easy to check that

$$\mathbf{A}(G^n) = [\mathbf{A}(G)]^n . \tag{4.A.1}$$

The following results deal with the strong connectivity of the n-th power of a digraph, which will be of interest in Section 5.1.

Proposition 4.A.1: Let $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$, where $\mathcal{B}$ is non-empty, denote a strongly connected, aperiodic digraph. Then G^n is strongly connected and aperiodic for any positive integer n.

Proof: The topology of G can be analyzed using its $N \times N$ adjacency matrix $\mathbf{A} = \mathbf{A}(G)$ where $N \triangleq |\mathcal{S}|$. The proposition holds trivially when $N = 1$ so that the case when $N \geq 2$ remains to be considered. Since G is strongly connected, $\mathbf{A}$ is irreducible. The Perron-Frobenius theorem [38, Thm. 15.4.2, p. 540] and the fact that G is aperiodic imply that $\mathbf{A}$ is *primitive*, i.e., $\mathbf{A}$ has one eigenvalue with largest magnitude. It follows from [38, Thm. 15.6.1, p. 546] that there is a positive integer i such that $\mathbf{A}^k > \mathbf{0}$ for $k \geq i$. Hence, there is a positive integer m such that $\mathbf{A}^{nm} > \mathbf{0}$, which implies that $\mathbf{A}^n$ is irreducible and thus that G^n is strongly connected. Moreover, $\mathbf{A}^n$ is primitive because $\mathbf{A}$ is primitive so that G^n is aperiodic. □

Proposition 4.A.2: Let $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$, where $\mathcal{B}$ is non-empty, denote a strongly connected digraph with period $P > 1$. Then G^P divides into P disjoint, strongly connected, and aperiodic subdigraphs.

Proof: $P > 1$ implies that G has at least two nodes. Applying the Perron-Frobenius Theorem [38, Thm. 15.4.2, p. 540] to $\mathbf{A} = \mathbf{A}(G)$ shows that P is the largest integer such that $\mathcal{S}$ can be partitioned into P disjoint subsets $\mathcal{S}_i$, $0 \leq i < P$, with the following property: For every branch b in G,

$$\sigma(b) \in \mathcal{S}_i \Rightarrow \epsilon(b) \in \mathcal{S}_{(i+1) \bmod P} . \tag{4.A.2}$$

Recall that the branches of G^P represent all length-P paths γ in G, i.e., for every such γ with $\sigma(\gamma) = s$ and $\epsilon(\gamma) = t$, there is a corresponding branch b in G^P with $\sigma(b) = s$ and $\epsilon(b) = t$. It follows from (4.A.2) and the definition of G^P that the start-node and end-node of any branch b in G^P is in the same subset $\mathcal{S}_i$, for some i. Therefore, G^P divides into

P disconnected subdigraphs H_i with node set $\mathcal{S}_i$. Every subdigraph H_i is strongly connected, for otherwise there wouldn't be a path from s to t in G for any nodes $s, t \in \mathcal{S}_i$. Moreover, each H_i is aperiodic because otherwise the period of G would exceed P. □

Proposition 4.A.1 is a special case of the following more general result[1].

Proposition 4.A.3: Let $G = (\mathcal{S}, \mathcal{B}, \sigma, \epsilon)$, where $\mathcal{B}$ is non-empty, denote a strongly connected digraph with period $P \geq 1$ and let n be a positive integer. Then G^n is strongly connected if and only if $\gcd(n, P) = 1$. Moreover, $\gcd(n, P) = 1$ implies that G^n has period P as well.

Proof: The proposition holds trivially when $N \triangleq |\mathcal{S}| = 1$ in which case $P = 1$. Consider now the case when $N \geq 2$. Suppose first that $\gcd(n, P) = 1$. This implies that $n = kP + r$, where $1 \leq r < P$ and $\gcd(r, P) = 1$. As in Proposition 4.A.2, $\mathcal{S}$ is partitioned into the disjoint subsets $\mathcal{S}_i$, $0 \leq i < P$. Since $\gcd(r, P) = 1$, r is a primitive element of the additive group of the ring of integers modulo P, i.e., $\mathbb{Z}_P = \{mr : 0 \leq m < P\}$. Hence, for any subsets $\mathcal{S}_i$ and $\mathcal{S}_j$ and any $s \in \mathcal{S}_i$, G^n contains a path from s to some $t \in \mathcal{S}_j$, since there is an integer m such that

$$mn \equiv mr \equiv j - i \pmod{P}.$$

Because $mn \equiv mr \equiv 0 \pmod{P}$ only if m is an integer multiple of P, G^n has period P. It remains to be shown that, for any subset $\mathcal{S}_j$ and any $t, u \in \mathcal{S}_j$, G^n contains a path from t to u. By Proposition 4.A.2, G^P consists of P disjoint, strongly connected, and aperiodic subdigraphs H_i. This fact and Proposition 4.A.1 imply that $(G^n)^P = (G^P)^n$ consists of P disjoint, strongly connected, and aperiodic subdigraphs H_i^n. Hence, G^n contains the desired path from t to u and is strongly connected.

Suppose now that $\gcd(n, P) > 1$. We have to show that G^n is not strongly connected. There is some positive integer m such that $n = km$ and $P = lm$. Hence, $G^n = (G^m)^k$. Recall the partitioning of $\mathcal{S}$ into P subsets $\mathcal{S}_i$ from Proposition 4.A.2. Since m divides P, $\mathcal{S}$ can be partitioned also into the m subsets $\tilde{\mathcal{S}}_i \triangleq \bigcup_{j=0}^{l-1} \mathcal{S}_{i+jm}$, $0 \leq i < m$, which have the following property: For every branch b in G,

$$\sigma(b) \in \tilde{\mathcal{S}}_i \Rightarrow \epsilon(b) \in \tilde{\mathcal{S}}_{(i+1) \bmod m}. \tag{4.A.3}$$

[1] We use the notation $\gcd(a, b)$ for the greatest common divisor of a and b.

It follows from (4.A.3) and the definition of G^m that the start-node and end-node of any branch $\tilde{b}$ in G^m is in the same subset $\tilde{\mathcal{S}}_i$, for some i. This implies that G^m divides into m disconnected subdigraphs (each of which has period l) and is not strongly connected. Hence $G^n = (G^m)^k$ is not strongly connected. □

Appendix 4.B Proof of Theorem 4.1

The proof of Theorem 4.1 relies on the following result.

Lemma 4.B.1: Let $G_c = (\mathcal{N}_c, \mathcal{B}_c)$ be the CRD of a partial K-ary STD and let G_c have $D = 0$ dead nodes, S source nodes, and R resource nodes. Then it is always possible to construct a cyclic path ω in G_c by a partial extension of its resource nodes such that

(i) A resource node is extended by at most one branch (and such a branch is always connected to a source node)

(ii) The path ω traverses m distinct source nodes and m distinct resource nodes, where $m \leq \min(S, R)$

(iii) There is a path from any source node $s \in \mathcal{N}_c$ to every node on ω

(iv) There is a path from every node on ω to any resource node $r \in \mathcal{N}_c$

A path ω as described in Lemma 4.B.1 will be called a *cyclic connector path.*

Proof: The existence of a cyclic connector path ω in G_c will be proved by first constructing a cyclic path γ in an auxiliary digraph $G_a = (\mathcal{N}_a, \mathcal{B}_a)$, defined as follows. Let the node set be $\mathcal{N}_a = \mathcal{S} \cup \mathcal{R}$, where $\mathcal{S}$ and $\mathcal{R}$ denote the source nodes and resource nodes of $\mathcal{N}_c$, respectively, and form a branch set $\mathcal{B}_a$ such that, for every source node s and every resource node r for which there is a path $s \rightarrow r$ in G_c, B_a contains a branch from s to r, i.e.,

$$\mathcal{B}_a \triangleq \{ (\sigma(\pi), \epsilon(\pi)) : \text{for all paths } \pi \text{ in } G_c \text{ with } d_{\text{in}}(\sigma(\pi)) = 0 \text{ and } d_{\text{out}}(\epsilon(\pi)) = 0 \}.$$

The nodes of G_a in $\mathcal{S}$ (or $\mathcal{R}$) will also be called source nodes (or resource nodes). For every resource node r in G_a, define $\mathcal{S}_r$ as the set of all source nodes, from which there is a path to r. Since $D = 0$, these R sets completely exhaust the source nodes; however, they are not disjoint in general. Similarly, for every source node s in G_a, define $\mathcal{R}_s$ as the set of all resource nodes, to which there is a path from s. Again, these S sets fully exhaust the resource nodes but are not disjoint in general.

The path γ in G_a is constructed according to Figure 4.B.1. It can

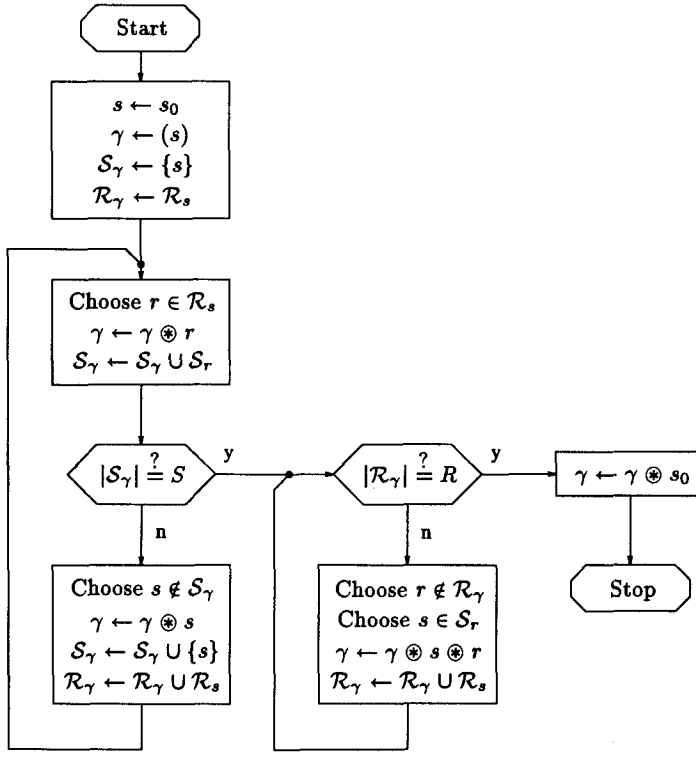


Figure 4.B.1: Algorithm for constructing the cyclic path γ in Lemma 4.B.1

be described by a sequence of nodes, since G_a has no parallel branches. Let $\mathcal{S}_\gamma$ be the set of all source nodes, from which there is a path (possibly of length zero) to a node in γ. Similarly, let $\mathcal{R}_\gamma$ be the set of

all resource nodes, to which there is a path (possibly of length zero) from a node in γ. The path γ is initialized with an arbitrary start node $s_0 \in \mathcal{S}$. The algorithm alternates between appending source nodes and resource nodes to γ, indicated by the concatenation operator '$\circledast$' in Figure 4.B.1. Appending a source node s to γ requires the insertion of one of K available branches from the current resource node $r = \epsilon(\gamma)$ to s.

Consider the first loop in Figure 4.B.1 and assume $|\mathcal{S}_\gamma| < S$, i.e., not all source nodes have a path to γ. Note that a source node cannot be appended to γ more than once because the algorithm chooses $s \notin \mathcal{S}_\gamma$ and updates $\mathcal{S}_\gamma \leftarrow \mathcal{S}_\gamma \cup \{s\}$. Moreover, by definition of $\mathcal{S}_\gamma$, the choice $s \notin \mathcal{S}_\gamma$ implies that no resource node $r \in \mathcal{R}_s$ can already be on γ. The first loop is left when $|\mathcal{S}_\gamma| = S$. In the second loop, assume $|\mathcal{R}_\gamma| < R$, i.e., not all resource nodes can be reached from γ. Clearly, the choice $r \notin \mathcal{R}_\gamma$ prevents us from choosing a resource node already on γ. Moreover, the choice $s \in \mathcal{S}_r$ implies that s is not already on γ for otherwise $r \in \mathcal{R}_\gamma$. In particular, this assures $s \neq s_0$. In the very last step, the path γ is made cyclic by appending s_0. Note that γ traverses m distinct source nodes and m distinct resource nodes, where $m \leq \min(S, R)$, and that at most one of K available branches is used per resource node during the construction of γ. Properties (i) to (iv) are therefore satisfied for the path γ in G_a. We now define a cyclic path ω in G_c by replacing every branch (s, r) in G_a by some path $s \rightarrow r$ in G_c. The proof is completed by noting that the properties of γ in G_a carry over to ω in G_c. □

Proof of Theorem 4.1: Recall that every resource node of G_c represents an unextended node of G. When we extend a node in G, we always extend the corresponding resource node in G_c. More precisely, if a branch b is added to G, a corresponding branch b_c is added to G_c such that the start-node (end-node) of b_c 'contains' the start-node (end-node) of b. Note that the modified G_c is not component-reduced in general. When we add a new node to G, we also add a corresponding node to G_c. The extension of all resource nodes yields RK new branches in G. Let us 'grow' the partial STD G to a complete STD with N nodes by adding $N-V$ new nodes. For every node added to G, exactly $K-1$ new branches are obtained. For instance, we can take any unextended node in G (or resource node in G_c) and build a K-ary tree rooted at that node using the $N-V$ new nodes. Thus, the extension of all unextended nodes and all new nodes results in a total number of $RK + (N - V)(K - 1)$ new branches.

Necessity: We have to show that G is not strongly N-connectable when neither (4.12) nor (4.13) is satisfied. Suppose first that $D = 0$ and $RK + (N - V)(K - 1) < S$. We have already seen that $RK + (N - V)(K - 1)$ is the total number of new branches. Therefore, not all source nodes can be reached and G is not strongly N-connectable. If $D = 1$ and $V < N$, then there is no way to reach any of the $N - V$ new nodes from a node in the 'dead component' of G, so G is not strongly N-connectable. Similarly, if $D = 1$, $V = N$, and $S > 0$, then the source node(s) of G_c cannot be reached from the dead node of G_c. Finally, when $D > 1$ it is not possible to get from one dead component to another dead component.

Sufficiency: Suppose that (4.12) is satisfied. Lemma 4.B.1 asserts the existence of a cyclic connector path ω. Assume that such a path is now constructed, so each of $m \leq \min(S, R)$ resource nodes is connected to one of m different source nodes. Since this requires m new branches, another $RK + (N - V)(K - 1) - m$ new branches are still available. But there remain only $S - m \leq RK + (N - V)(K - 1) - m$ unconnected source nodes, so there is at least one new branch to end at each source node. All other new branches are also connected to source nodes. In the following, we will use the notation $n \rightarrow n'$ if there is a path from node n to node n' [1], $n \rightarrow \omega$ if there is a path from node n to a node in path ω, and $\omega \rightarrow n$ if there is a path from a node in ω to node n. The choice of the new branches and (iii) and (iv) of Lemma 4.B.1 imply that, for any pair of nodes $n, n' \in \mathcal{N}_c$, there are resource nodes r, r' and source nodes s, s' such that

$$n \rightarrow r \rightarrow s \rightarrow \omega \rightarrow r' \rightarrow s' \rightarrow n'$$

and thus $n \rightarrow n'$. Therefore, since each new branch added to G_c corresponds to a new branch added to G, we have constructed a complete, strongly connected STD with N nodes. This proves that G is strongly N-connectable. Suppose now that (4.13) is satisfied. It follows from $S = 0$ that the dead node is the only node of G_c, so G is strongly connected. Moreover, G is a complete STD since $V = N$. But a complete, strongly connected STD with N nodes is trivially strongly N-connectable. □

[1] By way of convention, we let $n \rightarrow n$.

Chapter 5

On Trellis-Coded Data Transmission over Channels with Intersymbol Interference and White Gaussian Noise

In this chapter, we study trellis-coded data transmission over linear intersymbol-interference (ISI) channels with finite memory and additive white Gaussian noise (AWGN). Observing that the channel filter can be viewed as a rate-1 finite-state trellis encoder, it seems natural to employ an outer trellis encoder for improving the system's reliability. For certain partial-response channels [6], Wolf and Ungerboeck [15] have designed bipolar trellis codes that yield a large free Euclidean distance at the channel output and, at the same time, eliminate sequences with long runs of identical symbols. Also for partial-response channels, Karabed and Siegel [13] and independently Eleftheriou and Cideciyan [16] have investigated so-called *matched spectral-null* (MSN) trellis codes, which are characterized by the property that the frequencies at which the code power spectral density vanishes correspond precisely to the frequencies at which the channel transfer function is zero. Trellis codes designed for the AWGN channel are often used on ISI channels, especially when the channel unit-sample response is too long for code optimization, unknown to the transmitter, or time-varying.

The main objectives of this chapter are to characterize the composite trellis encoder formed by cascading the outer trellis encoder with the channel filter and to determine the performance of the composite

encoder with AWGN and maximum-likelihood (ML) decoding.

Two entirely different approaches for data transmission over ISI channels should be mentioned. For both methods, the ISI coefficients must be known to the transmitter. The method of *channel partitioning*[1] [52], [5] is motivated by the derivation of capacity for the DTGC with ISI (cf. Sections 2.3 and 3.1) where the ISI channel is decomposed into a bank of parallel, decoupled channels without ISI. Because the component channels are memoryless, codes designed for the AWGN channel can be used. The method suffers from the non-linearities of practical channels for which the assumption of decoupled channels is only an approximation. Moreover, it is limited to ISI channels that accept multilevel inputs. The method called *Tomlinson-Harashima precoding* [53], [54], [55] uses a (stable) inverse channel filter employing modulo arithmetic in the transmitter. Together with a modulo operation at the receiver input, this results in an ISI-free channel with a gain equal to the magnitude of the leading ISI coefficient. However, the noise process seen after the modulo operation in the receiver is neither independent of the data nor white Gaussian. Nevertheless, such an approximation is often made to justify the use of codes designed for the AWGN channel. Tomlinson-Harashima precoding has the advantage that the inverse channel filter can be easily adapted to different ISI channels. Channel partitioning and Tomlinson-Harashima precoding will not be considered further in this thesis.

Assuming that the modulator, physical channel and demodulator result in a discrete-time channel with ISI and AWGN, the communication system of interest can be drawn as in Figure 5.1. An (n, k) trellis encoder T encodes the binary information sequence $\{X_j\}$ into the $\mathcal{A}$-ary sequence $\{Y_l\}$, where the alphabet $\mathcal{A}$ is a finite subset of the real or complex numbers. The encoded sequence $\{Y_l\}$ is transmitted over an ISI channel with transfer function $H(z) = \sum h_m z^{-m}$ (referred to as the *channel filter*) and AWGN $\{W_l\}$, and the received process $\{R_l\}$ is fed to the decoder, whose outputs are the decisions $\{\hat{X}_{j-\delta}\}$, where δ is some positive delay. If the discrete-time channel in Figure 5.1 represents a physical baseband channel, the letters in the alphabet $\mathcal{A}$ and the coefficients h_m are real, and $\{W_l\}$ is a real AWGN process with $W_l \sim \mathcal{N}(0, N_0/2)$; if it represents a physical passband channel, the elements of $\mathcal{A}$ and the h_m *can* be complex, and $\{W_l\}$ is a proper complex AWGN process with $W_l \sim \mathcal{N}_p(0, N_0)$.

[1] Also known as multitone or multicarrier modulation.

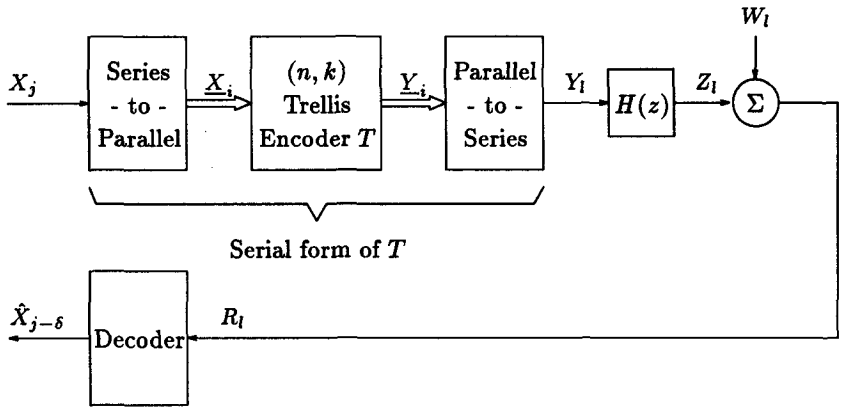


Figure 5.1: Trellis-coded communication system

In Section 5.1, an (n,k) trellis encoder T is defined as a parallel input - parallel output device, i.e., a device that encodes a sequence of binary k-tuples $\underline{x}_i$ into a sequence of $\mathcal{A}$-ary n-tuples $\underline{y}_i$. The binary information sequence $\{x_j\}$ is converted to the sequence $\{\underline{x}_i\}$ according to

$$\underline{x}_i \triangleq [x_{ik},\ x_{ik+1},\ \ldots,\ x_{(i+1)k-1}],\qquad i = \lfloor j/k \rfloor - 1, \tag{5.1}$$

and $\{\underline{y}_i\}$ is converted to the code sequence $\{y_l\}$ using

$$\underline{y}_i \triangleq [y_{in},\ y_{in+1},\ \ldots,\ y_{(i+1)n-1}],\qquad i = \lfloor l/n \rfloor. \tag{5.2}$$

It will be convenient to call the cascade of the series-to-parallel converter, the trellis encoder T and the parallel-to-series converter the *serial form* of the trellis encoder T. An (n,k) trellis encoder T or, more precisely, its serial form, has a *nominal rate*

$$R \triangleq k/n \quad \text{bits/symbol.} \tag{5.3}$$

The channel filter is assumed to have finite memory μ, i.e.,

$$H(z) = \sum_{m=0}^{\mu} h_m z^{-m}, \tag{5.4}$$

where $h_0 \neq 0$ and $h_\mu \neq 0$, and unit energy, i.e., $\sum_{m=0}^{\mu} |h_m|^2 = 1$.

As we will examine more closely in Section 5.1, the serial form of T followed by $H(z)$ can be viewed as the serial form of a composite (n, k) trellis encoder T_c with $N\,|\mathcal{A}|^\mu$ states, where N denotes the number of states of T. Since T_c may have *transient* or uncontrollable states, the *steady-state composite encoder* or simply the *steady-state encoder* T_c' will be defined as the encoder obtained from T_c by deleting the transient states. Thus, the communication system in Figure 5.1 is equivalent to the one in Figure 5.2.

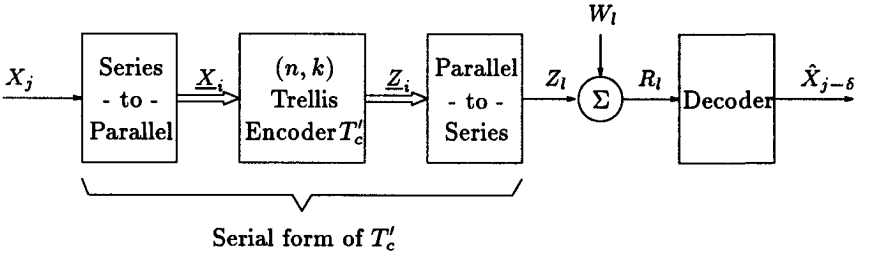


Figure 5.2: Equivalent communication system with steady-state composite encoder T_c'

The distance spectrum, which will be defined in Section 5.3, is generally non-uniform for a steady-state composite encoder, i.e., it depends on the reference path, the path in the trellis [56], [57] corresponding to the transmitted data sequence.

For *uncoded* data transmission over ISI channels, the ML decoder is known as the maximum-likelihood sequence estimator (MLSE) [7]. Its bit error probability can be upper-bounded using the method of *error sequences* [7], [26, Sec. 4.9], [2, Sec. 6.7.1], since the assumption of i.i.d. data symbols leads to a simple expression for the probability that an error sequence is allowable. Unfortunately, the method of error sequences does not apply to coded transmission over ISI channels since (i) not all error sequences are possible [26, Sec. 4.10] and (ii) the probability that an error sequence is admissible is not equal to the product of the probabilities that each component of the error sequence is admissible. We will therefore generalize the well-known upper bound on the bit error probability for Viterbi decoding [26, Sec. 4.4, pp. 242], [27, Sec. 6.E] to trellis encoders with a non-uniform distance spectrum such as the above

steady-state composite encoder. The generalized upper bound, which we will present in Section 5.3, involves the *average squared Euclidean distance spectrum* of the steady-state composite encoder T_c', where the average is taken over all possible reference paths and, for each of these reference paths, over all possible detours in the trellis of T_c'. In particular, it involves the minimum or *free* squared Euclidean distance Δ_f and the average number of bit errors over all detours at Δ_f, denoted $\bar{n}_{\Delta_f}$. An algorithm for the efficient evaluation of average distance spectra is described in Section 5.4 and applied to the analysis of bipolar trellis encoders for the dicode channel in Section 5.5.

The bit error probability will always be treated as a function of the signal-to-noise ratio

$$\rho \triangleq E_b/N_0, \tag{5.5}$$

where

$$E_b = \mathrm{E}\left[|Y_l|^2\right]/R \tag{5.6}$$

is the energy per bit and N_0 is the one-sided noise power spectral density. In general, an approximation of the generalized upper bound on bit error probability using only Δ_f and $\bar{n}_{\Delta_f}$ is not precise enough for low to medium E_b/N_0. Nevertheless, Δ_f and $\bar{n}_{\Delta_f}$ are the two most important parameters for encoder optimization.

It should be mentioned that ρ is a 'universal' parameter for the comparison of communication systems as in Figure 5.1, in a sense we describe in the following. Note that a complex channel with real ISI coefficients h_m and proper complex AWGN of sample variance N_0 can be decomposed into two real channels with the same coefficients and independent real AWGN processes with sample variance $N_0/2$. Clearly, two identical communication systems operating independently on the two real channels can be viewed as a combined communication system operating on the complex channel. Hence, both component systems and the combined system have the same bit error probability. We call ρ 'universal' since (in the case of real ISI coefficients) ρ is the same for the component systems and for the combined system. To see this, assume that each component encoder has a rate R and an average symbol energy E_s. Then the combined encoder has rate $2R$ and an average symbol energy $2E_s$, so

$$\rho = E_s/(RN_0)$$

for both component systems and the combined system, as was to be shown.

5.1 Characterization of Trellis Encoders

An (n, k) trellis encoder T encodes a sequence of binary k-tuples $\underline{x}_i$ into a sequence of $\mathcal{A}$-ary n-tuples $\underline{y}_i$, where $\mathcal{A}$ is a finite alphabet. It can be viewed as a Mealy machine [48] with parallel inputs and parallel outputs or, equivalently, as a labeled digraph[2]. Only finite-state, time-invariant trellis encoders will be considered in this thesis.

Definition 5.1: A finite-state, time-invariant (n, k) *trellis encoder* T is a labeled digraph $(\mathcal{A}, \mathcal{S}, \mathcal{B})$, where $\mathcal{A}$ (the *output alphabet*) and $\mathcal{S}$ (the *state set*) are finite non-empty sets and $\mathcal{B}$ (the labeled *branch set*) is a subset of $\mathcal{S} \times \{0,1\}^k \times \mathcal{A}^n \times \mathcal{S}$ such that, for any $s \in \mathcal{S}$, the projection of $\mathcal{B}_{\text{out}}(s)$ (the set of labeled branches emanating from state s) upon its second component is *onto* $\{0,1\}^k$. ———

Definition 5.1 was inspired by the notion of a transition graph in [58]. Unless stated otherwise, the output alphabet $\mathcal{A}$ is assumed to be a (finite) subset of the real or complex numbers. A branch $b = (s, \underline{x}, \underline{y}, s') \in \mathcal{B}$ means that, for a given start-state s and binary input k-tuple $\underline{x}$, the trellis encoder proceeds to state s', generating an $\mathcal{A}$-ary output n-tuple $\underline{y}$. The projections of $\mathcal{B}$ onto its four components will be denoted by σ, ϕ, π, and ϵ. For a branch $b \in \mathcal{B}$, $\sigma(b)$ and $\epsilon(b)$ denote the *start-node* and *end-node*, respectively. Moreover, $\underline{x} = \phi(b)$ and $\underline{y} = \pi(b)$ denote the *input label* (a binary k-tuple) and the *output label* (an $\mathcal{A}$-ary n-tuple), respectively, of the branch b. The condition on $\mathcal{B}$ in Definition 5.1 can be written as

$$\{\phi(b) : b \in \mathcal{B}_{\text{out}}(s)\} = \{0,1\}^k \quad \text{for all } s \in \mathcal{S} \tag{5.7}$$

and ensures that, for every $s \in \mathcal{S}$ and for every $\underline{x} \in \{0,1\}^k$, exactly one branch b labeled with $\phi(b) = \underline{x}$ starts at node s. Thus, an (n, k) trellis encoder can be obtained by assigning a pair of labels $(\underline{x}, \underline{y}) \in \{0,1\}^k \times \mathcal{A}^n$ to every branch of a 2^k-ary STD such that the labeled STD satisfies (5.7).

Defining a trellis encoder as a labeled digraph has the advantage that the pictorial language of graph theory becomes available. Thus, we may speak of an *aperiodic* trellis encoder. A trellis encoder is called *controllable* if it is a strongly connected digraph. A controllable trellis encoder can be driven to any state and, in particular, back to the initial

[2]The graph terminology was introduced in Section 4.1.

state. Controllability is usually desired to avoid unnecessary computations in the Viterbi algorithm [56], [57]. Another advantage of defining a trellis encoder as a labeled digraph is the fact that many concepts of graph theory have a counterpart in the rich theory of nonnegative matrices [38, Chap. 15].

When an (n,k) trellis encoder T is viewed as a finite-state Mealy machine, it is appropriately described by the output alphabet $\mathcal{A}$, the state set $\mathcal{S}$, the *next-state function*

$$\begin{array}{rccc} f: & \mathcal{S}\times\{0,1\}^k & \to & \mathcal{S} \\ & s,\underline{x} & \mapsto & f(s,\underline{x}), \end{array} \tag{5.8}$$

and the *output function*

$$\begin{array}{rccc} g: & \mathcal{S}\times\{0,1\}^k & \to & \mathcal{A}^n \\ & s,\underline{x} & \mapsto & g(s,\underline{x}). \end{array} \tag{5.9}$$

We will call $(\mathcal{A},\mathcal{S},f,g)$ the *Mealy representation* of T.

As its name suggests, a 'trellis encoder' naturally specifies an associated 'trellis', which is the key to understanding the Viterbi algorithm [56], [57].

Definition 5.2: An (L,τ) *trellis* of an (n,k) trellis encoder $T = (\mathcal{A},\mathcal{S},\mathcal{B})$ is a labeled digraph $\tilde{T} = (\mathcal{A},\tilde{\mathcal{S}},\tilde{\mathcal{B}})$ with nodes $\tilde{\mathcal{S}} = \bigcup_{i=0}^{L+\tau}\tilde{\mathcal{S}}_i$ and branches $\tilde{\mathcal{B}} = \bigcup_{i=0}^{L+\tau-1}\tilde{\mathcal{B}}_i$, where $\tilde{\mathcal{S}}_i \triangleq \{\tilde{s}=(s,i):\ s\in\mathcal{S}_i\}$ and $\tilde{\mathcal{B}}_i \triangleq \left\{\tilde{b}=((\sigma(b),i),\ \phi(b),\ \pi(b),\ (\epsilon(b),i+1)):\ b\in\mathcal{B}_i\right\}$ are the nodes and branches at trellis depth i, respectively, defined recursively by

$$\mathcal{S}_i = \begin{cases} \{s_0\} & \text{if } i=0, \\ \{\epsilon(b):\ b\in\mathcal{B}_{i-1}\} & \text{if } 1\le i\le L+\tau, \end{cases} \tag{5.10}$$

$$\mathcal{B}_i = \begin{cases} \bigcup_{s\in\mathcal{S}_i}\mathcal{B}_{\text{out}}(s) & \text{if } 0\le i<L, \\ \{b_i(s):\ s\in\mathcal{S}_L\} & \text{if } L\le i<L+\tau, \end{cases} \tag{5.11}$$

for a specified *initial state* $s_0\in\mathcal{S}$ and specified *tail paths* $\gamma(s) = \big(b_L(s),\ b_{L+1}(s),\ \ldots,\ b_{L+\tau-1}(s)\big)$, such that $\gamma(s)$ is a path in T from $s\in\mathcal{S}_L$ to s_0. ———

The trellis of an (n,k) trellis encoder will be referred to as a 2^k-ary trellis. Every path in an (L,τ) trellis of an (n,k) trellis encoder corresponds to the encoding of L binary k-tuples into a sequence of $L+\tau$

$\mathcal{A}$-ary n-tuples, whose serial form of length $(L+\tau)n$ is called a *codeword*. The encoder starts from some initial state s_0 and, after encoding Lk bits, returns to s_0 by following a specified tail path $\gamma(s_L)$, where s_L is the state at time L. Note that the binary k-tuples $\phi(b_i(s))$, $L \leq i < L+\tau$, $s \in \mathcal{S}_L$, are *dummy bits*. The nodes $(s_0, 0)$ and $(s_0, L+\tau)$ are called the *root node* and the *toor*[3] *node* of the trellis, respectively [57]. The tail paths ensure that all codewords end at the toor node. An (L, τ) *trellis code* is the list of all 2^{Lk} codewords generated by an (L, τ) trellis. In practice, one often assumes $L = \infty$, in which case we speak of an infinite trellis.

A *detour* of length l, $1 \leq l < \infty$, is defined as a path that diverges from a 'correct' or *reference path* at some trellis depth i and that remerges with the reference path at depth $i + l$ for the *first* time.

For a general trellis code, *free distance* is defined as the minimum distance[4] between the code sequences generated by a reference path and a detour with the same end-nodes as the reference path, where the minimum is over all reference paths of finite, nonzero length and over all corresponding detours in the infinite trellis.

The *detour memory* of an (n, k) trellis encoder T is simply defined as the detour memory of its underlying 2^k-ary STD and can be interpreted as $L_{\min} - 1$, where $L_{\min}$ denotes the length of the shortest detour in the trellis of T, over all possible reference paths. The detour memory of an (n, k) convolutional encoder with a polynomial encoding matrix is given by

$$M = \min_{1 \leq i \leq k} \max_{1 \leq j \leq n} \deg [g_{ij}(D)], \tag{5.12}$$

where the polynomial $g_{ij}(D)$ is the transfer function from the i-th input to the j-th output. It is an open question, under which conditions a polynomial encoding matrix and a rational encoding matrix that generate the same code give rise to the same detour memory. The detour memory should be distinguished from the *memory* of an (n, k) trellis or convolutional encoder whose state is given by the contents of k shift registers (each one bit wide) without feedback. For such an encoder, one defines the memory as the maximum length of all k shift registers. Hence, a convolutional encoder with a polynomial encoding matrix has

[3] 'toor' is 'root' spelled backwards.

[4] By 'distance', we mean the appropriate distance measure for the given channel (cf. Section 5.3).

memory[5]

$$m \triangleq \max_{1 \le i \le k} \max_{1 \le j \le n} \deg [g_{ij}(D)].$$

We warn the reader that a nonminimal trellis encoder with N states and detour memory M can be reduced to a trellis encoder with fewer states and possibly a smaller detour memory. Such a reduction of the detour memory occurs, for instance, when the encoder is based on a shift register, the last stage of which does not affect the encoder outputs.

We now present a simple upper bound on the free distance of a trellis code. This upper bound involves only the detour memory of an encoder for that code and the maximum distance between two elements of the output alphabet. The bound is often tight, particularly when the output alphabet has a small cardinality. For an (n, k) trellis encoder T with detour memory M, recall that the shortest detour in the trellis of T has length $M + 1$. Hence, the *free Hamming distance* of a binary (n, k) trellis code generated by an encoder with detour memory M satisfies

$$d_{H_f} \le (M + 1)\, n. \tag{5.13}$$

Similarly, the *free squared Euclidean distance* of a bipolar (± 1) trellis code generated by an encoder with detour memory M satisfies

$$d^2_{E_f} \le 4\,(M + 1)\, n. \tag{5.14}$$

Notice that a periodic encoder is nonminimal because the (∞, τ) trellis code generated by an encoder T with period $P > 1$ is generated also by the maximal strongly connected components of T^P, which have fewer states and are aperiodic (cf. Proposition 4.A.2).

One might ask whether there is a finite integer l_0 such that $\mathcal{S}_l = \mathcal{S}$ for all $l \ge l_0$ in Definition 5.2. We now show that such an l_0 exists for every controllable, aperiodic trellis encoder T. Since T has an irreducible, primitive adjacency matrix, there is a positive integer l_0 such that [38, Thm. 15.6.1, p. 546]

$$\mathbf{A}^l > \mathbf{0} \quad \text{for } l \ge l_0, \tag{5.15}$$

which implies that, for any $l \ge l_0$ and for any $s \in \mathcal{S}$, there is a path of length l in T from node s_0 to node s and a corresponding path in

[5] In [59, p. 294], m is called the *memory order*. The memory (order) is commonly used to express the constraint length as $(m + 1)n$.

the trellis from node $(s_0, 0)$ to node (s, l). Inequality (5.15) implies also the existence of tail paths $\gamma(s)$ of finite length. Summarizing, we have proved

Lemma 5.1: Let $T = (\mathcal{A}, \mathcal{S}, \mathcal{B})$ denote a controllable, aperiodic trellis encoder and let s_0 be an arbitrary initial state in $\mathcal{S}$. Then there is a finite integer l_0 such that the nodes at trellis depth $l \geq l_0$ satisfy $\mathcal{S}_l = \mathcal{S}$. Moreover, there is a finite integer τ such that, for any $s \in \mathcal{S}$, T contains a tail path $\gamma(s)$ of length τ from s to s_0 and the trellis has a corresponding tail path from node (s, L) to node $(s_0, L+\tau)$. ———

For the traceback operation in the Viterbi algorithm [26], it is often desirable to have a trellis encoder with uniform in-degree. It follows from Theorem 4.2 that trellis encoders based on a 2^k-ary STD with maximum detour memory have uniform in-degree. Löliger [58, p. 46] has shown recently that convolutional encoders (over groups, rings, or fields) obtained from linear transition graphs have uniform in-degree. However, not every trellis encoder has this property, even if it is controllable. Such asymmetric behavior is exhibited, for instance, by the cascade of some matched spectral-null encoders with a partial-response channel. Hence, the digraph obtained by reversing all branch directions of a 2^k-ary trellis is not necessarily another 2^k-ary trellis.

Trellis or convolutional encoders are usually called non-catastrophic if a finite number of channel errors can result only in a finite number of decoding errors [59, p. 308]. Notice that, for a finite output alphabet, a finite number of channel errors is equivalent to a finite distance between two codewords, where 'distance' can be Hamming distance or Euclidean distance, whichever is suitable for the given channel. However, the terminology 'channel errors' inappropriately anticipates making hard decisions on the received channel symbols. Therefore, we prefer to call a trellis encoder *non-catastrophic* if any decoded path that diverges from a reference path and that accumulates *finite distance* to that reference path causes a *finite number of bit errors*. It is well-known that a convolutional encoder with a polynomial encoding matrix is non-catastrophic if and only if there exists a feedforward inverse [60], [59, pp. 306-308]. For a general trellis encoder, no comparably simple criteria to test for catastrophicity have yet been found.

5.2 On the Cascade of a Trellis Encoder with a Finite-Impulse-Response Channel Filter

In this section, the cascade of a trellis encoder with a finite-impulse-response (FIR) filter is investigated.

Theorem 5.1: Let $T = (\mathcal{A}, \mathcal{S}, \mathcal{B})$ denote a controllable (n, k) trellis encoder with $N = |\mathcal{S}|$ states and let $H(z)$ be an FIR filter of order μ as defined in (5.4). Define $T_c = (\mathcal{A}_c, \mathcal{S}_c, \mathcal{B}_c)$ as that (n, k) trellis encoder whose serial form corresponds to the serial form of T, followed by $H(z)$. Then T_c contains exactly one maximal strongly connected sink component[6] $T_c' = (\mathcal{A}_c', \mathcal{S}_c', \mathcal{B}_c')$, which is a controllable (n, k) trellis encoder with

$$N_c' = |\mathcal{S}_c'| \leq \min\left(N\,|\mathcal{A}|^{\mu},\ N\,2^{kL}\right) \tag{5.16}$$

states, where $L \triangleq \lceil \mu/n \rceil$. Moreover, every state $s_c \in \mathcal{S}_c \backslash \mathcal{S}_c'$ is *transient* in the sense that any path of length at least L starting at s_c ends (and remains) in $\mathcal{S}_c'$.

The encoders T_c and T_c' in Theorem 5.1 will be referred to as the *composite encoder* and the *steady-state (composite) encoder*, respectively.

Proof: The state of the composite encoder T_c is defined as

$$s_{ci} \triangleq \left(s_i,\ y_{in-1},\ y_{in-2},\ \cdots,\ y_{in-\mu}\right), \tag{5.17}$$

where s_i is the state of T and y_{in-m}, $1 \leq m \leq \mu$, are the $\mathcal{A}$-ary symbols stored in the shift register of the filter. Thus, the cardinality of the state set $\mathcal{S}_c$ is $N\,|\mathcal{A}|^{\mu}$. The next-state function of T_c depends on the next-state function of T, the output function of T, and on the memory μ. The output alphabet of T_c is

$$\mathcal{A}_c \triangleq \left\{ \textstyle\sum_{m=0}^{\mu} h_m\, y_{\mu-m} :\ y_m \in \mathcal{A},\ 0 \leq m < \mu \right\}.$$

We proceed by defining the subset

$$\begin{aligned} \mathcal{S}_c' \triangleq \{\ & \left(s_L,\ y_{Ln-1},\ y_{Ln-2},\ \cdots,\ y_{Ln-\mu}\right): \\ & s_L = \epsilon(\gamma),\ (y_0, y_1, \cdots, y_{Ln-1}) = \pi(\gamma), \\ & \text{for all paths } \gamma = (b_0, b_1, \cdots, b_{L-1}) \text{ in } T\ \} \end{aligned} \tag{5.18}$$

[6] The nodes or states in a maximal strongly connected sink component are sometimes called *essential* [61].

of $\mathcal{S}_c$, where $L \triangleq \lceil \mu/n \rceil$ and where $\epsilon(\gamma)$ and $\pi(\gamma)$, respectively, denote the end-node of γ and the output sequence generated by γ. We further define the subset

$$\mathcal{B}'_c \triangleq \{b:\ b \in \mathcal{B}_c,\ \sigma(b) \in \mathcal{S}'_c\}$$

of $\mathcal{B}_c$. Since $Ln - \mu = \lceil \mu/n \rceil n - \mu \geq 0$, any path of length L in T will drive T_c to a unique state in $\mathcal{S}'_c$ independent of the initial state of the filter. Since there are exactly $N\, 2^{kL}$ paths of length L in T, it follows that $|\mathcal{S}'_c| \leq \min\left(N\,|\mathcal{A}|^{\mu},\ N\,2^{kL}\right)$.

Suppose that T_c is in some state

$$s_{cL} = \left(s_L,\ \ y_{Ln-1},\ \ y_{Ln-2},\ \ \cdots,\ \ y_{Ln-\mu}\right) \in \mathcal{S}'_c$$

at time L. By definition of $\mathcal{S}'_c$, there is a path $\gamma = (b_0, b_1, \ldots, b_{L-1})$ in T such that T_c ends in s_{cL}. Since the end-node $s_L = \epsilon(b_{L-1})$ is uniquely determined by s_{cL}, the 2^k successors of s_{cL} are obtained by appending a branch $b_L \in \mathcal{B}_{\text{out}}(s_L)$ to γ. Since $s_{cL} \in \mathcal{S}'_c$, the corresponding branch b_{cL} is in $\mathcal{B}'_c$. But every successor $s_{c,L+1}$ must be in $\mathcal{S}'_c$, since there is a path $\gamma' = (b_1, b_2, \ldots, b_L)$ in T such that T_c ends in $s_{c,L+1}$. This establishes the fact that $\mathcal{S}'_c$ is a sink in $\mathcal{S}_c$. Since T is controllable, it is possible to walk along any length-L path γ in T by first getting to its start-node. This implies that the labeled graph $(\mathcal{A}_c, \mathcal{S}'_c, \mathcal{B}'_c)$ is strongly connected. It follows from the definition of $\mathcal{B}'_c$ that, for any $s \in \mathcal{S}'_c$,

$$\{\phi(b) : b \in \mathcal{B}_{c\,\text{out}}(s)\} = \{0,1\}^k.$$

The output symbols on the branches in $\mathcal{B}'_c$ form a subset $\mathcal{A}'_c$ of $\mathcal{A}_c$. Summarizing, we have shown that T'_c is a controllable (n,k) trellis encoder.

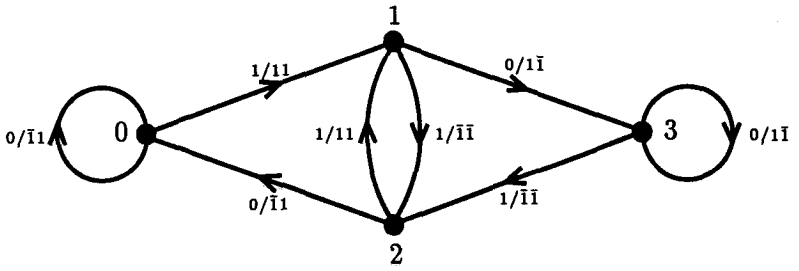
If $\mathcal{S}'_c$ is a proper subset of $\mathcal{S}_c$, suppose that T_c is in a state $s_c \notin \mathcal{S}'_c$. Starting from s_c, any path of length at least L must lead to a state $s'_c \in \mathcal{S}'_c$, by definition of $\mathcal{S}'_c$. This proves that the states in $\mathcal{S}_c \backslash \mathcal{S}'_c$ are transient. □

The following examples show that either one of the terms in the minimization of (5.16) can be smallest.

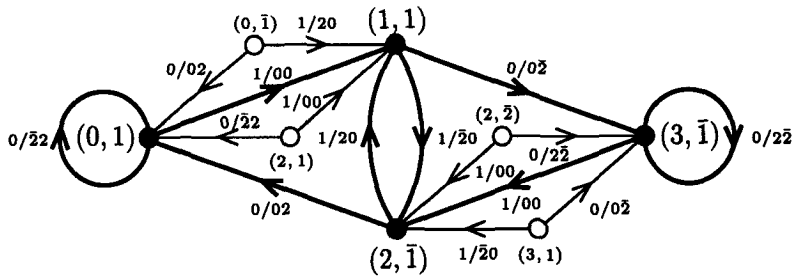
Example 5.1: For $k = 2$, $n = 3$, $\mu = 1$, and $|\mathcal{A}| = 2$, one obtains $L = 1$ and $N'_c \leq \min(2N,\ 4N) = 2N$.

Example 5.2: For $k = 1$, $n = 2$, $\mu = 3$, and $|\mathcal{A}| = 2$, we get $L = 2$ and $N'_c \leq \min(8N,\ 4N) = 4N$.

Figure 5.3 illustrates Theorem 5.1 for a bipolar $(2,1)$ trellis encoder T with $N = 4$ states and for the channel filter $H(z) = 1 - z^{-1}$ of the dicode channel[7]. The composite encoder T_c has $N_c = 8$ states



Trellis encoder T



Composite encoder T_c with steady-state encoder T_c' (bold) as subgraph

Figure 5.3: Cascade of a bipolar $(2,1)$ encoder T with an FIR filter $H(z) = 1 - z^{-1}$. The branch labels have the form $u_i/v_{2i}v_{2i+1}$, where u_i is an information bit and v_l is an encoder output.

labeled by $s_{ci} = (s_i, y_{in-1})$ (cf. (5.17)). The essential and transient states of T_c are marked with filled and hollow circles, respectively. The steady-state encoder $T_c' = (\mathcal{A}_c', \mathcal{S}_c', \mathcal{B}_c')$ (shown bold in Figure 5.3) has only $N_c' = 4$ states. This is due to the fact that, for every state s in T, the two branches ending at s end with the same code symbol, which implies that a unique *channel state* 1 or $\bar{1}$ can be assigned to every state of T. In our example, the channel states 1, 1, $\bar{1}$, and $\bar{1}$ are assigned to the states 0, 1, 2, and 3 of T, respectively. The steady-

[7] In the remainder of this section, we use $\bar{a}$ to denote $-a$.

state encoder in Figure 5.3 is an example used by Wolf and Ungerboeck [15, Fig. 3 and Fig. 11(b)] to illustrate code design for the dicode channel by means of set partitioning[8]. In the light of Theorem 5.1, their technique rests on the identification of the set of allowed output n-tuples for every branch in a 2^k-ary STD when channel states 1 or $\bar{1}$ have been assigned to the nodes of this STD.

The steady-state composite encoder T_c' can be determined efficiently by applying the *Tarjan algorithm* [50], [47], [51] to the composite encoder T_c. This algorithm determines the maximal strongly connected components of a digraph with a computational complexity that is linear in the number of nodes. Alternatively, $\mathcal{S}_c'$ and then T_c' can be determined by following all possible paths of length L in T, as suggested by (5.18).

It is worth mentioning that the total *memory*[9] m_{tot} of an (n, k) trellis encoder with memory m cascaded with an FIR filter of order μ is given by [26, p. 286]

$$m_{\text{tot}} = m + \lceil \mu/n \rceil.$$

[8] More will be said about this encoder in Section 5.5.

[9] See Section 5.1 for the definition of *memory* (as opposed to the *detour memory*).

5.3 An Upper Bound on the Bit Error Probability for Non-Uniform Trellis Encoders and Viterbi Decoding

In this section, the well-known upper bound on the bit error probability for Viterbi decoding [26, Sec. 4.4, pp. 242], [27, Sec. 6.E], is generalized to time-invariant trellis encoders (cf. Section 5.1) with a non-uniform distance spectrum and used on memoryless, output-symmetric [26, p. 242] channels. We assume that a controllable, aperiodic (n, k) trellis encoder and an associated (L, τ) 2^k-ary trellis are given. As opposed to the uniform case, an *assumption on the probability distribution for the information sequence* will be required in the case of a non-uniform distance spectrum. In general, the bound will be valid only for the chosen distribution. In order to separate the problems of source and channel coding, we assume that the information sequence is comprised of independent, equiprobable binary digits, i.e., it has the distribution that is approximated by the output of any good source encoder.

The upper bound we wish to generalize is based on the enumeration of detours (cf. Section 5.1) from a reference path in the trellis. For a given reference path, it is possible to count the number of detours with distance d, number of bit errors i, and length l. Here and hereafter, 'distance' refers to a *distance measure*, which (i) is used by the maximum-likelihood decoder to find the codeword closest to the received sequence and (ii) is an additive branch function in the sense that the distance of a subpath in the trellis from the corresponding subpath of the reference path must be the sum of the distances of each branch in that subpath from the corresponding branch in the reference path. For instance, Hamming distance is the appropriate measure when binary trellis codes are used on the binary symmetric channel [10, p. 17], and the *squared* Euclidean distance is appropriate when trellis codes are employed on the AWGN or proper complex AWGN channel. A collection of pairs (d, i) or triples (d, i, l) is commonly referred to as a *distance spectrum*. To obtain the generalization, one has to overcome the following difficulties:

- The distance spectrum generally depends on the reference path, i.e., it is conditioned on the initial state (where the detours begin) and on the reference information sequence.

- The stationary state-probability distribution need not be uniform.
- In a practical communication system, the trellis encoder starts from a specified initial state at time zero. Therefore, the encoder approaches the stationary state-probability distribution only asymptotically with time.
- It is desirable to obtain a tight upper bound on bit error probability, which is independent of the specified initial state. For a finite trellis, it appears to be difficult to derive such a bound since one might choose an initial state s_0 with particularly 'bad' distance spectra[10] for the reference paths starting at $s_{\text{root}} = (s_0, 0)$. This indicates further that an upper bound for the infinite trellis is generally not also an upper bound for a finite trellis.

A distance spectrum depending on [independent of] the reference path, as well as the corresponding *encoder*, will be called *non-uniform* [*uniform*]. It will be shown that the conditional distance spectra must be averaged according to the stationary state-probability distribution and the probability distribution of the reference information sequence. Since an infinitely long reference information sequence has zero probability, we are forced to begin with a finite trellis and to use a limiting argument for the infinite trellis.

Each encoder input is a binary k-tuple $\underline{x}$. We will represent $\underline{x}$ by the 2^k-ary number $u \triangleq \sum_{j=0}^{k-1} x_j 2^j$. If u and $\hat{u}$ represent $\underline{x}$ and $\hat{\underline{x}}$, respectively, it will be convenient to define $d_H(u, \hat{u})$ as the *Hamming distance* $d_H(\underline{x}, \hat{\underline{x}})$. Using the sequence notation $u_q^r = (u_q, u_{q+1}, \ldots, u_r)$, we define also $d_H(u_q^r, \hat{u}_q^r) \triangleq \sum_{j=q}^{r} d_H(u_j, \hat{u}_j)$.

Definition 5.3: We define $a_{d,i,l}^{L}(s, u_0^{L-1})$ as the number of detours in the (L, τ) trellis with respect to the reference path determined by the initial state s and the length-L sequence u_0^{L-1}, where each detour diverges at time zero, has distance d, causes i information bit errors, and has length[11] l. We further define $a_{d,i,l}^{\infty}(s, u_0^{l-1})$ as the number of detours in the *infinite* trellis with respect to the reference path determined by the initial state s and the length-l sequence u_0^{l-1}, where each detour diverges at time zero, has distance d, causes i information bit errors, and has length l.

[10] In practice, of course, one would choose an initial state with particularly good distance spectra for the reference paths starting at $s_{\text{root}} = (s_0, 0)$.

[11] Recall from Section 5.1 that a detour ends where it remerges for the *first time* with the reference path.

Note that $a_{d,i,l}^{L}(s, u_0^{L-1})$ vanishes for $l > L + \tau$. Let the random variable W_j be the number of information bit errors if the Viterbi decoder initiates a detour at time (or trellis depth) j, and let $W_j = 0$ if the decoder is already on a detour that began earlier or if it follows the reference path from time j to $j+1$. The total number of information bit errors is then given by $\sum_{j=0}^{L-1} W_j$. We wish to derive an upper bound on the *bit error probability*

$$P_b \triangleq \mathrm{E}\left[\sum_{j=0}^{L-1} W_j\right] / (kL). \tag{5.19}$$

Suppose that the Viterbi decoder leaves the reference path at time j for a detour of length l. The relevant part of the reference path is determined by the state at time j, s_j, and some sequence $u_j^{m(j,l)}$, where

$$m(j,l) \triangleq \min(j + l - 1, L - 1), \tag{5.20}$$

and the relevant part of the detour path is determined by $\hat{s}_j = s_j$ and some sequence $\hat{u}_j^{m(j,l)}$. The number of information bit errors on this detour is $d_H(u_j^{m(j,l)}, \hat{u}_j^{m(j,l)})$. The average number of bit errors over all possible detours diverging from the reference path at time j, $0 \le j < L$, is given by

$$\begin{aligned}
&\mathrm{E}\left[W_j \,|\, \hat{S}_j = S_j = s_j,\ U_j^{L-1} = u_j^{L-1}\right] \\
&= \sum_{l=1}^{L-j+\tau} \sum_{\substack{\hat{u}_j^{m(j,l)}:\\ \hat{s}_{j+t} \neq s_{j+t},\ \text{if } 1 \le t < l \\ \hat{s}_{j+l} = s_{j+l}}} d_H(u_j^{m(j,l)}, \hat{u}_j^{m(j,l)})\, P_{\hat{U}_j^{m(j,l)} | \hat{S}_j S_j U_j^{L-1}}(\hat{u}_j^{m(j,l)} | s_j, s_j, u_j^{L-1}).
\end{aligned} \tag{5.21}$$

But the probability that the Viterbi decoder chooses a specific detour for a given reference path can be upper-bounded by

$$P_{\hat{U}_j^{m(j,l)} | \hat{S}_j S_j U_j^{L-1}}(\hat{u}_j^{m(j,l)} | s_j, s_j, u_j^{L-1}) \le P(y_j^{j+l-1} \to \hat{y}_j^{j+l-1}), \tag{5.22}$$

where $P(y \to \hat{y})$ denotes the probability of error when the codeword $\hat{y}$ corresponding to the detour is the only alternative to the transmitted codeword y. The upper bound (5.22) follows from the fact that the decoding regions become larger in this two-codeword decision problem. Assuming that the codewords are transmitted over a memoryless,

output-symmetric channel [26, p. 242], $P(y \to \hat{y})$ depends only on the distance between the two codewords, viz.,

$$P(y \to \hat{y}) = P_2(\, d(y, \hat{y})\,) = P(\hat{y} \to y), \tag{5.23}$$

where $P_2(\,.\,)$ is called the *pairwise error probability*. We now substitute (5.22) and (5.23) into (5.21) and then use the fact that the length-l detours starting at time j can be enumerated by means of $a_{d,i,l}^{L-j}(\,.\,,\,.\,)$, the number of detours in the $(L-j, \tau)$ trellis. Hence

$$\begin{aligned}
&\mathrm{E}\left[W_j \,|\, \hat{S}_j = S_j = s_j,\ U_j^{L-1} = u_j^{L-1} \right] \\
&\quad \le \sum_{l=1}^{L-j+\tau} \sum_{\substack{\hat{u}_j^{m(j,l)}: \\ \hat{s}_{j+t} \neq s_{j+t},\ \text{if } 1 \le t < l \\ \hat{s}_{j+l} = s_{j+l}}} d_H(u_j^{m(j,l)}, \hat{u}_j^{m(j,l)})\, P_2(\, d(y_j^{j+l-1}, \hat{y}_j^{j+l-1})\,) \\
&\quad = \sum_{d=d_f}^{\infty} \sum_{i=1}^{\infty} \sum_{l=1}^{L-j+\tau} i\, a_{d,i,l}^{L-j}(s_j, u_j^{L-1})\, P_2(d), \quad 0 \le j < L, \qquad (5.24)
\end{aligned}$$

where d_f denotes the free distance defined in Section 5.1. By definition of W_j,

$$\mathrm{E}\left[W_j \,|\, S_j = s_j,\ \hat{S}_j \neq s_j,\ U_j^{L-1} = u_j^{L-1} \right] = 0. \tag{5.25}$$

By the assumption of independent, uniformly distributed binary inputs, $\{U_j\}$ is a sequence of independent, uniformly distributed 2^k-ary random variables. This implies further that both the encoder state S_j and the decoder state $\hat{S}_j$ are independent of U_j^{L-1}. Therefore,

$$\begin{aligned}
P_{S_j \hat{S}_j U_j^{L-1}}(s, s, u_j^{L-1}) &= P_{S_j \hat{S}_j}(s, s)\, P_{U_j^{L-1}}(u_j^{L-1}) = \\
&P_{S_j}(s)\, P_{\hat{S}_j | S_j}(s, s)\, 2^{-k(L-j)} \le P_{S_j}(s)\, 2^{-k(L-j)}, \qquad (5.26)
\end{aligned}$$

where the inequality is tight in the interesting range of operation because $P_{\hat{S}_j | S_j}(s, s)$ is well-approximated by 1 for medium to high signal-to-noise ratios. Applying the theorem on total expectation to (5.24)-(5.26) yields

$$\begin{aligned}
&\mathrm{E}[\, W_j \,] \le \sum_{s \in \mathcal{S}} P_{S_j}(s) \sum_{d=d_f}^{\infty} P_2(d) \sum_{i=1}^{\infty} i \cdot \\
&\sum_{l=1}^{L-j+\tau} 2^{-k(L-j)} \sum_{u_0^{L-j-1} \in \{0,1,\ldots,2^k-1\}^{L-j}} a_{d,i,l}^{L-j}(s, u_0^{L-j-1}), \quad 0 \le j < L, \qquad (5.27)
\end{aligned}$$

where we have replaced the summation over the dummy variables u_j^{L-1} by a summation over all u_0^{L-j-1}.

Unfortunately, it seems difficult to further upper-bound the right-hand side of (5.27) in such a way that (i) the upper bound remains tight and (ii) the dependence on L is eliminated. (This can be done for a *uniform* distance spectrum; in that case, the upper bound for the infinite trellis is also an upper bound for any finite trellis [27, Sec. 6.E].) Fortunately, however, we can (and will) derive a tight upper bound valid only for the *infinite* trellis since (5.27) holds also in the limit as L approaches infinity. In order to simplify the right-hand side of (5.27), we will break up the term on the second line into two sums over the range $1 \le l \le L-j$ and $L-j+1 \le l \le L-j+\tau$, respectively. The first sum satisfies

$$\sum_{l=1}^{L-j} 2^{-k(L-j)} \sum_{u_0^{L-j-1}} a_{d,i,l}^{L-j}(s, u_0^{L-j-1}) = \sum_{l=1}^{L-j} 2^{-kl} \sum_{u_0^{l-1}} a_{d,i,l}^{\infty}(s, u_0^{l-1}), \tag{5.28}$$

since the number of length-l detours does not depend on u_l^{L-j-1} and is the same for the $(L-j,\tau)$ trellis and the infinite trellis. For the second sum, recall that in the $(L-j,\tau)$ trellis the reference path is uniquely determined by the initial state s and the data sequence u_0^{L-j-1}. The dummy information sequence corresponding to the tail of this reference path will be denoted by $\tilde{u}_{L-j}^{L-j+\tau-1}(s, u_0^{L-j-1})$. Letting '$\circledast$' denote sequence concatenation, we can upper-bound the second sum by

$$\begin{aligned}
&\sum_{l=L-j+1}^{L-j+\tau} 2^{-k(L-j)} \sum_{u_0^{L-j-1}} a_{d,i,l}^{L-j}(s, u_0^{L-j-1}) \\
&\quad \le \sum_{l=L-j+1}^{L-j+\tau} 2^{-k(L-j)} \sum_{u_0^{L-j-1}} a_{d,i,l}^{\infty}(s, u_0^{L-j-1} \circledast \tilde{u}_{L-j}^{l-1}(s, u_0^{L-j-1})) \\
&\quad \le \sum_{l=L-j+1}^{L-j+\tau} 2^{k(l-L+j)}\, 2^{-kl} \sum_{u_0^{L-j-1}} \sum_{u_{L-j}^{l-1}} a_{d,i,l}^{\infty}(s, u_0^{l-1}) \\
&\quad \le 2^{k\tau} \sum_{l=L-j+1}^{L-j+\tau} \max_{u_0^{l-1}} a_{d,i,l}^{\infty}(s, u_0^{l-1}),
\end{aligned} \tag{5.29}$$

where the first inequality is due to the fact that in the infinite trellis

the detours need not follow the unique tail paths, the second inequality holds because $\tilde{u}_{L-j}^{l-1}(s, u_0^{L-j-1})$ is included in the sum over *all* u_{L-j}^{l-1}, and the last inequality is true because the average of 2^{kl} numbers cannot exceed their maximum. Combining (5.27), (5.28), and (5.29) yields

$$\mathrm{E}[W_j] \le \sum_{s \in S} P_{S_j}(s) \sum_{d=d_f}^{\infty} P_2(d) \sum_{i=1}^{\infty} i \cdot \left[\sum_{l=1}^{L-j} 2^{-kl} \sum_{u_0^{l-1}} a_{d,i,l}^{\infty}(s, u_0^{l-1}) + 2^{k\tau} \sum_{m=L-j+1}^{L-j+\tau} \max_{\bar{u}_0^{m-1}} a_{d,i,m}^{\infty}(s, \bar{u}_0^{m-1}) \right], \tag{5.30}$$

where $0 \le j < L$. The transition to the infinite trellis[12] will cause the unwanted second term in the square brackets to vanish. Thus, we take the limit of *both* sides of (5.30) as $L \to \infty$. Using the fact that the limit of the sum of a finite number of terms is the sum of their limits, we obtain

$$\mathrm{E}[W_j] \le \sum_{s \in S} P_{S_j}(s) \sum_{d=d_f}^{\infty} P_2(d) \sum_{i=1}^{\infty} i \cdot \left[\sum_{l=1}^{\infty} 2^{-kl} \sum_{u_0^{l-1}} a_{d,i,l}^{\infty}(s, u_0^{l-1}) + \tau\, 2^{k\tau} \lim_{m \to \infty} \max_{\bar{u}_0^{m-1}} a_{d,i,m}^{\infty}(s, \bar{u}_0^{m-1}) \right], \tag{5.31}$$

for all $j \ge 0$. If the trellis encoder T satisfies

$$\lim_{l \to \infty} a_{d,i,l}^{\infty}(s, u_0^{l-1}) = 0 \tag{5.32}$$

for all s, u_0^{l-1}, and for d in the range $d_f \le d < \infty$, the second term in the square brackets of (5.31) vanishes. Since the numbers $a_{d,i,l}^{\infty}(s, u_0^{l-1})$ are integers, (5.32) is equivalent to the following condition: For any finite $d \ge d_f$, there is a finite integer L_d such that

$$a_{d,i,l}^{\infty}(s, u_0^{l-1}) \equiv 0 \quad \text{for all } l > L_d. \tag{5.33}$$

The condition (5.32) or (5.33) states that two paths that diverge and never remerge must accumulate an unbounded distance. As a consequence, any two cyclic paths in the labeled graph T, which can be entered from a common node by two paths of the same length and produce

[12] Theoretically, this implies an infinite decoding delay. However, experience with convolutional codes suggests that the path memory can be limited to a finite window for general trellis codes without introducing a significant performance loss.

the same code sequence are prohibited, even if they correspond to the same information sequence! This shows that a non-catastrophic encoder does not necessarily satisfy (5.32). For instance, a non-catastrophic encoder having two self-loops labeled with the same input $\underline{x}$ and the same output $\underline{y}$ does not satisfy (5.32). Condition (5.33) is necessary for evaluating the upper bound on the computer since, for a given d, it guarantees a finite search depth for all detours at distance d. Using (5.32), (5.31) is simplified to $\mathrm{E}[W_j] \leq \beta_j$, $j \geq 0$, where

$$\beta_j \triangleq \lim_{d^* \to \infty} \sum_{s \in S} P_{S_j}(s) \sum_{d=d_f}^{d^*} P_2(d) \sum_{i=1}^{\infty} i \sum_{l=1}^{L_d} 2^{-kl} \sum_{u_0^{l-1}} a_{d,i,l}^{\infty}(s, u_0^{l-1}).$$

Hence, the bit error probability for the infinite trellis satisfies

$$\begin{aligned} P_b &= \lim_{m \to \infty} \frac{1}{km} \mathrm{E}\left[\sum_{j=0}^{m-1} W_j\right] \\ &= \lim_{m \to \infty} \frac{1}{km} \sum_{j=0}^{m-1} \mathrm{E}[W_j] \leq \lim_{m \to \infty} \frac{1}{km} \sum_{j=0}^{m-1} \beta_j. \end{aligned} \tag{5.34}$$

Observe that β_j is a function of the state-probability distribution $P_{S_j}(.)$. As j approaches infinity, $P_{S_j}(\,.\,)$ converges to a limit $P_S(\,.\,)$, the *stationary* state-probability distribution. The existence of this limit is a consequence of [38, Thm. 15.8.1, p. 552] since, by the assumption of independent, uniformly distributed inputs U_j and of an aperiodic trellis encoder, one obtains a time-invariant, primitive transition-probability matrix

$$\mathbf{\Pi} = 2^{-k}\mathbf{A}, \tag{5.35}$$

where $\mathbf{A}$ is the adjacency matrix of T. (In fact, the stationary state-probability distribution is given by the left eigenvector of $\mathbf{\Pi}$ corresponding to the unit eigenvalue, normalized such that its components sum to one.) Therefore, the sequence $\beta_0, \beta_1, \ldots$ has also a limit, denoted β. It is well-known that the sequence of arithmetic averages $\alpha_0, \alpha_1, \ldots$, where $\alpha_m \triangleq \frac{1}{m} \sum_{j=0}^{m-1} \beta_j$, has the same limit β [62, Kap. 3.1.3]. Hence, the right-hand side of (5.34) equals β/k. Summarizing, we have proved the following result.

Theorem 5.2: Let T denote a controllable, aperiodic (n, k) trellis encoder satisfying (5.33), i.e., with the property that two paths with a finite distance between them have a finite unmerged span. Let the encoder input be a right-sided, infinite sequence of independent, uniformly distributed 2^k-ary random variables and let $P_S(\,.\,)$ denote the stationary state-probability distribution. Assume further that the encoded sequence is transmitted over a memoryless, output-symmetric channel, on which the pairwise error probability for two codewords at distance d is given by $P_2(d)$. Then the bit error probability for Viterbi decoding is upper-bounded by

$$P_b \leq \frac{1}{k} \lim_{d^* \to \infty} \sum_{d=d_f}^{d^*} P_2(d)\, \bar{n}_d, \tag{5.36}$$

where d_f denotes free distance and

$$\bar{n}_d = \sum_{s \in S} P_S(s) \sum_{i=1}^{\infty} i \sum_{l=1}^{L_d} 2^{-kl} \sum_{u_0^{l-1} \in \{0,1,\ldots,2^k-1\}^l} a_{d,i,l}^{\infty}(s, u_0^{l-1}), \tag{5.37}$$

where L_d is the maximum length of a detour at distance d. ———

The set

$$\mathcal{D} \triangleq \{(d, \bar{n}_d) : \text{there exists a detour at } d \geq d_f\},$$

will be called the *average distance spectrum.* It should be pointed out that, for a *uniform* distance spectrum, (5.37) reduces to

$$\bar{n}_d = \sum_{i=1}^{\infty} i \sum_{l=1}^{L_d} a_{d,i,l}^{\infty}(0, 0_0^{l-1})$$

since $a_{d,i,l}^{\infty}(s, u_0^{l-1}) \equiv a_{d,i,l}^{\infty}(0, 0_0^{l-1})$. In general, the numbers $\bar{n}_d$ are *rational;* for a uniform encoder, they are *integers.* As we will see in Section 5.5, the possibility of making the parameter $\bar{n}_{d_f}$ smaller than one makes it even more important for nonlinear trellis encoders than for linear ones. The upper bound (5.36) can be approximated by extending the sum only to a parameter $d_{\max} < \infty$.

The following property is often useful for evaluating the average distance spectrum.

Property 5.1: The stationary state-probability distribution of a controllable, aperiodic (n, k) trellis encoder T with uniform in-degree and independent, uniformly distributed 2^k-ary inputs is the uniform distribution.

Proof: We have already seen that the stationary state-probability distribution is given by the left eigenvector of the transition-probability matrix (5.35) corresponding to the unit eigenvalue, normalized such that its components sum up to one. Since T has uniform in-degree 2^k, every column-sum of its adjacency matrix $\mathbf{A}$ equals 2^k, and every column-sum of $\mathbf{\Pi}$ equals one. Hence, the desired left eigenvector of $\mathbf{\Pi}$ is $2^{-N}[1, 1, \ldots, 1]$, where N is the number of states of T. □

Of particular interest is the real or proper complex AWGN channel, for which the *squared* Euclidean distance $d_E^2(\,.\,,\,.\,)$ is the appropriate distance measure. In order to avoid confusion of squared Euclidean distance with Euclidean distance, we shall replace the symbol d in Theorem 5.2 by Δ to denote squared Euclidean distance. The pairwise error probability for two codewords at squared Euclidean distance Δ is given by

$$P_2(\Delta) = Q\left(\sqrt{\Delta/(2N_0)}\right), \tag{5.38}$$

where

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du.$$

Further, (5.36) becomes

$$P_b \leq \frac{1}{k} \lim_{\Delta^* \to \infty} \sum_{\Delta=\Delta_f}^{\Delta^*} P_2(\Delta)\, \bar{n}_\Delta, \tag{5.39}$$

where Δ_f denotes the free squared Euclidean distance and

$$\bar{n}_\Delta = \sum_{s \in S} P_S(s) \sum_{i=1}^{\infty} i \sum_{l=1}^{L_\Delta} 2^{-kl} \sum_{u_0^{l-1} \in \{0,1,\ldots,2^k-1\}^l} a^{\infty}_{\Delta,i,l}(s, u_0^{l-1}). \tag{5.40}$$

Correspondingly, the *average squared Euclidean distance spectrum* is defined as

$$\mathcal{D}_E \triangleq \{\, (\Delta, \bar{n}_\Delta) : \text{ there exists a detour at squared Euclidean distance } \Delta \geq \Delta_f \,\}. \tag{5.41}$$

5.4 Efficient Evaluation of Average Distance Spectra

It is well-known that both trellis-based and tree-based decoding algorithms can be used for calculating distance spectra of convolutional encoders [59, p. 376]. On the other hand, the computation of average distance spectra (cf. Section 5.3) for non-uniform trellis encoders is generally considered infeasible because the average is to be taken over all possible reference paths. This has led to the investigation of regular trellis codes [63], [64], for which the distance spectrum can be evaluated assuming an arbitrary reference path[13]. A less stringent condition called quasi-regularity was introduced in [64], which is satisfied for a broader class of trellis codes. It has been claimed recently that the distance spectrum of quasi-regular codes used on ISI channels can be found assuming an arbitrary reference path [65, p. 629], regardless of the ISI coefficients. However, this contradicts our experience with binary convolutional codes[14], which yield a uniform distance spectrum on the 'straight-wire' channel (as they should), but a highly non-uniform distance spectrum on some partial-response channels. For further comments on [65], the reader is referred to [66].

In this section, a *modified Viterbi algorithm* (MVA) for computing average distance spectra of non-uniform trellis encoders is described, with emphasis on an efficient implementation. We start by rewriting (5.37) as

$$\bar{n}_d = \sum_{s_0 \in \mathcal{S}} P_S(s_0)\, \bar{n}_d(s_0),$$

where

$$\bar{n}_d(s_0) \triangleq \sum_{i=1}^{\infty} i \sum_{l=1}^{L_d} 2^{-kl} \sum_{u_0^{l-1} \in \{0,1,\ldots,2^k-1\}^l} a^{\infty}_{d,i,l}(s_0, u_0^{l-1}) \tag{5.42}$$

is the average number of bit errors over all detours at distance d that diverged from a reference path at trellis node $(s_0, 0)$ and over all reference paths starting from that trellis node. Hereafter, reference paths are assumed to start at trellis node $(s_0, 0)$. Moreover, we consider only those

[13] To be precise, the 'regularity' defined in [63] and [64] is actually a property of trellis *encoders*.

[14] Assuming binary antipodal signaling.

detours, which depart from the reference path at node $(s_0, 0)$. Equation (5.42) suggests that $\bar{n}_d(s_0)$ can be obtained iteratively by setting it to zero initially and by an update

$$\bar{n}_d(s_0) \leftarrow \bar{n}_d(s_0) + i\, 2^{-kl} \tag{5.43}$$

for every detour of length l with a codeword distance d and an information distance i from the reference path. In (5.43), the factor 2^{-kl} is the probability of a reference path of length l.

The key idea of the MVA is the assignment of a list $\mathcal{L}_l(s)$ to every node (s, l) of the trellis, where s is in the state set $\mathcal{S}$ and $l \geq 0$, in order to keep track of accumulated distances. A list $\mathcal{L}_l(s)$ is either empty or contains entries of the form (d, i), where d and i is the accumulated codeword distance and information distance, respectively, between a path from $(s_0, 0)$ to (s, l) and the reference path from $(s_0, 0)$ to trellis depth l. A list may contain repeated entries and need not be ordered with respect to codeword distance. Let $\underline{x}/\underline{y}$ and $\underline{x}_{\text{ref}}/\underline{y}_{\text{ref}}$ be the labels of some branch in the trellis and of the branch on the reference path at the same depth, respectively. Then the codeword and information distance between these branches is denoted by $d(\underline{y}, \underline{y}_{\text{ref}})$ and $d_H(\underline{x}, \underline{x}_{\text{ref}})$, respectively, where $d(\,.\,,\,.\,)$ is a distance measure as described in Section 5.3 and $d_H(\,.\,,\,.\,)$ is the Hamming distance.

To simplify the description of the MVA, the nodes of the trellis (except for the root node) are assumed to have uniform in-degree 2. Suppose that two branches merge at trellis node (s, l) and that none of these branches belongs to the reference path, as depicted in Figure 5.4. As the MVA proceeds from depth $l-1$ to depth l, the list $\mathcal{L}_l(s)$ is

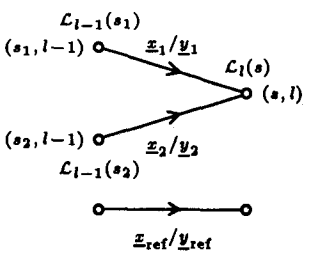


Figure 5.4: Two branches merging at trellis node (s, l)

determined by updating and then concatenating the lists $\mathcal{L}_{l-1}(s_1)$ and

$\mathcal{L}_{l-1}(s_2)$, viz.,

$$\mathcal{L}_l(s) = \Big(\mathcal{L}_{l-1}(s_1) + (\,d(\underline{y}_1, \underline{y}_{\text{ref}}), d_H(\underline{x}_1, \underline{x}_{\text{ref}})\,)\Big) \circledast \Big(\mathcal{L}_{l-1}(s_2) + (\,d(\underline{y}_2, \underline{y}_{\text{ref}}), d_H(\underline{x}_2, \underline{x}_{\text{ref}})\,)\Big), \qquad (5.44)$$

where '$\mathcal{L}_j(s) + (d,i)$' denotes the list obtained by adding (d,i) component-wise to every element of $\mathcal{L}_j(s)$ and '$\circledast$' denotes list concatenation. Note that all detours, which traverse trellis node (s,l) and which have accumulated a codeword distance d from the reference path at that node, can be represented by a single list entry (d,i) for computing $\bar{n}_d(s_0)$. Thus, the list $\mathcal{L}_l(s)$ may optionally be sorted and condensed into the form

$$\{(d_1,i_1), (d_2,i_2), \ldots\}, \text{ where } d_1 < d_2 < \ldots \text{ and } i_m \triangleq \sum_{(d_m,i) \in \mathcal{L}_l(s)} i.$$

As an example, we consider the computation of the average squared Euclidean distance spectrum for a bipolar trellis encoder cascaded with the dicode channel. Figure 5.5 shows the trellis of the steady-state composite encoder[15] for three different reference paths of length 3 and illustrates the computation of the numbers $\bar{n}_\Delta(s_0)$ for $s_0 = 0$ and $\Delta \leq \Delta_{\max} = 56$. Note that all detours must follow the branch from node $(0,0)$ to node $(1,1)$ and that no detour starts at node $(1,0)$. This is achieved by letting $\mathcal{L}_0(0) \leftarrow \{(0,0)\}$ and $\mathcal{L}_0(1) \leftarrow \{\}$. The entry $\{(0,0)\}$ acts as a 'seed' for the generation of the lists $\mathcal{L}_l(s)$. For an efficient implementation of the MVA, it is desirable to perform operation (5.44) (optionally followed by a sorting operation) for every pair of merging branches, whether or not one of them belongs to the reference path. If we blindly use (5.44) for $\mathcal{L}_1(0)$ in Figure 5.5 (a), we get $\mathcal{L}_1(0) = \{(0,0)\}$. We must remove the entry $(0,0)$ from $\mathcal{L}_1(0)$ to prevent new detours from starting at depth $l = 1$. The operation of removing the 'null-entry' $(0,0)$ is denoted by "$\stackrel{\text{N}}{\rightarrow}$" in Figure 5.5 (a). At depth 2 in Figure 5.5 (a), the semi-bold detour remerges with the reference path. According to (5.43), the entry $(\Delta,i) = (24,2)$ of $\mathcal{L}_2(0)$ results in

$$\bar{n}_{24}(0) \leftarrow \bar{n}_{24}(0) + 0.5.$$

In general, an operation (5.43) must be performed for every list entry representing a merged detour and, after these updates, the list must

[15] This encoder will be considered also in Section 5.5, Figure 5.8.

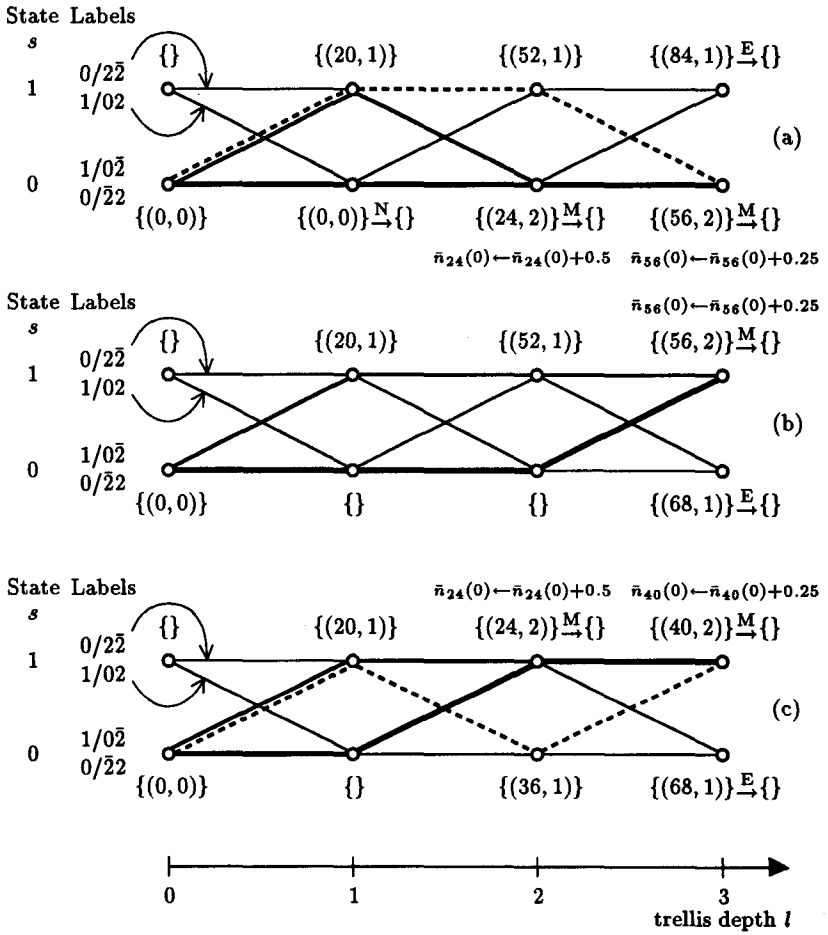


Figure 5.5: Computation of the average squared Euclidean distance spectrum ($\Delta_{\max}=56$) for the steady-state encoder obtained from the cascade of a bipolar trellis encoder with the dicode channel. (Reference paths are shown bold; detours are drawn semi-bold or dashed. A list $\mathcal{L}_l(s)$ with entries of the form (Δ, i) is attached to every trellis node (s, l).)

be emptied. The operation of removing entries that represent merged detours is indicated by "$\underset{\rightarrow}{\text{M}}$" in Figure 5.5. At depth 3 in Figure 5.5 (a), the dashed detour remerges with the reference path, which results in an update of $\bar{n}_{56}(0)$. Moreover, the entry $(84, 1)$ is removed from list $\mathcal{L}_3(1)$ since 84 exceeds $\Delta_{\max}$. This operation is denoted by "$\underset{\rightarrow}{\text{E}}$" in Figure 5.5. Notice that the processed lists at depth 3 are all empty in Figure 5.5 (a). Therefore, the MVA returns to depth 2 and increments $x_{\text{ref}\,2}$ by one. This results in the reference path shown in Figure 5.5 (b). It is essential that the lists $\mathcal{L}_j(s)$, $0 \leq j \leq 2$, need not be recomputed! Proceeding from depth 2 to depth 3, the MVA generates the lists $\mathcal{L}_3(0)$ and $\mathcal{L}_3(1)$. After processing these lists, it returns to depth 1, from where it follows the reference path shown in Figure 5.5 (c).

It is easy to see that the reference paths are generated by walking along a binary tree[16] in a depth-first manner. The beginning of this tree-walk is illustrated in Figure 5.6. Every terminal node in this tree

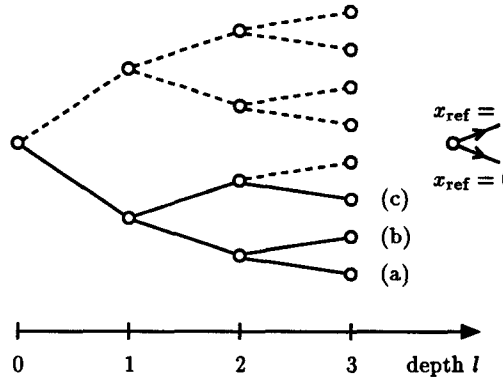


Figure 5.6: Tree-walk for generating reference paths. The labels (a), (b), and (c) correspond to the reference paths $(0, 0, 0)$, $(0, 0, 1)$, and $(0, 1, 0)$ and refer to Figure 5.5 (a), (b), and (c), respectively.

represents a reference information sequence $(x_{\text{ref}\,0}, x_{\text{ref}\,1}, \ldots, x_{\text{ref}\,l-1})$ such that the lists $\mathcal{L}_l(s)$ in the MVA trellis are empty for all $s \in \mathcal{S}$ and at least one list $\mathcal{L}_{l-1}(s)$ for $s \in \mathcal{S}$ is non-empty. It should be pointed

[16] In the general case, the walk is along a 2^k-ary tree.

out that the terminal nodes are not necessarily all at the same depth, as it is the case for our simple example.

We conclude this section with some remarks on the performance and the computational costs of the MVA. For a given parameter $d_{\max}$, one usually specifies an upper limit for the depth of the MVA trellis. If the depth exceeds that limit, this is an indication for a catastrophic trellis encoder. However, even if the encoder is catastrophic, the MVA may terminate without exceeding the limit! The time required for the evaluation of an average distance spectrum grows exponentially with $d_{\max}$ since the maximum over all numbers L_d in (5.42) for $d \leq d_{\max}$ grows linearly (or faster) with $d_{\max}$ and since the number of reference paths grows exponentially with L_d. As a final comment, the MVA can be accelerated if the distance spectrum is known to be uniform. In that case, replacing the factor 2^{-kl} in (5.43) by one allows us to stop the algorithm the first time it reaches a depth l with $\mathcal{L}_l(s) = \{\}$ for all $s \in \mathcal{S}$.

5.5 Analysis of Bipolar Trellis Encoders for the Dicode Channel

In this section, some bipolar trellis encoders for the dicode channel [6] $H(z) = 1 - z^{-1}$ will be analyzed and compared using the modified Viterbi algorithm (MVA) of Section 5.4. For calculating distance spectra, it is convenient to assume ± 1 channel inputs and to use unnormalized, i.e., integer-valued, partial-response channel coefficients [15]. Hence, the Euclidean distances at the channel output will be based on a unit average symbol energy at the channel input and on a channel energy $\| h \|^2 = \sum_{m=0}^{\mu} |h_m|^2 \geq 1$. Therefore, replacing the unit average symbol energy by $\mathrm{E}\left[|Y_l|^2 \right] = R E_b$ (cf. (5.6)) and normalizing the channel coefficients to unit energy will scale the squared Euclidean distances by a factor $R E_b / \| h \|^2$. This is taken into account by modifying (5.38) as

$$P_2(\Delta) = Q\left(\sqrt{\frac{\Delta \cdot R}{2 \| h \|^2} \frac{E_b}{N_0}} \right). \tag{5.45}$$

We begin with a fairly simple (but remarkably good) $(2, 1)$ code for the dicode channel, the so-called *biphase code* [13, p. 821], which is a block code with the two codewords[17] $[\bar{1}1]$ and $[1\bar{1}]$. Observe that the code words are *balanced*, i.e., their components sum up to zero. Hence, the biphase code has a first-order spectral null at zero frequency, i.e., it is a simple MSN code. Assuming the input assignment $0/\bar{1}1$ and $1/1\bar{1}$, the steady-state encoder obtained from the cascade of the biphase encoder with the dicode channel has the trellis[18] shown in Figure 5.7. Notice that a channel state 1 or $\bar{1}$ is assigned to every state of the steady-state composite encoder, according to the bipolar symbol stored in the delay cell of the dicode channel. The branch labels have the form $x_i / z_{2i}\, z_{2i+1}$, where x_i is a binary information digit and z_l is an output of the channel filter. For the trellis in Figure 5.7, the MVA yields the average squared Euclidean distance spectrum

$$\mathcal{D}_E = \{(24, 1),\ (40, 1),\ (56, 1.75),\ (72, 2),\ \ldots\}. \tag{5.46}$$

The steady-state encoder is clearly non-catastrophic since the code 2-tuples on its four branches are distinct. However, it suffers from the following *node-synchronization problem.* Since a Viterbi decoder cannot

[17] For branch labels, we use $\bar{a}$ to denote $-a$.

[18] More precisely, Figure 5.7 shows a section of the infinite trellis.

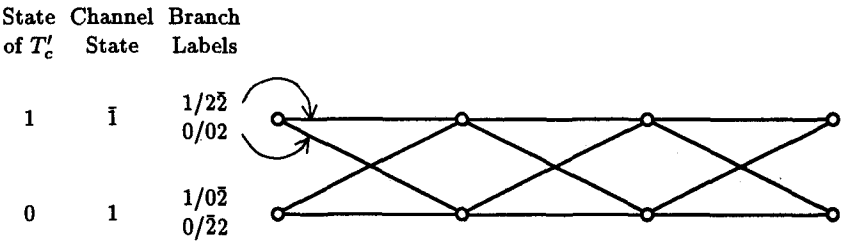


Figure 5.7: Trellis of the steady-state encoder obtained from the cascade of the biphase encoder with the dicode channel

generally be assumed to observe the entire transmitted codeword, an arbitrarily long sequence

$$\ldots 2\bar{2}2\bar{2}2\bar{2}2\bar{2}2 \ldots$$

will be decoded either as

$$\ldots 1111 \ldots$$

or as

$$\ldots 0000 \ldots,$$

depending on the timing offset. This defect can be eliminated by modifying the trellis as shown in Figure 5.8. Notice that the modified trellis is now based on a bipolar $(2,1)$ encoder with *two* states. The average

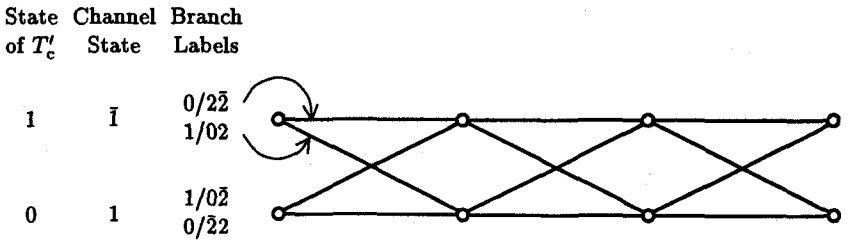


Figure 5.8: Trellis obtained by modifying the input assignment in Figure 5.7

squared Euclidean distance spectrum of the modified trellis is given by

$$\mathcal{D}_E = \{(24, 2),\ (40, 1),\ (56, 1.5),\ (72, 1.25),\ \dots\},$$

which yields twice the bit error probability of (5.46) at high E_b/N_0. One might therefore prefer the biphase encoder since a well-designed source encoder is very unlikely to produce long runs of identical information digits.

Among those bipolar $(2, 1)$ encoders, which yield a steady-state composite encoder with $N \leq 8$ states in cascade with the dicode channel, we have not been able to find a better encoder than the biphase encoder and, for the best encoders found, free squared Euclidean distance at the channel output is 24 for $N = 2$ and $N = 8$, but only 16 for $N = 4$. It appears that the dicode channel is responsible for this surprising behavior.

The following two examples will demonstrate the importance of the average number of bit errors over all detours at free distance for trellis codes with a non-uniform distance spectrum. Consider the trellis in Figure 5.9 of Wolf and Ungerboeck's four-state encoder [15, Fig. 11 (b)], for which $\Delta_f = d_{Ef}^2 = 16$. Notice that all detours at free distance have length $M + 1 = 3$, where $M = 2$ is the detour memory. Moreover, for each of the $2^{M+1} = 8$ reference paths of length 3 with the same start-node, there is exactly one detour of length 3, and every such detour is at free distance and causes two bit errors, regardless of the start-node. Hence, the average number of bit errors at free distance is also two. Indeed, the MVA yields

$$\mathcal{D}_E = \{(16, 2),\ (24, 3),\ (32, 3.5),\ (40, 8.5),\ \dots\}.$$

The trellis in Figure 5.9 should be compared[19] to the one in Figure 5.10, where again $\Delta_f = 16$ and all detours at free distance have length $M + 1 = 3$. For each start-node, however, there is only one pair of parallel paths at free distance (shown bold in Figure 5.10 for the start-node $(0, 0)$), and the corresponding information sequences differ in two bits. Since the probability of choosing one of these paths as the reference path is $2/2^{M+1} = 2^{-M} = 0.25$, the average number of bit errors over all detours at free distance is $0.25 \cdot 2 = 0.5$. The average squared Euclidean distance spectrum now becomes

$$\mathcal{D}_E = \{(16, 0.5),\ (24, 1.75),\ (32, 3.875),\ (40, 9.875),\ \dots\}.$$

[19]The comparison is fair since both encoders use that input assignment, which avoids the above node-synchronization problem.

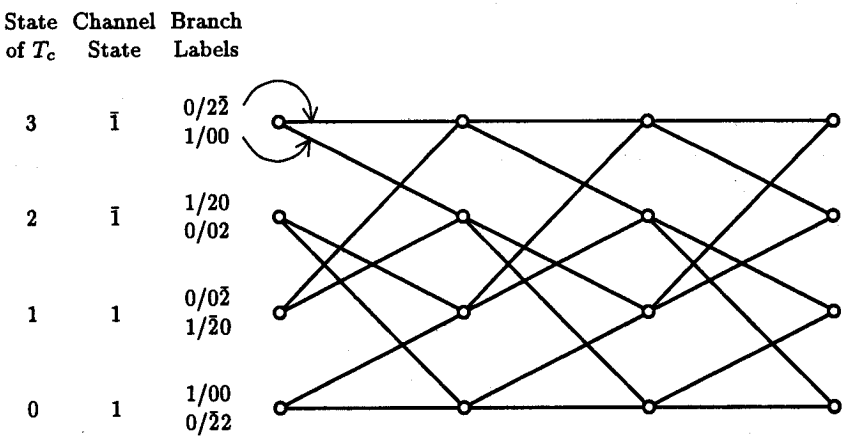


Figure 5.9: Trellis of a four-state encoder by Wolf and Ungerboeck that can be interpreted as a steady-state composite encoder according to Theorem 5.1. For every reference path of length 3, there is one detour of length 3 and every such detour is at free distance.

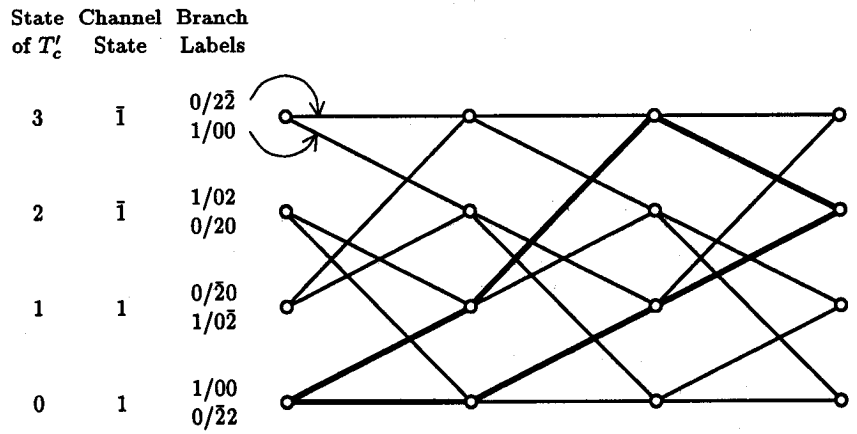


Figure 5.10: Trellis of a steady-state composite encoder with only one pair of parallel paths at free distance per start-node (shown bold for the start-node $(0,0)$).

Thus, the bit error probability at high E_b/N_0 has been reduced by a factor of $2^M = 4$.

For coded transmission over the dicode channel, we have already noted the superiority of the simple biphase encoder over other bipolar rate-1/2 encoders when the steady-state composite encoder has no more than eight states. In Table 5.1, we have listed the most important parameters of four bipolar rate-1/2 trellis encoders. Three of these

Encoder T	(n,k)	N_T	$N_{T'_c}$	Δ_f at encoder output	Δ_f at channel output	$\bar{n}_{\Delta_f}$ at channel output
MSN 1	(2,1)	1	2	8	24	1.00000
MSN 2	(2,1)	8	12	16	24	1.03125
MSN 3	(6,3)	8	16	16	24	0.28125
OFD-FC	(2,1)	8	16	24	16	0.06250

Table 5.1: Comparison of four bipolar rate-1/2 trellis encoders for the dicode channel. N_T and $N_{T'_c}$ denote the number of states of the trellis encoder T and the steady-state composite encoder T'_c, respectively. The squared Euclidean distances at the channel output are based on *unnormalized* channel coefficients.

trellis encoders are MSN encoders and one is a convolutional encoder that was not designed for the dicode channel:

- MSN 1 is the biphase encoder described above.
- MSN 2 is the best encoder (w.r.t. free Euclidean distance at the encoder output) obtained from a search based on the assignment of code 2-tuples to the branches of 2^k-ary state-transition diagrams. In the Mealy representation $(\{1,\bar{1}\}, \{0,1,\ldots,7\}, f, g)$, the next-state function is given by $f(s,x) = f_{s,x}$, $0 \le s < 8$, $x \in \{0,1\}$, where

$$\mathbf{F} = [f_{s,x}] = [0,1;\ 2,3;\ 4,5;\ 6,7;\ 0,1;\ 3,6;\ 4,5;\ 2,7],$$

  and the output function is given by $g(s,x) = \underline{v}(g^0_{s,x})$, $0 \le s < 8$, $x \in \{0,1\}$, where

$$\mathbf{G}^0 = [g^0_{s,x}] = [2,3;\ 1,2;\ 1,3;\ 3,0;\ 0,2;\ 0,2;\ 0,1;\ 3,1]$$

and $\underline{v}(\,.\,)$ is given by the table

$g^0_{s,x}$	$\underline{v}(g^0_{s,x})$
0	$\bar{1}\,\bar{1}$
1	$1\,\bar{1}$
2	$\bar{1}\,1$
3	$1\,1$

.

According to [21], MSN 2 generates a first-order spectral null at zero frequency, because there is a mapping ψ from the state set $\mathcal{S} = \{0, 1, 2, \ldots, 7\}$ to the set of real numbers $\{0, 2, 4\}$ such that, for any path γ in the labeled-digraph representation of MSN 2 (cf. Definition 5.1), the so-called *running digital sum* of γ is given by $\psi(\epsilon(\gamma)) - \psi(\sigma(\gamma))$, where $\epsilon(\gamma)$ and $\sigma(\gamma)$ denote the end-node and the start-node of γ, respectively.

- The $(n, k) = (6, 3)$ encoder MSN 3 was obtained from a standard binary convolutional encoder with $(n, k) = (4, 3)$ and $m = 2$ [59, Fig. 10.3, p. 292 and Table 11.1(e), p. 331] using the 'Hamming distance preserving mapping' introduced by Ferreira [67]. At the price of a small rate reduction, this mapping yields *balanced* code 6-tuples, which property implies a spectral null at zero frequency.

- Finally, OFD-FC is a standard binary convolutional encoder with $(n, k) = (2, 1)$ and $m = 3$ [59, Table 11.1(c), p. 330] (designed for optimum free distance on a flat channel) with a binary-to-bipolar output mapping.

Among these encoders, MSN 1 has by far the smallest (Viterbi) decoding complexity. Observe that the dicode channel increases free Euclidean distance when one of the MSN encoders is used. An interesting property of encoder MSN 2 and encoder MSN 3 should also be mentioned, namely, that the free squared Euclidean distance both at the encoder output and at the channel output *exceeds* the lower bound reported in [13, Prop. 6], [16, Thm. 6], which has been observed to hold with equality in most cases [68]. (For encoders with a first-order spectral null at zero frequency and the dicode channel, the lower bound at the encoder output and at the channel output is $\Delta_f \geq 8$ and $\Delta_f \geq 16$, respectively.) Because of the large free squared Euclidean distance at the encoder output and at the channel output, MSN 2 and MSN 3 are useful for both the flat channel and the dicode channel with AWGN!

Leer - Vide - Empty

Chapter 6

Conclusions

The results obtained in this dissertation can be summarized as follows.

In Chapter 2, we have introduced the notions of proper complex random variables and processes for statistical communication theory.

- For proper complex random variables and processes, which are defined by a vanishing pseudo-covariance, the second-order statistics are specified completely by the mean and the covariance.
- The differential entropy of a complex random vector with a specified correlation matrix has been shown to take its maximum if and only if the random vector is zero-mean Gaussian and *proper*.
- For a bandpass communication channel with additive, wide-sense stationary noise, the noise process in the equivalent baseband channel has been shown to be proper complex.
- A DFT correspondence has been derived relating circular stationarity in the time domain to uncorrelatedness in the frequency domain for sequences of proper complex random variables.

In Chapter 3, the capacity of ISI channels with AWGN and the information rate for i.i.d. inputs on such channels have been dealt with.

- The derivation of the capacity of the ISI channel with AWGN has been simplified and the results have been generalized to channels with a complex unit-sample response and proper complex AWGN.

- Allpass-equivalent ISI channels, which have the same transfer function up to a suitable allpass factor, have been shown to be equivalent also with respect to mutual information.
- The derivation of a lower bound on the information rate for i.i.d. inputs has been simplified by using the information-theoretic equivalence of certain allpass-transformed ISI channels.

In Chapter 4, we have investigated K-ary state-transition diagrams (STD's) for trellis encoders.

- An algorithm for the construction of all non-isomorphic K-ary STD's with given topological constraints has been derived.
- It has been shown that K-ary STD's with maximum detour memory, i.e., K-ary STD's with $N = K^M$ nodes and detour memory M, are strongly connected and have uniform in-degree.
- All non-isomorphic binary STD's with maximum detour memory and $N = 1, 2, 4, 8$, and 16 nodes have been found.

In Chapter 5, we have studied trellis-coded data transmission over ISI channels.

- A simple upper bound on the free distance of an (n, k) trellis code has been derived, which is proportional to the detour memory of a 2^k-ary STD.
- The serial form of a controllable trellis encoder followed by an FIR filter has been viewed as the serial form of a composite trellis encoder that contains exactly one maximal strongly connected component, referred to as the steady-state encoder.
- The well-known upper bound on bit error probability for convolutional encoders and maximum-likelihood decoding has been generalized to time-invariant trellis encoders with a non-uniform distance spectrum.
- The generalized upper bound on bit error probability involves an average distance spectrum, which we have shown can be evaluated efficiently using a modified Viterbi algorithm.

- It has been shown that the average number of bit errors over all detours at free distance can be smaller than one for trellis encoders with a non-uniform distance spectrum and is therefore a more important parameter for non-uniform encoders than for uniform ones.

Leer - Vide - Empty

Abbreviations

AWGN	Additive White Gaussian Noise
CRD	Component-Reduced Digraph
c.w.s.s.	circularly wide-sense stationary
DFT	Discrete Fourier Transform
DTGC	Discrete-Time Gaussian Channel
ETH	Eidgenössische Technische Hochschule (Swiss Federal Institute of Technology)
FIR	Finite Impulse Response
HDSL	High-rate Digital Subscriber Line
IDFT	Inverse Discrete Fourier Transform
i.i.d.	independent and identically distributed
IIR	Infinite Impulse Response
ISI	Intersymbol interference / Institut für Signal- und Informationsverarbeitung (Signal and Information Processing Laboratory)
MGC	Memoryless Gaussian Channel
ML	Maximum Likelihood
MLSE	Maximum-Likelihood Sequence Estimation
MSN	Matched Spectral Null
MVA	Modified Viterbi Algorithm
NCGC	N-Circular Gaussian Channel
OFD	Optimum Free Distance
PAM	Pulse Amplitude Modulation

p.d.f.	probability density function
QAM	Quadrature Amplitude Modulation
STD	State-Transition Diagram
VA	Viterbi Algorithm
WGN	White Gaussian Noise
w.s.s.	wide-sense stationary

Bibliography

[1] H. Nyquist, "Certain factors affecting telegraph speed," *Bell Syst. Tech. J.*, p. 324, Apr. 1924.

[2] J. G. Proakis, *Digital Communications.* New York: McGraw-Hill, 1983.

[3] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the I.R.E.*, vol. 37, pp. 10–21, Jan. 1949.

[4] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering.* New York: Wiley, 1965.

[5] S. Kasturia, J. T. Aslanis, and J. M. Cioffi, "Vector coding for partial response channels," *IEEE Trans. Inform. Theory*, vol. IT-36, pp. 741–762, July 1990.

[6] P. Kabal and S. Pasupathy, "Partial response signaling," *IEEE Trans. Commun.*, vol. COM-23, pp. 921–934, Sept. 1975.

[7] G. D. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363–378, May 1972.

[8] I. N. Andersen, "Sample-whitened matched filters," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 653–660, Sept. 1973.

[9] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-34, pp. 380–388, May 1988.

[10] R. G. Gallager, *Information Theory and Reliable Communication.* New York: Wiley, 1968.

[11] K. J. Kerpez, "Viterbi receivers in the presence of severe intersymbol interference," in *IEEE Global Telecomm. Conf. GLOBECOM '90*, pp. 2009–2013, 1990.

[12] G. Ungerboeck, "Adaptive maximum-likelihood receiver for carrier-modulated data-transmission systems," *IEEE Trans. Commun.*, vol. COM-22, pp. 624–636, May 1974.

[13] R. Karabed and P. H. Siegel, "Matched spectral-null codes for partial-response channels," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 818–855, May 1991.

[14] A. V. Oppenheim and R. W. Schafer, *Digital Signal Processing.* Englewood Cliffs, NJ: Prentice Hall, 1975.

[15] J. K. Wolf and G. Ungerboeck, "Trellis coding for partial-response channels," *IEEE Trans. Commun.*, vol. COM-34, pp. 765–773, Aug. 1986.

[16] E. Eleftheriou and R. Cideciyan, "On codes satisfying M-th order running digital sum constraints," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1294–1313, Sept. 1991.

[17] B. H. Marcus and P. H. Siegel, "On codes with spectral nulls at rational submultiples of the symbol frequency," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 557–568, July 1987.

[18] C. M. Monti and G. L. Pierobon, "Codes with a multiple spectral null at zero frequency," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 463–472, Mar. 1989.

[19] R. L. Adler, D. Coppersmith, and M. Hassner, "Algorithms for sliding block codes: An application of symbolic dynamics to information theory," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 5–22, Jan. 1983.

[20] B. H. Marcus, P. H. Siegel, and J. K. Wolf, "Finite-state modulation codes for data storage," *IEEE J. Sel. Areas Commun.*, vol. SAC-10, pp. 5–37, Jan. 1992.

[21] G. L. Pierobon, "Codes for zero spectral density at zero frequency," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 435–439, Mar. 1984.

[22] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," to appear, *IEEE Trans. Inform. Theory*, vol. IT-39, July 1993.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: Wiley, 1991.

[24] W. Hirt, *Capacity and Information Rates of Discrete-Time Channels with Memory.* Ph.D. thesis, Swiss Federal Institute of Technology (ETH), Zürich, 1988.

[25] S. Shamai (Shitz), L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 1527–1539, Nov. 1991.

[26] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding.* New York: McGraw-Hill, 1979.

[27] J. L. Massey, *Applied Digital Information Theory I (Lecture Notes).* ETH-Zürich, ISI, 1992.

[28] S. Haykin, *Adaptive Filter Theory.* Englewood Cliffs, NJ: Prentice Hall, 1986.

[29] J. L. Doob, *Stochastic Processes.* New York: Wiley, 1953.

[30] R. A. Wooding, "The multivariate distribution of complex normal variables," *Biometrika*, vol. 43, pp. 212–215, 1956.

[31] W. Feller, *An Introduction to Probability Theory and its Applications*, vol. II. New York: Wiley, 2nd ed., 1966.

[32] W. A. Gardner, *Introduction to Random Processes.* New York: McGraw-Hill, 2nd ed., 1990.

[33] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution," *Ann. Math. Statist.*, vol. 34, pp. 152–176, 1963.

[34] E. J. Kelly and I. S. Reed, "Some properties of stationary Gaussian processes," tech. rep. TR-157, MIT Lincoln Lab, June 1957.

[35] R. Arens, "Complex processes for envelopes of normal noise," *IRE Trans. Inform. Theory*, vol. IT-3, pp. 204–207, Sept. 1957.

[36] I. S. Reed, "On a moment theorem for complex Gaussian processes," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 194–195, Apr. 1962.

[37] R. Bellman, *Introduction to Matrix Analysis.* New York: McGraw-Hill, 1960.

[38] P. Lancaster and M. Tismenetsky, *The Theory of Matrices.* New York: Academic Press, 1985.

[39] J. Dugundji, "Envelopes and pre-envelopes of real waveforms," *IRE Trans. Inform. Theory*, vol. IT-4, pp. 53–57, Mar. 1958.

[40] M. Zakai, "Second-order properties of the pre-envelope and envelope processes," *IRE Trans. Inform. Theory*, vol. IT-6, pp. 556–557, Dec. 1960.

[41] M. Zakai, "The representation of narrow-band processes," *IRE Trans. Inform. Theory*, vol. IT-8, pp. 323–325, July 1962.

[42] T. Kailath, *Linear Systems.* Englewood Cliffs, NJ: Prentice Hall, 1980.

[43] S. Verdú, "The capacity region of the symbol-asynchronous Gaussian multiple-access channel," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 733–751, July 1989.

[44] R. S. Cheng and S. Verdú, "Gaussian multiaccess channels with ISI: Capacity region and multiuser water-filling," *IEEE Trans. Inform. Theory*, vol. IT-39, pp. 773-785, May 1993.

[45] L. A. Zadeh and C. A. Desoer, *Linear System Theory: The State Space Approach.* New York: McGraw-Hill, 1963.

[46] W. T. Tutte, *Graph Theory*, Encyclopedia of Mathematics and its Applications, vol. 21. Reading, MA: Addison-Wesley, 1984.

[47] S. Even, *Graph Algorithms.* Rockville, Maryland: Computer Science Press, 1979.

[48] Z. Kohavi, *Switching and finite automata theory.* New York: McGraw-Hill, 2nd ed., 1978.

[49] R. W. Robinson, "Counting strongly connected finite automata," in *Graph Theory with Appl. to Algorithms and Comp. Science*, Y. Alavi, Ed., pp. 671–685. New York: Wiley, 1985.

[50] R. Tarjan, "Depth-first search and linear graph algorithms," *SIAM J. Control*, vol. 1, pp. 146–160, 1972.

[51] R. Sedgewick, *Algorithms in C++*. Reading, MA: Addison-Wesley, 1992.

[52] J. S. Chow, J. C. Tu, and J. M. Cioffi, "A discrete multitone transceiver system for HDSL applications," *IEEE J. Sel. Areas Commun.*, vol. SAC-9, pp. 895–907, Aug. 1991.

[53] M. Tomlinson, "New automatic equaliser employing modulo arithmetic," *Electronics Letters*, vol. 7, pp. 138–139, Mar. 1971.

[54] M. Miyakawa and H. Harashima, "A method of code conversion for a digital communication channel with intersymbol interference," *Trans. Inst. Electron. Commun. Eng. Jap.*, vol. 52-A, pp. 272–273, June 1969.

[55] M. V. Eyuboğlu and G. D. Forney, Jr., "Trellis precoding: Combined coding, precoding and shaping for intersymbol interference channels," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 301–313, Mar. 1992.

[56] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, pp. 268–276, Mar. 1973.

[57] J. L. Massey, "Foundation and methods of channel encoding," in *Proc. Int. Conf. Inform. Theory and Systems, NTG-Fachbericht*, vol. 65, pp. 148–157, Sept. 1978.

[58] H.-A. Löliger, *On Euclidean-Space Group Codes.* Ph.D. thesis, Swiss Federal Institute of Technology (ETH), Diss. ETH No. 9720, Zürich, 1992.

[59] S. Lin and D. J. Costello, *Error Control Coding: Fundamentals and Applications.* Englewood Cliffs, NJ: Prentice Hall, 1983.

[60] J. L. Massey and M. K. Sain, "Inverses of linear sequential circuits," *IEEE Trans. on Computers*, vol. C-17, pp. 330–337, Apr. 1968.

[61] R. Ash, *Information Theory.* New York: Wiley, 1965.

[62] I. N. Bronstein and K. A. Semendjajew, *Taschenbuch der Mathematik.* Thun und Frankfurt a. Main: Harri Deutsch, 1981.

[63] A. R. Calderbank and N. J. A. Sloane, "New trellis codes based on lattices and cosets," *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 177–195, Mar. 1987.

[64] M. Rouanne and D. J. Costello, "An algorithm for computing the distance spectrum of trellis codes," *IEEE J. Sel. Areas Commun.*, vol. SAC-7, pp. 929–940, Aug. 1989.

[65] C. Schlegel, "Evaluating distance spectra and performance bounds of trellis codes on channels with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-37, pp. 627–634, May 1991.

[66] C. D. Frank, "Comments on 'Evaluating distance spectra and performance bounds of trellis codes on channels with intersymbol interference'," *IEEE Trans. Inform. Theory*, vol. IT-38, pp. 1623–1625, Sept. 1992.

[67] H. C. Ferreira, D. A. Wright, and A. L. Nel, "Hamming distance preserving mappings and trellis codes with constrained binary symbols," *IEEE Trans. Inform. Theory*, vol. IT-35, pp. 1098–1103, Sept. 1989.

[68] E. Eleftheriou, private communication, 1992.

Index

An *italic* page number directs to the definition of the index entry.

Leer - Vide - Empty

Curriculum Vitae

Fredy Daniel Neeser, geboren am 27. September 1960 in Zürich.

1967-1973 Besuch der Primarschule in Adliswil.

1973-1980 Ausbildung am stadtzürcherischen Literargymnasium Freudenberg. Abschluss mit eidgenössischer Matura, Typus A.

1980-1986 Studium der Elektrotechnik an der ETH Zürich mit Vertiefung in Nachrichtentechnik (Informationstheorie und Netzwerktheorie). Studienarbeit über Partial-Response Signaling. Abschluss als diplomierter Elektroingenieur ETH.

1987-1989 Nach dem Eintritt ins Institut für Signal- und Informationsverarbeitung (ISI) der ETH Zürich, Bearbeitung des Projektes 'Adaptive Gabelschaltung' in Zusammenarbeit mit der Alcatel in Zürich und der Arbeitsgemeinschaft für Industrie und Forschung (AFIF). Erfahrungen mit adaptiven Filtern (Fast Least Squares) zur Echokompensation und mit dem Gleitkomma-Signalprozessor DSP32-C von AT&T.

1989-1993 Doktorand am ISI unter der Leitung von Prof. Dr. J. L. Massey. Mitwirkung an einem Mobilfunk-Projekt der ETH und der schweizerischen PTT. Design von Empfängern für Direct-Sequence Spread-Spectrum Systeme (US Pat. No. 5,181,225). Vereinfachung der Kapazitätsberechnung von Kanälen mit Intersymbol-Interferenz mittels Diskreter Fourier-Transformation durch Einführung komplexer Kanalmodelle und Anwendung der Methode zur Bestimmung der Kapazitätsregion von asynchronen Vielfachzugriffskanälen. Untersuchung trellis-codierter Datenübertragung auf Kanälen mit Intersymbol-Interferenz.