Two topics on intersymbol interference channels

Tax, Stefan J.E.T.

Award date:
1988

EINDHOVEN UNIVERSITY OF TECHNOLOGY

Department of Electrical Engineering

Information and Communication Theory Group

graduate report

TWO TOPICS ON INTERSYMBOL INTERFERENCE CHANNELS

by

Stefan J.E.T. Tax

coach: Dr.Ir. A.J. Vinck

EINDHOVEN, January 1988

PREFACE

This graduate report is the result of the research I conducted from november 1987 until december 1988 at the Group Information and Communication Theory of the Technical University Eindhoven. Part of the research was done at the Deutsche Forschungs- und Versuchsanstalt fuer Luft- und Raumfahrt (DFVLR) in Oberpfaffenhofen. The report consists of two parts that both treat channels with intersymbol interference (ISI).

In the first part the ideal bandlimited channel is considered. The conditions for ISI-free communication and the effects of ISI on the performance of bit-by-bit detection are treated. This part of the report was meant partly as an exercise to get acquantaince with intersymbol interference.

In the second part the (digital) magnetic recording channel is considered. Aim of this part is a clear presentation of different approaches and models for the magnetic recording channel. The need for such an overview came from difficulties in comparing results from various publications on coding for the magnetic recording channel (in particular [IMM87] and [CAL]). Attention is also paid to the difference in view on the magnetic recording system between its user and the designer of codes for the system. The presented theory serves as a basis for the design of codes for the magnetic recording channel.

Summaries are included in the two parts.

ACKNOWLEDGEMENTS

I wish to thank the members of the Information and Communication Theory Group for the instructive and pleasant time I had during my stay. Specials thanks are due to my coach Dr. Ir. Han Vinck who supported me not only in the theoretical aspects of my research. I also want to thank Dr. Robert Schweikert and Dipl. Ing. Hans Peter Foerster of the DFVLR for the opportunity to conduct part of my research at their group and the many fruitful discussions on intersymbol interference. Finally I want to thank Dr. Ir. Kees A. Schouhamer Immink of the Philips Research Laboratories for the stimulating discussions on magnetic recording.

TABLE OF CONTENTS

PART II : ON MODELS AND CODING FOR THE MAGNETIC RECORDING CHANNEL

PART I

COMMUNICATION OVER AN IDEAL BANDLIMITED CHANNEL

SUMMARY

In this part of the report we consider baseband communication over an ideal bandlimited additive white Gaussian noise channel. The signalling method is binary pulse amplitude modulation. Bit-by-bit detection is employed. It is well known that for this channel intersymbol interference (ISI) free communication is possible for rates up to the Nyquist rate, being twice the channelbandwidth W. We'll show that using raised cosine spectrum pulses that occupy the total available bandwidth W the error performance is equal at all rates $W<R<2W$. The following problem is considered: having designed an ISI-free system for a certain channel rate R, this rate must be increased. (For instance to be able to apply coding at a constant source rate.) With the former result it is seen instantaneously that this increase can be accomplished without degradation of the errorperformance. Modification of the transmitter and receiver filters is necessary. Another possibility is to increase the rate without system modification. The loss due to the arising intersymbol interference is described. It is computed for an example.

As this report is meant as an exercise and introduction to intersymbol interference the analysis is sometimes rather extensive.

1 INTRODUCTION

This part of the report is about communication over ideal bandlimited channels. It deals with several interrelated problems. The first problem can be stated as follows:

> Consider an ideal bandlimited channel with bandwidth W. It has frequency transfer function:
>
> $$C(f)= \begin{cases} 1 & |f| \leq W \\ 0 & \text{otherwise} \end{cases}$$
>
> At which rates can we communicate over this channel, using binary pulse amplitude modulation, under the condition that the pulses do not overlap at the detection moments? In other words: the communication must be intersymbol interference (ISI) free.

In chapters 2 and 3 this question is answered. In chapter 2 the conditions for ISI-free communication at a certain rate R are derived. This gives the well known Nyquist criterion. Raised cosine spectrum pulses satisfy the Nyquist criterion. In chapter 3 it will be shown that using these pulses it is possible to communicate at rates $W \leq R \leq 2W$ over an ideal bandlimited channel, with equal errorperformance at all rates.

With this result we can formulate another problem, that has resemblance to practical situations:

> Assume we have a communicationsystem that is designed for ISI-free communication at channel rate R. It uses raised

cosine spectrum pulses. Bit-by-bit detection is employed. This is possible because of the ISI-freeness. The bandwidth occupied by the signals exactly equals the maximum allowed bandwidth. (E.g for frequency division multiplexing).
Now for some reason the channel rate must be increased. (E.g. because a channel code has to be applied at constant source rate). The question is how to achieve this increase in rate without violencing the bandwidth constraint.

With the former result it can be seen immediately that it is possible to increase the channelrate up to 2W without deterioration of the errorperformance. This however requires a redesign of the system, especially of the transmitter and channel filters. This is described in chapter 3.

It is also possible to increase the channel rate without modification of the filters. Using the system at a higher rate than it was designed for causes intersymbol interference. The deterioration of the errorperformance due to the ISI is decribed in chapter 4.

In this report we only consider the theoretical aspects of intersymbol interference. Much attention will be paid to the frequency domain interpretation of the processes. We explicitly do not take into account practical problems like synchronisation, clock recovery and realizability of the filters. Starting point for the analysis is the communication theory as it is presented by Wozencraft and Jacobs [WOZ65]. The reader is assumed to be familiar with this theory.

2 DESIGN OF A BANDLIMITED ISI-FREE SYSTEM

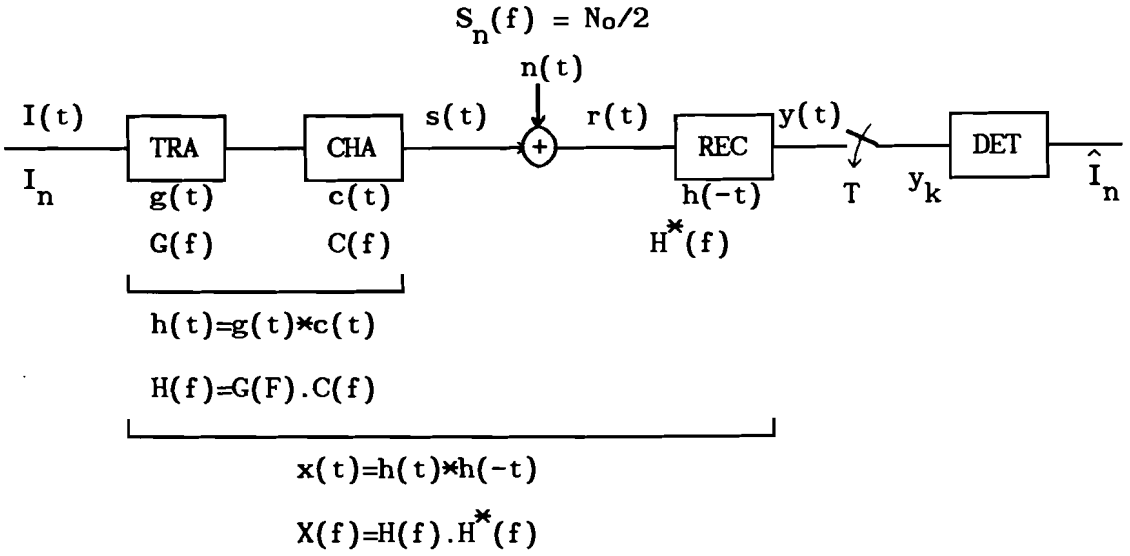
We will describe the design of a system for ISI-free communication at channel rate R, using bandlimited pulses. First the communication system will be presented. Then the conditions for ISI-free communication will be derived. The raised cosine spectrum pulses that satisfy these conditions will be presented.

2.1 Communication system model

We restrict ourselves to baseband communication using binary pulse amplitude modulation (PAM). The system (fig. 2.1) consists of a transmitter filter G(f), a channel filter C(f), a receiver filter $H^*(f)$, a sampler and a (bit-by-bit) threshold detector. Transmitter and channel together have transfer function H(f)=G(f).C(f). The receiver filter is matched to the concatenation of transmitter and channel filter.

In this chapter the transmitter and channel filter will not be considered separate. The condition for ISI-free communication will be given in terms of $X(f)=H(f).H^*(f)$. The actual bandwidth of the channel will therefore not be specified in this chapter. The H(f) though, must have some bandlimitation. In chapter 3 the possibilities to communicate over an ideal channel with specified bandwidth W will be presented.

The signals are corrupted by zero mean additive white Gaussian noise n(t) with power density spectrum $S_n(f)=N_0/2$. The channel symbol



$$I(t) = \sum_{n=0}^{\infty} I_n.\delta(t-nT) \qquad I_n \in \{ -\sqrt{E_c}, +\sqrt{E_c} \}$$

$$s(t) = \sum_{n=0}^{\infty} I_n.h(t-nT)$$

$$r(t) = \sum_{n=0}^{\infty} I_n.h(t-nT) + n(t) \qquad S_n(f) = N_o/2$$

$$y(t) = \sum_{n=0}^{\infty} I_n.x(t-nT) + z(t) \qquad S_z(f) = (N_o/2).|H^*(f)|^2$$

Fig. 2.1 The binary PAM communication system

sequence $\underline{I}=(I_0,I_1,\ldots)$ is uncoded. The symbols $I_i \in \{-\sqrt{E_c},+\sqrt{E_c}\}$ are taken from a symmetric memoryless source.

Pulse amplitude modulation

Binary PAM is a form of bit-by-bit signalling [WOZ65,p.290]. For symbol k the receiver decides between two antipodal signals $-\sqrt{E_c}h(t-kT)$ and $+\sqrt{E_c}h(t-kT)$, each of energy E_c (fig. 2.2). The signal out of the channel (not corrupted by noise) is given by:

$$s(t)=\sum_{n=0}^{\infty} I_n.h(t-nT) \qquad I_n \in \{-\sqrt{E_c},+\sqrt{E_c}\}. \qquad (2.1)$$

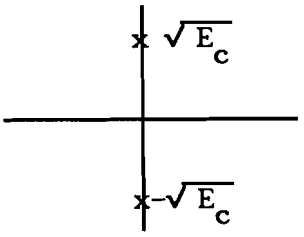


Fig. 2.2 Signal constellation of binary PAM

In an infinite bandwidth system the pulse h(t) can have finite duration T_h. By sending the channelsymbols at intervals $T \geq T_h$ we accomplish that the pulses do not overlap at reception. This is a way to make the building block waveform h(t) orthogonal to a time translate h(t-nT):

$$\int_{-\infty}^{\infty} h(t).h(t-nT).dt = 0 \quad \text{for } n \neq 0. \qquad (2.2)$$

When this condition is met the optimal maximum a posteriori probability (MAP) receiver reduces to a matched filter followed by a sampler and a threshold detector taking bit-by-bit decisions.

In an bandlimited system the pulses have infinite duration. This complicates the design of a pulse that is orthogonal to all its T-sec time translates. Before describing the design of these pulses, the function of the receiver and its implementation will be discussed briefly.

Receiver structure

Let h(t) be a pulse that is orthogonal to all its T-sec time translates. To detect the k-th pulse the received signal r(t)=s(t)+n(t) is "projected" on the one-dimensional space spanned by h(t-kT). The for the decision relevant data vector is

$$\begin{aligned} y_k \triangleq y(kT) &= \int_{-\infty}^{\infty} r(t).h(t-kT).dt \\ &= \int_{-\infty}^{\infty} \Big(\sum_{n=0}^{\infty} I_n.h(t-nT) + n(t) \Big).h(t-kT).dt \\ &= \sum_{n=0}^{\infty} I_n . \int_{-\infty}^{\infty} h(t-nT).h(t-kT).dt + \int_{-\infty}^{\infty} n(t).h(t-kT).dt \\ &= I_k . \int_{-\infty}^{\infty} h^2(t).dt + z_k . \end{aligned} \tag{2.3}$$

The last step follows from the orthogonality. By choosing

$$\int_{-\infty}^{\infty} h^2(t).dt = 1 \tag{2.4}$$

we make the pulses orthonormal. The energy sent over the channel per channel bit is equal to

$$E_c = E\{I_k^2\} . \int_{-\infty}^{\infty} h^2(t).dt = E\{I_k^2\} . \tag{2.5}$$

where E{.} is expectation.

So $I_k = \pm \sqrt{E_c}$ and

$$y_k = I_k + z_k . \tag{2.6}$$

The noise term z_k is the projection of the noise n(t) upon the signal space spanned by h(t-kT). With the theorem of irrelevance [WOZ65,p.229] it follows that the part of the noise outside of this signal space is irrelevant for optimal detection. In the sequel we will show that the noise samples z_k are uncorrelated and independent (§2.3). This is a consequence of the fact that h(t) is orthogonal to

all h(t-nT).

A receiver that computes the y_k by multiplying r(t) with h(t-kt) and integrating the result is called a correlation receiver. Normally, when the pulse h(t) is time limited, this can be implemented by a filter with impulse response h(-t) matched to h(t) followed by a sampler. Although h(t) has infinite duration here, we will assume this matched filter receiver structure. In practice namely, the pulse h(t) has to be time limited, and is an approximation of the ideal infinite duration pulse.

2.2 Pulse design for ISI-free communication

To determine an infinite response pulse h(t) that is orthonormal to all its time translates h(t-kT), we consider the total system of transmitter, channel and receiver with transfer function

$$X(f)=H(f).H^{*}(f) \qquad (2.7a)$$

and impulse response

$$x(t)=h(t)*h(-t). \qquad (2.7b)$$

At the output of the receiver filter we have

$$y(t)=\sum_{n=0}^{\infty} I_n.x(t-nT) + z(t) \qquad (2.8)$$

with z(t) coloured zero mean Gaussian noise. From the orthonormality of h(t) and h(t-kT) it follows that

$$x(kT) = \begin{cases} 1 & \text{for } k=0 \\ 0 & \text{for } k\neq 0 \end{cases} \qquad (2.9)$$

Discrete time system

As we consider the signal values at times kT only, we can introduce the discrete time model of the system. Using an obvious notation we get:

$$\begin{aligned} y_k &= \sum_{n=0}^{\infty} I_n . x_{k-n} + z_k \\ &= I_k . x_0 + z_k \\ &= I_k + z_k \end{aligned} \qquad (2.10)$$

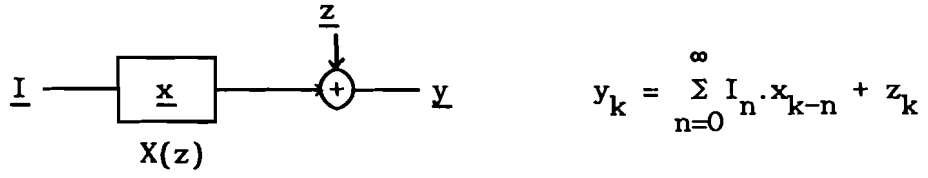


Fig. 2.3 Equivalent discrete time channel

For the z-transform of the system we have [PAP77,p31]:

$$X(z) \triangleq \sum_{n=-\infty}^{\infty} x_n . z^{-n} = x_0 . z^0 = 1. \qquad (2.11)$$

With $z=e^{j2\pi fT}$ we get the frequency response:

$$X(e^{j2\pi fT}) = 1. \qquad (2.12)$$

This is the condition on the frequency response of the discrete time system for ISI-free communication. Now we proceed to find the equivalent condition to be imposed on the continuous time frequency response.

If a continuous time system with bandlimited impulse response x(t) and Fourier transform X(f) is sampled at times kT, it is equal to a

discrete time system having frequency response [PAP77,p.26]

$$X(e^{j2\pi fT}) = \frac{1}{T} \sum_{n=-\infty}^{\infty} X(f+\frac{n}{T}) \triangleq \overline{X}(f) \qquad (2.13)$$

This implies that a whole class of continuous time systems has the same discrete time response. All continuous time systems X(f) with equal $\overline{X}(f)$ have the same discrete time response when sampled at rate 1/T.

We call the unique continuous time system that has frequency response limited to the interval $[-\frac{1}{2T}, \frac{1}{2T}]$ the equivalent Nyquist system of this class of continuous time systems

$$X_{eq}(f) = \begin{cases} \overline{X}(f) & |f| \leq \frac{1}{2T} \\ 0 & \text{otherwise} \end{cases} \qquad (2.14)$$

The Nyquist criterion

Now we find the condition on X(f) for ISI-free communication. It is called the Nyquist criterion (e.g. [LUC68,p.48]).

> Let a continuous time system with bandlimited frequency response X(f) be used at rate R=1/T. It is intersymbol interference free iff
>
> $$\overline{X}(f) \triangleq \frac{1}{T} \sum_{n=-\infty}^{\infty} X(f+\frac{n}{T}) = \text{constant} \qquad (2.15)$$

The simplest choice for X(f) is

$$X(f) = \begin{cases} T & |f| \leq \frac{1}{2T} \\ 0 & \text{otherwise} \end{cases} \qquad (2.16a)$$

with impulse response the sinc-pulse:

$$x(t) = \frac{\sin(\pi t/T)}{\pi t/T} \quad . \tag{2.16b}$$

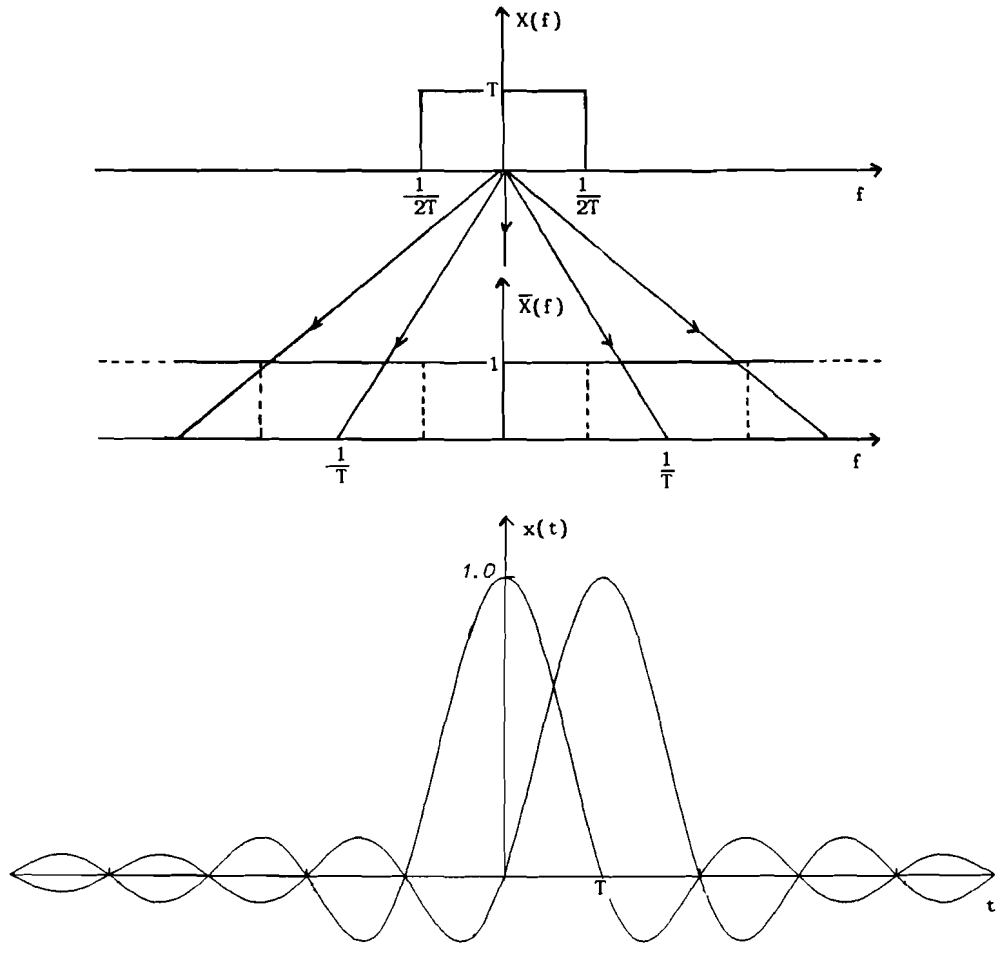


Fig. 2.4 Sinc-pulse satisfying the Nyquist criterion

Raised cosine spectrum pulses

The sinc-pulse is only one out of a class known as the raised cosine spectrum pulses. They satisfy the Nyquist criterion (fig. 2.5) and have frequency response [LUC68,p.50]:

$$X(f) = \begin{cases} T & 0 \leq |f| \leq \frac{1-\beta}{2T} \\ \frac{T}{2}\cdot\left[1-\sin(\pi T\frac{f-1/2T}{\beta})\right] & \frac{1-\beta}{2T} \leq |f| \leq \frac{1+\beta}{2T} \end{cases} \tag{2.17a}$$

$$x(t) = \frac{\sin(\pi t/T)}{\pi t/T} \cdot \frac{\cos(\beta\pi t/T)}{1-4\beta^2 t^2/T^2} \qquad 0 \leq \beta \leq 1 \qquad (2.17b)$$

β is called the roll-off factor of the spectrum (fig. 2.6).

The sinc pulse is a raised cosine spectrum pulse with roll-off factor β=0. It will be clear that the pulses with small roll-off factors are difficult to implement, because of the fast decay of the frequency response.

With Parseval's theorem it is easily checked that (2.17a) satisfies the normalisation (2.4):

$$\int_{-\infty}^{\infty} h^2(t)dt = \int_{-\infty}^{\infty} |H(f)|^2 df = \int_{-\infty}^{\infty} H(f).H^*(f).df = \int_{-\infty}^{\infty} X(f)df$$

Because the surface under X(f) is equal to the surface under $X_{eq}(f)$ we find:

$$= \int_{-\infty}^{\infty} X_{eq}(f)df = \int_{-1/2T}^{1/2T} X_{eq}(0)df = \frac{1}{T}.X_{eq}(0) = \frac{1}{T}.X(0) = \frac{1}{T}.T = 1 \qquad (2.18)$$

We used the fact that $X_{eq}(f)=X_{eq}(0)=X(0)=T$ for $|f| \leq 1/2T$.

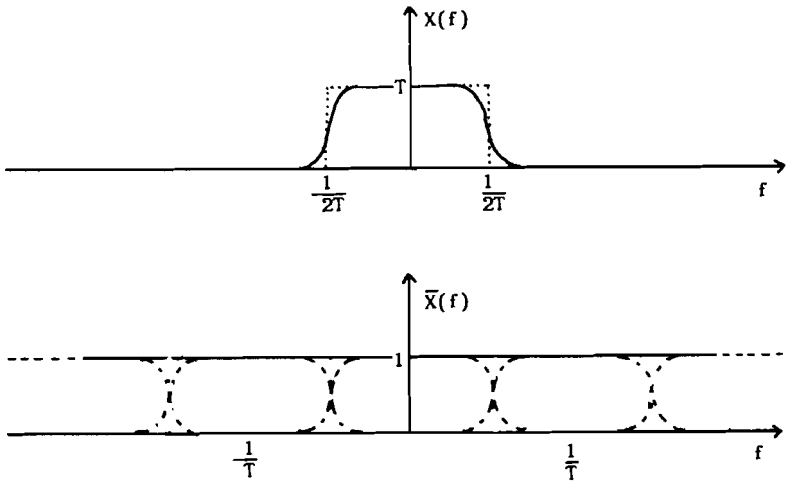


Fig. 2.5 Raised cosine spectrum pulse satisfying the Nyquist criterion

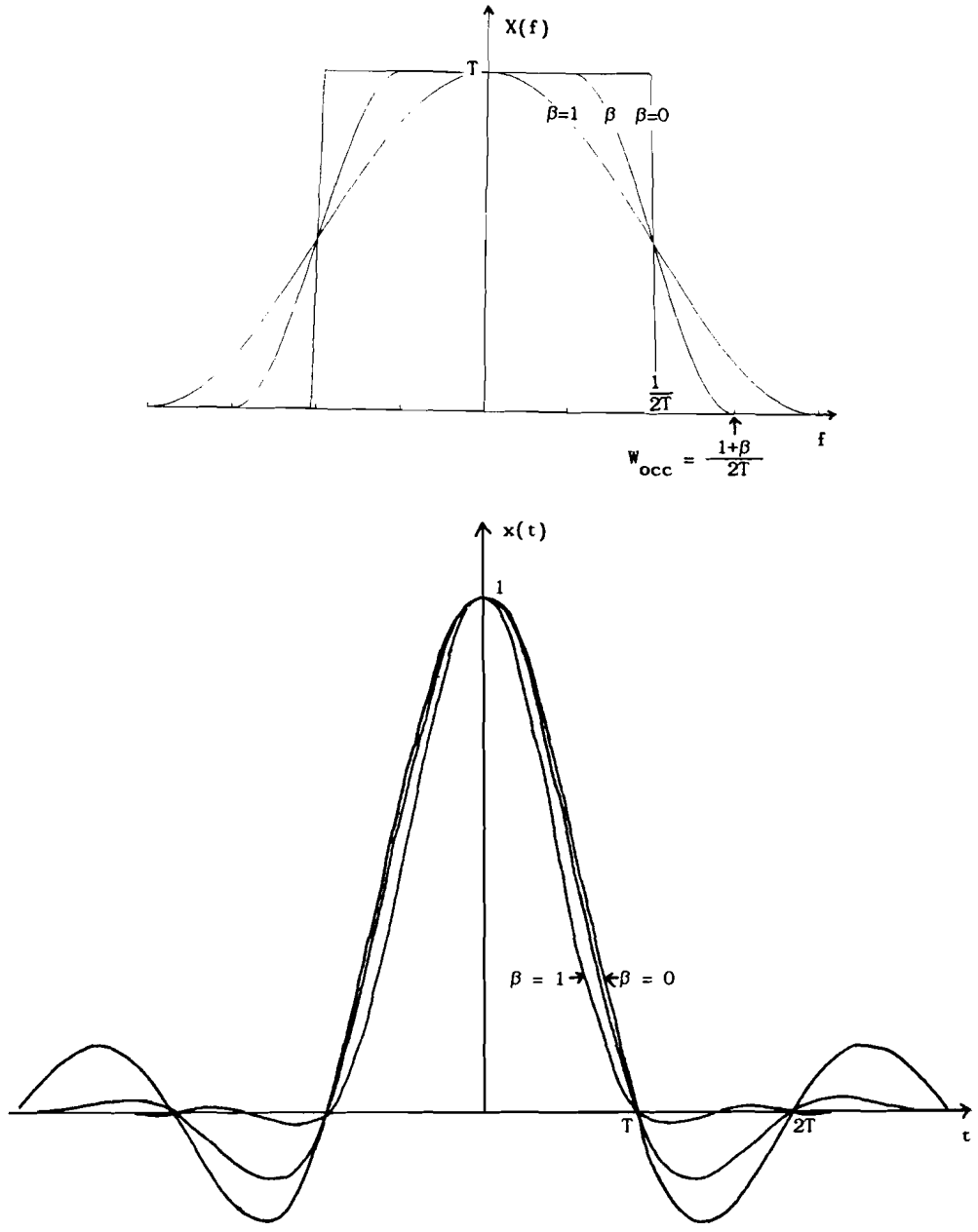


Fig. 2.6 Raised cosine spectrum pulses

Although there are more pulses that satisfy the Nyquist criterion we will restrict ourselves to raised cosine spectrum pulses. In the sequel X(f) will mean a raised cosine spectrum.

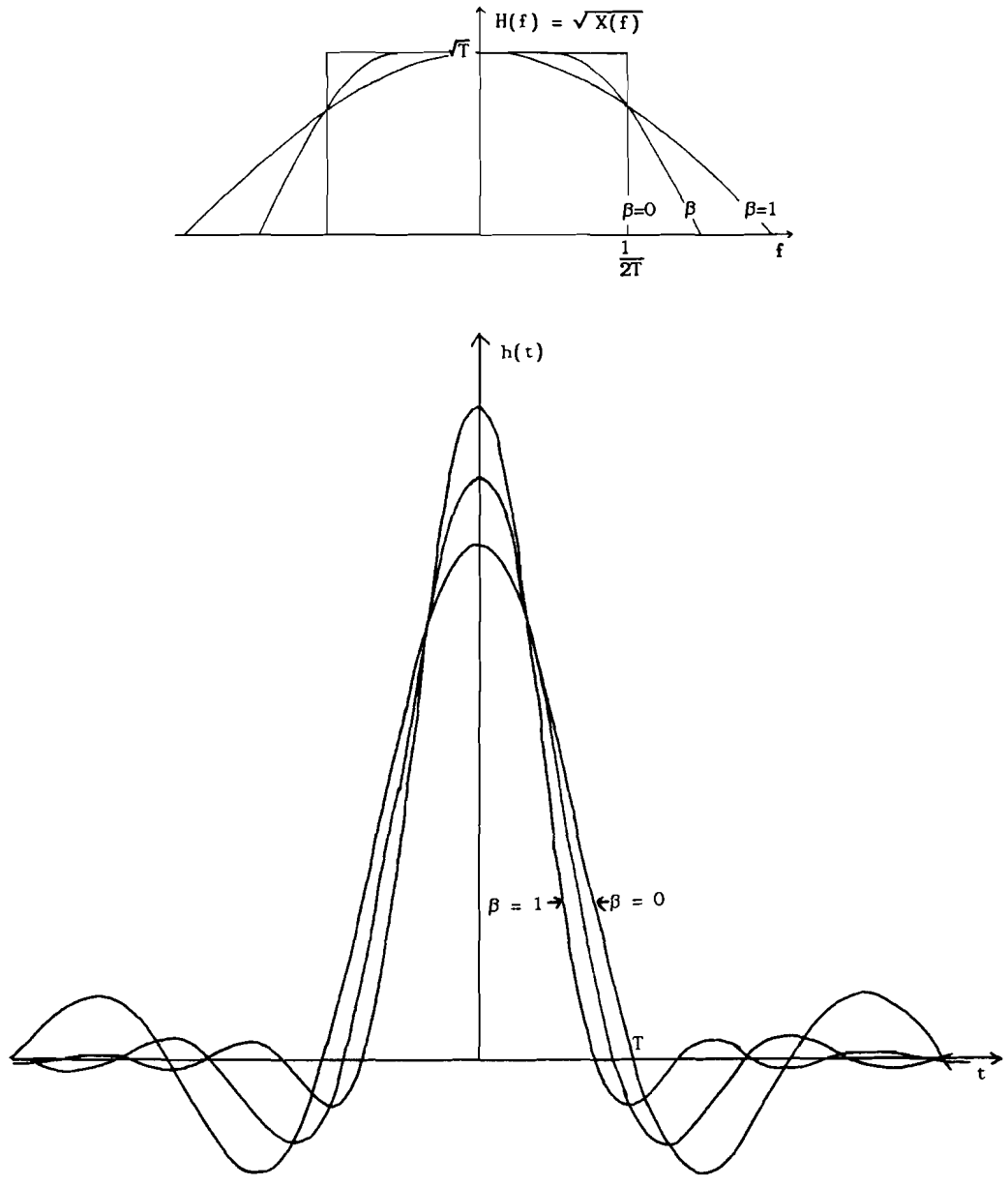


Fig. 2.7 Filter design and impulse responses

Filter design

Having chosen an X(f), the transmitter-channel en receiver filter are determined uniquely by the fact that these are matched:

$$X(f) = H(f).H^{*}(f) \tag{2.18}$$

Because X(f) is real we can write:

$$H(f) = \sqrt{X(f)} = H^{*}(f). \tag{2.19}$$

Note that we have not considered any restrictions on H(f) as they can be imposed by C(f). This is subject of investigation in the next chapter.

Note from fig 2.7 that (unlike x(t)) the impulseresponses h(t) do not have zero crossings at times kT.

2.3 Remarks

Some remarks can be made on the function of the matched filter and about the noise.

Function of the matched filter

Having entered the frequency domain, another view on the function of the matched filter will be presented. The additive Gaussian noise at the input of the receiver is modeled as zero mean additive white Gaussian noise. The receiver filter reduces the average energy in a noise sample from infinite to a finite value. It bandlimits the noise. Before filtering the noise n(t) has power density spectrum

$$S_n(f) = N_0/2 \tag{2.21}$$

and the average energy in a noise sample is

$$E\{n^2(t)\} = \int_{-\infty}^{\infty} S_n(f)df = \infty \qquad (2.22)$$

After bandlimiting with the filter $H^*(f)$ the noise z(t) is coloured and has power density spectrum

$$\begin{aligned} S_z(f) &= S_n(f).|H^*(f)|^2 \\ &= (N_o/2).H(f).H^*(f) \\ &= (N_o/2).X(f) \quad [W/Hz]. \quad \text{(fig. 2.7.c)} \end{aligned} \qquad (2.23)$$

The average energy in a noise sample is

$$E\{z^2(t)\} = \int_{-\infty}^{\infty} S_z(f)df = (N_o/2).\int_{-\infty}^{\infty} X(f)df = N_o/2 \quad [V^2] \qquad (2.24)$$

Here we used

$$\int_{-\infty}^{\infty} |H^*(f)|^2 df = \int_{-\infty}^{\infty} X(f)df = \int_{-1/2T}^{1/2T} X_{eq}(f)df = \int_{-1/2T}^{1/2T} T.df = 1. \qquad (2.25)$$

Note that the fact that the mean energy in a noise sample z(t) equals the spectral power density $S_n(f)$ might cause confusion. Note also that the power density function $S_z(f)$, which is a characteristic of a stochastic process, is expressed as a Fourier transform $(N_o/2).X(f)$ of a (deterministic) impulse response.

Independency of the noise samples

From the autocorrelation function of the noise z(t) we can learn that, although the noise z(t) is coloured, the noise samples z(kT) are statistical independent. This result is valid for all ISI-free systems (with white Gaussian input noise).

The autocorrelation function $R_{zz}(\tau)$ of the stochastic process z(t) is given by the Fourier transform of the power density spectrum $S_z(f)$.

(Actually $S_z(f)$ is defined as the Fourier transform of $R_{zz}(\tau)$ [PAP77,p.303].)

$$S_z(f) = \int_{-\infty}^{\infty} R_{zz}(\tau).e^{-j2\pi f\tau}d\tau \qquad (2.26a)$$

$$R_{zz}(\tau) = \int_{-\infty}^{\infty} S_z(f).e^{j2\pi f\tau}df. \qquad (2.26b)$$

Here we find:

$$S_z(f) = (N_o/2).X(f) \qquad (2.27a)$$

$$R_{zz}(\tau) = (N_o/2).x(\tau) \qquad (2.27b)$$

Because of the ISI-freeness (2.9) the function $x(\tau)$ has zero crossings at $\tau=kT$ ($k\neq0$) (fig. 2.8d). This means the noise samples at times kT are mutually uncorrelated if the system is ISI-free. For samples from a Gaussian noise process this means these samples are statistical independent [WOZ65, p.155].

Note that from the impulse response h(-t) of the receiver filter, which has no zero crossings at times t=kT, one might abusively think that the noise samples are correlated or dependent (fig.2.8e).

The uncorrelatedness of the noise samples is also reflected by the power density spectrum and autocorrelation function of the (discrete time) sampled noise process. The power density spectrum of this process z_k is given by:

$$\begin{aligned}\overline{S}_z(f) &= \frac{1}{T}\sum_{n=-\infty}^{\infty} S_z(f+\tfrac{n}{T}) \\ &= \frac{1}{T}\sum_{n=-\infty}^{\infty} (N_o/2).X(f+\tfrac{n}{T}) \\ &= (N_o/2).\overline{X}(f) \\ &= N_o/2. \qquad \text{(fig.2.8f).}\end{aligned} \qquad (2.28)$$

For the autocorrelation function we find:

$$R_{zz}[k] = R_{zz}(\tau=kT) = (N_o/2).x(kT)$$
$$= \begin{cases} N_o/2 & k = 0 \\ 0 & \text{elsewhere} \end{cases} . \qquad (2.29)$$

The Fourier transform relations between $\bar{S}_z(f)$ and $R_{zz}[k]$ are given by [PAP77,p.319]:

$$\bar{S}_z(f) = \sum_{k=-\infty}^{\infty} R_{zz}[k].e^{-j2\pi kTf} \qquad (2.30a)$$

$$R_{zz}[k] = T. \int_{-1/2T}^{1/2T} \bar{S}_z(f).e^{j2\pi kTf} df. \qquad (2.30b)$$

Spectral characterisation of the stochastic processes

The white input noise n(t) is a (wide sense) stationary process [WOZ65, p.183]. Filtered with $H^*(f)$ is gives another stationary process z(t). These processes can be characterized by their autocorrelationfunctions $R_{nn}(\tau)$, $R_{zz}(\tau)$ or power density spectra $S_n(f)$, $S_z(f)$.

The digitally modulated signals

$$s(t) = \sum_{n=0}^{\infty} I_n h(t-nT)$$

and

$$\sum_{n=0}^{\infty} I_n x(t-nT)$$

are cyclostationary processes [PRO83,p.119]. This means they have periodic (T) mean and autocorrelation function. The analysis of these time-dependent stochastic processes is difficult. As the impulse responses of the filters give enough insight in the problem of ISI, the spectral analysis of the digitally modulated signals is left out of the analysis here.

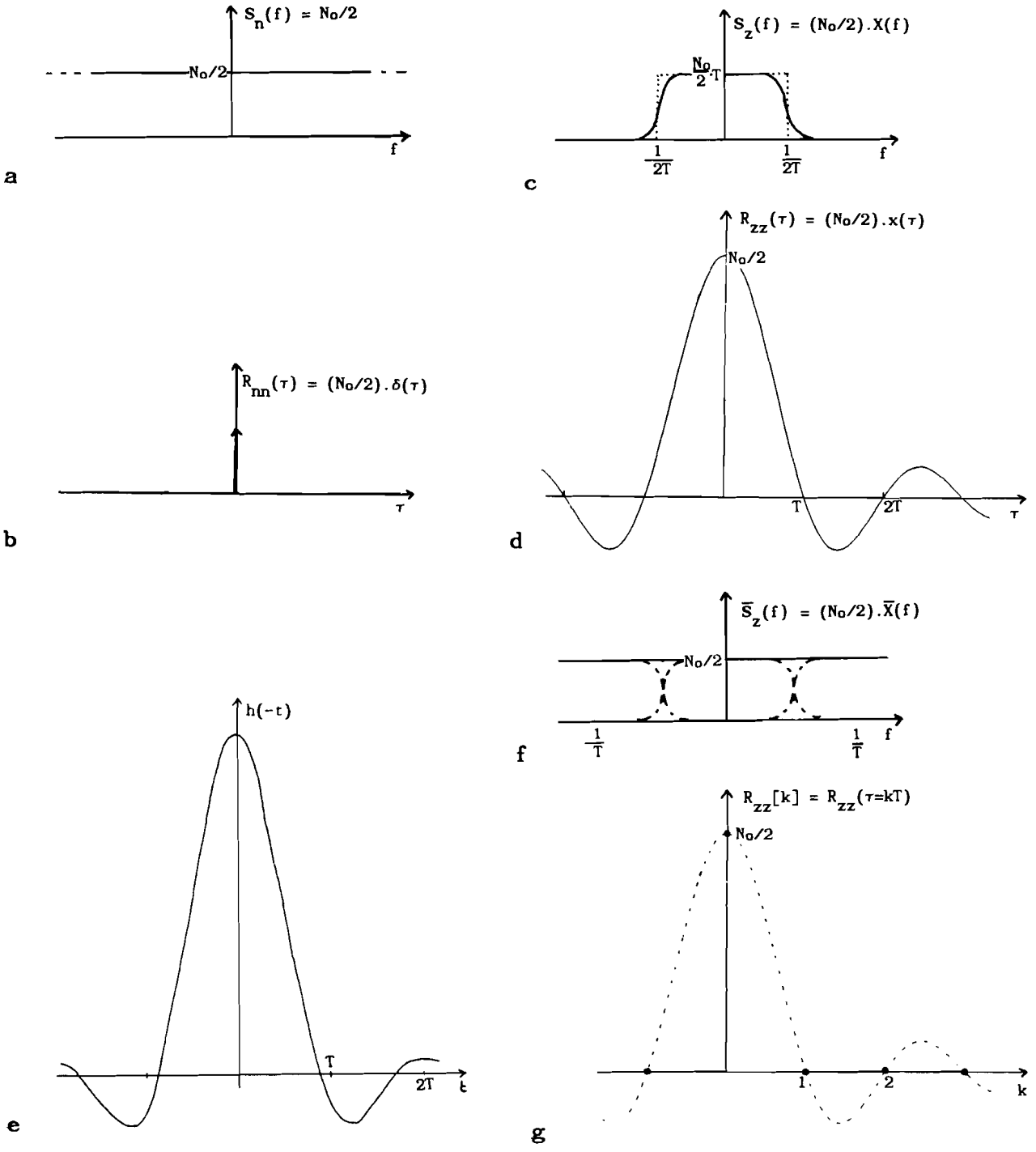


Fig. 2.8 Power density spectra and autocorrelations of the noise

a,b: P.d.s and a.c. of the input noise process

c,d: P.d.s and a.c. of the coloured output noise process

e: Impulse response of receiving filter

f,g: P.d.s. and a.c. of the sampled noise process

2.4 Conclusion

We found the Nyquist criterion (2.15) as the condition for bandlimited ISI-free communication. Raised cosine spectrum pulses (2.17c) satisfy this criterion.

ISI-free-ness of the communication implies uncorrelatedness of the noise samples after matched filtering (2.27). In case of Gaussian noise this means statistical independency. The bit-by-bit detection method is optimal because of the ISI-free-ness and the implied statistical independency of the noise samples.

3. ISI-FREE COMMUNICATION AT VARIOUS RATES

Now we know how to design a bandlimited system for ISI-free communication at a specific rate, let us proceed to the case where the ideal channel filter restricts the signal spectrum to [-W,W]:

$$C(f) = \begin{cases} 1 & |f| \leq W \\ 0 & \text{otherwise} \end{cases} . \qquad (3.1)$$

This situation can occur for instance in a large bandwidth system, where the band is divided in separate channels. We will describe how the system has to be changed to increase the channel rate without violence of the bandwidth constraint.

3.1 System design

Using raised cosine spectrum pulses with roll-off factor β for ISI-free communication at rate R=1/T the occupied bandwidth (fig. 2.6)

$$W_{occ} = \frac{1+\beta}{2T} = \frac{1+\beta}{2}.R \qquad 0 \leq \beta \leq 1. \qquad (3.2)$$

This follows from the definition of the raised cosine spectrum pulse (2.17a). The channel restricts the bandwidth to W:

$$W_{occ} = \frac{1+\beta}{2}.R \leq W. \qquad (3.3)$$

So using a system with roll-off factor β we can send ISI-free at rates

$$0 \leq R \leq W.\frac{2}{1+\beta} \qquad (3.4)$$

where equality at the right hand side implies that the maximum bandwidth W is used.

In the sequel we will asume total bandwidth occupation. This gives the relation between R,W and β:

$$R = \frac{2W}{1+\beta} \qquad 0 \leq \beta \leq 1. \qquad (3.5)$$

So we can send ISI-free at rates R=1/T, with

$$W \leq R \leq 2W \tag{3.6a}$$

using a system with transfer function

$$X(f) = \begin{cases} T & 0 \leq |f| \leq \frac{1-\beta}{2T} \\ \frac{T}{2}\cdot\left[1-\sin\left(\pi T.\frac{f-1/2T}{\beta}\right)\right] & \frac{1-\beta}{2T} \leq |f| \leq \frac{1+\beta}{2T} \end{cases} \tag{3.6b}$$

and roll-off factor (fig. 3.1)

$$\beta = \frac{2W}{R} - 1. \tag{3.6c}$$

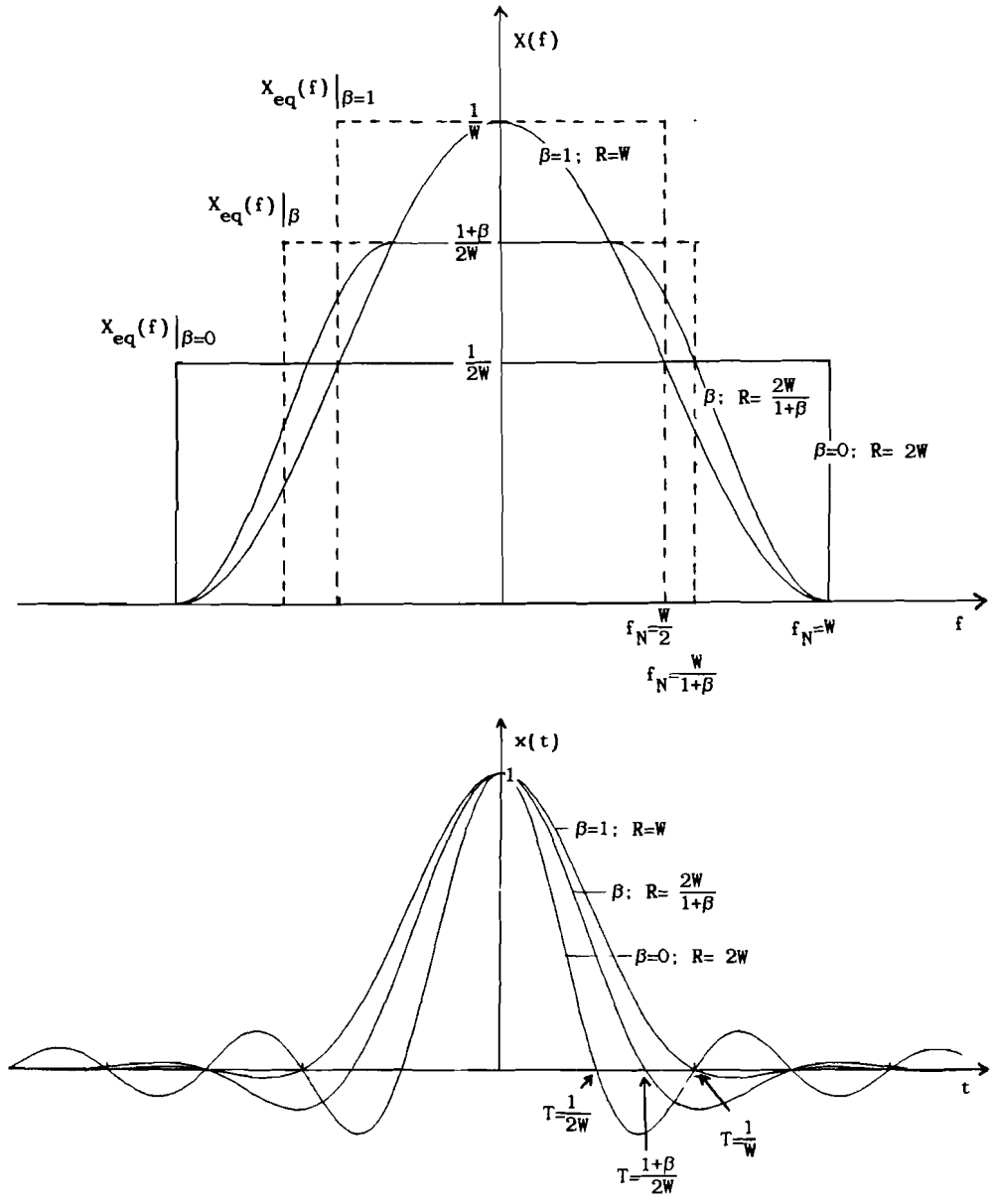


Fig. 3.1 Transmission at rates $W \leq R \leq 2W$, *keeping* E_c *constant*

The transmitter and receiver filter are determined uniquely by the fact that the receiver filter is matched to the transmitter-channel filter:

$$H(f) = H^{*}(f) = \sqrt{X(f)} \tag{3.7}$$

$\underline{E_c \text{ constant for all rates}}$

To be able to compare the performance of the system at various rates the energy sent over the channel per channelbit must be equal for all rates. From (2.18) we know that this is realized by the normalization:

$$\int_{-\infty}^{\infty} h^2(t)dt = \int_{-\infty}^{\infty} X(f)df = 1. \tag{3.8}$$

What does an increase of the rate mean for the system transfer function? Because the rate increases, the symmetrypoint f=1/2T of the flank moves to a higher frequency. To keep the surface under X(f) constant the total height of the transfer function is diminished. This is accomplished by the factor T in the transfer function.

3.2 Performance at various rates.

For bit-by-bit detection the performance is measured by the signal to noise ratio at the sample moments. Because we have kept the E_c constant for all rates we can compare the performances. The received symbols are

$$\begin{aligned} y_k &= \sum_{n=0}^{\infty} I_n x_{k-n} + z_k \\ &= I_k x_0 + z_k \\ &= I_k + z_k. \end{aligned} \tag{3.9}$$

So the output signal to noise ratio is given by

$$SNR = \frac{E\{I_k^2\}}{E\{z_k^2\}} = \frac{E_c}{E\{z_k^2\}} \tag{3.10}$$

Average energy in a noise sample

The average energy in a noise sample z_k

$$E\{z_k^2\} = E\{z^2(t)\} = \int_{-\infty}^{\infty} S_z(f)df = (N_o/2).\int_{-\infty}^{\infty} X(f)df = N_o/2 \tag{3.11}$$

is constant for all rates, because of the normalization (2.4).

With (2.24b) one can also express $E\{z_k^2\}$ in the autocorrelation function $R_{zz}(\tau)$:

$$E\{z^2(t)\} = \int_{-\infty}^{\infty} S_z(f).e^{j2\pi f0}df = R_{zz}(0) = (N_o/2).x(0) = N_o/2 \tag{3.12a}$$

or, in the discrete time:

$$E\{z_k^2\} = T\int_{-1/2T}^{1/2T} \bar{S}_z(f).e^{j2\pi 0Tf}df = R_{zz}[0] = N_o/2. \tag{3.12b}$$

So by normalizing

$$\int_{-\infty}^{\infty} h^2(t)dt = 1$$

we keep E_c and $E\{z^2(t)\}$ constant for all rates.

With the notion of noise equivalent bandwidth [CAR68,p.125] it can also be explained what happens when increasing the rate from $R_1=1/T_1$ to $R_2=1/T_2$. The noise equivalent bandwidth B_N of a filter H(f) is defined as

$$B_N = \frac{1}{|H(f)|^2_{max}} . \int_0^{\infty} |H(f)|^2 df. \tag{3.13}$$

Here this reduces to

$$B_N = \frac{1}{X(0)}.\int_0^{\infty} X(f)df = \frac{1}{2.X(0)} = \frac{1}{2T}. \tag{3.14}$$

The noise equivalent bandwidth can be interpreted as the bandwidth of the ideal rectangular filter with the same maximum gain $X(f)_{max}$ that passes as much noise power as the filter in question.

When increasing the rate from R_1 to R_2 the equivalent noise bandwidth increases from $1/2T_1$ to $1/2T_2$. This would mean that the average energy in a noise sample becomes bigger. But because the DC-gain $\sqrt{X(0)} = \sqrt{T}$ of the filter $H^*(f)$ decreases at the same time, the average energy in a noise sample remains the same (fig. 3.2).

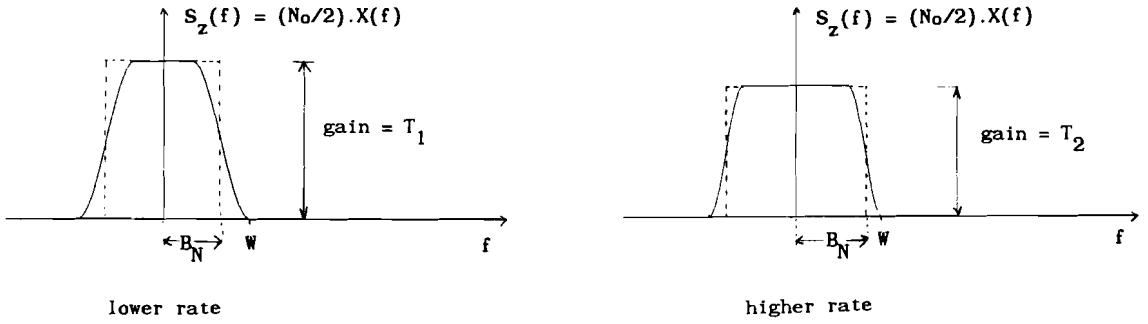


Fig. 3.2 The energy in a noise sample is equal for all rates

Performance

We now find that the output signal to noise ratio is equal for all rates:

$$SNR = \frac{E\{I_k^2\}}{E\{z_k^2\}} = \frac{E_c}{N_o/2} \qquad (3.15)$$

The error probability is given by [LUC68,p.55]

$$P[E] = Q(\sqrt{2E_c/N_o}) \qquad (3.16a)$$

where Q(.) is the tail of the Gaussian distribution

$$Q(\alpha) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^{\infty} e^{-\tau^2/2} d\tau. \qquad (3.16b)$$

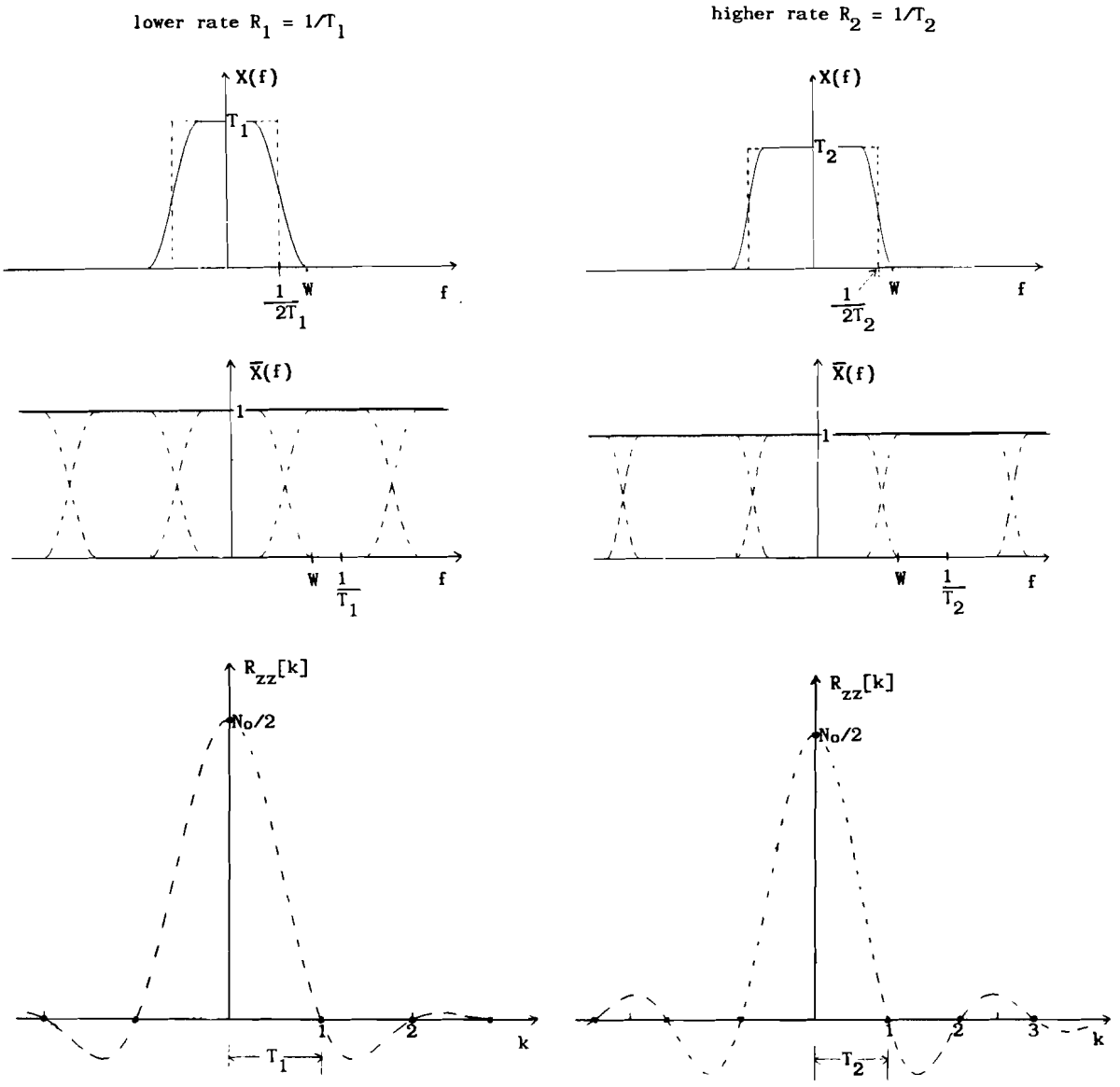


Fig. 3.3 Overview of communication at two rates R_1 *and* R_2

3.3 Conclusion

Using raised cosine spectrum pulses with frequency transfer function (3.6b) and roll-off factor $\beta = \frac{2W}{R} - 1$ we can send ISI-free with rate

$R=1/T$ ($W \leq R \leq 2W$) over an ideal channel with bandwidth W. The performance is equal at all rates. This means that an increase of the channel rate is possible without loss of performance. The transmitter and receiver filter have to be adapted though.

4 CHANGE OF RATE WITHOUT SYSTEMMODIFICATION: ISI

Assume we have designed an ISI-free raised cosine spectrum system for nominal rate $R_0=1/T_0$. It has roll-off factor $\beta_0 =(2W/R_0)-1$. For some reason the channel rate must be increased to $R>R_0$, without changing the transmitter and receiver filters. For instance because changing the filters is too expensive, or, as it is in a satellite, it is impossible too change them. Intersymbol interference will occur as a consequence of the increase of the rate. The bit-by-bit detection method that was optimal in the ISI-free case is now not optimal. The penalty on the errorperformance due to the ISI will be described and calculated for an example.

4.1 Description of the ISI

Let us take a closer look at what happens when the channel rate is increased. The transmitter filter has an impulseresponse h(t) that is orthogonal to all its time translates $h(t-kT_0)$. Because of this orthogonality it was possible to project the received signal r(t) upon the space spanned by $h(t-kT_0)$ to get sufficient statistics to detect the k'th symbol $I_k.h(t-kT_0)$ optimally. The projections of all the other symbols $I_l.h(t-lT_0)$ ($l\neq k$) upon this space was zero. But at a higher rate $R=1/T$ the waveforms h(t-kT) loose their orthogonality (fig. 4.1). Thus the projections of $I_l.h(t-lT)$ upon the space spanned by h(t-kT) are unequal to zero.

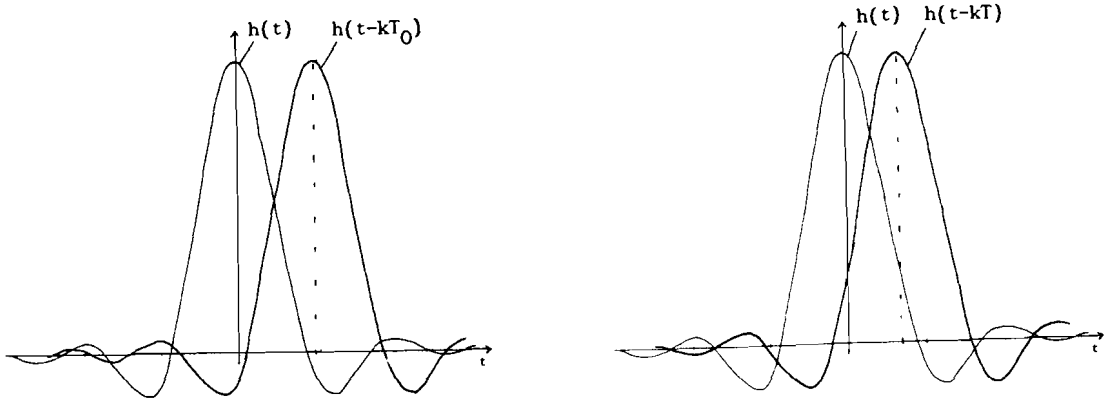


Fig. 4.1. Loss of orthogonality at higher rate R=1/T.

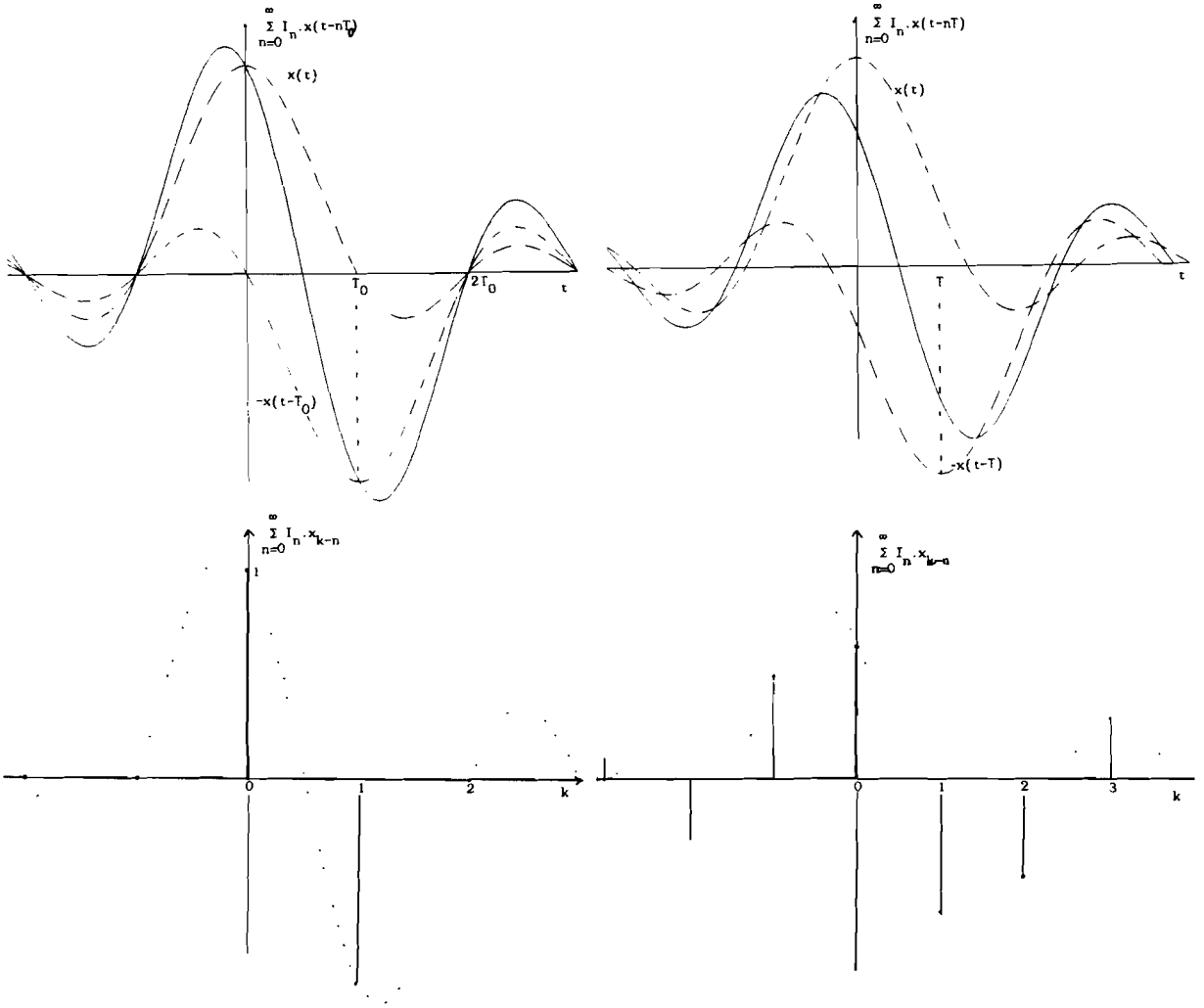


Fig. 4.1. Intersymbol interference

One usually considers the total system response when describing the ISI:

$$y_k = \sum_{n=0}^{\infty} I_n x_{k-n} + z_k \qquad (4.1)$$

$$= I_k + \sum_{\substack{n=0 \\ n\neq k}}^{\infty} I_n x_{k-n} + z_k .$$

Here the second term on the right hand side is called the intersymbol interference. It is a consequence of the fact that $x_k \neq 0$ for $k \neq 0$. So the received symbol I_k is corrupted by an ISI-term and a noise sample (figs 4.2 and 4.3).

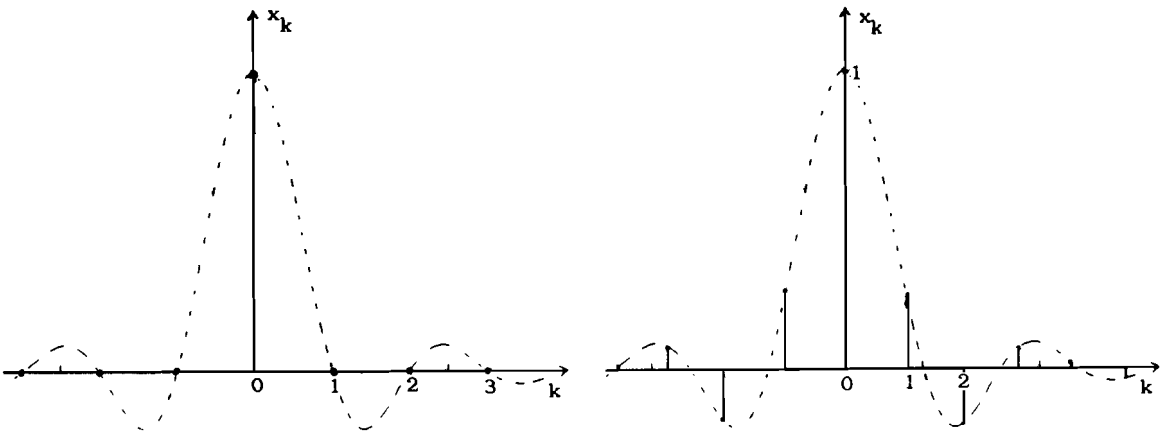


Fig. 4.3 Discrete time impulse responses at R_0 and R

ISI-term

The ISI is best decribed in the time domain. The coefficients of the discrete time impulse response x_k are computed from

$$x_k = x(kT) = \frac{\sin(\pi kT/T_0)}{\pi kT/T_0} \cdot \frac{\cos(\beta\pi kT/T_0)}{1-(2\beta kT/T_0)^2} . \qquad (4.2)$$

The discrete time impulse response x_k is (in principle) unequal zero for an infinite number of terms, because x(t) has infinite duration. This means that an infinite number of symbols I_l interfere with I_k ($l \neq k$). For roll-off factors $\beta \neq 0$ the value of x_k converges to zero for $k \to \pm\infty$. For bigger β the value of x_k diminishes faster than for smaller β. In practice one can approximate the ISI well by truncating the infinite impulse response $\underline{x}$ to $\underline{x}'$ with finite length 2L+1: $\underline{x}' = (x_{-L}, \ldots\ldots, x_0, \ldots\ldots x_L)$. Criterium for truncation can be: take L the smallest value for which 99.9 percent of the energy of the impulse response is in the truncated response:

$$\sum_{n=-L}^{L} x_n^2 \geq 0.999 \sum_{n=-\infty}^{\infty} x_n^2. \qquad (4.3)$$

Here 0.999 is an arbitrary factor, assuring that the energy in the neglected ISI-terms is relatively small. Note that because x(t) is even $x_k = x_{-k}$.

Noise term

The noise samples z_k are now correlated, and therefore statistical dependent. This can be seen from the autocorrelation function of the sampled noise:

$$R_{zz}[k] = R_{zz}(kT) = (N_o/2).x(kT) = \begin{cases} N_o/2 & k=0 \\ \neq 0 & k \neq 0 \end{cases}.$$

The mean energy in an noise sample still equals

$$E\{z_k^2\} = R_{zz}[0] = N_o/2.$$

The correlatedness of the z_k is also reflected by the power density spectrum of the sampled noise

$$\bar{S}_z(f) = (N_o/2).\bar{X}(f) = (N_o/2).\frac{1}{T}.\sum_{n=-\infty}^{\infty} X(f+\tfrac{n}{T}) \neq \text{constant} .$$

The surface included by $f=\pm\frac{1}{2T}$, the f-axis and $T.\bar{S}_z(f)$ is equal for all

rates (fig. 4.4):

$$R_z[0] = T. \int_{1/2T}^{1/2T} \bar{S}_z(f).df = N_o/2.$$

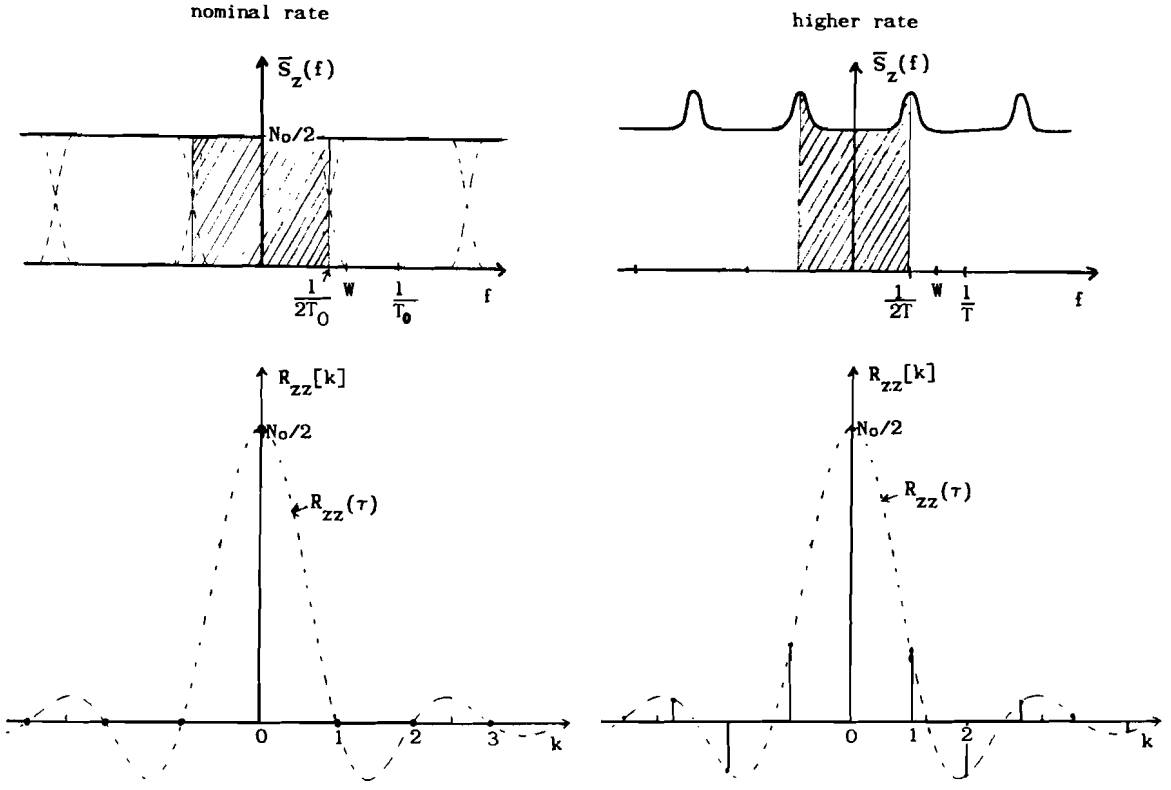


Fig. 4.4. Noise samples are correlated for $R \neq R_0$

4.2 Performance of the bit-by-bit detection method

The bit-by-bit detector consists of a threshold detector with threshold at zero (fig.2.2). It decides $\hat{I}_k = \sqrt{E_c}$ if $y_k > 0$, and $\hat{I}_k = -\sqrt{E_c}$ otherwise. The probability of a decision error is given by the probability that the ISI-term and the noise together drive the

received signal to the other side of the threshold:

$$\begin{aligned} P_e &= P(I_k=\sqrt{E_c}).P(\hat{I}_k=-\sqrt{E_c}\,|\,I_k=\sqrt{E_c}) + P(I_k=-\sqrt{E_c}).P(\hat{I}_k=\sqrt{E_c}\,|\,I_k=-\sqrt{E_c}) \\ &= P(\hat{I}_k=-\sqrt{E_c}\,|\,I_k=\sqrt{E_c}) = P(y_k \leq 0 \mid I_k=\sqrt{E_c}) \\ &= P(\sqrt{E_c} + \sum_{n\neq k} I_n x_{k-n} + z_k \leq 0) \\ &= P(z_k \leq -\sqrt{E_c} - \sum_{n\neq k} I_n x_{k-n}) \\ &= P(z_k \geq \sum_{n=0}^{\infty} I_n x_{k-n} \mid I_k=\sqrt{E_c}). \end{aligned}$$

To be able to compute this expression we must take the truncated impulse response (4.3) $\underline{x}'=(x_{-L},\ldots,x_0,\ldots,x_L)$. If we take for simplicity k=L+1, then we have 2L interfering symbols I_l (l=0,..,L,L+2,..2L+1). Let us define $\underline{I}'$ as the set of sequences $\underline{I}'=(I_0,\ldots,I_{2L+1})$ with $I_{k=L+1}=\sqrt{E_c}$. Here, in the uncoded case, all $\underline{I}'\in\underline{I}'$ are equiprobable with probability $P(\underline{I}') = 2^{-2L}$. With the notation

$$a(\underline{I}') = \sum_{n=0}^{L} I'_n x'_{k-n} = \sum_{n=0}^{L} I'_n x_{k-n}$$

we find for the bit (symbol) error probability

$$\begin{aligned} P_e &= \sum_{\underline{I}'\in\underline{I}'} P(\underline{I}').P(\, z_k \geq a(\underline{I}')\,) \\ &= 2^{-2L}. \sum_{\underline{I}'\in\underline{I}'} Q\left[\sqrt{(2.a^2(\underline{I}')/N_0)}\,\right], \end{aligned}$$

where Q(.) is the tail of the Gaussian distribution (3.17b). Here we used $E\{z_k^2\}=N_0/2$. The computation of P_e thus involves the computation of the levels $a(\underline{I}')$ for all 2^{2L} sequences $\underline{I}'\in\underline{I}'$.

4.3 Remarks and an example

Dependency of the errors

Note that the statistical dependency of the noise samples z_k plays no role in the computation of the errorprobability. This is because the bit-by-bit detector takes independent decisions about the symbols. This does not mean that the errors are statistical independent! In the case of coding this makes the analysis of the deterioration due to the ISI complicated.

Performance measure

Notice also that with ISI there is no meaningful definition of signal to noise ratio at the sample moments as a precise measure of performance. The computation of the mean squared value of the received term minus noise $E\{(y_k-z_k)^2\}$ would give a value larger than E_c in the case of ISI. With equal $E\{z_k^2\}=N_0/2$ this would imply that the signal to noise ratio in the case of ISI is bigger than in the ISI-free case.

As the computation of the P_e is quite tedious, measures of ISI are defined to approximate or bound the role of the ISI. An overvieuw of some of the most used measures is given in [LUC68,p.79].

The oversampling mistake

A final remark can be made about the interpretation of the discrete time impulse response. One might think that apart from $I_k x_0$ the

responses $I_k x_{l-k}$ of symbol I_k at other times $l \neq k$ give extra information for the decision about I_k (fig. 4.3). If this were so, one could detect with arbitrary reliability, by just increasing the sampling rate and thus gathering more information. This is not valid because the projection of a signal s(t) on the space spanned by h(t-lT) only gives new information if this space is orthogonal to the space spanned by h(t-kT) ($l \neq k$). This is evidently not the case here.

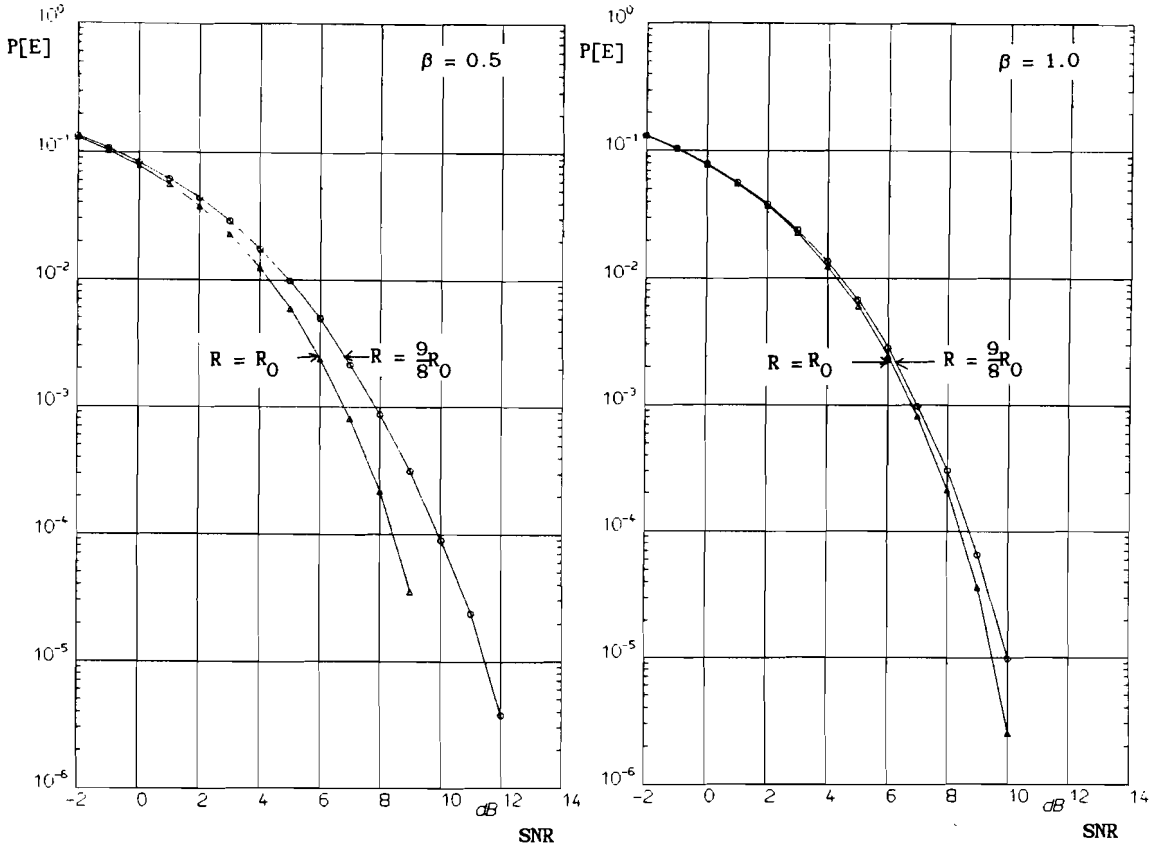


Fig. 4.5. Performance for β=0.5 and β=1.0, R=R_0 and R=$\frac{9}{8}R_0$

Example

We computed the errorperformance for systems with β=0.5 and 1.0, at a rate $R=\frac{9}{8}.R_0$, where R_0 is the ISI-free rate. For these β and R the impulse response $\underline{x}$ can be truncated to 2L+1=5 terms $\underline{x}'$, to have 99.9% of the energy of the discrete time impulseresponse. For comparison the errorprobability in the ISI-free case (3.17) is also depicted .

Note that for smaller roll-off factors the deterioration due to the ISI is larger than for larger roll-off factors. This is a consequence of the slower decay of the response x(t).

4.4 Conclusion

When increasing the rate in an ISI-free system without adaptation of the transmitter and receiver filters ISI occurs. Bit-by-bit detection is not optimal anymore Computation of the errorprobability is possible, but tedious. The noise samples and the decision errors are statistical dependent. In the uncoded case with bit-by-bit decisions this plays no role, but when coding is applied this complicates the analysis of the error performance. For smaller roll-off factors the deterioration due the ISI is more serious than for larger roll-off factors.

5 CONCLUSIONS

Using raised cosine spectrum pulses with frequency transfer function (3.6b) and roll-off factor $\beta = \frac{2W}{R} - 1$ we can send ISI-free with rate R=1/T ($W \leq R \leq 2W$) over an ideal channel with bandwidth W. The performance is equal at all rates. This means that an increase of the rate up to 2W is possible without loss of performance. The transmitter and receiver filter have to be adapted though.

The ISI-free-ness of the communication implies that the noise samples are statistical independent. This makes it possible to compute the error performance of the bit-by-bit detection method in the presence of intersymbol interference.

When increasing the rate in an ISI-free system without adaptation of the transmitter and receiver filter ISI occurs. Bit-by-bit detection is now not optimal anymore. Computation of the error probability is possible, but tedious. The noise samples and decision errors are statistical dependent. For smaller roll-off factors the deterioration due to the ISI is more serious than for larger roll-off factors.

REFERENCES

[CAR68] A.Bruce Carlson

Communication Systems

McGraw-Hill - Kogakusha Ltd, Tokyo, 1968

[LUC68] R.W. Lucky, J. Salz and E.J. Weldon, Jr.

Principles of Data Communication

John Wiley, New York, 1968

[PAP77] A. Papoulis

Signal Analysis

McGraw-Hill, New York, 1977

[PRO83] J.G. Proakis

Digital Communications

McGraw-Hill, New York, 1983

[WOZ65] J.M. Wozencraft and I.M. Jacobs

Principles of Communication Engineering

John Wiley, New York 1965

PART II

ON MODELS AND CODING FOR THE MAGNETIC RECORDING CHANNEL

SUMMARY

Coded high density digital magnetic recording is considered. A transition is made from a for the user of the system relevant problem statement to manageable design problems. The model that is necessary for this translation can be formulated in two ways. The two approaches lead to different receivers. The resulting systems though are equivalent. The two approaches give rise to different views on the magnetic recording system. The report serves as a basis for the design of codes for the magnetic recording channel.

In magnetic recording systems intersymbol interference (ISI) occurs. Then maximum likelihood sequence estimation (MLSE) is the optimal detection method. The system is analyzed for MLSE detection. The minimum squared Euclidian distance δ^2_{min} is key parameter for the error performance. A useful division of δ^2_{min} in parts is given. With this division the effects of coding and changes in information density can be made clear. It gives rise to a systematic approach to the problem of coding for the magnetic recording channel. Coding can be used to improve error performance or increase information density.

The general methods are applied to the Lorentzian stepresponse model, which is a good approximation of reality. An example gives insight in the design of codes.

For low information densities the Lorentzian stepresponse can be approximated by the finite duration triangular stepresponse model.

1 INTRODUCTION

This part of the report is about high density digital magnetic recording. When reading back digital data on tape, intersymbol interference occurs. With recording codes (e.g. [IMM87], [KOB70], [KOB71] and [WOL86]) the error performance can be improved. A good knowledge of the magnetic recording model is necessary to be able to design a code for a specific system. The characteristics of the system change with the storage density. At high densities the ISI is more severe.

In chapter 2 different models for the magnetic recording channel and system are presented. The relation between physical parameters, channel characteristics and error probability is discussed. In chapter 3 these relations are discussed for a special model: the Lorentzian step-response model. An example gives more insight into the design of codes for this channel. The usefulness of the simplified triangular stepresponse model will be discussed.

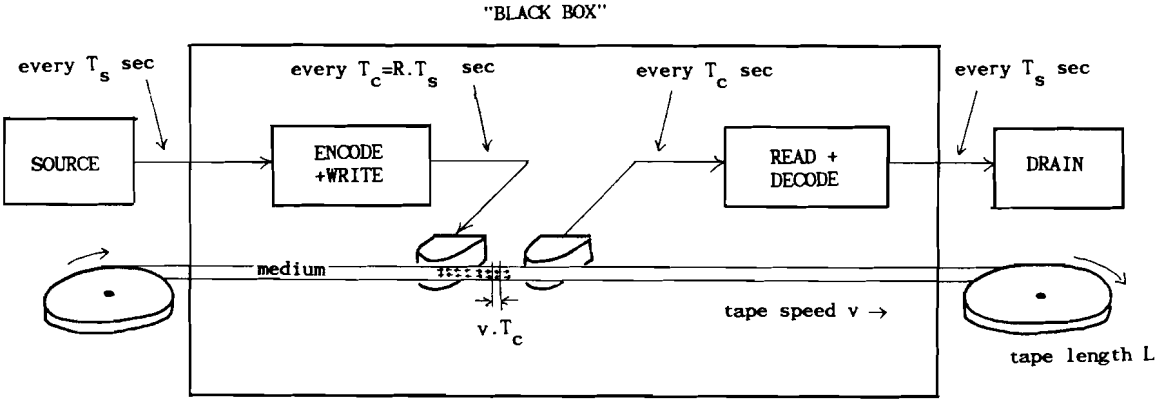


Fig. 1.1 The magnetic recording system

Important in the analysis is the difference in view on the magnetic recording system between user and code/system designer.

The users "black-box" view of the system can be stated as follows:

A binary source generates symbols every T_s seconds. These are stored in the recording device, and read back after some time. The intersymbol time at reading back is also T_s. The bit error probability is P_e. The user storage density D_s [bits/m] determines the total number of source bits N_s that is stored on L meters of medium (tape):

$$N_s = D_s.L. \qquad (1.1)$$

The user sees a relation between D_s, T_s and P_e. The problem can be stated as follows:

> Is there a system that can store information at user storage density D_s and inter sourcesymbol time T_s, with bit error probabilty $P_e \leq P_0$?

The code/system designer sees the system as follows:

The source symbols have intersymbol time T_s. The required user storage density is $D_s = N_s/L$ [bits/m]. First the tape speed v is determined. With tape speed v the total recording time of L meters of tape is L/v. The total number of stored source symbols $N_s = L/(v.T_s)$. With $N_s = D_s.L$ we find v determined uniquely by D_s and T_s:

$$v = 1/(T_s.D_s). \qquad (1.2)$$

Now the code rate R must be chosen and the code structure designed. With rate R the time between channel symbols is

$$T_c = R.T_s. \qquad (1.3)$$

The tape speed v and the inter channel symbol time T_c determine the shape and energy of the read-back pulses, and thereby the channel characteristics. The code structure and channel characteristics determine the channel bit error probability that equals the source bit error probability P_e. As T_c is given by $R.T_s$ the code designer has the rate R and the code structure as design parameters. This gives the problem statement:

> Given D_s and T_s. Is there a rate R and code with rate R for which a system with tape speed $v = 1/(D_s.T_s)$ and inter channelsymbol time $T_c = R.T_s$ has bit error probability $P_e \leq P_0$?

In the sequel we will see that this is a complex question. To gain insight into the problem some less complicated problems can be derived. But first the mathematical model for the magnetic recording system and the optimal detection method will be presented.

Two topics that recur throughout the analysis are the comparison of the two different models for the magnetic recording channel on the one hand, and the transition from the users' problem statement to manageable design problems on the other hand.

2 GENERAL SYSTEM DESCRIPTION

In this chapter the general characteristics of a magnetic recording channel embedded in a system will be described. The differentiating nature of the read-back process gives rise to differences with (normal) non-differentiating communication systems. Two equivalent models for the recording channel are presented. They give different insights in the channel model.

2.1 Recording channel

Two equivalent channel models can be derived. The first one is based on the physical processes in the magnetic recording system. In the second model these processes are more abstrahated.

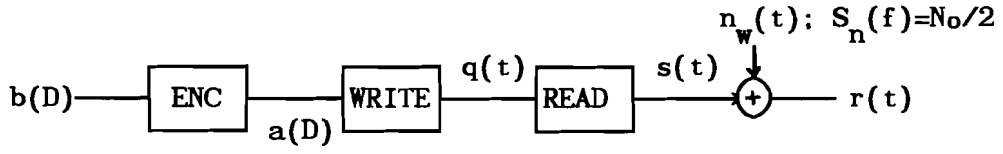


Fig. 2.1 General model for the recording channel

2.1.1 Two models for the recording channel

The stream of binary source symbols (with inter sourcesymbol time T_s)

$$b(D) = b_0 + b_1.D + b_2.D^2 + b_3.D^3 + \ldots \qquad b_i \in \{-1.1\} \qquad (2.1)$$

(with D the delay operator) is encoded with the rate $R = T_c/T_s$ code to the stream of binary channel symbols (with inter channelsymbol time T_c)

$$a(D) = a_0 + a_1.D + a_2.D^2 + a_3.D^3 + \ldots \qquad a_i \in \{-1,1\}. \qquad (2.2)$$

The set of all possible encoded sequences a(D) is denoted with $\mathbf{S}_c$. The encoded symbols a_i are written on the tape (medium) as magnetic domains with positive or negative polarization. The stored signal q(t) is modeled as a sequence of full-T_c pulses:

$$q(t) = \sum_{n=0}^{\infty} a_i.p_{Tc}(t-iT_c) \qquad (2.3)$$

with

$$p_{Tc}(t) = \begin{cases} \sqrt{E_b} & |t| \le T_c/2 \\ 0 & \text{otherwise} \end{cases} . \qquad (2.4)$$

The contstant E_b expresses the possibility to change the value of the magnetization level. Note that, unlike many other communication systems, the impulse response $p_{Tc}(t)$ of the write (or transmit) device depends on the inter channelsymbol time T_c.

When reading back the tape signal q(t) the read head detects changes in the magnetic field. This is modeled by the stepresponse of the read device $\sqrt{E_h}.h_v(t)$. Here $h_v(t)$ is even ($h_v(-t) = h_v(t)$) and real and has normalized energy :

$$\int_{-\infty}^{\infty} h_v^2(t).dt = 1. \qquad (2.5)$$

The energy E_h and form $h_v(t)$ of the stepresponse depend on the tape speed v. The actual function $\sqrt{E_h}.h_v(t)$ will not be specified until chapter 3, because the results in this chapter are valid for all even real $\sqrt{E_h}.h_v(t)$. In the illustrations of this chapter the Lorentzian stepresponse (chapter 3) will be used.

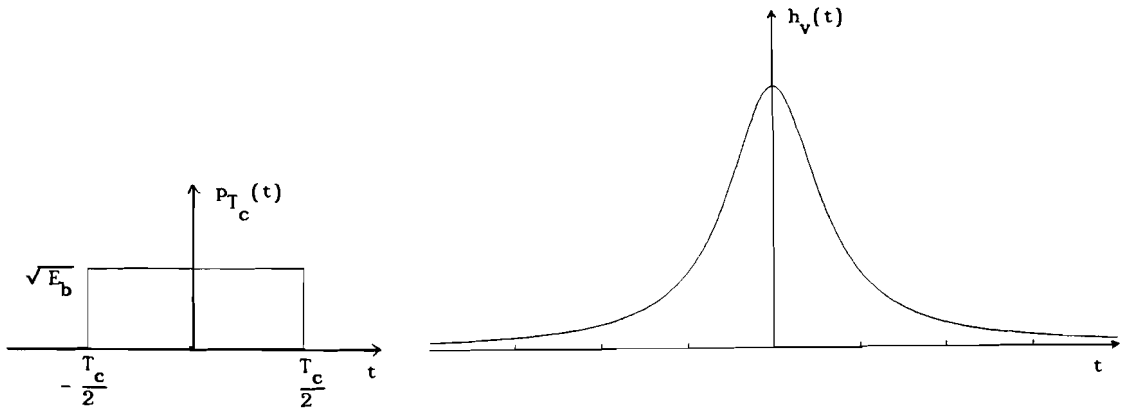


Fig. 2.2 Write pulse and typical step response $h_v(t)$

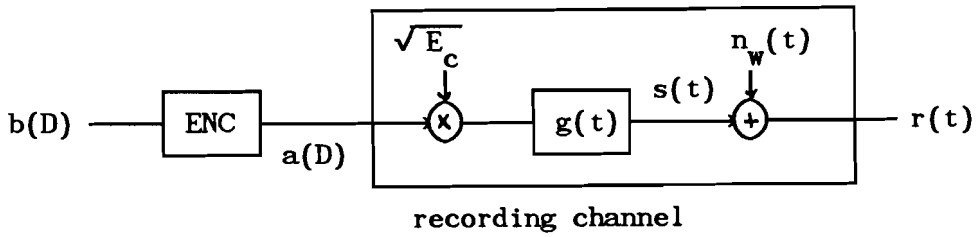


Fig. 2.3 First channel model

First channel model

Every pulse $p_{Tc}(t)$ contains two steps. The impulse response of the recording channel is the concatenation of the write- and read operations:

$$\sqrt{E_c}.g(t) = \sqrt{E_b}.\sqrt{E_h}.\{h_v(t+T_c/2)-h_v(t-T_c/2)\} \qquad (2.6)$$

where g(t) is the normalized impulseresponse with energy 1:

$$\int_{-\infty}^{\infty} g^2(t).dt = 1. \qquad (2.7)$$

For the received energy per channel symbol a_i we have

$$E_c = E_b.E_h.\int_{-\infty}^{\infty} \{h_v(t+T_c/2)-h_v(t-T_c/2)\}^2.dt. \qquad (2.8)$$

Note that both energy E_c and form $g(t)$ of the channel impulse response depend on the tape speed v and the inter channelsymbol time T_c.

The signal

$$s(t) = \sqrt{E_c}.\sum_{n=0}^{\infty} a_i.g(t-iT_c) \qquad (2.9)$$

is corrupted by additive zero mean white Gaussian noise $n_w(t)$ with power density spectrum $S_n(f)=N_o/2$. So all the noise that occurs in the write device, tape and read head is modeled as white noise, independent of any parameter. For the received signal we thus have

$$r(t) = s(t) + n_w(t)$$

$$= \sqrt{E_c}.\sum_{n=0}^{\infty} a_i.g(t-iT_c) + n_w(t) \qquad (2.10)$$

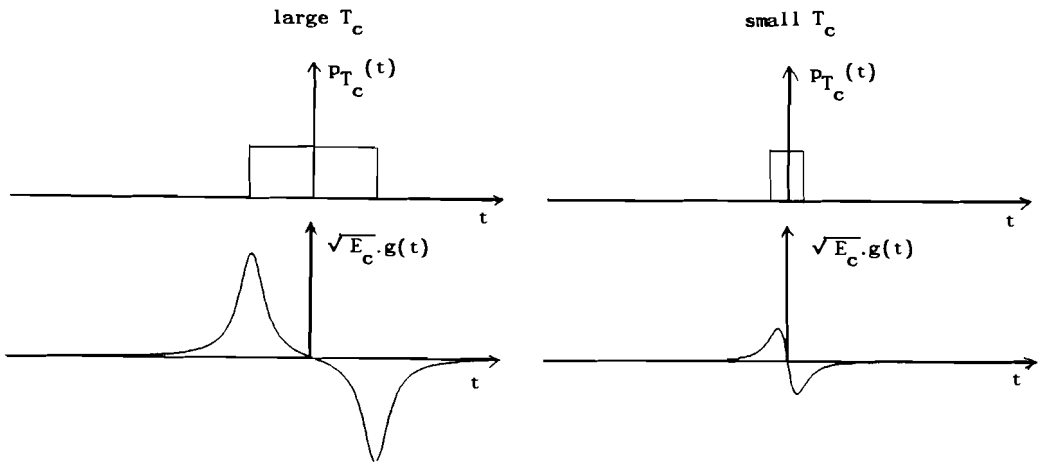


Fig. 2.4 Write pulse and channel response for two values of T_c

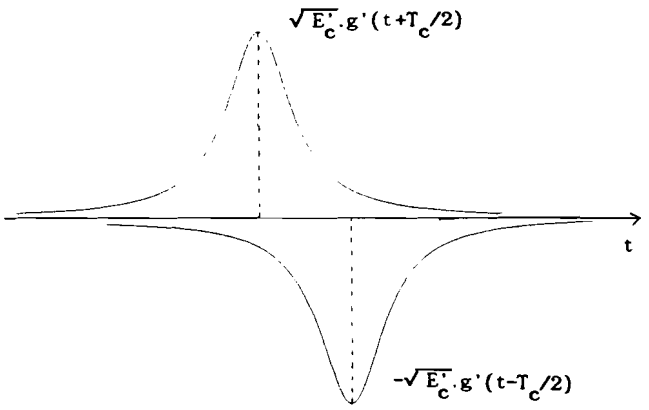


Fig. 2.5 Two responses for two level changes

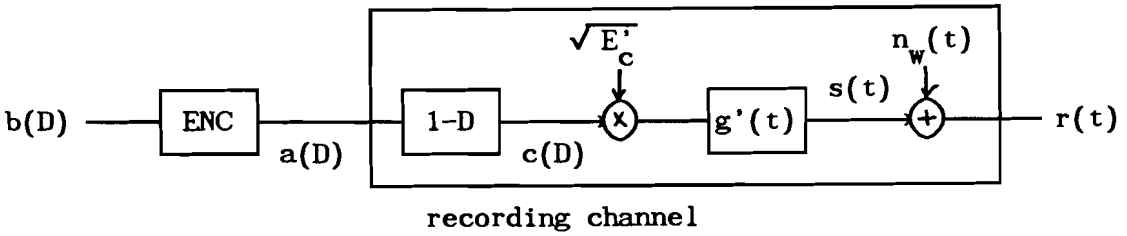


Fig. 2.6 Second channel model

Second channel model

An other channel model less reflects the physical processes, but accurately describes the differentiating nature of the magnetic recording process. One recorded pulse $a_i.p_{Tc}(t-iT_c)$ generates at reading out two pulses $a_i.\sqrt{E_b}.\sqrt{E_h}.h_v(t+T_c/2)$ and $-a_i.\sqrt{E_b}.\sqrt{E_h}.h_v(t-T_c/2)$ time T_c later. We can associate the first pulse with time i and the second with time i+1. Then at time i we have

actually written the (ternary) symbol

$$c_i \triangleq a_i - a_{i-1} \qquad c_i \in \{-2,0,2\} \tag{2.11a}$$

or $c(D) = a(D).(1-D).$ (2.11b)

The symbol stream c(D) is fed into the channel filter with impulse response equal to the stepresponse

$$\sqrt{E'_c}.g'(t) = \sqrt{E_b}.\sqrt{E_h}.h_v(t+T_c/2). \tag{2.12}$$

With

$$\int_{-\infty}^{\infty} g'^2(t).dt = 1 \tag{2.13}$$

we find for the received energy per symbol c_i

$$E'_c = E_b.E_h \tag{2.14}$$

and for the normalized channel response

$$g'(t) = h_v(t+T_c/2). \tag{2.15}$$

Note that for this channelfilter the form g'(t) and energy E'_c only depend on the tape speed v.

For the received signal we have

$$\begin{aligned} r(t) &= s(t) + n_w(t) \\ &= \sqrt{E'_c}.\sum_{n=0}^{\infty}(a_i-a_{i-1}).g'(t-iT_c) + n_w(t). \end{aligned} \tag{2.16}$$

2.1.2 Remarks

One could define the energy put into the channel per channelbit as

$$E_p \triangleq \int_{-\infty}^{\infty} p_{Tc}^2(t).dt = E_b.T_c. \tag{2.17}$$

The channel though only responds to changes in the signal q(t), for it is defined by its stepresponse $h_v(t)$. This makes the definition E_p

useless. Also E_b can not be interpreted as the energy put into the channel per sourcebit. Actually, proper energy conservation is not main concern in magnetic recording. The value of the magnetization levels $\pm\sqrt{E_b}$ determine the effort to magnetize the medium. Further one is only interested in the inter channelsymbol time T_c, as it determines the number of channelbits that can be stored on the medium.

2.2 Receiver and discrete time system

2.2.1 Maximum likelihood sequence estimation

The pulses g(t) and its time translates $g(t-kT_c)$ are not orthonormal. This means that intersymbol interference (ISI) occurs. Bit by bit detection is not optimal. Forney [FOR72] describes the maximum likelihood sequence estimator (MLSE) for optimal detection of digital sequences in the presence of ISI. The estimator chooses the channel sequence $\hat{a}(D) \in \mathbf{S}_c$ that best explains the received waveform r(t):

$$P[r(t)\,|\,a(D)=\hat{a}(D)] \geq P[r(t)\,|\,a(D)\neq\hat{a}(D)] \quad \text{for all } a(D) \in \mathbf{S}_c. \quad (2.18)$$

The estimator consists of a so-called whitened matched filter followed by a recursive nonlinear processor , the Viterbi algorithm. The whitened matched filter comprises a matched filter followed by a sampler and a discrete time noise whitening filter. The noise samples are whitened (made statistical independent) to facilitate the analysis of the performance of the Viterbi algorithm.

As the following analysis leans heavily on the contents of [FOR72] the reader is assumed to be familiar with it.

2.2.2 Whitened matched filter and equivalent discrete time system

Primary question in the design of the whitened matched filter is: "Where must it be matched to ? To the stepresponse of the the channel (second model) or to the total response (first model) ?" The whitened matched filter is derived for the two recording channel models of §2.1. The resulting systems will appear essentially equal. For resasons of analysis first the second model is considered.

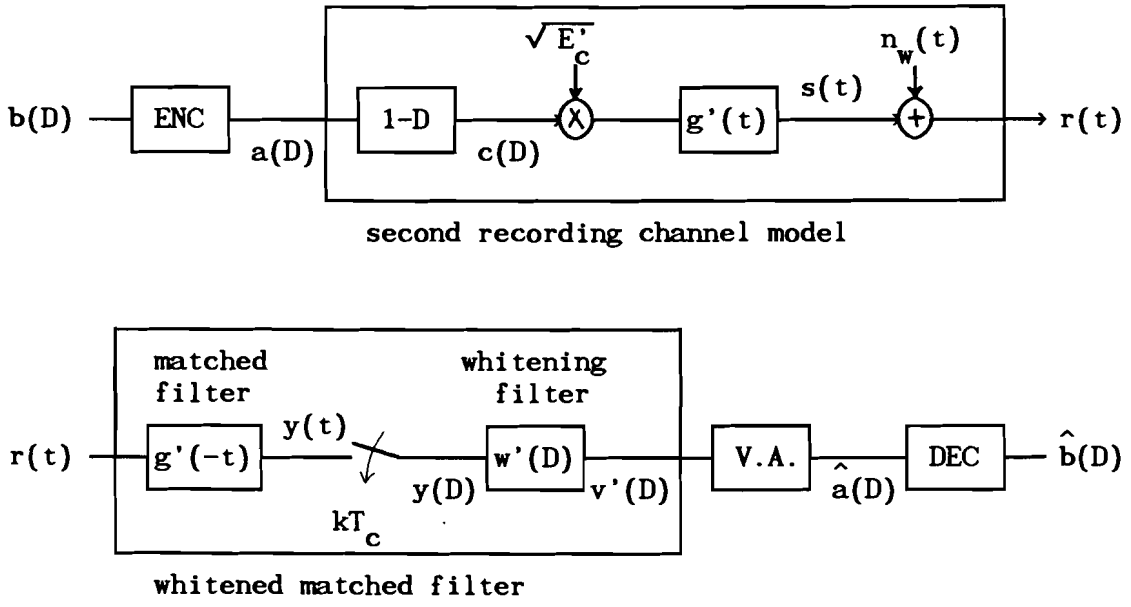


Fig. 2.7 Second channel model embedded in the system

Second channel model

The receiving filter g'(-t) is matched to the normalized impulse response $g'(t) = h_v(t+T_c/2)$. So

$$g'(-t) = h_v(t-T_c/2). \tag{2.19}$$

To determine the discrete time transfer function between c(D) and y(D) we define

$$x'(t) \triangleq g'(t)*g'(-t)$$
$$= h_v(t+T_c/2)*h_v(t-T_c/2). \tag{2.20}$$

Sampling x'(t) at times kT_c we associate with it the polynomial

$$x'(D) = x'_\ell D^{-\ell} + \ldots + x'_1 D^{-1} + x'_0 + x'_1 D + \ldots + x'_\ell D^\ell \tag{2.21}$$

with $x'_k \triangleq x'(kT_c)$ and ℓ possibly infinite. In practice ℓ can always be chosen finite. The polynomial x'(D) is real and even ($x'_{-k} = x'_k$) and can be split in two polynomials

$$x'(D) = f'(D).f'(D^{-1}) \tag{2.22a}$$

where

$$f'(D) = f'_0 + f'_1 D + \ldots + f'_\ell D^\ell \tag{2.22b}$$

is chosen to have all roots inside the unit circle. Then $f'(D^{-1})$ has all roots outside the unit circle. The polynomial f'(D) has normalized sum of the squared coefficients

$$\sum_{i=0}^{\ell} f'^2_i = 1. \tag{2.23}$$

This follows from

$$x'_0 = x'(0) = \int_{-\infty}^{\infty} g'^2(\tau).d\tau = 1 \tag{2.24}$$

and $x'(D) = f'(D^{-1}).f'(D) = (f'_\ell D^{-\ell} + \ldots + f'_0).(f'_0 + \ldots + f'_\ell D^\ell)$

$$= f'_\ell.f'_0.D^{-\ell} + \ldots + (f'^2_0 + f'^2_1 + \ldots + f'^2_\ell) + \ldots + f'_0.f'_\ell.D^\ell. \tag{2.25}$$

The sampled noise z(D) at the output of the matched filter is coloured

and has autocorrelation function

$$R_{zz}(D) = (N_o/2).x(D). \tag{2.26}$$

This means that the average energy in a noise sample

$$E\{z_k^2\} = (N_o/2).x(0) = N_o/2, \tag{2.27}$$

due to the normalization (2.13)

$$\int_{-\infty}^{\infty} g'^2(t).dt = 1.$$

The noise z(D) is decorrelated with the transversal filter

$$w'(D) = 1/f'(D^{-1}) \tag{2.28}$$

which is realizable (the sum of the coefficients is square summable) because $f'(D^{-1})$ has all roots outside the unit circle. The whitened noise $\eta(D)$ has autocorrelation function

$$\begin{aligned} R_{\eta\eta}(D) &= R_{zz}(D).w'(D).w'(D^{-1}) \\ &= (N_o/2).x(D).w'(D).w'(D^{-1}) \\ &= (N_o/2).f'(D).f'(D^{-1}).\frac{1}{f'(D^{-1})}.\frac{1}{f'(D)} \\ &= N_o/2. \end{aligned} \tag{2.29}$$

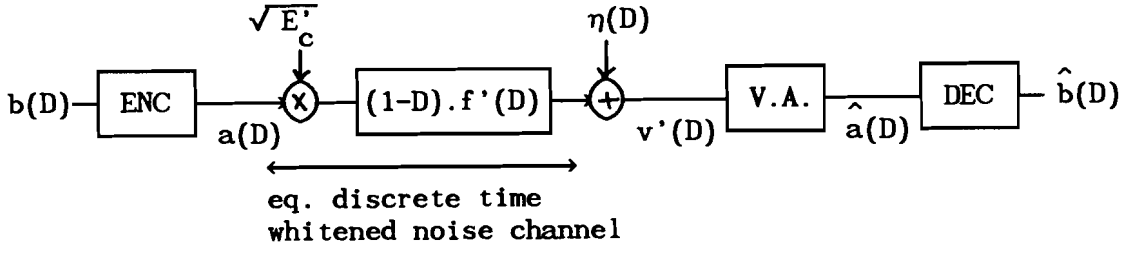


Fig. 2.9 Equivalent discrete time whitened noise model in the system

For the signal at the output of the whitening filter we find

$$v'(D) = a(D).\sqrt{E'_c}.(1-D).x'(D).w'(D) + \eta(D)$$

$$= a(D).\sqrt{E'_c}.(1-D).f'(D) + \eta(D) \qquad (2.30a)$$

with $E'_c = E_b.E_h$. (2.30b)

$\sqrt{E'_c}.(1-D).f'(D)$ is called the <u>equivalent discrete time whitened noise channel</u>.

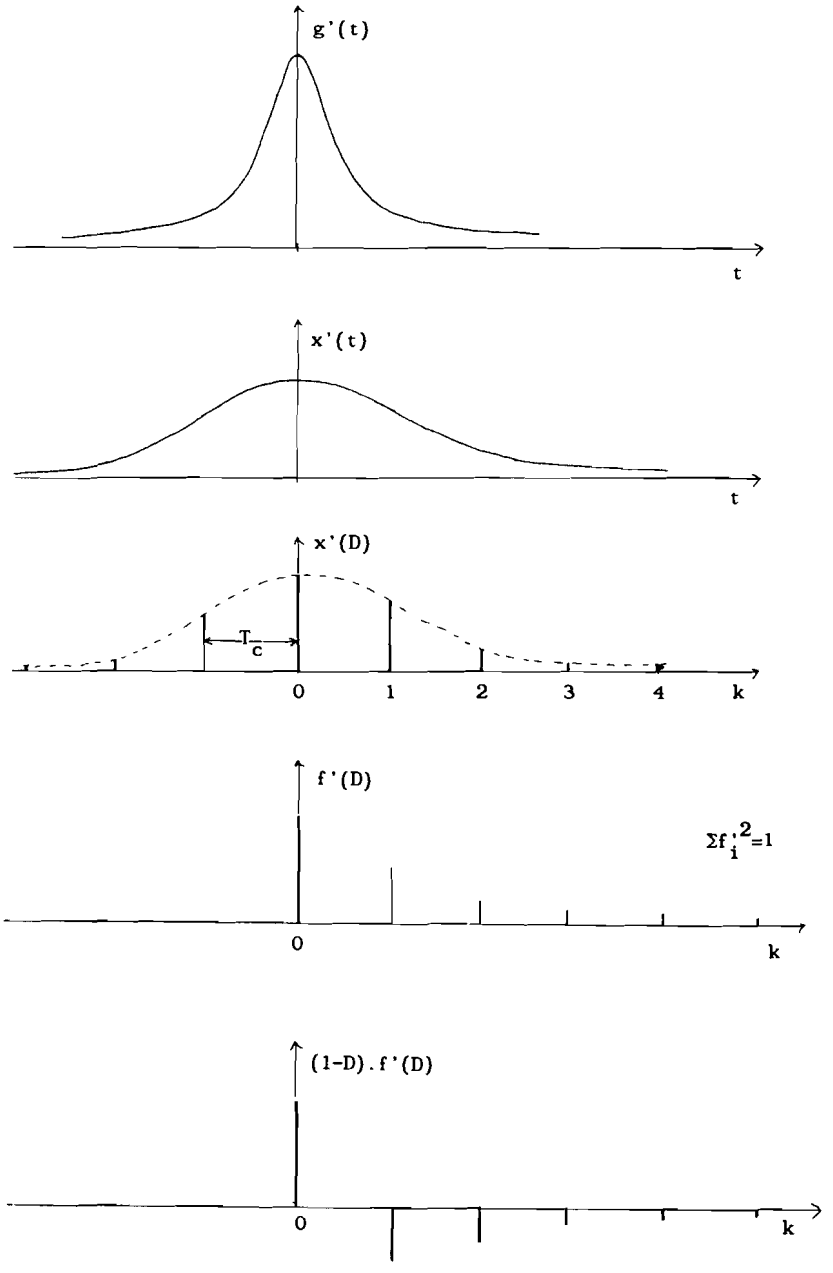


Fig. 2.8 Continuous time and discrete time responses

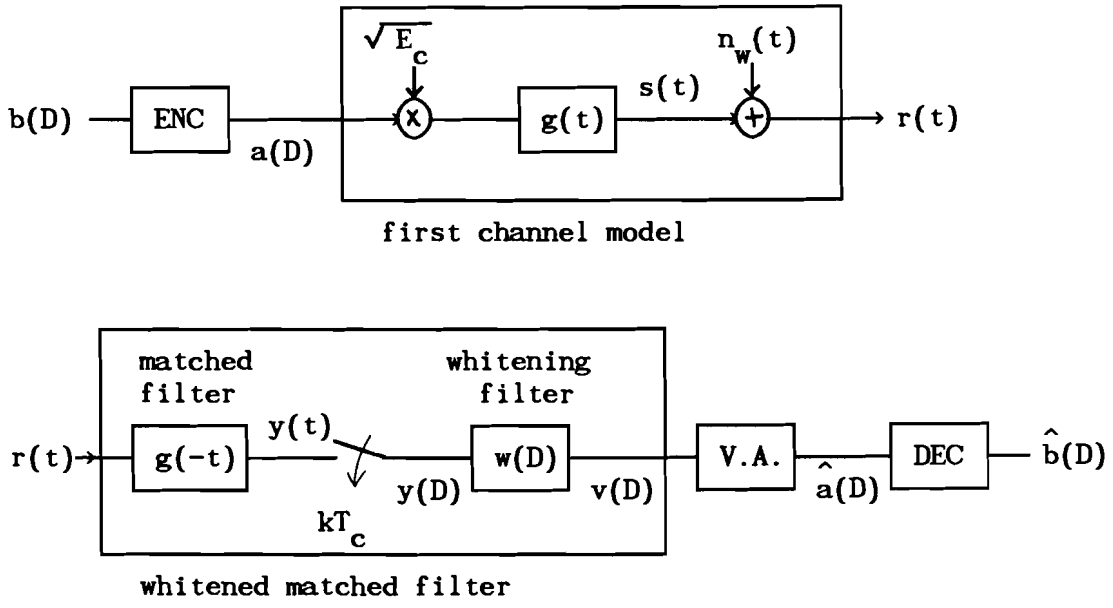


Fig. 2.10 First channel model embedded in the system

First channel model

The receiving filter g(-t) is matched to the normalized impulse response g(t). We define

$$x(t) \triangleq g(t)*g(-t) \tag{2.31}$$

and with $x_k \triangleq x(kT_c)$

$$x(D) = x_{\ell+1}D^{-\ell-1} + .. + x_1D^{-1} + x_0 + x_1D^1 + .. + x_{\ell+1}D^{\ell+1}. \tag{2.32}$$

We will show that the polynomial x(D) can be split in two polynomials

$$x(D) = f(D).f(D^{-1}) \tag{2.33a}$$

where $f(D) = \sqrt{E_c'/E_c}.(1-D).f'(D)$

and $f(D^{-1}) = \sqrt{E_c'/E_c}.(1-D^{-1}).f'(D^{-1}).$ (2.33b)

Note that due to the former relations x(D) has length $2(\ell+1)+1$,

whereas x'(D) has length $2\ell+1$.

Writing out x(t) we get

$$x(t) = g(t)*g(-t)$$

$$= E_b E_h/E_c.\{ h_v(t+T_c/2)-h_v(t-T_c/2)\}*\{h_v(-t+T_c/2)-h_v(-t-T_c/2) \}$$

$$= E_b E_h/E_c.\{ h_v(t+T_c/2)*h_v(-t+T_c/2) - h_v(t-T_c/2)*h_v(-t+T_c/2)$$

$$- h_v(t-T_c/2)*h_v(-t+T_c/2) + h_v(t-T_c/2)*h_v(-t-T_c/2) \}$$

With the even-ness of $h_v(t)$ and $x'(t) = h_v(t+T_c/2)*h_v(t-T_c/2)$ we find

$$x(t) = E_b E_h/E_c.\{ x'(t) - x'(t-T_c) - x'(t+T_c) +x'(t) \}$$

$$= E_b E_h/E_c.\{ -x'(t+T_c) + 2.x'(t) - x'(t-T_c) \}. \qquad (2.34a)$$

So $x(kT_c) = (E_c'/E_c).\{ -x'((k+1)T_c) +2x'(kT_c) -x'((k-1)T_c)\}$ (2.34b)

$$x_k = (E_c'/E_c).\{ -x'_{k+1} +2x'_k -x'_{k-1}\}. \qquad (2.34c)$$

$$x(D) = (E_c'/E_c).\{ -D^{-1} + 2 - D \}.x'(D)$$

$$= (E_c'/E_c).(1-D).f'(D).(1-D^{-1}).f'(D^{-1})$$

$$= \sqrt{E_c'/E_c}.(1-D).f'(D).\sqrt{E_c'/E_c}.(1-D^{-1}).f'(D^{-1})$$

$$= f(D).f(D^{-1}). \qquad (2.34d)$$

Note that also for f(D)

$$\sum_{i=0}^{\ell+1} f_i^2 = 1. \qquad (2.35)$$

For the whitening filter we find

$$w(D) = \frac{1}{f(D^{-1})} = \frac{1}{1-D^{-1}}.\frac{1}{f'(D^{-1})}$$

$$= (\ldots + D^{-2}+ D^{-1}+1).\frac{1}{f'(D^{-1})} . \qquad (2.36)$$

It can be implemented as a (discrete time) integrator followed by w'(D), the transversal whitening filter of the second channel model. Note that the zero $\beta=1$ of $f(D^{-1})$ lies on the unit circle. This means

that the coefficients of w(D) are not square summable: the integrator should summate from time k=$-\infty$. But because there is no signal a(D) before k=0 it suffices to integrate from k=0.

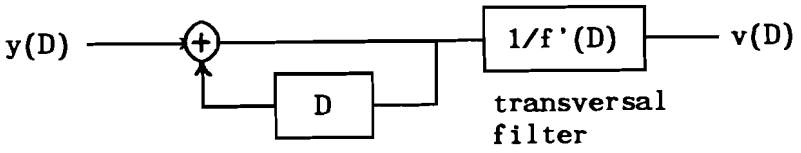


Fig. 2.11 Whitening filter w(D)

For the signal at the output of the whitening filter we have

$$v(D) = a(D).\sqrt{E_c}.x(D).w(D) + \eta(D)$$

$$= a(D).\sqrt{E_c}.f(D) + \eta(D) \qquad (2.37a)$$

with $R_{\eta\eta}(D) = N_o/2.$ (2.37b)

It is readily shown that the equivalent discrete time whitened noise channel $\sqrt{E_c}.f(D)$ equals $\sqrt{E_c'}.(1-D).f'(D)$:

$$\sqrt{E_c}.f(D) = \sqrt{E_c}.\sqrt{E_c'/E_c}.f'(D) = \sqrt{E_c'}.(1-D).f'(D). \qquad (2.38)$$

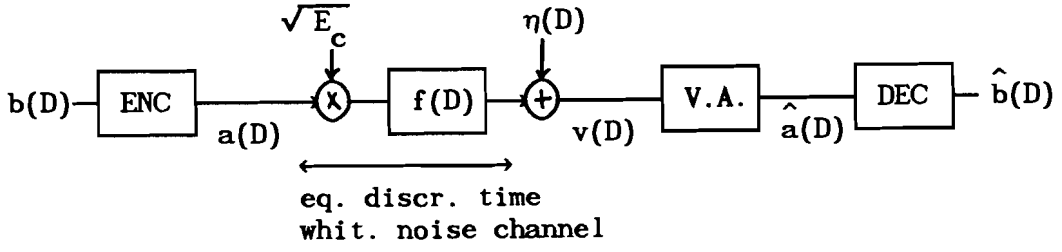


Fig. 2.13 Equivalent discrete time whitened noise model in the system

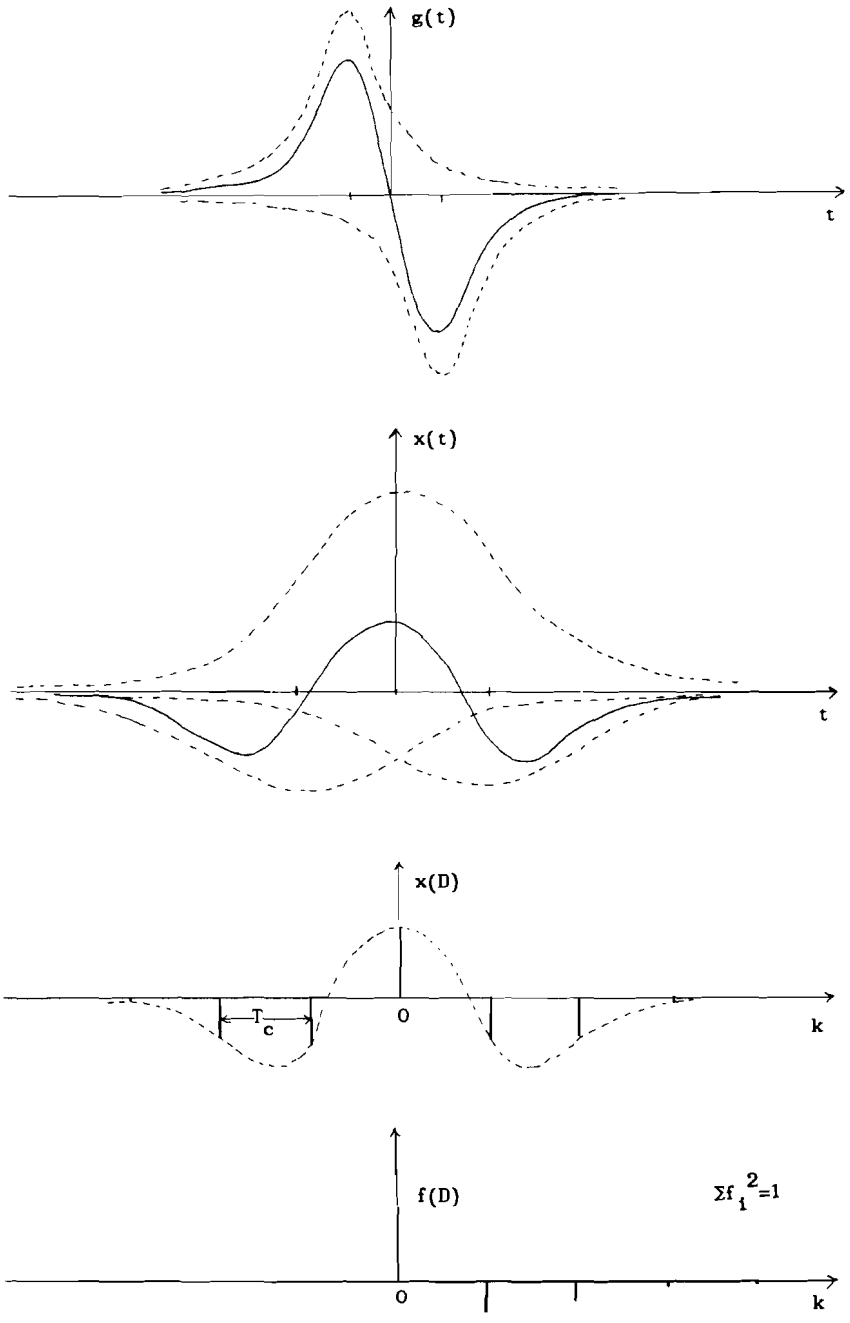


Fig. 2.12 Continuous time and discrete time responses

2.2.3 Interpretation of the two models

At this time a reflection on the two models is at place. The main

results for the continuous time and the equivalent discrete time channel are summarized in the scheme. Some results to be derived in this section are added.

<u>First channel model</u>

$$r(t) = \sqrt{E_c}.\sum_{n=0}^{\infty} a_i.g(t-iT_c) + n_w(t)$$

$$\sqrt{E_c}.g(t) = \sqrt{E_b}.\sqrt{E_h}.\{h_v(t+T_c/2)-h_v(t-T_c/2)\}$$

$$F_{dif} \triangleq \int_{-\infty}^{\infty} \{h_v(t+T_c/2)-h_v(t-T_c/2)\}^2 dt$$

$$E_c = E_b.E_h.F_{dif}$$

$$v(D) = \sqrt{E_c}.f(D) + \eta(D)$$

$$\sum_{i=0}^{\ell+1} f_i^2 = 1 \qquad \sum_{i=0}^{\ell+1} f_i = 0$$

<u>Second channel model</u>

$$r(t) = \sqrt{E'_c}.\sum_{n=0}^{\infty} (a_i-a_{i-1}).g'(t-iT_c) + n_w(t)$$

$$\sqrt{E'_c}.g'(t) = \sqrt{E_b}.\sqrt{E_h}.h_v(t+T_c/2)$$

$$E'_c = E_b.E_h$$

$$v(D) = \sqrt{E'_c}.(1-D).f'(D) + \eta(D)$$

$$\sum_{i=0}^{\ell} f'^2_i = 1 \qquad \sum_{i=0}^{\ell+1} [(1-D).f'(D)]_i^2 = E_c/E'_c = F_{dif}$$

<u>Equality of two models</u>

$$\sqrt{E_c}.f(D) = \sqrt{E'_c}.(1-D).f'(D) \quad \text{or} \quad \sqrt{F_{dif}}.f(D) = (1-D).f'(D)$$

Scheme 2.1 Summary of the two models

In the first channel model the differentiation is incorporated in the continuous time response $\sqrt{E_c}.g(t)$. The received energy per channelbit is given by

$$E_c = E_b.E_h.\int_{-\infty}^{\infty}\{h_v(t+T_c/2)-h_v(t-T_c/2)\}^2dt. \qquad (2.39)$$

Here the integral represents the effect of differentiation and dispersion on the energy. We define the differentiation factor as

$$\begin{aligned} F_{dif} &\triangleq \int_{-\infty}^{\infty}\{h_v(t+T_c/2)-h_v(t-T_c/2)\}^2dt \\ &= \int_{-\infty}^{\infty}\{h_v^2(t+T_c/2)+h_v^2(t-T_c/2)-2.h_v(t+T_c/2).h_v(t-T_c/2)\}dt \\ &= 2.\int_{-\infty}^{\infty}h_v^2(t).dt - 2.\int_{-\infty}^{\infty}h_v(t+T_c/2).h_v(t-T_c/2).dt \\ &= 2 - 2.\int_{-\infty}^{\infty}h_v(t+T_c/2).h_v(t-T_c/2).dt. \qquad (2.40) \end{aligned}$$

Clearly $F_{dif}=0$ for $T_c=0$. F_{dif} obtains its maximal value 2 if $h_v(t+T_c/2)$ and $h_v(t-T_c/2)$ are orthonormal (e.g. when $T_c = \infty$).
So with

$$E_c = E_b.E_h.F_{dif} \qquad (2.41)$$

we have

$$\sqrt{E_c}.f(D) = \sqrt{E_b}.\sqrt{E_h}.\sqrt{F_{dif}}.f(D). \qquad (2.42)$$

Due to the normalisation of g(t) we can split the equivalent discrete time channel in an amplifying part $\sqrt{E_c} = \sqrt{E_b}.\sqrt{E_h}.\sqrt{F_{dif}}$ and a non-amplifying dispersive part f(D) that smears out the symbols a(D). The fact that f(D) does not amplify is reflected by (2.35)

$$\sum_{i=0}^{\ell+1} f_i^2 = 1.$$

From $f(D) = \sqrt{E_c'/E_c}.(1-D).f'(D)$ we also find

$$\sum_{i=0}^{\ell+1} f_i = \sqrt{E_c'/E_c}.[f_0' + (f_1'-f_0') + \ldots + (f_\ell'-f_{\ell-1}') - f_\ell'] = 0. \qquad (2.43)$$

This is a consequence of the differentiating nature of the magnetic recording channel.

With the notion of F_{dif} the transition from the first to the second channel model is easily made:

$$\sqrt{E_c}.f(D) = \sqrt{E_b}.\sqrt{E_h}.\sqrt{F_{dif}}.f(D)$$
$$= \sqrt{E_b}.\sqrt{E_h}.(1-D).f'(D). \qquad (2.44)$$

Where in the first model the loss in energy due to the differentiation is represented by $\sqrt{F_{dif}}$, in the second model it is discounted in $(1-D).f'(D)$:

$$\sum_{i=0}^{\ell+1} [(1-D).f'(D)]_i^2 = \sum_{i=0}^{\ell+1} [\sqrt{F_{dif}}.f(D)]_i^2$$
$$= F_{dif}.\sum_{i=0}^{\ell+1} f_i^2 = F_{dif}. \qquad (2.45)$$

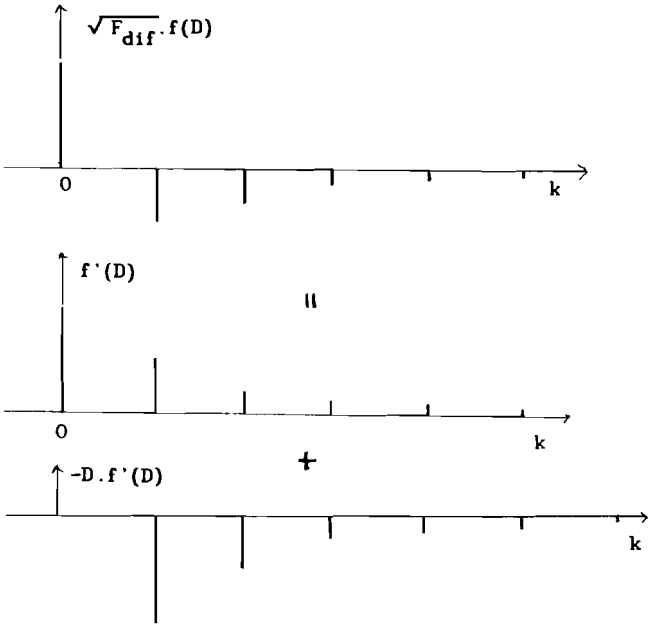


Fig. 2.14 $\sqrt{F_{dif}}.f(D)$ *dissected in* $f'(D)$ *and* $-D.f'(D)$

Remarks on the implementation of the receiver

For the first channel model the matched filter g(-t) depends on both v and T_c. In the second model g'(-t) only depends on v. This can be advantageous when the sample time T_c is not constant (e.g. during experimentation). It should also be noted that the whitening filter does not have to be implemented. The Viterbi algoritm can operate directly on the outputs of the matched filter [VIT79, §4.9]. The whitening filter is only introduced to facilitate the analysis (because the noise samples at its output are statistical independent).

2.3 Euclidian distance as performance parameter

Forney [FOR72 and FOR72*] derives bounds on the bit error probability P_e for MLSE detection. The minimum squared Euclidian distance is key parameter in the bounds.

2.3.1 Euclidian distance and error probability

Let a(D) and $\tilde{a}(D)$ be distinct encoded sequences $\in \mathbf{S}_c$ that differ in a finite number of positions $a_0 \ldots\ldots a_\lambda$. The ternary error sequence e(D) $\in \mathbf{S}_e$ is defined as

$$e(D) \triangleq (a(D)-\tilde{a}(D))/2$$
$$= e_0 + e_1 D + e_2 D^2 + \ldots + e_\lambda D^\lambda \qquad e_i \in \{-1, 0, 1\} \qquad (2.46)$$

with $\mathbf{S}_e$ the set of all possible error sequences. The code set $\mathbf{S}_c$ determines $\mathbf{S}_e$ uniquely (but not vice versa).

The signal error sequence $\epsilon(D)$ is defined as the output of the noiseless discrete time channel with input e(D):

$$\begin{aligned}\epsilon(D) &= e(D).\sqrt{E_c}.f(D)\\ &= e(D).\sqrt{E_c'}.(1-D).f'(D)\\ &= \epsilon_0 + \epsilon_1 D + \epsilon_2 D^2 + \ldots + \epsilon_{\lambda+\ell+1} D^{\lambda+\ell+1} \qquad \epsilon_i \in (-\infty, \infty). \end{aligned} \tag{2.47}$$

The squared Euclidian distance between a(D) and $\tilde{a}(D)$ is defined as the energy in the associated signal error sequence:

$$\begin{aligned}\delta^2(e(D)) &\triangleq \|\epsilon(D)\|^2\\ &= \sum_{i=0}^{\lambda+\ell+1} \epsilon_i^2. \end{aligned} \tag{2.48}$$

The energy in the signal error sequence is a measure for the bit error probability. For small P_e the error sequences e(D) with small $\delta^2(e(D))$ dominate the error performance. The minimum squared Euclidian distance δ^2_{min} is defined as

$$\delta^2_{min} \triangleq \min_{e(D)\in S_e} \delta^2(e(D)) \tag{2.49}$$

and the set of all error sequences e(D) with $\delta^2(e(D))=\delta^2_{min}$ is denoted with $E_{\delta min}$. The bit error probability is bounded by [FOR72]:

$$K_o.Q\left[\sqrt{2.\delta^2_{min}/N_o}\right] \leq P_e \leq K_2.Q\left[\sqrt{2.\delta^2_{min}/N_o}\right] \tag{2.50}$$

where K_0 and K_2 depend on the set $E_{\delta min}$ but are independent of the noise power density $N_o/2$. We will not actually calculate the bounds. It suffices to see the key role of δ^2_{min} in the determination of P_e. Let's say that there is a K (depending on $E_{\delta min}$) for which:

$$P_e \simeq K.Q\left[\sqrt{2.\delta^2_{min}/N_o}\right]. \tag{2.51}$$

Computation of δ^2_{min}

For the first channel model we have for the squared Euclidian distance:

$$\begin{aligned}\delta^2(e(D)) &= \|\epsilon(D)\|^2 \\ &= \|e(D).\sqrt{E_c}.f(D)\|^2 \\ &= E_c.\|e(D).f(D)\|^2. \end{aligned} \tag{2.52a}$$

With

$$d^2(e(D)) \triangleq \|e(D).f(D)\|^2 \tag{2.52b}$$

the $\delta^2(e(D))$ is divided in an amplification factor E_c and a part $d^2(e(D))$ that depends on the non-amplifying dispersive part f(D) and the code set S_c. E_c is divided further in $E_b.E_h.F_{dif}$. For the minimum squared Euclidian distance we analogously have:

$$\delta^2_{min} = E_c.d^2_{min} \tag{2.53a}$$

$$\text{with } d^2_{min} \triangleq \min_{e(D)\in S_e} d^2(e(D)) \tag{2.53b}$$

the normalized squared Euclidian distance.

The straightforward computation of the normalized minimum squared Euclidian distance d^2_{min} thus involves the generation of all possible error sequences $e(D) \in S_e$, the computation of $\epsilon(D) = e(D).f(D)$ and the calculation of $\Sigma \epsilon_i^2$.

For the second channel model we have

$$\delta^2_{min} = E'_c.d'^2_{min} \tag{2.54a}$$

$$\text{with } d'^2_{min} \triangleq \min_{e(D)\in S_e} \|e(D).(1-D).f'(D)\|^2. \tag{2.54b}$$

Here the effect of the differentiation is incorporated totally in d'^2_{min}. Note that

$$d'^2_{min} = F_{dif} \cdot d^2_{min}. \qquad (2.55)$$

It must be noted that for the determination of $\delta^2(e(D))$ the polynomial x(D) does not necessarily have to be split in $f(D).f(D^{-1})$ [FOR72, formula 66]:

$$\begin{aligned} d^2(e(D)) &= \|e(D).f(D)\|^2 \\ &= [e(D^{-1}).f(D^{-1}).f(D).e(D)]_0 \\ &= [e(D^{-1}).x(D).e(D)]_0. \end{aligned} \qquad (2.56)$$

With the matrix-vector notation

$$e = [e_0, e_1, \ldots, e_\lambda] \qquad (2.57a)$$

and **X** the $(\lambda+1)*(\lambda+1)$ matrix with entries

$$X_{i,j} = x_{i-j} = x_{j-i} \qquad (2.57b)$$

we have the neat result

$$d^2(e) = e.X.e^T. \qquad (2.57c)$$

In [PRO83,p.405..408] an other method for the computation of $d^2(e(D))$ (using matrices) is presented.

2.3.2 Dependency of δ^2_{min} on parameters; coding

With the dissection of δ^2_{min} in components the influence of the user parameters D_s, T_s and the designer variables R and S_c can be made clear. Let the user information density D_s and the inter sourcesymbol time T_s be given. The tape speed v is determined uniquely by D_s and T_s (1.2):

$$v = 1/(D_s.T_s).$$

The variable R determines T_c (1.3):

$$T_c = R.T_s.$$

The code S_c only has influence on the dispersive part of δ^2_{min}.

First channel model

$$\delta^2_{min} = E_c.d^2_{min}$$
$$= E_b.E_h.F_{dif}.d^2_{min}. \qquad (2.58)$$

E_b is an independent parameter.

E_h depends on v, thus on $D_s.T_s$: $E_h(D_s.T_s)$.

$F_{dif} = \int_{-\infty}^{\infty}\{h_v(t+T_c/2)-h_v(t-T_c/2)\}^2dt$, and depends on v and T_c, thus on D_s,T_s and R: $F_{dif}(D_s,T_s,R)$.

The f(D) depends on v and T_c, so on D_s,T_s and R. The code S_c influences the d^2_{min} via S_e: $d^2_{min}(D_s,T_s,R,S_c)$.

Note that always $\Sigma f_i^2=1$. So the influence of R on the δ^2_{min} is divided in an energetic part F_{dif} and a dispersive part d^2_{min}.

Summarized: (2.59)

$$\delta^2_{min}(D_s,T_s,R,S_c) = E_b.E_h(D_s.T_s).F_{dif}(D_s,T_s,R).d^2_{min}(D_s,T_s,R,S_c).$$

Second channel model

With $d'^2_{min} = F_{dif}.d^2_{min}$ we see immediately

$$\delta^2_{min}(D_s,T_s,R,S_c) = E_b.E_h(D_s.T_s).d'^2_{min}(D_s,T_s,R,S_c). \qquad (2.60)$$

In this approach the role of the rate R on δ^2_{min} is all incorporated in d'^2_{min}.

So in the first model the separate influence of R and S_c on δ^2_{min} is more obvious.

2.4 Resulting problem statement

Now we can translate the design problem from chapter 1 to a more specific design problem. From $P_e \leq P_0$ and

$$P_e \simeq K.Q\left[\sqrt{2\delta^2_{min}/N_0}\right] \tag{2.61}$$

we have the requirement:

$$\delta^2_{min} \geq \frac{N_0}{2}.\left[Q^{inv}(P_e/K)\right]^2 \triangleq \delta^2_0. \tag{2.62}$$

The task of the designer can be stated as:

> Given user information density D_s and inter sourcesymbol time T_s. Is there a rate R and code S_c with rate R for which the sequences a(D) have
>
> $$\delta^2_{min}(D_s,T_s,R,S_c) \geq \delta^2_0$$
>
> over a channel with transfer function $\sqrt{E_c}.f(D)$?

This can be made more specific if the stepresponse of the channel is given. In the next chapter the system is observed closely for the Lorentzian stepresponse.

2.5 Conclusions

Two models for the differentiating magnetic recording channel were analyzed and appear to result in equivalent system models when a whitened matched filter receiver is used. The differentiation factor F_{dif} represents the effect of differentiation and dispersion on the received energy E_c per channelbit.

With maximum likelihood sequence estimation the key parameter for the bit error probability P_e is the minimum squared Euclidian distance δ^2_{min}. The δ^2_{min} can be divided in an amplification part E_c and d^2_{min} that depends on the non-amplifying dispersive part of the channel. This division is very useful in the interpretation of the effects of coding.

A design problem is stated as the translation of the problem formulated by the user of the magnetic recording system.

3 THE LORENTZIAN STEPRESPONSE MODEL

The former analysis is applied to the Lorentzian stepresponse model. Main goal is to add some insight to (coding for) the magnetic recording model. Most results in this part are derived straight (or rewritten) from [IMM 87]. Attention will be paid to the derivation of simpler coding problems and associated gains.

3.1 System description

3.1.1 The Lorentzian stepresponse

The write pulse $p_{Tc}(t)$ is the full-T_c pulse

$$p_{Tc}(t) = \begin{cases} \sqrt{E_b} & |t| \leq T_c/2 \\ 0 & \text{otherwise} \end{cases}. \qquad (3.1)$$

Notice the difference with [IMM87], that is explained by the fact that the inter sourcesymbol time T_s is not necessarily constant here.

The Lorentzian stepresponse is given by

$$\sqrt{E_h}.h_v(t) = \frac{vK}{pw_{50}}.\frac{1}{1 + (2vt/pw_{50})^2}, \qquad (3.2)$$

where v is the medium to head speed, K a constant and pw_{50} a constant that determines the dispersivity of the recording channel. The normalized response $h_v(t)$ has unit energy:

$$\int_{-\infty}^{\infty} h_v^2(t).dt = 1. \qquad (3.3)$$

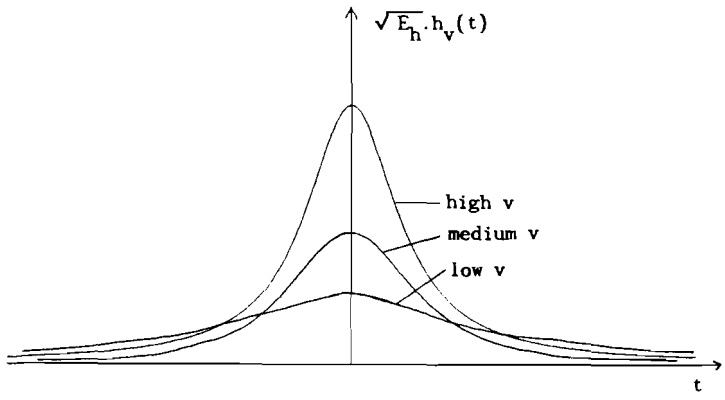


Fig. 3.1 Lorentzian step response for two values of the tape speed v

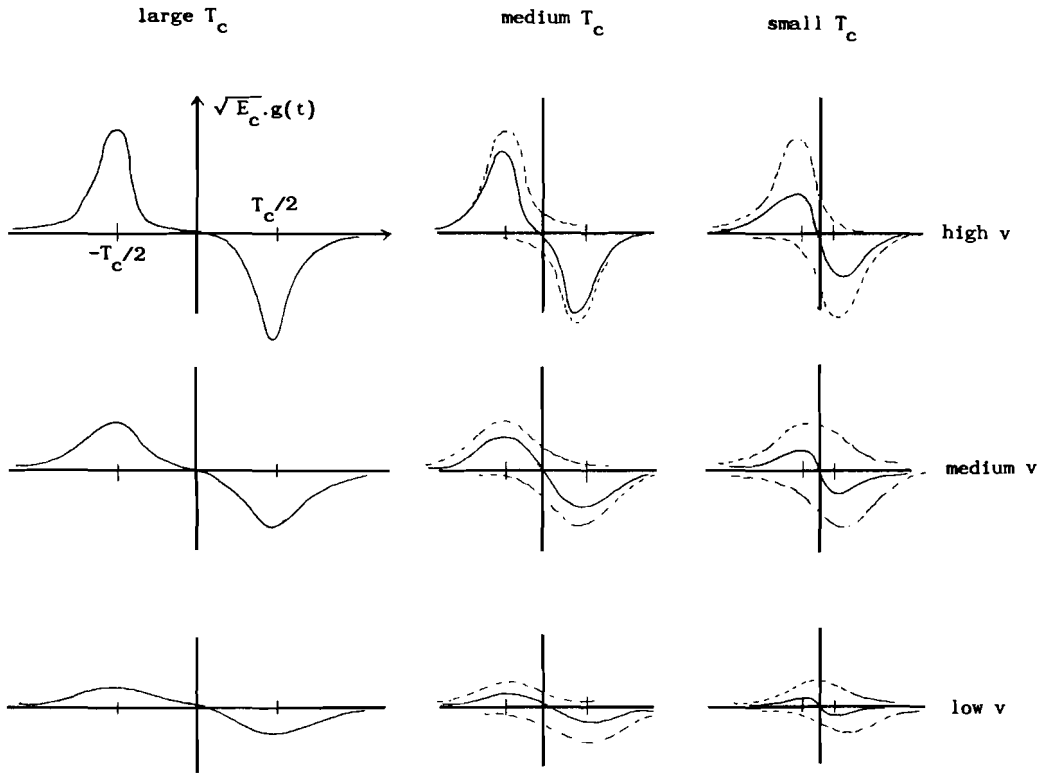


Fig. 3.2 Channel response $\sqrt{E_c}.g(t)$ for various v and T_c

The following equality is very helpful in the analysis of the Lorentzian stepresponse channel:

$$\int_{-\infty}^{\infty} \sqrt{E_h}.h(t-\tau/2).\sqrt{E_h}.h(t+\tau/2).d\tau = \frac{\pi}{4}.\frac{vK^2}{pw_{50}}.\frac{1}{1+(v\tau/pw_{50})^2} \quad (3.4)$$

With $\tau=0$ we find that the energy E_h in the stepresponse is proportional to the tape speed v:

$$E_h = \frac{\pi}{4}.\frac{vK^2}{pw_{50}}. \quad (3.5)$$

Dividing both sides of (3.4) by E_h we have the neat and useful result:

$$\int_{-\infty}^{\infty} h_v(t-\tau/2).h_v(t+\tau/2).dt = \frac{1}{1+(v\tau/pw_{50})^2} \quad (3.6)$$

3.1.2 Discrete time responses

The discrete time response x(D) is derived, using x'(D) as an intermediate result. The continuous time response x'(t) can be derived in a closed analytic form:

$$\begin{aligned} x'(t) &= g'(t)*g'(-t) \\ &= \int_{-\infty}^{\infty} g'(\tau).g'(\tau-t).d\tau \\ &= \int_{-\infty}^{\infty} h_v(\tau+T_c/2).h_v(\tau-t+T_c/2).d\tau. \end{aligned}$$

With $\lambda = \tau + T_c/2 - t/2$ $(d\lambda = d\tau)$ and (3.6) we have:

$$\begin{aligned} x'(t) &= \int_{-\infty}^{\infty} h_v(\lambda+t/2).h_v(\lambda-t/2).d\lambda \\ &= \frac{1}{1+(vt/pw_{50})^2}. \end{aligned} \quad (3.7)$$

Here the usefulness of the Lorentzian pulse $h_v(t)$ becomes clear: The convolution of two pulses can be determined explicitly. For the coefficients of the sampled response x'(D) we have

$$x'_k \triangleq x'(kT_c) = \frac{1}{1+(kvT_c/pw_{50})^2}. \quad (3.8)$$

Apparently x'(D) only depends on $v.T_c$, the physical distance between the symbols on the tape. Defining the physical density

$$S_p \triangleq \frac{pw_{50}}{vT_c} \quad [m^{-1}] \tag{3.9}$$

we have the simple expression

$$x'_k = \frac{1}{1 + (k/S_p)^2} . \tag{3.10}$$

Now the coefficients x_k can be determined with

$$x_k = (E'_c/E_c).\{ -x'_{k+1} + 2x_k - x'_{k-1}\}. \tag{3.11}$$

$E'_c/E_c = 1/F_{dif}$, with

$$\begin{aligned} F_{dif} &\triangleq \int_{-\infty}^{\infty} \{h_v(t+T_c/2)-h_v(t-T_c/2)\}^2.dt \\ &= 2 - 2.\int_{-\infty}^{\infty} h_v(t+T_c/2).h_v(t-T_c/2).dt \\ &= 2 - 2.\frac{1}{1 + (vT_c/pw_{50})^2} \\ &= \frac{2}{S_p^2 + 1} . \end{aligned} \tag{3.12}$$

Note that from (3.12) we have the generally valid expression

$$F_{dif} = 2 - 2.x'_1 \tag{3.12a}$$

So for the coefficients x_k we have the expression (3.13)

$$x_k = \frac{S_p^2 + 1}{2}.\left[- \frac{1}{1 + ((k+1)/S_p)^2} + \frac{2}{1 + (k/S_p)^2} - \frac{1}{1 + ((k-1)/S_p)^2} \right]$$

We do not explicitly split the polynomial x(D) in $f(D).f(D^{-1})$, because this is analytically difficult. As already noted in §2.3.1 the important parameter δ^2_{min} can be computed straight from x(D).

3.1.3 Dependency of δ^2_{min} on parameters

Before presenting results similar to §2.3.2 we introduce the normalized information density

$$S \triangleq \frac{pw_{50}}{vT_s} = S_p.R. \tag{3.14}$$

It has the simple relation to the user storage density D_s

$$D_s = \frac{1}{vT_s} = \frac{S}{pw_{50}}. \tag{3.15}$$

In the sequel only S and S_p will be used.

Let the normalized information density S and inter sourcesymbol time T_s be specified. The tape speed v is determined by $v = pw_{50}/ST_s$. The influence of R and S_c on δ^2_{min} is given with:

$$\delta^2_{min}(S,T_s,R,S_c) = E_b.E_h(S.T_s).F_{dif}(\frac{S}{R}).d^2_{min}(\frac{S}{R},S_c) \tag{3.16a}$$

with

$$E_h(S.T_s) = \frac{\pi}{4}.\frac{pw_{50}}{ST_s}.\frac{K^2}{pw_{50}} = \frac{\pi}{4}.\frac{K^2}{ST_s} \tag{3.16b}$$

and $$F_{dif}(\frac{S}{R}) = \frac{2}{(S/R)^2+1}, \tag{3.16c}$$

and $$d^2_{min}(\frac{S}{R},S_c) = \min_{e(D)\in S_e} [e(D).x(D).e(D^{-1})]_0 \tag{3.16d}$$

where x(D) is a function of $S_p = \frac{S}{R}$.

Note the key role S_p has in both F_{dif} and d^2_{min}.

3.1.4 Some analysis

Before analyzing the design and effect of codes, the uncoded situation is observed more closely. The dependency of δ^2_{min}(uncoded) on S and T_s is analyzed (R=1 and S_p=S) :

$$\delta^2_{min}(\text{uncoded}) = \delta^2_{min}(S.T_s) = E_b.E_h(S.T_s).F_{dif}(S).d^2_{min}(S). \tag{3.17}$$

$\underline{E_c \text{ as a function of S and } T_s}$

$$E_C(S,T_S) = E_b.E_h(S.T_s).F_{dif}(S)$$
$$= E_b.\frac{\pi}{4}.\frac{K^2}{ST_s}.\frac{2}{S^2+1}. \qquad (3.18)$$

So $E_c(S,T_s)$ is a monotoneous decreasing function in both T_s and S. When S increases, E_c decreases slowly for small S, but fast for higher S. Let us analyze the effect of an increase in S by a factor $\alpha > 1$. Observe that $S = pw_{50}/vT_s$ can be increased to αS by

i) decreasing v to v/α (at constant T_s)

ii) decreasing T_s to T_s/α (at constant v)

iii) mixed change of v and T_s.

We will compute the change in E_c for i) and ii). See also fig. 3.2.

i) Increase $S \rightarrow \alpha S$ by $v \rightarrow v/\alpha$ (constant T_s)

$$\frac{E_c(\alpha S,T_s)}{E_c(S,T_s)} = \frac{S^2+1}{\alpha(\alpha^2S^2+1)} = \begin{cases} \simeq \frac{1}{\alpha} & \text{for } S << 1 \\ \simeq (\frac{1}{\alpha})^3 & \text{for } S >> 1 \end{cases} . \qquad (3.19)$$

ii) Increase $S \rightarrow \alpha S$ by $T_s \rightarrow T_s/\alpha$ (at constant v)

Note that with v the E_h remains constant. So only F_{dif} changes:

$$\frac{E_c(\alpha S,T_s/\alpha)}{E_c(S,T_s)} = \frac{S^2+1}{\alpha^2S^2+1} = \begin{cases} \simeq 1 & \text{for } S << 1 \\ \simeq (\frac{1}{\alpha})^2 & \text{for } S >> 1 \end{cases} . \qquad (3.20)$$

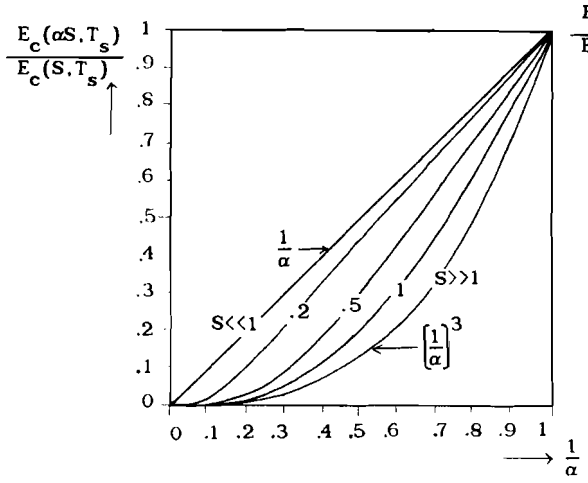


Fig 3.3 Increase S at constant T_s

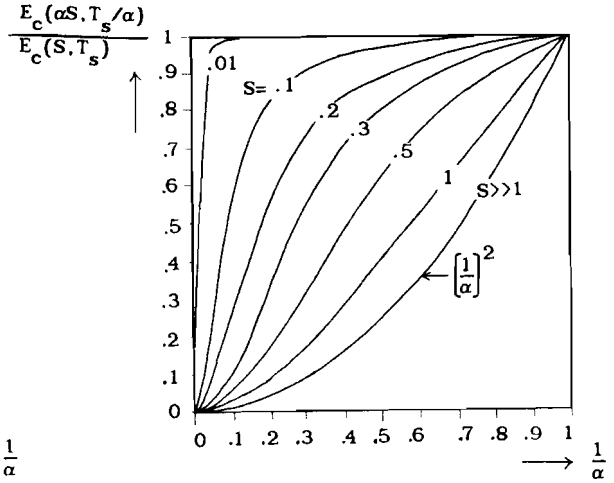


Fig. 3.4 Increase S at constant v

So it is favourable to keep v constant. In audio and video recording however, it is likely that T_s is specified. Then S can only be changed via v. For data storage applications one generally has more freedom to change T_s.

Discrete time responses as a function of S_p

The responses x(D) and f(D) are a function of S_p=S/R. In the uncoded case S_p=S. In fig. 3.5 the responses are given for low (S_p=0.1) medium (S_p=1) and high (S_p=3) physical densities. The f(D) are computed with a root-splitting algorithm.

d^2_{min}(uncoded) as a function of S_p

The sets of error sequences $E_{\delta min}$ that dominate the error performance are given for $0 \leq S_p \leq 5$ [KER87].

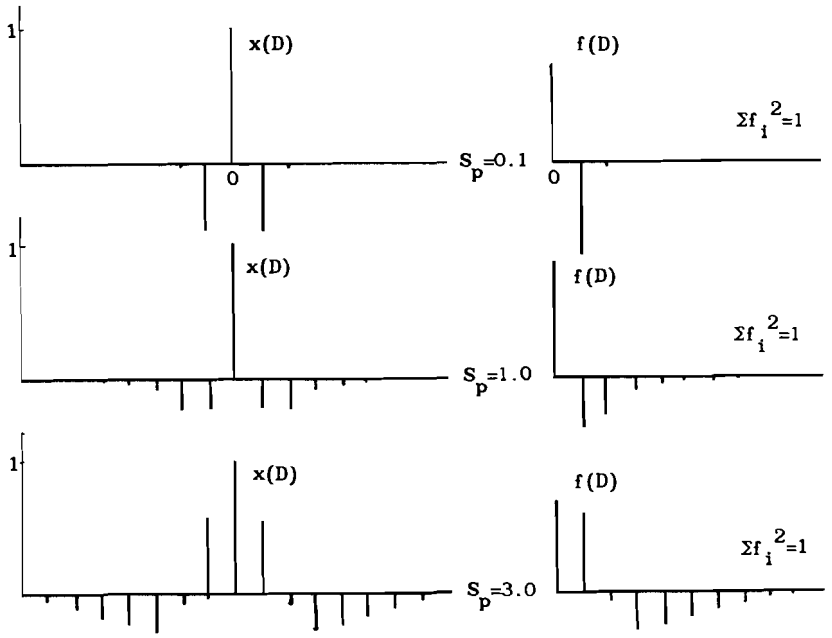


Fig. 3.5 x(D) and f(D) for various S_p

S_p	$E_{\delta min}$
0-2.4	$1 \quad , -1$
2.4-4	$1-D+D^2 \quad , -1+D-D^2$
4-5	$1-D+D^2-D^3+D^4 \quad , -1+D-D^2+D^3-D^4$

Table 3.1 Dominating error sequences

The error sequences $e(D)=\pm 1$ give $d^2_{min}=1$. For the other sequences the d^2_{min} depends on S_p. Although it is possible to derive an analytic relation between d^2_{min} and S_p for a e(D) [IMM87] we do not present it here.

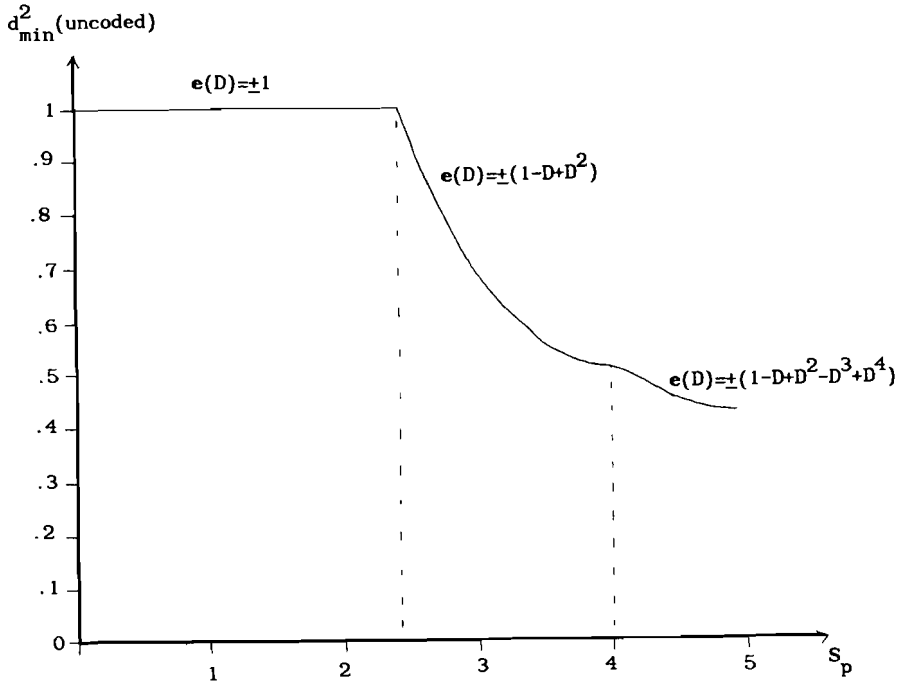


Fig. 3.6 d^2_{min}(uncoded) as a function of S_p

3.2 Problem statements

The problem as stated in §2.4 is too general: The noise power N_0 is not given and can not be estimated theoretically for real systems. The demanded error probability P_e is also different for various applications. This makes it impossible to give a theoretical answer on the question if there exists a system with a certain error performance. But using the model of the magnetic recording channel we can solve the question if we can expect any benefit from the application of codes. This question and derived problems are subject of investigations in the following paragraphs.

3.2.1 Coding for certain S

Question of primary interest is: Can we expect any merit from coding? Or: is it possible to improve the error performance by coding, at a certain constant S. We can define the <u>netto gain</u> $G_n(S,R,\mathbf{S}_c)$ that depends on S and the applied code:

$$G_n(S,R,\mathbf{S}_c) \triangleq \frac{\delta^2_{min}(\text{coded})}{\delta^2_{min}(\text{uncoded})}$$

$$= \frac{E_b.E_h(S.T_s).F_{dif}(\frac{S}{R}).d^2_{min}(\frac{S}{R},\mathbf{S}_c)}{E_b.E_h(S.T_s).F_{dif}(S).d^2_{min}(S,\text{uncoded})}$$

$$= \frac{S^2+1}{(S/R)^2+1}.\frac{d^2_{min}(\frac{S}{R},\mathbf{S}_c)}{d^2_{min}(S,\text{uncoded})}. \qquad (3.21)$$

So the problem can be stated as:

> Let S be given. Is there a rate R and code $\mathbf{S}_c$ with rate R, for which the netto coding gain
>
> $$G_n(S,R,\mathbf{S}_c) \geq 1?$$

Notice that T_s has no influence on G_n. Futher it must be noted that G_n is not a coding gain in the usual sense of the word. As the channel changes through the application of the code, and G_n depends on S, it is a netto gain resulting from a deterioration of the channel and an improvement by the code. It can be divided in an <u>energy loss</u>

$$L_E(S,R) \triangleq \frac{F_{dif}(\frac{S}{R})}{F_{dif}(S)} = \frac{S^2+1}{(S/R)^2+1} \qquad (3.22)$$

that must be compensated by a <u>distance gain</u>

$$G_d(S,R,S_c) \triangleq \frac{d^2_{min}(\frac{S}{R},S_c)}{d^2_{min}(S,\text{uncoded})} \tag{3.23}$$

to have a netto gain $G_n = L_E(S,R).G_d(S,R,S_c) \geq 1$.

For $L_E(S,R)$ we find

$$L_E(S,R) = \begin{cases} \simeq 1 & \text{for } S \ll 1 \\ \simeq R^2 & \text{for } S \gg 1 \end{cases}. \tag{3.24}$$

So at high densities the loss due to the rate is higher than for low densities. This makes it harder to obtain a netto gain by coding..

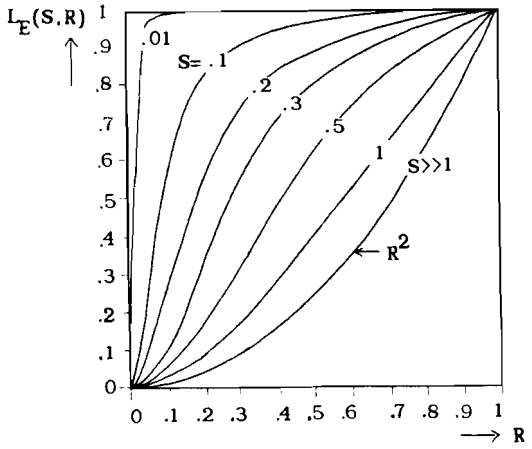


Fig. 3.7 Energyloss due to coding, for various S

For $S \leq 2.4$ we have $d^2_{min}(S,\text{uncod}) = 1$, so $G_d(S \leq 2.4,R) = d^2_{min}(\frac{S}{R},S_c)$.

Notice that for a given S, the channel f(D) changes with every choice of R, as $S_p = \frac{S}{R}$ determines f(D). It is better to analyze the channel f(D) for several S_p, design rate R codes, and calculate the netto gain $G_n(S,R,S_c)$ for the resulting $S = S_p.R$. An example of coding at S_p=2.0 will be given in §3.2.4.

3.2.2 Coding at a certain S_p

Applying a rate R code at a certain S_p, we can specify at forehand the required $d^2_{min}(S_p,S_c)$ to have a netto gain $G_n(S,R,S_c) \geq 1$ at $S=S_p.R$:

$$G_n = L_E(S_p.R,R).G_d(S_p.R,R,S_c)$$

$$= \frac{(S_p.R)^2+1}{S_p^{\,2}+1}.\frac{d^2_{min}(S_p,S_c)}{d^2_{min}(S_p.R,\text{uncoded})}. \qquad (3.25)$$

$G_n \geq 1$ gives the requirement:

$$d^2_{min}(S_p,S_c) \geq d^2_{min}(S_p.R,\text{uncoded}).\frac{S_p^{\,2}+1}{(S_p.R)^2+1}. \qquad (3.26)$$

When $S = S_p.R \leq 2.4$ this reduces to

$$d^2_{min}(S_p,S_c) \geq \frac{S_p^{\,2}+1}{(S_p.R)^2+1}. \qquad (3.27)$$

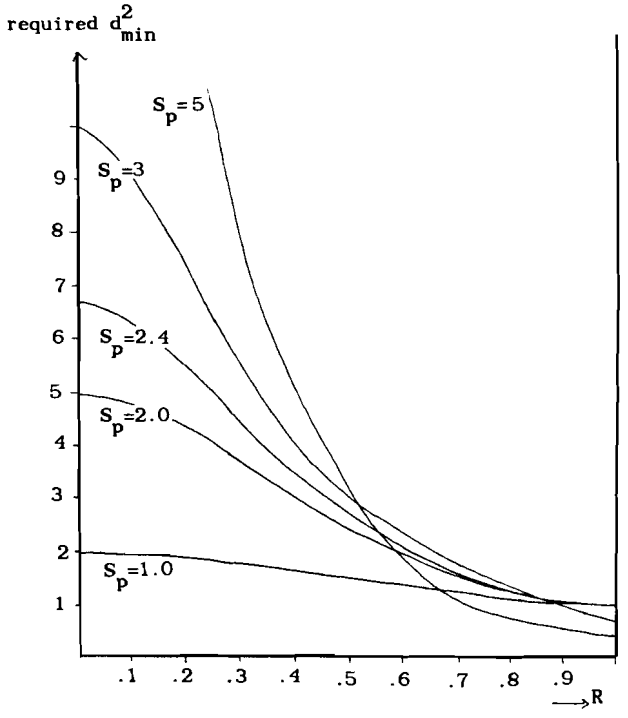


Fig. 3.8 Required d^2_{min}(coded) for netto gain

3.2.3 Increasing S at equal performance

An other parameter that can be held constant is the bit error probability P_e. The question arises: "Is it possible to increase the normalized information density S, without deterioration of the error performance?" Clearly P_e increases when S is increased to αS ($\alpha > 1$) (see §3.1.4). So coding must be applied to compensate for the loss due to the increase of S. A code must be designed for αS, that has a netto gain $G_n(\alpha S, R, \mathbf{S}_c)$ that equals the loss due to the increase of S. This brings us back to the problem of §3.2.1.

3.2.4 Example: coding for S_p=2.0

Consider S_p=2.0. Although not necessary for the analysis, we work with f(D), computed with a rootsplitting algorithm. The infinite length f(D) can be restricted to finite length.

S_p = 2.0	
i	f_i
0	0.85989
1	0.19975
2	- 0.34603
3	- 0.26500
4	- 0.14542
5	- 0.07898
6	- 0.04508
7	- 0.02717
8	- 0.01716
9	- 0.01124
10	- 0.00754
11	- 0.00517
12	- 0.00350
13	- 0.00248
14	- 0.00141
15	- 0.00131

Table 3.2 Coefficients of f(D)

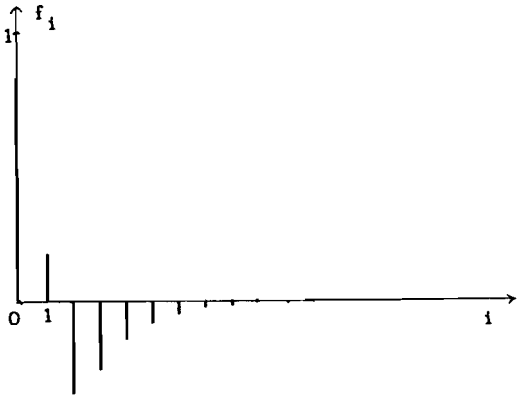


Fig. 3.9 Response f(D)

We can make a distance profile of the channel by generating all error sequences e(D) and computing the resulting $d^2(e(D))$. The number of error sequences that have to be generated can be restricted because sets of sequences have the same $d^2(e(D))$. These equivalency sets E_d can be generated with the following rules:

i) $d^2(e(D)) = d^2(-e(D))$.

This is seen instantaneously from

$$d^2(e(D)) = \|e(D).f(D)\|^2 = \|-e(D).f(D)\|^2.$$

ii) $d^2(e_0+e_1D+\ldots+e_\lambda D^\lambda) = d^2(e_\lambda+e_{\lambda-1}D+\ldots+e_0D^\lambda)$.

This follows from

$$d^2([e_0,e_1,\ldots,e_\lambda]) = [e_0,e_1,\ldots,e_\lambda].\mathbf{X}.[e_0,e_1,\ldots,e_\lambda]^T$$

and the symmetry of $\mathbf{X}$: $X_{i,j} = X_{j,i}$ (or $x_{i-j} = x_{j-i}$).

In table 3.4 and figure 3.8 each equivalency set is denoted with one representant **e** of the set. N_e is the size of the equivalency set.

λ	e	N_e	$d^2(e)$	equivalent sequences N_d
0	1	2	1.0000	-1
1	1-1	2	1.5000	-1 1
	1 1	2	2.5000	-1-1
2	1-1 1	2	1.4615	-1 1-1
	1 0 1	2	1.4616	-1 0-1
	1 0-1	2	2.5384	-1 0 1
	1 1 1	2	3.4615	-1-1-1
	1 1-1	4	3.5384	-1-1 1, -1 1 1, 1-1-1
3	1 0 0 1	2	1.5770	-1 0 0-1
	1-1 1-1	2	1.8461	-1 1-1 1
	1 0 1-1	4	2.3846	-1 0-1 1, -1 1 0 1, 1-1 0-1
	1 0 0-1	2	2.4228	-1 0 0 1
	1 0 1 1	4	2.5385	-1 0-1-1, 1 1 0 1, -1-1 0-1
	1 0-1 1	4	2.6153	-1 0 1-1, 1-1 0 1, -1 1 0-1
	1 1-1 1	4	3.0769	-1-1 1-1, 1-1 1 1, -1 1-1-1
	1 1 1 1	2	4.0000	-1-1-1-1
	1-1-1 1	2	4.1538	-1 1 1-1
	1 1 0-1	4	4.4614	-1-1 0 1, -1 0 1 1, 1 0-1-1
	1 1 1-1	4	4.9231	-1-1-1 1, -1 1 1 1, 1-1-1-1
	1 1-1-1	2	6.0000	-1-1 1 1

Table 3.4 Distance profile at S_p=2.0

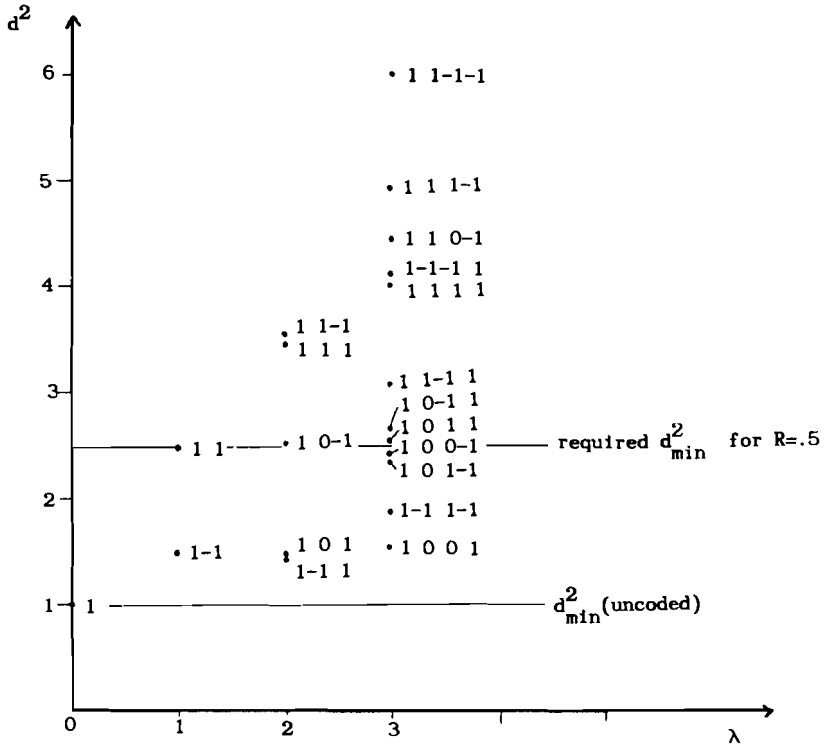


Fig. 3.10 Distance profile at S_p=2.0

Coding

Let us design a rate R=0.5 code. From $G_n \geq 1$ we have the requirement

$$d^2_{min}(S_p=2,S_c) \geq d^2_{min}(S=2*0.5=1).\frac{2^2+1}{(2*0.5)^2+1} = 1.\frac{5}{2} = 2.5.$$

So all error sequences e(D) with $d^2(e(D))$ must be deleted by the code. The simple repetition code

$$-1 \rightarrow -1-D$$

$$1 \rightarrow 1+D$$

deletes all error sequences where the errors do not occur in pairs $\pm(1+D)$. This gives a new distance profile where $d^2_{min}=2.5$, for e(D)=1+D. So we have exactly no netto gain: $G_n(S=1,R=0.5,S_c) = 1$.

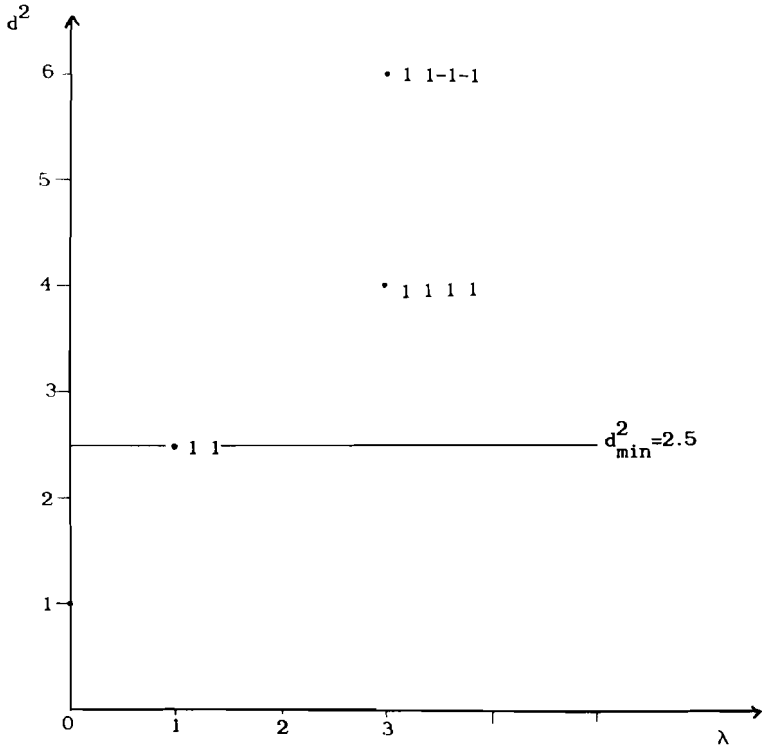


Fig. 3.11 Distance profile after coding

Remark

From this example we can also learn not to over-estimate the role of δ^2_{min}. Although the error performance is dominated by $e(D)=\pm 1$, the error sequences $e(D)=\pm(1-D)$, $\pm(1-D+D^2)$ and $\pm(1+D^2)$ also have small $\delta^2(e(D))$, and therefore great influence on P_e. Certainly at the densities S_p where the sets $E_{\delta min}$ change, at least two kinds of error sequences dominate the error performance.

Forney gives an exact formula of P_e as a function of all error sequences [FOR72, formula 82]. With this formula on can have a better idea of the really obtained gain.

3.3 Conclusions

With the approach presented in chapter 2 the Lorentzian stepresponse model was analyzed. Restricted research problems were formulated. Several gains and losses were defined. It appears to be best to design codes for various physical densities S_p and then translate the results back to the normalized information density S.

For the design of codes for a certain S_p the distance profile is indispensable. From this profile one can make a better estimate of the error performance and (netto) coding gain. This relativates the force of δ^2_{min} as a parameter for the error probability.

4 THE TRIANGULAR STEPREPONSE MODEL

For ease of analysis one can approximate the real stepresponse of the channel by an analytically simple response. The Lorentzian stepresponse gave rise to computable results, but its infinite length is a drawback. One also encounters the (finite length) triangular stepresponse as a model (e.g. [CAL]). With the notion of F_{dif} the Lorentzian and triangular stepresponse are readily compared.

4.1 Channel and system model

The normalized steprsponse h(t) has duration 1 and unit energy:

$$h(t) = \begin{cases} \sqrt{3}.(1-2|t|) & |t| \leq 0.5 \\ 0 & \text{otherwise} \end{cases} . \qquad (4.1)$$

E_h is not specified further, but is taken 1 here, so $\sqrt{E_h}.h(t) = h(t)$. Note that because E_h and h(t) do not depend on v, this model is only suited for analysis of the real system at constant v.

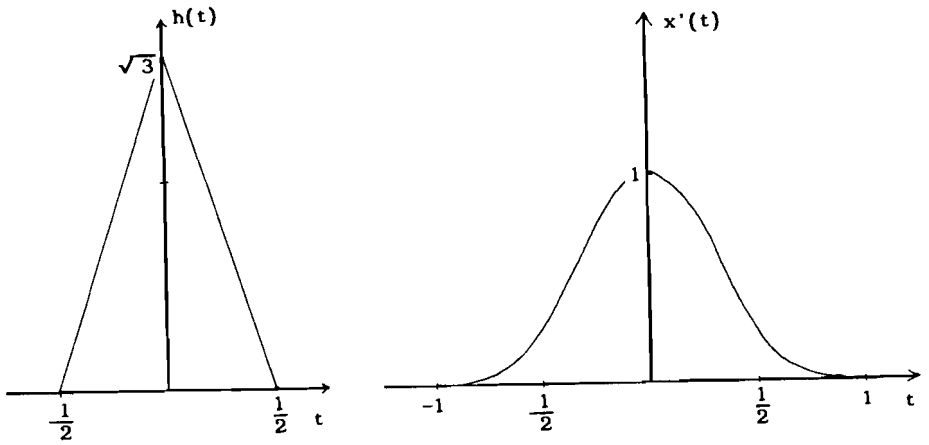


Fig. 4.1 Responses h(t) and x'(t)

We determine x(D) from x'(D). With (3.7) we have

$$x'(t) = g'(t) * g'(-t)$$
$$= \int_{-\infty}^{\infty} h(\lambda+t/2).h(\lambda-t/2).d\lambda \qquad (4.2)$$

Straightforward analysis gives [CAL]

$$x'(t) = \begin{cases} 6|t|^3-6|t|^2+1 & 0 \le |t| \le 0.5 \\ 2(1-|t|)^3 & 0.5 \le |t| \le 1 \\ 0 & \text{otherwise} \end{cases} . \qquad (4.3)$$

The coefficients $x'_k \triangleq x'(kT_c)$ are given simply by:

$$x'_k = \begin{cases} 6|kT_c|^3-6|kT_c|^2+1 & 0 \le |kT_c| \le 0.5 \\ 2(1-|kT_c|)^3 & 0.5 \le |kT_c| \le 1 \\ 0 & \text{otherwise} \end{cases} . \qquad (4.4)$$

The piecewise expression for x'_k makes a general expression for

$$x_k = (1/F_{dif}).\{-x'_{k-1} + 2.x'_k - x'_{k+1}\} \qquad (4.5)$$

very complicated. For a certain T_c though x(D) is easily determined.

From $F_{dif} = 2 - 2.x'_1$ (3.12a) we find

$$F_{dif} = \begin{cases} 12(T_c^2-T_c^3) & 0 \le T_c \le 0.5 \\ 2 - 4(1-T_c)^3 & 0.5 \le T_c \le 1 \\ 2 & T_c \ge 1 \end{cases} . \qquad (4.6)$$

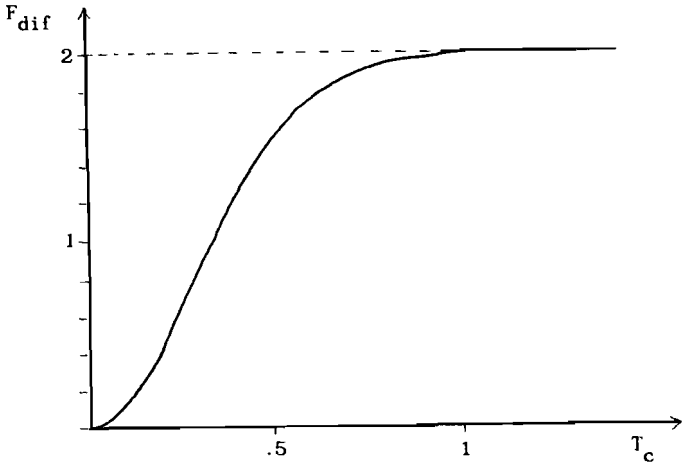


Fig. 4.2 F_{dif} as a function of T_c

The discrete time response is equal for all $T_c \geq 1$. From (4.4) we have $x'(D) = 1$ and with $F_{dif} = 2$

$$\begin{aligned} x(D) &= (1/F_{dif}).(-D^{-1} + 2 - D).x'(D) \\ &= -\tfrac{1}{2}D^{-1} + 1 - \tfrac{1}{2}D \\ &= \tfrac{1}{2}\sqrt{2}.(-D^{-1}+1).\tfrac{1}{2}\sqrt{2}.(1-D) \\ &= f(D^{-1}).f(D). \end{aligned} \qquad (4.7)$$

So for $T_c \geq 1$ $f(D) = \frac{1}{2}\sqrt{2}.(1-D)$ is a differentiation.

For $\frac{1}{2} \leq T_c \leq 1$ $x'(D)$ has three nonzero coefficients:

$$x'_0 = 1$$

$$x'_1 = x'_{-1} = 2(1-T_c)^3.$$

With $F_{dif} = 2-4(1-T_c)^3$ we have for the nonzero coefficients of $x(D)$:

$$x_0 = 1$$

$$x_1 = x_{-1} = \{4(1-T_c)^3-1\}/\{2-4(1-T_c)^3\}$$

$$x_2 = x_{-2} = \{-2(1-T_c)^3\}/\{2-4(1-T_c)^3\}.$$

For other T_c similar expressions can be derived.

The received energy per channelbit E_c is given by

$$E_c = E_b.E_h.F_{dif} = E_b.F_{dif} \qquad (4.8)$$

and the expression for F_{dif} (4.6).

4.2 Dependency of δ^2_{min} on T_c

In the uncoded case the minimum sqared Euclidian distance is given by

$$\delta^2_{min}(T_c) = E_b.F_{dif}(T_c).d^2_{min}(T_c). \qquad (4.9)$$

For $T_c \geq 1$ the δ^2_{min} is easily determined. $F_{dif} = 2$. Because $f(D)$ has length 2 and $\Sigma f_i^2 = 1$ the $d^2_{min} = 1$ [PRO83, p.405]. The set $E_{\delta min}$ of

errorsequences with minimum distance is given by

$$E_{\delta min} = \{ \pm 1,\ \pm(1+D),\ \pm(1+D+D^2),\ \ldots \}. \qquad (4.10)$$

$T_c = 0.5$; $f(D) = 0.7887 - 0.5774D - 0.2113D^2$

λ	e	N_e	$d^2(e)$	equivalent sequences E_d
0	1	2	1.0001	-1
1	1 1	2	1.3334	-1-1
	1-1	2	2.6670	-1 1
2	1 1 1	2	1.3334	-1-1-1
	1 0 1	2	1.6669	-1 0-1
	1 0-1	2	2.3335	-1 0 1
	1 1-1	4	3.3336	-1-1 1, -1 1 1, 1-1-1
	1-1 1	2	4.0005	-1 1-1

Table 4.1 Distance profile for $T_c = \frac{1}{2}$

$T_c = 0.3333$; $f(D) = 0.7953 - 0.0099D - 0.5901D^2 - 0.1049D^3$

λ	e	N_e	$d^2(e)$	equivalent sequences E_d
0	1	2	1.0000	-1
1	1-1	2	1.9167	-1 1
	1 1	2	2.0832	-1-1
2	1 0 1	2	1.0833	-1 0-1
	1-1 1	2	1.9167	-1 1-1
	1 1 1	2	2.2499	-1-1-1
	1 0-1	2	2.9166	-1 0 1
	1 1-1	4	3.9165	-1-1 1, -1 1 1, 1-1-1

Table 4.2 Distance profile for $T_c = \frac{1}{3}$

For $T_c = \frac{1}{2}$ and $T_c = \frac{1}{3}$ distance profiles are computed (tables 4.1 and 4.2). From these profiles it can be concluded that for $\frac{1}{3} \leq T_c < 1$

errorperformance is dominated by single errors e(D) = $\pm$1. For smaller T_c initially e(D) = $\pm(1+D^2)$ will dominate the rror performance. These calculations are not executed here. We can conclude that

$$d^2_{min}(T_c \geq \frac{1}{3}) = 1. \tag{4.11}$$

For δ^2_{min} we thus have the results

$$\delta^2_{min}(T_c \geq 1) = F_{dif}(T_c \geq 1) = 2. \tag{4.12a}$$

$$\delta^2_{min}(\frac{1}{2} \leq T_c < 1) = F_{dif}(\frac{1}{2} \leq T_c < 1) = 2-4(1-T_c)^3. \tag{4.12b}$$

4.3 Comparison with the Lorentzian stepresponse model

How can the triangular and Lorentzian stepresponse model be compared, when their definitions are so different? We have already noted that the triangular model is only valid for analysis at constant v, because of the absence of v in E_h. The models are primarily used to determine the effects of coding. The tape speed v is kept constant in the comparison of uncoded and coded transmission (§3.2.1), so one can compare the effects of coding for both models.

For the Lorentzian stepresponse we had for the energyloss due to coding with rate R (3.22):

$$L_E(S,R) \triangleq \frac{F_{dif}(S_p = \frac{S}{R})}{F_{dif}(S_p = S)} = \frac{S^2+1}{(S/R)^2+1}. \tag{4.13}$$

In the triangular stepresponse model we have

$$L_E(T_s,R) \triangleq \frac{F_{dif}(T_c = T_s.R)}{F_{dif}(T_c = T_s)}. \tag{4.14}$$

One usually takes T_s=1 for the uncoded case, as this is the smallest T_s for which the channel is a pure differentiation and the energy

$E_c=2.E_b$. So we get:

$$L_E(T_s=1,R) \triangleq \frac{F_{dif}(T_c=R)}{F_{dif}(T_c=1)} = \begin{cases} 6(R^2-R^3) & 0 \le R \le 0.5 \\ 1-2(1-R)^3 & 0.5 \le R \le 1 \end{cases} . \quad (4.15)$$

Comparison of the two models learns that for S≃0.3 the energyloss of the Lorentzian model is similar to the triangular model (fig. 4.3). For higher S the penalty on coding is less for the triangular model.

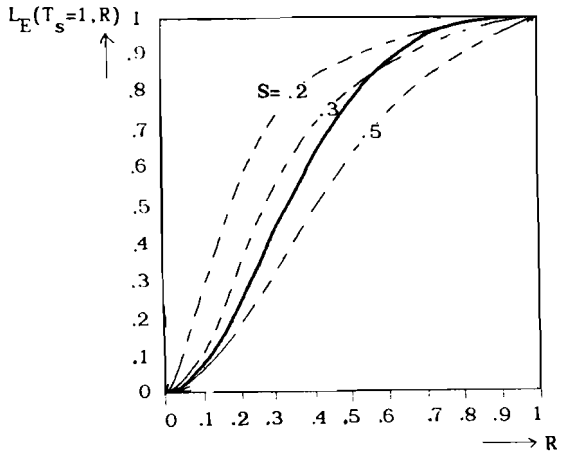


Fig. 4.3 Comparison of L_E for Lorentzian and triangular stepresponse

Comparison of the d^2_{min} for the two channels can give more insight in the similarity of the models. For the Lorentzian model $d^2_{min}=1$ for $0 \le S_p \le 2.4$. So at S=0.3 for codes with rates $\frac{1}{8} \le R \le 1$. The triangular model has $d^2_{min}=1$ for $T_c \ge \frac{1}{3}$. So at $T_s=1$ for codes with rates $\frac{1}{3} \le R \le 1$. This means that for rates $\frac{1}{3} \le R \le 1$ the two models have equal d^2_{min}. For lower rates the two models differ too much.

From this one can conclude that only for normalized information densities S≃0.3 and medium to high coding rates the triangular model

can possibly be used as an approximation for the Lorentzian model. A direct comparison of the responses (fig 4.4) shows that is true. The restricted validity of the triangular model (premising that the Lorentzian model is realistic) means that it is useless for high density magnetic recording.

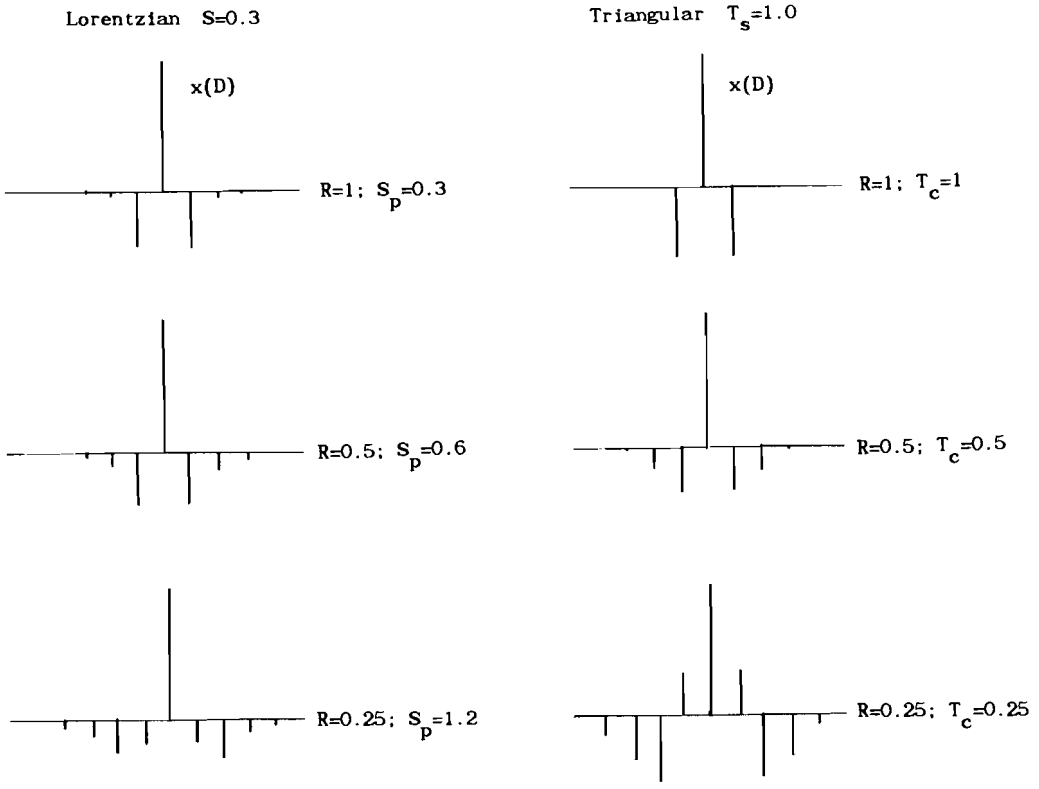


Fig. 4.4 Comparison of Lorentzian and triangular responses

5 REMARKS AND CONCLUSIONS

We analyzed the magnetic recording channel for the maximum likelihood sequence estimation method. Two approaches for the recording channel model gave rise to two different receivers, and different insights into the channel. Both approaches resulted in the same system model, and can therefore be used interchangeably. The differentiation factor F_{dif} that represents the effects of differentiation and dispersion on the received energy per channelbit appears to be a useful quantity.

A transition from the users' view on the magnetic recording system to manageable design problems was made. In the case of the Lorentzian stepresponse model it appeared to be best to design codes for various physical densities S_p and translate the results back to the for the user relevant normalized information density S.

The triangular stepresponse model can be used as an approximation of the Lorentzian stepresponse only for low ($S \simeq 0.3$) normalized information densities and medium to high code rates ($\frac{1}{2} \leq R \leq 1$). This restricted validity makes it useless as a model for high density magnetic recording ($S \geq 2$).

It must be stressed here that all analysis was done for optimal (MLSE) detection. Complexity and implementability of receivers and detection methods were not taken into account. Certainly for higher physical densities the response length becomes so long that even the detector for uncoded transmission becomes impractically complicated. The design of simpler practical detectors and codes for these detectors are

research topics. To gain insight into the design of codes one can initially design codes for optimal detectors.

REFERENCES

[CAL] A.R. Calderbank, L.H. Ozarow, C. Heegard, "A New Approach to High Density Magnetic Recording"
Bell Laboratories, Murray Hill, New Jersey.

[FOR72] G.D. Forney, Jr. "Maximum Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference"
I.E.E.E. Transactions on Information theory, Vol. IT-18, No 3, may 1972, pp. 363-378

[FOR72*] G.D. Forney, Jr. "Lower bounds on Error Probability in the Presence of Large Intersymbol Interference"
I.E.E.E Transactions on Communications, Vol COM20 , feb. 1972, pp76-77

[IMM87] K.A. Schouhamer Immink, "Coding Techniques for the Noisy Magnetic Recording Channel: A State of the Art Report"
Philips Research Laboratories, Eindhoven, june 1987

[KER87] G. Kerpen
Graduate report. Eindhoven University of Technology-
Philips Research Laboratories To appear.

[KOB70] H. Kobayashi, "Coding Schemes for Reduction of Intersymbol Interference in Data Transmission Systems"
IBM J. Res. Develop., Vol. 14, july 1970, pp 434-353.

[KOB71] H. Kobayashi, "A Survey of Coding Schemes for Transmission or Recording of Digital Data"
I.E.E.E. Trans. Comm., Vol. COM-19, No.6, Dec.1971, pp 1087-2100.

[PRO83] J.G. Proakis
Digital Communications
Mc.Graw-Hill, New York, 1983

[VIT79] A.J. Viterbi, J.K. Omura,
Principles of Digital Communications and Coding
Mc.Graw-Hill, New York, 1979

[WOL86] J.K. Wolf, G. Ungerboeck, "Trellis Coding for Partial-Response Channels"
I.E.E.E. Trans. Comm., Vol. COM-34, No.8, Aug. 1986, pp. 765-773