

MASTER

Spectrum shaping with Markov chains

de With, P.H.N.

Award date:
1984

[Link to publication](#)

Disclaimer

This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain

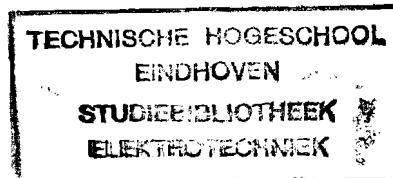
M. G. van

NAT.LAB.

REPORT No. 5926

P.H.N. de With

SPECTRUM SHAPING
WITH MARKOV CHAINS



M.S.-thesis.

Research supported by:

ir. K.A. Schouhamer Immink,

Philips Research Laboratories Eindhoven;

dr.ir. A.J. Vinck,

University of Technology Eindhoven;

and supervised by:

Prof.dr.ir. J.P.M. Schalkwijk,

University of Technology Eindhoven.

300013

ABSTRACT

The power spectrum of a code can be shaped in a preferable form by adding redundant information. The obtained spectrum properties are used to improve the performance of a communication system. In this Report we consider performance criteria for dc-free codes. Maximum entropies of binary sequences subject to a particular spectral constraint are presented. It is shown that codes with nearly optimum rate can have considerable differences in dc content. In order to improve the code spectrum, sequences are introduced with a bound on the running digital sum (RDS) and on the sum of the RDS. It is shown that, besides a dc-free power spectrum, these sequences have a zero second derivative at zero frequency, resulting in a substantial decrease of dc content.

Period of work: May 1983 - April 1984.

Notebook(s): none.

CONTENTS

1.0	<i>Introduction</i>	3
1.1	Scope.	3
1.2	Application of modulation systems.	4
1.3	Statement of the problem.	5
2.0	<i>Markov source</i>	7
2.1	Preliminary and restrictions.	7
2.2	Source model.	8
2.3	Asymptotic behaviour of Q.	12
2.4	First-order power spectrum.	13
2.5	Maxentropic case.	16
3.0	<i>Power Spectrum</i>	18
3.1	Correlation function.	18
3.2	Spectrum.	19
3.3	Correlation function for even N.	22
4.0	<i>Measure of dc content</i>	25
4.1	Cut-off frequency.	25
4.2	Second derivative.	27
4.3	Low-frequency power.	29
4.4	Mean error rate.	31
4.4.1	Channel model.	31
4.4.2	Error rate.	34
5.0	<i>Maximum entropy</i>	39
5.1	Optimization problem.	39
5.2	Sources with a weight constraint.	40
5.3	Analogy to rate distortion theory.	43
5.4	Source with a variance constraint.	45
5.5	Source with an error rate constraint.	46

5.6 Results.	48
6.0 <i>Zero second derivative</i>	53
6.1 Additional constraint.	53
6.2 Source model.	54
6.3 Computational results.	57
7.0 <i>Conclusions</i>	61
Acknowledgments	62
References	63

1.0 INTRODUCTION

1.1 Scope.

In the majority of practical communication systems, data is not transmitted in its original form, but redundant information is added by coding.

In most cases, coding implies two steps in information processing. The first step is widely known as source coding: the transformation of the original message into a string of binary digits. An important aim here is to translate the message in the most efficient way or with an allowable degree of distortion. The second step is called channel coding and establishes a 'matching' of the binary stream to the transmission channel. In this case, the transmitted data should satisfy requirements such as a power spectrum that is well-fitted to the transfer function of the channel and/or a protection against transmission errors.

This report is primarily related to binary spectrum shaping codes.

A special class of spectrum shaping codes is known as dc-free codes, which are codes suitable for communication channels that have a poor response in the low-frequency range. The spectrum of a code can be made zero at other frequencies using interleaving techniques [1] and therefore dc-free codes are an interesting class to study. An introduction on dc-free codes may be the survey by Kobayashi [2].

A binary spectrum shaping code maps sequences onto other sequences; the mapping transfers power in the frequency domain to a prescribed range. Because of this 'modulating' property, the code mapping is called a modulation system.

Digital modulation in a transmission system is done in the transmitter after the source message is encoded and usually a protection against errors is coincided. In the receiver, demodulation is the first action that must be

undertaken before error correction and transformation to the original message can take place.

1.2 Application of modulation systems.

Usually, the power spectral density function of a code is made zero at a given frequency and small around that frequency. The code spectrum generally has its minimum at zero frequency and the Nyquist bandwidth (half of the channel symbol rate). This implies that the coded signal will pass through transformers and a carrier or timing wave can be supplemented for receiver synchronization.

We require that a spectrum shaping code has:

- a favourable power spectrum.

The code spectrum must be dc-free and almost zero in a substantial range around zero frequency.

- an acceptable information rate.

Obviously, this property is complementary with the demand for a favourable power spectrum. The shaping is obtained by adding redundant information resulting in a reduced information rate. The exchange between 'bandwidth' and information rate is an important relation which will appear many times in this report.

Application of codes with the above properties are found in various fields.

Some codes are referred to as line codes [3,4], and they are often used in telephone systems, where a dc-free spectrum is a necessity. Ternary codes are often used in cable systems whereas in optical fibre communication only binary mappings are applied [5].

Another class is known as run-length-limited codes [6,7]. In this case the number of consecutive zeros between two ones lies between a minimum and a maximum value. The minimum value ensures that transitions are sufficiently spaced, thus reducing the intersymbol interference. The maximum value

guarantees that enough transitions are made to allow the extraction of a timing reference at the receiver.

In optical disc recording systems, a power spectrum with a low dc content is less vulnerable to high-pass filtering. High-pass filtering is necessary to avoid wander of the decision level when dirty discs are read out. The interference with servo signals, which are of course relatively slowly varying signals with considerable dc content, moreover is decreased [8].

Besides the application in optical recording, the modulating property is also used in magnetic recording [9] where it is difficult to read very low-frequency or dc waveforms with sufficiently signal-to-noise ratio.

1.3 Statement of the problem.

The reader may have noticed that the examples in the preceding section describe communication systems with channels that have a poor response in the low-frequency range. In the following, such a channel will be called an ac-coupled channel; a model is depicted in Fig. 1.1.

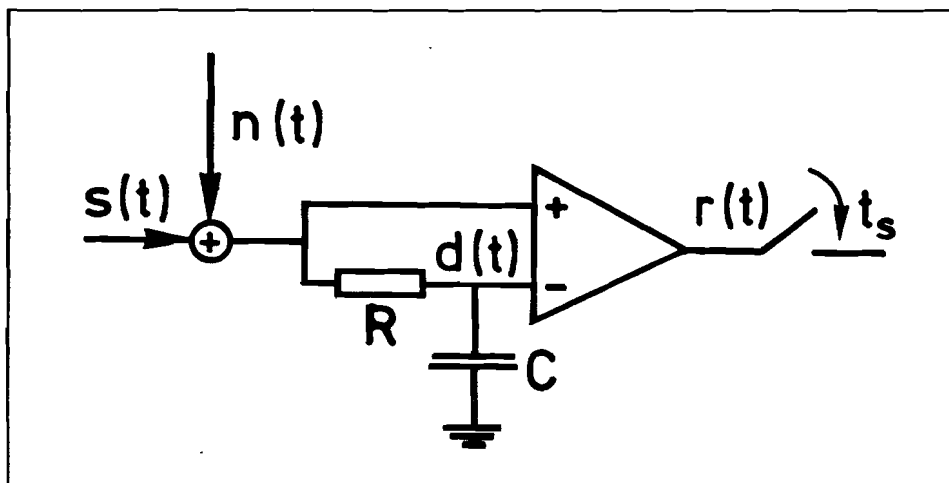


Fig. 1.1. An ac-coupled channel model.

Let us assume $s(t)$ to be a stream of rectangular pulses with amplitude $\pm A$, and consider the case where a long string

of ones with amplitude $+A$ is transmitted. The capacitor C will be charged and the decision level of the voltage detector is raised thus increasing the chance of misdetections.

The above example illustrates why long sequences of the same polarity are not desirable and implicates the following more general question: which statistical properties must be satisfied by a transmitted sequence to provide a favourable power spectrum and what information rate can be reached in this case?

The report is organized as follows. Chapter 2 describes the source model that is used to generate the channel symbol stream. Chapter 3 shows how the power spectral density function of a source can be calculated. Chapter 4 mentions some performance parameters in which the dc content of a code can be expressed. Asymptotic rates when an extra spectral constraint is set are derived in chapter 5. Chapter 6 describes an extension of the source model with improved spectral properties. Chapter 7 presents the conclusions.

2.0 MARKOV SOURCE

In this chapter we describe a binary Markov information source which creates the code output sequence. After the introduction of some important parameters, a formal description of the source is given. The transition probability matrix of the source generating a first-order power spectrum and that with maximum entropy are derived.

2.1 Preliminary and restrictions.

The source model is used to observe the statistics of the symbol stream that is transmitted over the channel. The design of dc-free codes in general is not discussed.

There is only a small amount of literature about the relation between the structure of a Markov source and the effects of this structure on the power spectral density function. Yasuda and Inose [10] prove that if a source model is 'loop-sum-zero', the resulting code has an essential dc-free power spectrum. Loop-sum-zero means that for any closed path in the state transition diagram of the encoder, the number of zeros equals the number of ones.

From practical applications, it is known that by balancing the number of ones and zeros in codewords, one can achieve a power spectrum that has real zeros at certain frequencies [1,11]. This agrees with the demand for a loop-sum-zero model proposed by Yasuda and Inose.

It is obvious that an important parameter of a dc-free code, will be the running digital sum (RDS) of the code sequence. In the sum, a '0' is weighted as -1; this can be made plausible by noticing that a '0' is often transmitted as the inverted waveform corresponding with a '1'. The RDS is specified by:

$$RDS = \sum_t X(t) .$$

It is assumed here that the source output is a sequence of dependent variables ..., X(t-T), X(t), X(t+T), ... , where $X(t) \in \{-1,+1\}$. The period T denotes the channel symbol duration. The dependency in the variables X(t) should yield a promising power spectrum, at the cost of a reduced information rate.

A second constraint is related with the composal of the Markov source. Because the RDS is a fundamental parameter for acquiring a favourable spectrum, sources that have one state for each value of the RDS are considered. In practice the number of states is finite; as a result the RDS is discrete-valued and bounded in a finite interval.

From the preceding it may be concluded that the magnitude of the RDS interval is another important parameter because the bound on the RDS accomplishes the absence of dc power. Consequently, it could be assumed that a less strong bound increases the dc content. As a second parameter we thus take the maximum variation of the RDS, i.e. the difference between its maximum and minimum value. This parameter is called the digital sum variation (DSV), which can be written as:

$$DSV = \max_{t_1, t_2} \left\{ \sum_{t_1}^{t_2} X(t) \right\} ,$$

where max { . } symbolizes a function that searches the maximum of all appearing values of the variation of the RDS. Thus, the DSV is the number of RDS states minus one.

2.2 Source model.

As already stated, this report deals with RDS-bounded sources with one state for each value of the RDS. The RDS changes with each transmitted symbol. So we can define, analogous to the code sequence X(t), a sequence of RDS

values ..., $Z(t-T)$, $Z(t)$, $Z(t+T)$, ... , where $Z(t)$ is uniquely related to $X(t)$ by:

$$Z(t) = Z(t-T) + X(t) , \quad Z(0) = a . \quad (2.1)$$

From the Markov property it follows that the probability

$$P(Z(t)|Z(t-T)Z(t-2T)Z(t-3T)\dots Z(0)) = P(Z(t)|Z(t-T)) .$$

If the code sequence $X(t)$ consists of symbols from the alphabet $\{+1,-1\}$, then the RDS takes values from the set $\{\dots, a+2, a+1, a, a-1, a-2, \dots\}$. In most cases, this set lies symmetrically around zero and consequently, a symmetric source has $a = 0$ for an odd number of states and $a = 1/2$ for an even number. The number of elements in the set is equal to the number of RDS states N . We thus define a row vector \underline{z} of permitted RDS values $\underline{z} = (z(1), z(2), \dots, z(i), \dots, z(N))$.

Furthermore, the source is defined by its transition probability matrix Q with the elements $p(j|i)$ as the probability of going from state i to state j .

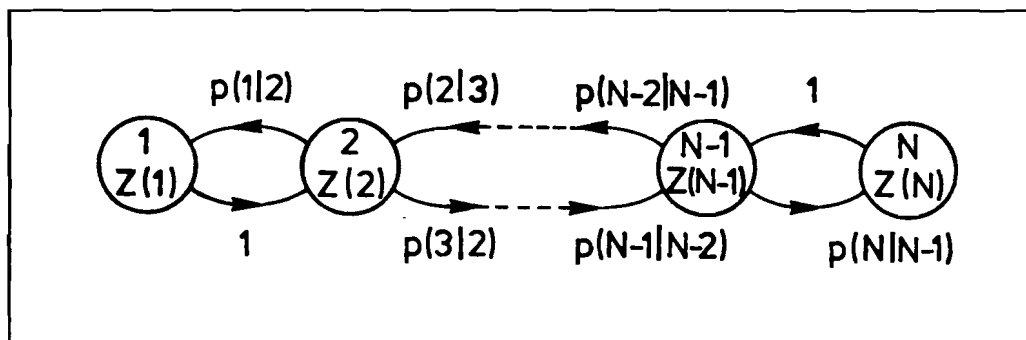


Fig. 2.1. RDS-bounded Markov source model.

Fig. 2.1 shows an N -state source; each state corresponds with a particular value of the RDS. The model starts in a given state i with RDS value $z(i)$ and after time T the actual state becomes $i+1$ with probability $p(i+1|i)$ and a '0' is transmitted. For a probability $p(i-1|i)$, a '1' is transmitted resulting in an RDS state $i-1$. So each period T

a transition is made as a function of the transition probabilities and the corresponding symbol is transmitted.

Before we mention some general formulas related to the Markov chain, we state some properties that can be noticed from Fig. 2.1.

- Irreducibility.

There is no other closed set of states than the set of all states [12].

- Periodicity.

It is impossible to return to a particular state except after h, 2h, 3h, ... transitions, with h = 2 in our case. The chain is cyclic [12] and can be divided in two cyclic classes: the odd and the even numbered states.

The adjacency matrix A of the source is the Toeplitz matrix and has the form:

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & \dots & \dots & 0 \\ 1 & 0 & 1 & 0 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & \dots & 0 \\ \cdot & & & & & & & \cdot \\ \cdot & & & & & & & \cdot \\ 0 & \dots & \dots & 0 & 0 & 1 & 0 & 1 \\ 0 & \dots & \dots & 0 & 0 & 1 & 0 & \end{bmatrix} .$$

The transition matrix Q is of the same form with the unity symbols replaced by the transition probabilities p(j|i). So the elements of Q are defined by Q(i,j) = A(i,j).p(j|i). As a result, the transition matrix Q has unit row sums.

When the transition probabilities are specified, the unique stationary probability distribution p_s(i) is given by the solution of:

$$\underline{p}_s \cdot Q = \underline{p}_s , \tag{2.2}$$

and:

$$\sum_{i=1}^N p_s(i) = 1 , \tag{2.3}$$

where \underline{p}_s stands for the row vector with elements $p_s(i)$. The entropy of the source is given by a sum of conditional entropies:

$$H(X) = \sum_{i=1}^N p_s(i) H(X|s=i) = - \sum_{i=1}^N p_s(i) \sum_{j=1}^N p(j|i) \log p(j|i) . \quad (2.4)$$

In the binary case, equation (2.4) is reduced to:

$$H(X) = \sum_{i=1}^N p_s(i) h(p(i+1|i)) , \quad (2.5)$$

where $h(p(i+1|i))$ stands for the binary entropy function with argument $p(i+1|i)$. As we shall see later, the variance of the sum (RDS) values σ_z^2 is of great importance, and is expressed by:

$$\sigma_z^2 = \sum_{i=1}^N p_s(i) (z(i))^2 . \quad (2.6)$$

Example: Consider the three-state source from Fig. 2.2 for $DSV = 2$ and RDS values $\underline{z} = (+1, 0, -1)$. This source has maximum entropy for $p(3|2) = p(1|2) = 1/2$; the stationary probabilities are $\underline{p}_s = (1/4, 1/2, 1/4)$ and the entropy $H(X) = 1/2$.

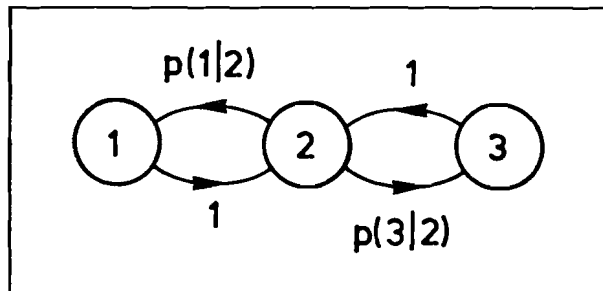


Fig. 2.2. $N=3$ source.

$$p_s(i) = \lim_{k \rightarrow \infty} (p(j|i)^{(k)} + p(j|i)^{(k+1)}) / 2, \quad (2.7)$$

for arbitrary j . The higher order transition probability $p(j|i)^{(k)}$ is defined by $p(j|i)^{(k)} = Q^k(i,j)$. Equation (2.7) can be used to calculate the stationary probabilities without the need to solve (2.2) and (2.3). The probability $p(j|i)^{(k)}$ can be written as [14]:

$$p(j|i)^{(k)} = p_s(j) + d(j|i)^{(k)} + e(j|i)^{(k)}, \quad (2.8)$$

where $d(j|i)^{(k)}$ jumps between $-p_s(j)$ and $+p_s(j)$ for consecutive values of k taking into account that the period of the source is two. The term $e(j|i)^{(k)}$ vanishes when k becomes larger because the moduli of most eigenvalues of Q are less than one.

According to the theorems of Perron-Frobenius [13], we conclude that:

- Q has two eigenvalues with modulus 1, $+1$ is related to $p_s(j)$ in (2.8), and -1 corresponding with $d(j|i)^{(k)}$;
- all other eigenvalues of Q have a modulus smaller than one and are real. They occur in pairs (period 2) with elements of opposite sign and together they form the term $e(j|i)^{(k)}$ in equation (2.8).

Because the eigenvalues of Q are real, higher order transition probabilities can be expressed as a summation of geometrical series.

2.4 First-order power spectrum.

In this section we derive the transition probabilities which generate a first-order power spectrum. The term first-order power spectrum is explained and some results from filter and prediction theory are used.

Upon encoding, the spectrum of a random stream of bits is shaped into a suitable form. It is therefore plausible to represent the encoding process by a linear time-invariant

of Justesen [16] are used. For rational power spectra the best linear predictor $\hat{X}(t)$ can easily be derived from the transfer function (2.9) and may be expressed as:

$$\hat{X}(t) = \hat{X}(t-T) - (1-r)X(t-T) . \quad (2.12)$$

Combining equation (2.12) with (2.1), yields:

$$\hat{X}(t) = -(1-r)Z(t-T) , \quad (2.13)$$

which states that the best linear predictor is proportional to the last RDS value. In other words, the predictor is dependent on the actual state of the Markov source. The construction of the source probabilities is therefore based on the assumption that the predictor coincides with the conditional expectation of the output when starting from the actual state:

$$\hat{X}(t) = E\{X(t)|s=i\} = \sum_{j=1}^N p(j|i)x(j|i) , \quad (2.14)$$

where $x(j|i) = z(j) - z(i)$. Equation (2.14) can be seen as a constraint on the transition probabilities. If it is possible to obtain a set of probabilities that satisfies (2.14), then the spectrum is a rational function, since the predictor has the proper form. A binary source has $|\hat{X}(t)| \leq 1$ and consequently $Z(t-T)$ is limited. From (2.12) follows that the outer states have $z(1) = 1/(1-r)$ and $z(N) = -1/(1-r)$. As a result we have a number of states N equal to $1 + 2/(1-r)$. Because the number of states N is discrete-valued, the parameter r can assume only certain values:

$$r = (N-3)/(N-1) , \quad N \geq 3 . \quad (2.15)$$

The vector \underline{z} is defined by:

$$z(i) = - (i-1) + 1/(1-r) , \quad i=1,2,\dots,N ,$$

and from (2.13) we have:

$$\hat{X}(t) = -1 + (1-r)(i-1) . \quad (2.16)$$

Rewriting (2.14) leads to a second expression for $\hat{X}(t)$:

$$\hat{X}(t) = (+1)p(i-1|i) + (-1)p(i+1|i) = 2p(i-1|i) - 1 . \quad (2.17)$$

The last step is valid because the sum of $p(i-1|i)$ and $p(i+1|i)$ is one. Combining (2.16) and (2.17) results in the solution for the transition probabilities:

$$p(i-1|i) = (i-1)(1-r)/2 = (i-1)/(N-1) , \quad (2.18)$$

for $i=1,2,\dots,N$. By solving (2.2) and (2.3) the stationary probabilities are obtained:

$$p_s(i) = 2^{-(N-1)} \binom{N-1}{i-1} , \quad (2.19)$$

and the entropy of the source can be expressed as:

$$H(X) = 2^{-(N-1)} \sum_{i=1}^N \binom{N-1}{i-1} h((i-1)/(N-1)) . \quad (2.20)$$

By taking a closer look at the calculated transition probabilities it is concluded that they are the same as the probabilities used in the Ehrenfest model of diffusion mentioned by Feller [12]. The chain with the derived distribution forces the transitions towards the centre of the source resulting in improved spectral properties.

2.5 Maxentropic case.

By assuming an RDS-bounded source, it is ensured that the power spectrum has no dc component. This section presents how the transition probabilities must be chosen to achieve maximum entropy when no further constraints are set.

3.0 POWER SPECTRUM

In this chapter attention is paid to the calculation of the power spectrum. In many cases, the power spectrum consists of a discrete and a continuous part. The continuous part is caused by a pure stochastic process, whereas the discrete part of the spectrum is caused by the deterministic component of the process.

It is shown how the correlation function and the power spectrum can be determined. For an even number of RDS states, the deterministic component of the output sequence is obtained and separated from the nondeterministic process.

3.1 Correlation function.

If the transition probabilities and the vector of possible RDS values are known, the power spectrum can be determined from the autocorrelation function $R_Z(k)$ of the sequence $Z(t)$, which is defined by:

$$R_Z(t, t+k) = E \{ Z(t)Z(t+k) \} = R_Z(k) , \quad (3.1)$$

where the function $E\{.\}$ stands for the expectation, and the time shift $k = k'T$. The relation between the power spectrum of the code sequence $X(t)$ and the correlation function of $Z(t)$ is described in the next section. The last step in (3.1) is valid in the ergodic case [18], where time averages may be replaced by ensemble averages and as a result, $R_Z(k)$ may be expressed as:

$$\begin{aligned} R_Z(k) &= \sum_{i=1}^N \sum_{j=1}^N z(i) z(j) p(i, j)^{(k)} = \\ &= \sum_{i=1}^N \sum_{j=1}^N z(i) z(j) p_s(i) p(j|i)^{(k)} = \underline{z}_p Q^k \underline{z}^T, \quad (3.2) \end{aligned}$$

It should be noticed that the same equality (3.4) holds for the correlation $R_Z(k)$ and spectrum $S_Z(f)$.

First, the spectrum of the sum sequence $Z(t)$ is calculated and then it is transformed to the spectrum of the code sequence $X(t)$, because it is easier to determine the correlation function of the sum sequence.

The combination of equations (3.2) and (3.4) -indexed with Z - results in the spectrum of the sum sequence. The transformation to the spectrum of the code sequence $X(t)$ can easily be derived by using the relation:

$$S_X(f) = 2 (1 - \cos 2\pi fT) S_Z(f) . \quad (3.5)$$

Equation (3.5) is derived by combining (2.11) with the z -transform of (2.1). Similarly, the correlation sequence $R_X(k)$ can be expressed as a function of $R_Z(k)$:

$$R_X(k) = \begin{cases} 2 R_Z(k) - R_Z(k-1) - R_Z(k+1) , & k \neq 0 ; \\ \sigma_x^2 = 1 , & k = 0 . \end{cases} \quad (3.6)$$

Summarizing, for calculation of the power spectrum it is necessary to calculate:

- powers of the transition matrix Q ;
- the stationary probabilities $p_s(i)$;
- the correlation function $R_Z(k)$ from (3.2);
- the power spectrum of $Z(t)$ with (3.4) and the transformation to the code spectrum (3.5).

Example: Consider the source for $N = 5$, with a centrosymmetric transition matrix. In this case all probabilities are determined except $p(3|2)$ which we shall assign the value $q = 1-p$. The RDS vector $\underline{z} = (+2, +1, 0, -1, -2)$. The characteristic equation of the matrix Q is:

$$(\lambda^2 - 1)(\lambda^2 - p) = 0 ,$$

and we conclude that $R_Z(k)$ vanishes with powers of p . According to (3.2) the correlation function of $Z(t)$ equals:

$$R_z(2k-1) = 2 p^k, \quad k \geq 1 ;$$

$$R_z(2k) = [2p + 1/2] p^k, \quad k \geq 0 .$$

After a straightforward but tedious calculation, where equations (3.4) and (3.6) are used, the spectrum is expressed by:

$$S_x(f) = 2q (1 - \cos 2\pi fT) \frac{(2p+1/2)(2-q) + 4p \cos 2\pi fT}{1+p^2 - 2p \cos 4\pi fT} .$$

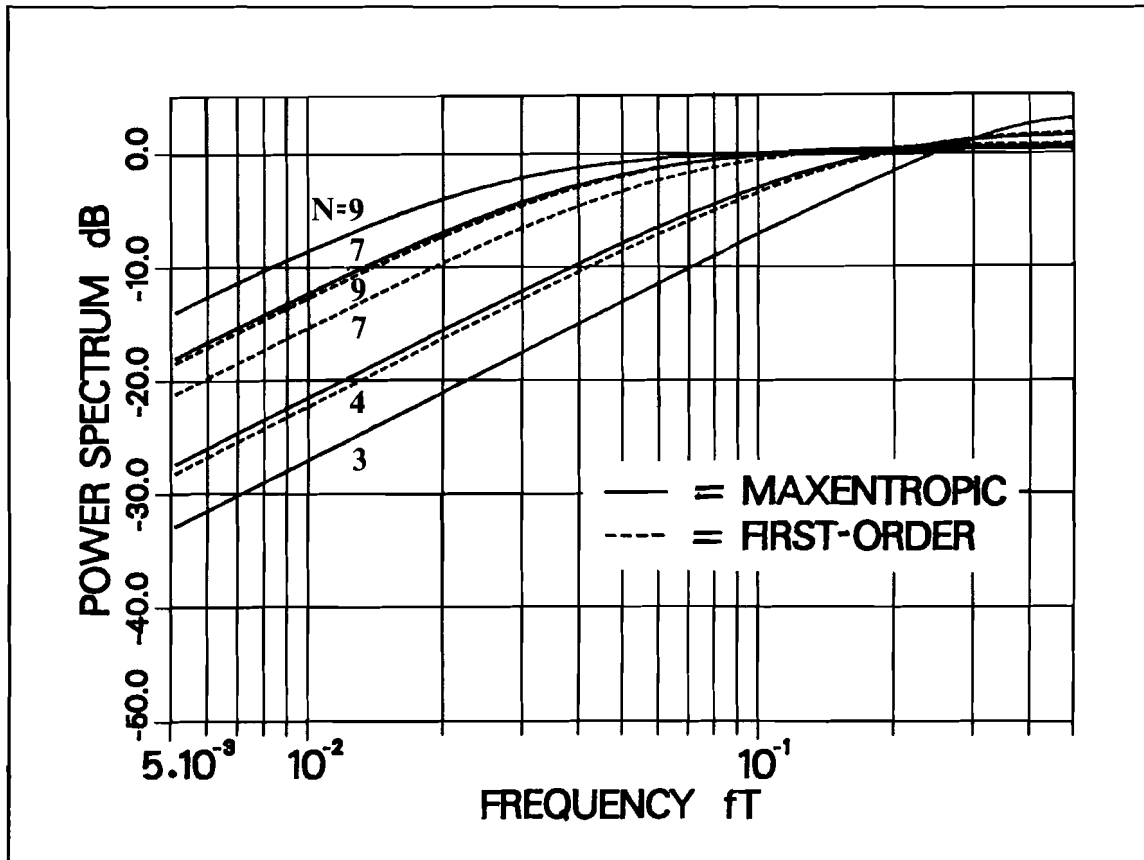


Fig. 3.1. Power spectra for maxentropic and first-order sources.

In Fig. 3.1 various power spectral density functions are plotted as a function of the relative frequency fT . It can be noticed that first-order sources generate sequences with less power in the low-frequency range than maxentropic sources.

3.3 Correlation function for even N.

Special attention must be paid to the case where the source has an even number of states. As already mentioned, the source consists of two cyclic classes: the odd and even numbered states. All states together have $E\{Z\} = 0$, but a source with an even number of states has two cyclic classes with each nonzero expectation for Z, and consequently $R_Z(k)$ doesn't vanish for large k.

As a matter of fact, $R_Z(k)$ is an alternating function for large k and has the following form:

$$\lim_{k \rightarrow \infty} R_Z(k) = (-1)^k C, \quad C \neq 0, \quad (3.7)$$

where C is a constant and as a result, the power spectrum has a line at $fT = 1/2$.

First, a general expression for the value of the constant C is derived. Secondly, it is proved that the 'first-order' transition probabilities from section 2.4, applied in a source with an even number of states, doesn't generate a line in the power spectral density function and thus C becomes equal to zero.

We assume a symmetric source and N to be even:

$$p_s(j) = p_s(N-j), \quad 1 \leq j \leq N/2, \quad (3.8)$$

$$\begin{cases} z(j) = (N+1-2j)/2, \\ z(N-j) = -z(j). \end{cases} \quad 1 \leq j \leq N/2, \quad (3.9)$$

It follows that:

$$\begin{aligned} E\{Z(2t)\} &= \sum_{\substack{j=2 \\ j \text{ even}}}^N z(j)p_s(j) = \\ &= \sum_{\substack{j \text{ even} \\ j=2}}^{N/2} z(j)p_s(j) + \sum_{\substack{j \text{ even} \\ N/2+1}}^N z(j)p_s(j). \end{aligned} \quad (3.10)$$

use the fact that $p_s(j) = p_s(N-j)$. It can be verified that the solution agrees with the first-order probability distribution from section 2.4.

For larger N and N even, it can be proved that a first-order distribution forces C to be zero. Let $p_s(j)$ be given by (2.19) :

$$p_s(j) = 2^{-(N-1)} \binom{N-1}{j-1},$$

and C by (3.12). If we substitute the $p_s(j)$ in the expression for C , it can be proved that:

$$\sum_{j=1}^{N/2} (-1)^j (N+1-2j) \binom{N-1}{j-1} = 0. \quad (3.13)$$

From [20] we have the following results:

$$\sum_{j=0}^m (-1)^j \binom{n}{j} = (-1)^m \binom{n-1}{m}, \quad (3.14)$$

$$\sum_{j=1}^n (-1)^{j+1} j \binom{n}{j} = 0. \quad (3.15)$$

Substitution of (3.14) in (3.13) yields:

$$(N-1)(-1)^{N/2} \binom{N-2}{N/2-1} + 2 \sum_{j=1}^{N/2-1} (-1)^j j \binom{N-1}{j} = 0. \quad (3.16)$$

It can be verified that using (3.15), the second term in (3.16) may be expressed as:

$$-(-1)^{N/2} N/2 \binom{N-1}{N/2}.$$

Substitution of this expression in (3.16) indeed makes certain that $C = 0$. However, it is reasonable to expect that more solutions will exist that generate no lines in the power spectral density function.

4.0 MEASURE OF DC CONTENT

Some performance parameters in which the dc content of a code can be assessed are mentioned in this chapter. In our case dc content stands for the low-frequency power around zero frequency.

From literature the cut-off frequency and the second derivative are known as important parameters. We introduce a frequency related to the low-frequency power, and a mean error rate as other relevant parameters.

To simplify formulas in this chapter, the rotational frequency ω is often used instead of f .

4.1 Cut-off frequency.

The most common performance parameter used in the literature [11,16] is the cut-off frequency. This is the frequency where the spectrum $S_X(f) = \sigma_x^2/2$. An approximation for the relative cut-off frequency f_0T is derived, assuming first-order transition probabilities. Afterwards, the validity is shown by comparing it with the frequency where the actual spectrum has value $\sigma_x^2/2$. Moreover, it is shown that the approximation formula agrees very well with the results if we use the transition probabilities from the maxentropic case.

According to the above statements, the cut-off frequency is derived from:

$$S_X(f_0) = \sigma_x^2/2 , \tag{4.1}$$

together with equation (2.10) it can be concluded that:

$$(1-\cos \omega_0) = (1-r)^2/2 , \tag{4.2}$$

where $\omega_0 = 2\pi f_0T$. For small values of $(1-r)$ we may use the approximation:

$$\omega_0 \approx 1-r . \tag{4.3}$$

From [16] we have the expression for the sum variance:

$$\sigma_z^2 = \sigma_x^2 / (2(1-r)) = \sigma_x^2 (N-1) / 4 . \tag{4.4}$$

Comparison of equations (4.3) and (4.4) results in the important relation between the cut-off frequency and the sum variance:

$$\omega_0^2 \approx \sigma_x^2 / (2 \sigma_z^2) . \tag{4.5}$$

In Table 4.1 we have compared the approximation formula (4.5) with the actual values of f_0T where $S_x(f_0) = \sigma_x^2/2$, both for first order and maxentropic probabilities.

N	first-order			maxentropic		
	σ_z^2	approx.	exact	σ_z^2	approx.	exact
3	0.50	.15915	.16667	0.5000	.15915	.16667
4	0.75	.10610	.10817	0.8028	.09913	.10065
5	1.00	.07958	.08043	1.1667	.06821	.06879
6	1.25	.06366	.06409	1.5940	.04992	.05020
7	1.50	.05305	.05330	2.0858	.03815	.03832
8	1.75	.04547	.04563	2.6424	.03012	.03024
9	2.00	.03979	.03989	3.2639	.02438	.02448

Table 4.1. Comparison of sum variances and cut-off frequencies (f_0T).

- The following conclusions can be drawn from Table 4.1:
- The approximation for f_0T shows a very good agreement with the actual values of f_0T (within 1% for $N > 5$). The accuracy of (4.5) increases with N because the assumption (4.3) is more accurate for large N (see equation (2.15)).
 - The relation between f_0T and σ_z^2 is also useful for

where the last step is valid if we define a second derivative frequency as $(\omega_0')^2 = 1/S^{(2)}(0)$. However, this 'cut-off' frequency is less valuable than the approximated cut-off frequency (4.5) from the preceding section.

The second derivative at zero frequency can easily be calculated with:

$$S_X^{(2)}(0) = 2 S_Z(0) , \quad (4.9)$$

which is derived by double differentiation of (3.5) and substitution of $f = 0$. Implementation of equation (4.9) is easier than (4.8) and less accuracy is needed.

Example: The power spectrum of the sum sequence $Z(t)$ with first-order probabilities can be derived by combining (3.5) and (2.10):

$$S_Z(f) = \frac{\sigma_x^2 (1+r)}{2(1+r^2 - 2r \cos 2\pi fT)} . \quad (4.10)$$

As a result, the second derivative is calculated with (4.9) and simply becomes:

$$S_X^{(2)}(0) = \frac{(1+r)}{(1-r)^2} = \frac{(N-1)(N-2)}{2} , \quad (4.11)$$

where the relation (2.15) between the parameter r and the number of RDS states N from section 2.4 is used. Eventually, by using (4.11), we find an expression for $S_X(\omega)$ for low frequencies:

$$S_X(\omega) \approx [(N-1)(N-2)/4] \omega^2 . \quad (4.12)$$

In Table 4.2 the numerical values of the second derivatives and the corresponding frequencies (equation (4.8)) are presented.

From Table 4.2 we note that the second derivative grows rapidly with increasing N . Furthermore, in the maxentropic case the second derivative reaches higher values than when

N	first-order		maxentropic	
	$S^{(2)}(0)$	$f_0'T$	$S^{(2)}(0)$	$f_0'T$
3	1.00	0.1592	1.00	0.1592
4	3.00	0.0919	3.58	0.0841
5	6.00	0.0650	8.67	0.0541
6	10.00	0.0503	17.42	0.0381
7	15.00	0.0411	31.21	0.0285
8	21.00	0.0347	51.62	0.0222
9	28.00	0.0301	80.43	0.0178

Table 4.2. Second derivatives and frequencies.

using first-order probabilities. Both statements agree with the conclusions from the preceding section.

4.3 Low-frequency power.

A third criterion that possibly indicates the dc content of a code may be the actual power present in a prescribed low-frequency range.

The frequency ω_p , where $(-\omega_p; +\omega_p)$ specifies the area that contains a certain percentage of the total power, is a measure that is closely related to the power in the low-frequency range. The power can be written as:

$$P = \int_{-\omega_p}^{+\omega_p} S(\omega) d\omega \quad (4.13)$$

Numerical calculations show that 0.1% or less of the power is a sensible measure. In this case the calculated frequencies are of the same order of magnitude as the cut-off frequencies from section 4.1. We have taken 0.01% of

4.4 Mean error rate.

4.4.1 Channel model.

In this section a new parameter is considered providing a favourable power spectrum when dealing with an RDS-bounded Markov source. The foregoing parameters are all closely related to the power spectrum of the code sequence. The mean error rate, however, does not show this close relationship. The mean error rate is primarily related to restrictions on the output sequences, especially as long sequences of consecutive zeros or ones are not desirable. If these restrictions are taken into account, then in turn they guarantee a power spectrum with small power in the low-frequency range.

Before deriving an expression for the mean error rate, the ac-coupled channel model from section 1.3 is introduced. A channel may be the ordinary telephone line or the read-out system of optically stored information.

Our aim is to transmit sequences over the channel with a minimum of transmission errors. The signal $s(t)$ is assumed to be a binary sequence of rectangular pulses, each of duration T and amplitude $\pm A$; a typical sequence is shown in Fig. 4.1a. The noise $n(t)$ is independent of the signal $s(t)$ with a zero-mean Gaussian probability density function, thus:

$$p(n) = 1 / (\sigma \sqrt{2\pi}) \exp(-n^2 / 2 \sigma^2) . \quad (4.15)$$

A third assumption concerns the ratio T/RC . In practical applications the ratio $T/RC \ll 1$. This assumption has two consequences. First, the capacitor C averages the noise. Therefore the voltage $d(t)$ on the inverting input of the amplifier is almost noiseless. Secondly, the voltage $d(t)$, which is the decision level of the system, is dependent on

For example, $d(i+1|i)$ means that we make the transition to state $i+1$, which has a lower RDS value (in our case) corresponding to output zero. The factor $1/2$ is introduced because samples are taken at $t_s = (k+1/2)T$, which is normally the case when amplitude detection is chosen at the centre of the pulse. From (4.16) and (4.17) we conclude that the decision level not only depends on the actual encoder state, but also on the transition.

Let us now proceed with the consequences of the base-line wander. This phenomenon decreases the distance between the received pulse height $r(t)$ and the decision level $d(t)$ and therefore increases the chance of misdetections. An example of decreasing distance is shown in Fig. 4.1c. After a long sequence of zeros or ones the amplitude of the received signal is diminished considerably and therefore sequences with sufficient transitions are desired.

From the communication point of view we want the RC-product to be as large as possible so that a small base-line wander is guaranteed. This also corresponds to the spectral properties of the ac-coupled channel. Whereas the decision level is the low-pass filtered version of the incoming signal, the channel as a whole acts as a high-pass filter with a cut-off frequency proportional to $1/RC$. A large RC-product results in less power loss in the low-frequency range.

It is also possible that the noise is frequency dependent, e.g. a channel with large noise power near dc. In this case ac-coupling is recommendable and we have to seek an optimum between noise influence and signal power loss. In the latter case ac-coupling can be applied although it must be emphasized that in general this is not possible. In practice it is the channel that is given and by coding we try to use the channel in the most reliable way.

4.4.2 Error rate.

We now calculate the mean error rate. First, we define the state oriented error rate, and secondly, we obtain the mean error rate by averaging over all possible states.

The received signal density functions are plotted in Fig. 4.2. For the sake of clearness the Gaussian functions are drawn on different heights to emphasize that the 'tails' determining the error probability are of different length.

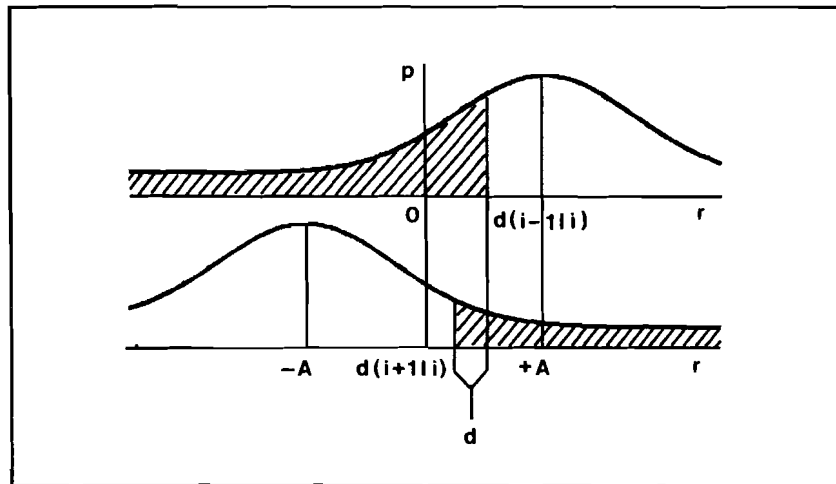


Fig. 4.2. Received signal density functions.

This difference is caused by the fact that the decision level depends on the actual state and on the transition. This situation differs from the usual one where the decision level is fixed or shifted by a fixed amount [22].

Let us now define the conditional error probability $P(E|i, '1')$ which stands for the probability of detecting a '0' when leaving state i after a '1' is transmitted. From Fig. 4.2 and (4.15) we immediately note that:

$$P(E|i, '1') = \int_{-\infty}^{d(i-1|i)} \frac{1}{\sigma \sqrt{2\pi}} \exp\left[-\frac{(n-A)^2}{2 \sigma^2} \right] dn . \quad (4.18)$$

Substitution of $x=(n-A)/\sigma$ in (4.18) yields:

$$P(E|i, '1') = Q [(A-d(i-1|i))/\sigma] , \quad (4.19)$$

where we have used the Q function, which is defined as:

$$Q(x) = \frac{1}{\sqrt{(2\pi)}} \int_x^{\infty} \exp(-a^2/2) da . \quad (4.20)$$

Similarly, the probability of detecting a '1' if a '0' is transmitted when leaving state i is given by:

$$P(E|i, '0') = Q [(A+d(i+1|i))/\sigma] . \quad (4.21)$$

Now the state oriented bit error rate BER(i) is defined by combining (4.19) and (4.21):

$$\begin{aligned} BER(i) &= p(i, '0') P(E|i, '0') + p(i, '1') P(E|i, '1') = \\ & p(i+1|i) Q [\frac{A+d(i+1|i)}{\sigma}] + p(i-1|i) Q [\frac{A-d(i-1|i)}{\sigma}] . \end{aligned} \quad (4.22)$$

This is obviously the sum of two Q functions, each weighted with the corresponding transition probabilities. Finally, the mean error rate E{BER} is obtained by averaging the BER(i)'s over all possible states, so that:

$$E\{BER\} = \sum_{i=1}^N BER(i) p_s(i) , \quad (4.23)$$

where we have used the stationary probabilities $p_s(i)$.

Example: Consider the 5-state source depicted in Fig. 4.3. Because of the symmetry the only parameter that can be varied is $p(3|2) = q$.

The computation of the mean error rate E{BER} is rather straightforward. The vector of RDS values is $\underline{z} = (+2, +1, 0, -1, -2)$, and the DSV = 4. According to (4.22) we obtain:

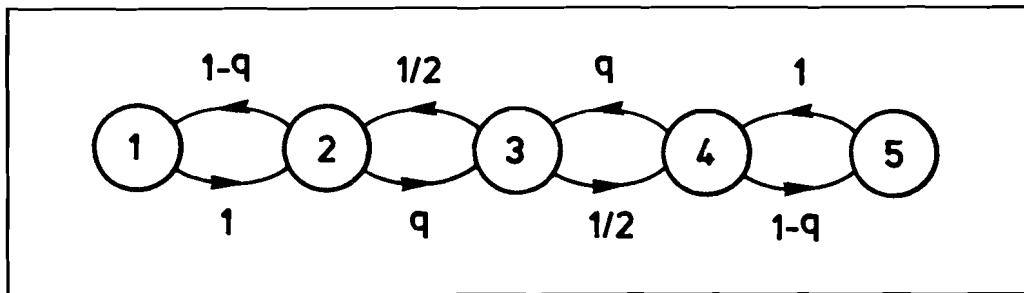


Fig. 4.3. Source with 5 states.

$$BER(1) = Q \left[\frac{A}{\sigma} \left(1 + \frac{3T}{2RC} \right) \right] ;$$

$$BER(2) = (1-q) Q \left[\frac{A}{\sigma} \left(1 - \frac{3T}{2RC} \right) \right] + q Q \left[\frac{A}{\sigma} \left(1 + \frac{T}{2RC} \right) \right] ;$$

$$BER(3) = Q \left[\frac{A}{\sigma} \left(1 - \frac{T}{2RC} \right) \right] .$$

From the symmetry it follows that $BER(4) = BER(2)$ and $BER(5) = BER(1)$. Consequently, we use (4.23) to calculate the mean error rate $E\{BER\}$:

$$E\{BER\} = \frac{q}{2} \left(Q \left[\frac{A}{\sigma} \left(1 + \frac{T}{2RC} \right) \right] + Q \left[\frac{A}{\sigma} \left(1 - \frac{T}{2RC} \right) \right] \right) + \frac{1}{2} (1-q) \left(Q \left[\frac{A}{\sigma} \left(1 + \frac{3T}{2RC} \right) \right] + Q \left[\frac{A}{\sigma} \left(1 - \frac{3T}{2RC} \right) \right] \right) .$$

By inspecting the arguments of the function Q in the above example it can be noticed that we have introduced two new parameters: the ratio A/σ , or the signal-to-noise ratio (SNR) if we express it in dB's, and the symbol time-to-time constant ratio T/RC .

The mean error rate is plotted in Fig. 4.4 for two different sources and two values of T/RC . Both sources have the same sum variance.

From Fig. 4.4 it can be concluded that:

- the mean error rate decreases for a larger value of the SNR as it could be expected;
- in general the mean error rate increases with a larger number of states, because the source generates sequences

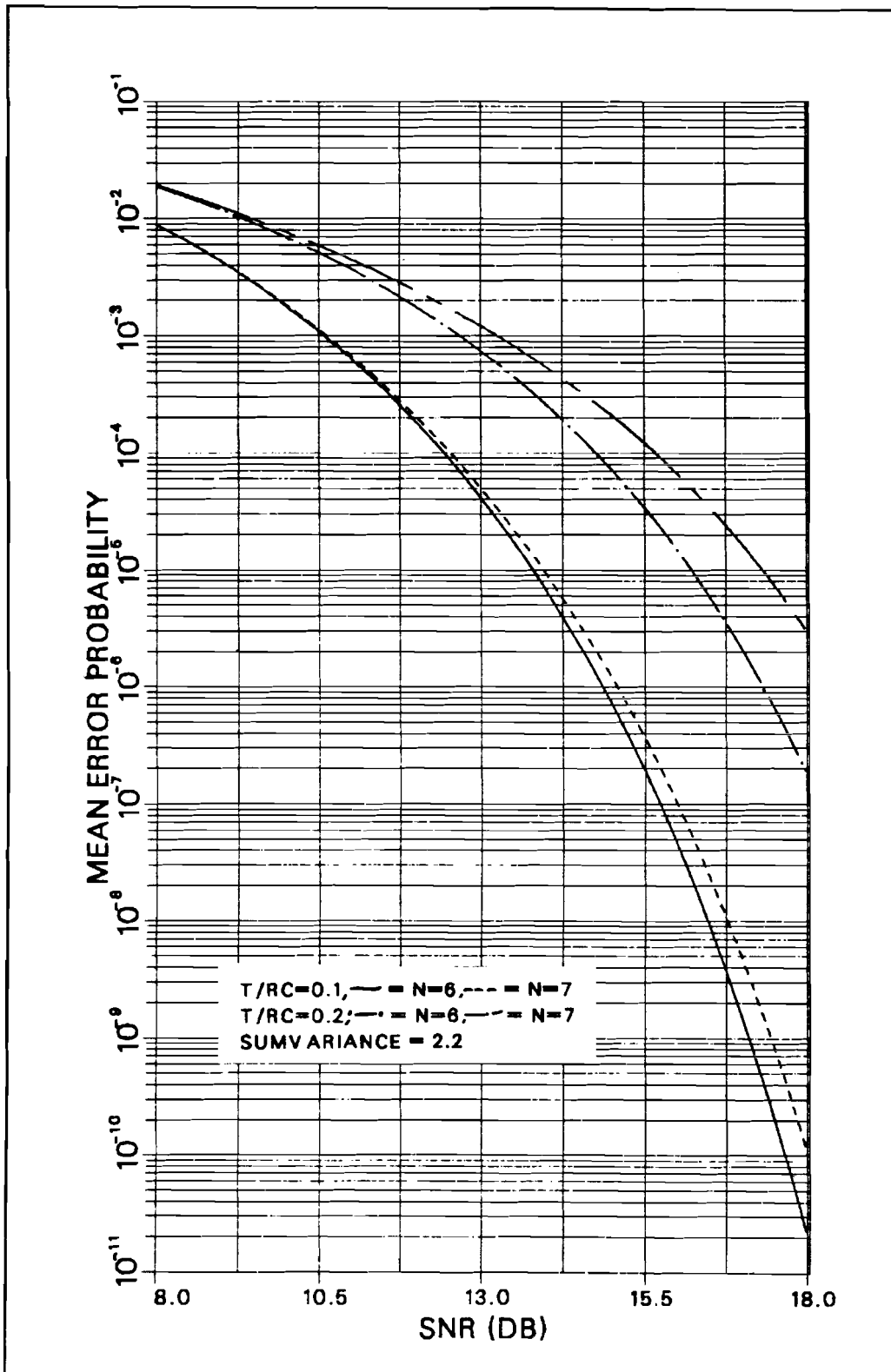


Fig. 4.4. Mean error rate for different sources and equal sum variance.

5.0 MAXIMUM ENTROPY

This chapter deals with the following question: if we demand that a certain performance criterion - besides the bound on the RDS - must be satisfied, how do the source probabilities have to be chosen in order to achieve the maximum entropy?

Before the derivations start the problem is formulated precisely. Then the maxentropic chain with a general weight constraint is derived. The developed theory is applied to the case of a variance and an error rate constraint. It is shown that there is some similarity between the maximum entropy and the rate distortion function. Finally, some computational results are presented.

5.1 Optimization problem.

We want to maximize the entropy $H(X)$ of the signal. Since the relationship (2.1) between the sequences $X(t)$ and $Z(t)$ is one to one, it is clear that $H(X) = H(Z)$. The problem can thus entirely be stated in terms of the sequence $Z(t)$.

From [23] we can see that conditioning never augments the entropy, so the sum sequence $Z(t)$ must satisfy the Markov property and consequently, it must be generated by a Markov source. This reasoning fits very well with our assumptions stated earlier. Note that two consecutive values of Z can only differ ± 1 because $X(t)$ is a binary sequence.

We now will focus on the performance criterion. Let the restriction have a certain value W , which we shall call the average weight W . Since it is possible to let W be the result of an average of transitions, we assign to each transition a certain weight $w(j|i)$. At a later stage, we replace the average W by one of the parameters mentioned in the preceding chapter.

If we want a power spectrum with a low dc content, then the transitions in the outer states must be of heavier

weight than the transitions between the inner states, because the former give the largest contribution to the dc content.

The problem posed here is to maximize the entropy:

$$H = \sum_{i=1}^N p_s(i) \sum_{j=1}^N - p(j|i) \ln p(j|i) , \quad (5.1)$$

subject to the constraints:

$$\sum_{j=1}^N p(j|i) = 1 , \quad \text{for all } i , \quad (5.2)$$

$$\sum_{i=1}^N p_s(i) = 1 , \quad (5.3)$$

$$\sum_{i=1}^N p_s(i) p(j|i) = p_s(j) , \quad \text{for all } j , \quad (5.4)$$

$$\sum_{i=1}^N p_s(i) \sum_{j=1}^N p(j|i) w(j|i) = W . \quad (5.5)$$

Equations (5.2), (5.3) and (5.4) are general constraints related to the regular definition of the Markov source. Equation (5.5) is an important constraint since it is related to the desired spectrum.

5.2 Sources with a weight constraint.

This section presents a maxentropic Markov source that satisfies a weight constraint. The derivation is based on results from Justesen [24], Berger [25] and Shannon [26].

A well-known method to solve the problem of maximizing a function subject to several constraints is the use of Lagrange multipliers. We will use this approach here. First,

we form the Lagrangian function I, which contains the constraints from the preceding section:

$$\begin{aligned}
 I = H + s \left(\sum_{i=1}^N p_s(i) \sum_{j=1}^N p(j|i) w(j|i) \right) + \\
 \sum_{j=1}^N t(j) \left(\sum_{i=1}^N p_s(i) p(j|i) - p_s(j) \right) + \\
 \sum_{i=1}^N r(i) \left(\sum_{j=1}^N p(j|i) \right) + u \sum_{i=1}^N p_s(i) , \quad (5.6)
 \end{aligned}$$

where $r(i)$, s , $t(j)$ and u are all Lagrange multipliers, and H is given by equation (5.1).

Secondly, the stationary point of the function I is found by solving:

$$\partial I / \partial p(j|i) = 0 , \quad (5.7)$$

$$\partial I / \partial p_s(i) = 0 . \quad (5.8)$$

Differentiation of I with respect to $p(j|i)$ or equivalently, applying (5.7) to (5.6) yields:

$$\begin{aligned}
 \sum_{i=1}^N p_s(i) \sum_{j=1}^N (-1 - \ln p(j|i)) + \sum_{i=1}^N p_s(i) \sum_{j=1}^N s w(j|i) + \\
 \sum_{j=1}^N t(j) \sum_{i=1}^N p_s(i) + \sum_{i=1}^N r(i) \sum_{j=1}^N 1 = 0 ,
 \end{aligned}$$

and by rewriting we obtain:

$$-\ln p(j|i) + s w(j|i) + t(j) + r(i)/p_s(i) = 1 . \quad (5.9)$$

The second equation in $p_s(i)$ and $p(j|i)$ comes from the differentiation with respect to $p(i)$, so applying (5.8) to (5.6):

$$\sum_{i=1}^N \sum_{j=1}^N -p(j|i) \ln p(j|i) + s \sum_{i=1}^N \sum_{j=1}^N p(j|i) w(j|i) + \sum_{i=1}^N \sum_{j=1}^N t(j) p(j|i) - \sum_{i=1}^N t(i) \cdot 1 + u \sum_{i=1}^N 1 = 0 ,$$

where we have used $\sum_j t(j) p_s(j) = \sum_i t(i) p_s(i)$ before differentiation. If we replace the number 1 by equation (5.2) we find:

$$\ln p(j|i) = s w(j|i) + t(j) - t(i) + u . \tag{5.10}$$

We now have two equations from which $p(j|i)$ and $p_s(i)$ can be solved. The transition probabilities are specified by:

$$p(j|i) = \exp(sw(j|i)) \frac{v(j)}{\lambda v(i)} , \tag{5.11}$$

where we have substituted $v(i) = \exp(t(i))$ and $\lambda = \exp(-u)$ in (5.10).

Before continuing let us examine (5.11). The constraint (5.2) can be expressed as:

$$\sum_{j=1}^N \exp(sw(j|i)) v(j) = \lambda v(i) . \tag{5.12}$$

We conclude that \underline{v} with elements $v(i)$ represents the right eigenvector of the matrix $Q_w = [\exp(s w(j|i))]$ and λ the corresponding eigenvalue. Similarly, (5.4) may be expressed as:

$$\sum_{i=1}^N \frac{p(i)}{v(i)} \exp(sw(j|i)) = \lambda \frac{p(j)}{v(j)} , \tag{5.13}$$

and thus $p(i)/v(i)$ are the elements of the left eigenvector with eigenvalue λ . Thus, the left eigenvector \underline{v}' is the eigenvector of the transposed matrix of Q_w . Consequently, the stationary probabilities are given by:

$$p_s(i) = c v(i) v'(i) , \quad (5.14)$$

where c stands for an arbitrary constant determined by (5.3). The symbol $v'(i)$ represents the i -th element of the left eigenvector \underline{v}' .

On the basis of the equations (5.1) and (5.11), we can derive an expression for the entropy as a function of the average weight W , called $H(W)$:

$$H(W) = \ln \lambda - s W \quad (\text{nats}). \quad (5.15)$$

This important relation indicates how the best possible rate may be obtained, as the maximum entropy is reached at maximum eigenvalue.

The reader should notice that for $w(j|i) = 0$ (for all i, j) equations (5.11), (5.14) and (5.15) transform into the corresponding relations (2.23), (2.24) and (2.22), respectively. This means that without the spectral constraint, the maxentropic chain from section 2.5 is derived.

The optimum Markov source for an average weight constraint can be determined as follows:

- form the weight matrix $Q_w = [\exp(sw(j|i))]$;
- calculate the maximum eigenvalue λ_m and its corresponding left and right eigenvectors \underline{v}' and \underline{v} ;
- derive the transition probabilities, which are specified by (5.11), and the stationary probabilities of (5.14);
- calculate the maximum entropy by using (5.15), where the variable s is determined by W .

5.3 Analogy to rate distortion theory.

In the preceding section the maximum entropy was derived as a function of an average weight. Because the resemblance between the entropy function $H(W)$ and a rate distortion

function $R(D)$ is strong, it is interesting to pay more attention to it.

In the rate distortion theory we are primarily interested in the minimum information that must be conveyed to the user in order to achieve a prescribed fidelity [25]. In our case, the objective is to maximize the entropy satisfying a spectral constraint.

Secondly, the conformity between the average weight W and the average distortion D can be noticed. Both parameters are set before the variational problem is solved. The distortion D is the outcome of averaging all single letter distortions $d(j|i)$ (i.e. the distortion between source and representation letter). Similarly, we have dealt with weight elements $w(j|i)$.

Thirdly, a rate distortion function can be determined if the source alphabet and its probability distribution are known. In our case, source and representation alphabet are the same, e.g. the set of Z values, and consequently they have the same probability distribution.

The fourth and last item concerns the parameter s . The rate distortion function $R(D)$ is the envelope of a set of straight lines with slope s in $(D, R(D))$. Similarly, the maximum entropy function $H(W)$ is the envelope of lines through $(W, H(W))$ with slope $-s$ (see Fig. 5.1).

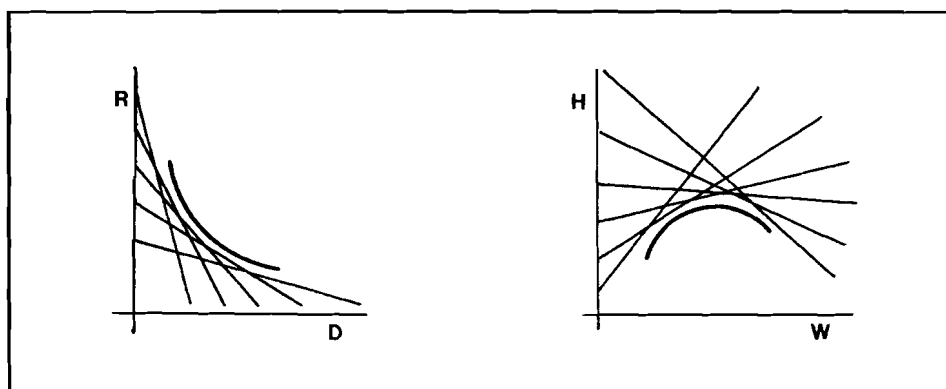


Fig. 5.1. Comparison of $R(D)$ and $H(W)$ curves.

$$Q = \begin{bmatrix} 0 & t^{5/2} & 0 & 0 & 0 \\ t^{5/2} & 0 & t^{1/2} & 0 & 0 \\ 0 & t^{1/2} & 0 & t^{1/2} & 0 \\ 0 & 0 & t^{1/2} & 0 & t^{5/2} \\ 0 & 0 & 0 & t^{5/2} & 0 \end{bmatrix},$$

where $t = \exp(s)$. The matrix has a largest eigenvalue equal to $\lambda_m = \sqrt{(2t+t^5)}$, with corresponding eigenvector:

$$v = (t^2, \sqrt{(2+t^4)}, 2, \sqrt{(2+t^4)}, t^2).$$

It can be verified that $p_s(i) = (t/4 \lambda_m^2)v(i)^2$ and the most important transition probability $p(3|2)$ may be expressed as $2/(2+t^4)$. As a result, the entropy equals:

$$H = \frac{1}{2} \ln (2t + t^5) - s \frac{2 + 5t^4}{4 + 2t^4}.$$

For example, if $\sigma_2^2 = 1$, then $t^4 = 2/3$ and $s = 1/4 \ln (2/3)$. Accordingly, we find $H = 0.541$ nats or equivalently, 0.781 bits.

5.5 Source with an error rate constraint.

A second application of the theory of section 5.2 is found by using the mean error rate as the average weight. It immediately follows from equation (4.22) that in this case the weights are:

$$w(j|i) = Q [(A+d(j|i))/\sigma] . \tag{5.18}$$

In spite of a quadratic weighting in section 5.4 we have a more or less 'exponential' weighting, by taking the weights as Q functions.

It should be noticed that the weight matrix Q_w is not symmetric in the case of a mean error rate constraint.

Example: Let us consider the source of section 5.4. The weight matrix Q_w now is:

$$Q_w = \begin{bmatrix} 0 & t^d & 0 & 0 & 0 \\ t^b & 0 & t^c & 0 & 0 \\ 0 & t^a & 0 & t^a & 0 \\ 0 & 0 & t^c & 0 & t^b \\ 0 & 0 & 0 & t^d & 0 \end{bmatrix},$$

where $t = \exp(s)$ and the maximum eigenvalue equals:

$$\lambda_m = \sqrt{(t^{b+d} + 2t^{a+c})}.$$

The constants may be expressed as:

$$a = Q [A/\sigma (1 - T/2RC)], \quad b = Q [A/\sigma (1 - 3T/2RC)], \\ c = Q [A/\sigma (1 + T/2RC)], \quad d = Q [A/\sigma (1 + 3T/2RC)].$$

The right and left eigenvector are specified by:

$$\underline{v} = (t^d, \sqrt{t^{b+d} + 2t^{a+c}}, 2t^a, \sqrt{t^{b+d} + 2t^{a+c}}, t^d), \\ \underline{v}' = (t^b, \sqrt{t^{b+d} + 2t^{a+c}}, 2t^c, \sqrt{t^{b+d} + 2t^{a+c}}, t^b),$$

and $p_s(i) = k v(i)v(i)'$, where $k = 1/(4 \lambda_m^2)$. Finally, we derive the mean error rate $E\{BER\}$:

$$E\{BER\} = \frac{(b+d) + 2(a+c)t^{a+c-b-d}}{2 + 4t^{a+c-b-d}}.$$

The entropy can be written as:

$$H = 1/2 \ln (t^{b+d} + 2t^{a+c}) - s (E\{BER\}).$$

The transition probability $p(3|2)$ equals:

$$p(3|2) = q = 2t^{a+c} / (t^{b+d} + 2t^{a+c}).$$

Note that for $a = c = 1/2$ and $b = d = 5/2$ the mean error rate constraint transforms into the variance constraint

discussed in the preceding section; the latter is the special symmetric case of the general, asymmetric criterion.

5.6 Results.

As described in section 5.2, it is possible to derive the maximum entropies for various numbers of states subject to a particular spectral constraint. We have calculated the maximum entropy for three cases: a sum variance, a second derivative and a mean error rate constraint. The curves are plotted on the following pages.

The functions can be regarded as theoretical limits for the entropy. A good code has a rate that approaches the limit; there are even codes that lie on the bound (e.g. biphase).

From Fig. 5.2 it can be noticed that when the number of RDS states N is large, two completely different values of the sum variance can be found for the same entropy. This means that for codes with rates approaching the maximum entropy, the dc content can differ significantly. Hence it must be possible to design a code with a sufficiently low dc content at the cost of only a small reduction of the maximum entropy.

From Figs. 5.2-5.4 it can be concluded that a strong constraint (e.g. a small variance) lowers the entropy considerably. This results in sequences that can be generated by a Markov source with a smaller number of RDS states.

The black points in the maxima of the curves represent the maximum entropy for RDS-bounded sequences only. These values have been derived in section 2.5.

Let us now have a closer look at Fig. 5.4. Surprisingly, the curves intersect and as a result, a higher entropy can be reached with a smaller DSV. An explanation for this phenomenon may be the weighting method: a heavier weight (in

comparison with the sum variance) is assigned to the transitions between the outer RDS states. Because the mean error rate constraint must be met, the outer states are seldom reached, resulting in a substantial loss of entropy. For example, if we take the fourth moment as a criterion, in spite of the variance, the same phenomenon will be observed. Moreover, the presented channel model is based on bit-by-bit detection, a method that detects a received symbol independent of the preceding symbols, whereas the Markov source generates dependent sequences.

The problem of deriving analytical expressions in case of a second derivative constraint has not been solved. The requirement for a prescribed sum of correlation terms (a second derivative constraint) is a less strong claim than the prescription of several terms of the correlation function [27]. However, the analytic solution of the optimization of the entropy subject to the first constraint seems to be even more complex than the solution in case of the second constraint. In the latter case the problem has not been solved either.

Fig. 5.2. Maximum entropy / sum variance.

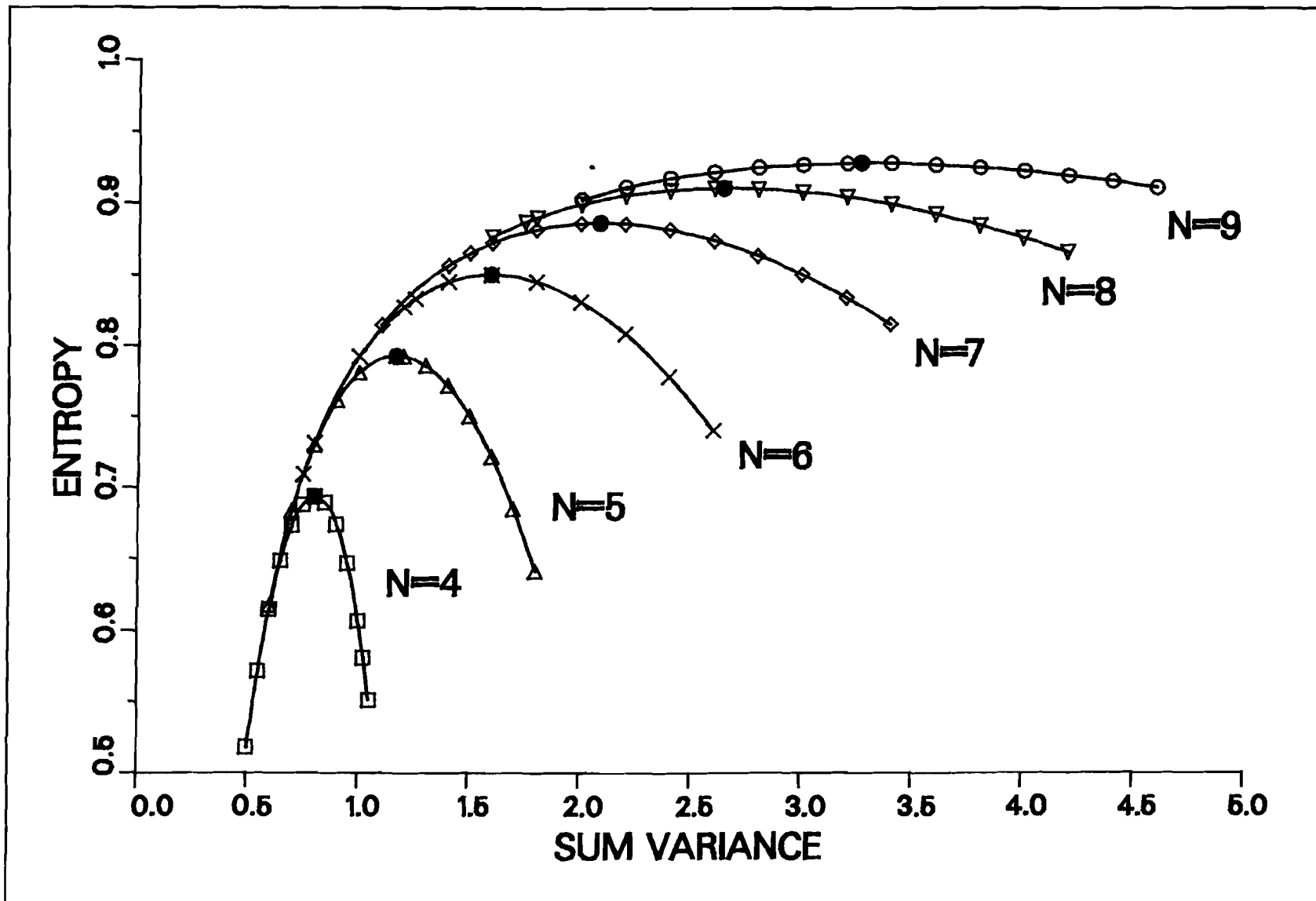
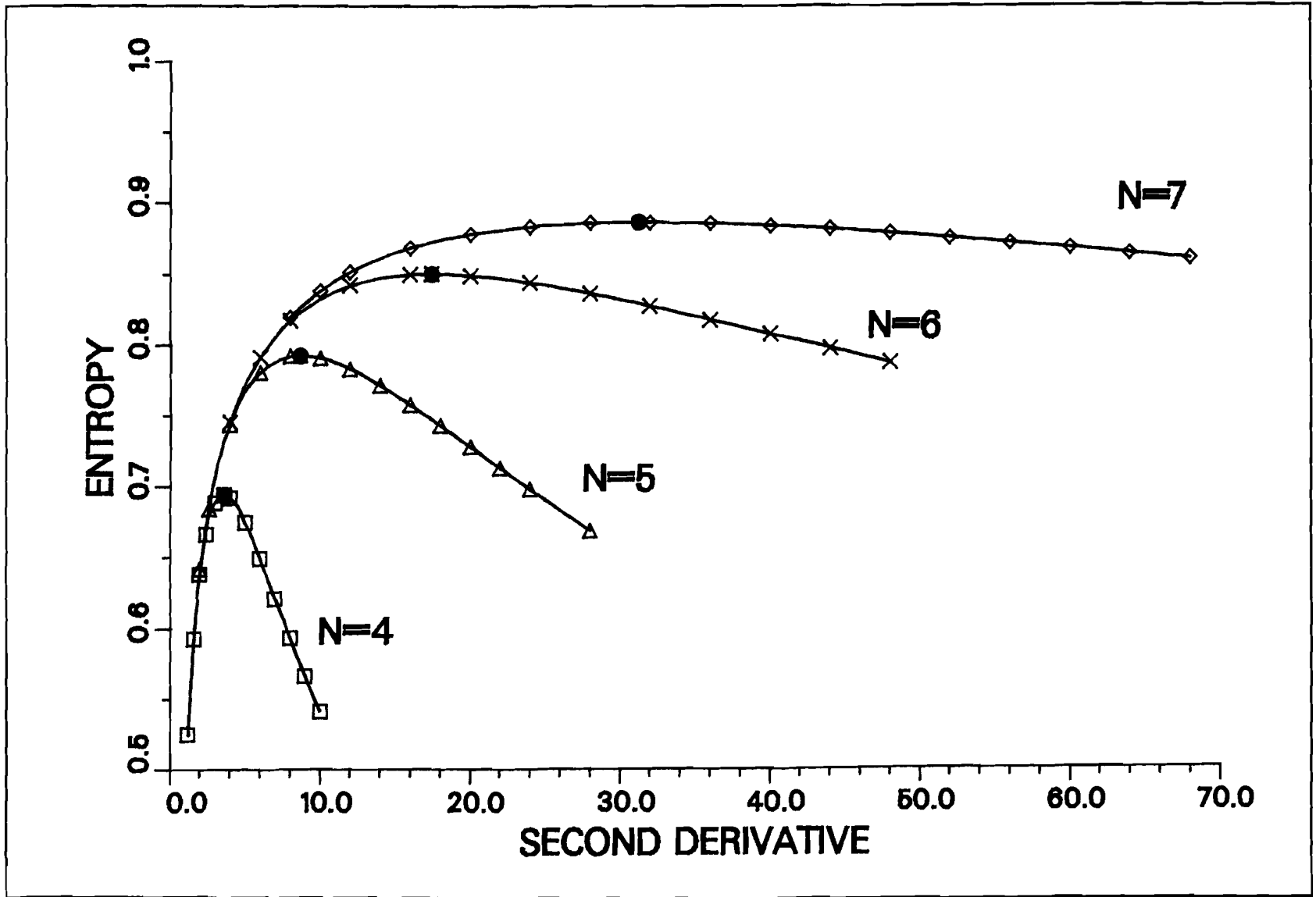


Fig. 5.3. Maximum entropy / second derivative.



6.0 ZERO SECOND DERIVATIVE

One of the parameters mentioned in chapter 4 is the second derivative. It is shown that for low frequencies the power spectrum can be approximated by a parabolic expression.

The present chapter describes a source model that generates sequences with a zero second derivative power spectrum. Maxentropic sources for several cases are considered.

6.1 Additional constraint.

A power spectrum with zero second derivative at zero frequency can be reached by imposing an extra constraint on RDS-bounded sequences. We introduce this constraint in the present section.

It was shown in the second chapter that a dc-free code spectrum is accomplished by a bound on the RDS. Analogous to this result we conjecture that a bound on the sum of the running digital sums (SRDS) should accomplish a dc-free spectrum of the sum sequence $Z(t)$. Consequently, the code spectrum should have zero second derivative, resulting in a considerable lower dc content than RDS-bounded sequences only. This conclusion is based on equation (4.9) which states that the second derivative for zero frequency equals two times the dc component of the RDS spectrum.

We define the sum of the running digital sums as:

$$\text{SRDS} = \sum_t Z(t) . \quad (6.1)$$

In accordance with section 2.2, the SRDS may be expressed as a sequence $Y(t)$:

$$Y(t) = Y(t-T) + Z(t) , \quad Y(0) = b ; \quad (6.2)$$

where b is the initial value of the sum-sum sequence $Y(t)$.

Furthermore, to achieve a dc-free code spectrum we demand the RDS to be bounded within the DSV. Similarly, a dc-free sum spectrum is reached by a bound on the SRDS, which we shall call the sum digital sum variation (SDSV):

$$\text{SDSV} = \max_{t_1, t_2} \left\{ \sum_{t_1}^{t_2} Z(t) \right\} . \quad (6.3)$$

It should be noticed that the process $Y(t)$ is non-Markovian [12], and a single bound on the SRDS would not suffice a dc-free power spectrum of $X(t)$.

6.2 Source model.

This section presents a new source model that includes both DSV and SDSV constraints.

Our aim is to design a source that includes both RDS and SRDS bounds. Instead of choosing one state for each value of the RDS, it is now proposed to assign a state to each allowable (RDS, SRDS) pair. As a result, we deal with a two-dimensional random walk. The coordinates (RDS, SRDS) define the several possible states and the RDS and SRDS bounds (DSV and SDSV constraint) specify the borders of the allowable region. The number of possible RDS values is assumed to be N and the number of possible SRDS values M .

Example 1: Consider the simple case where $N = 3$ and $M = 3$ (DSV = 2 and SDSV = 2). The vectors of RDS and SRDS values are $\underline{y} = \underline{z} = (-1, 0, +1)$. The source model is sketched in Fig. 6.1. There are nine states that satisfy the DSV and SDSV constraints. However, state 3 and 7 are rejected since we prefer an irreducible set of states. For this reason state 3 and 7 are drawn with broken lines.

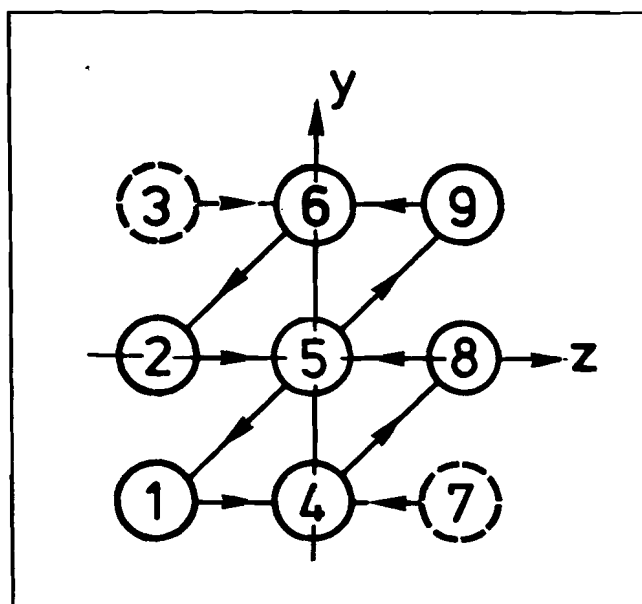


Fig. 6.1. Source with $DSV = SDSV = 2$.

The maxentropic chain has $p(1|5) = p(9|5) = 1/2$ with entropy $1/4$. It follows by definition that $R_z(0) = 1/2$ and $R_z(2) = -1/4$. Hence the spectrum of the code sequence $X(t)$ is:

$$S_X(f) = (1 - \cos 2\pi fT)(1 - \cos 4\pi fT).$$

The practical implementation of this source may be a biphase variant where the following mapping is proposed: $1 \rightarrow 1001$ and $0 \rightarrow 0110$.

- Example 2:* Consider all chains with $N = 3$ and $M > 3$ (M odd). These sources have the following properties:
- the sum variance is $1/2$ because the boundary states are reached half of the time. The sum-sum variance has the same value as the maxentropic sum variance (with M states) from section 4.1.
 - $R_z(2k+1) = 0$ for all k .
 - the maxentropic chain with $N = 3$ and variable M has transition probabilities:

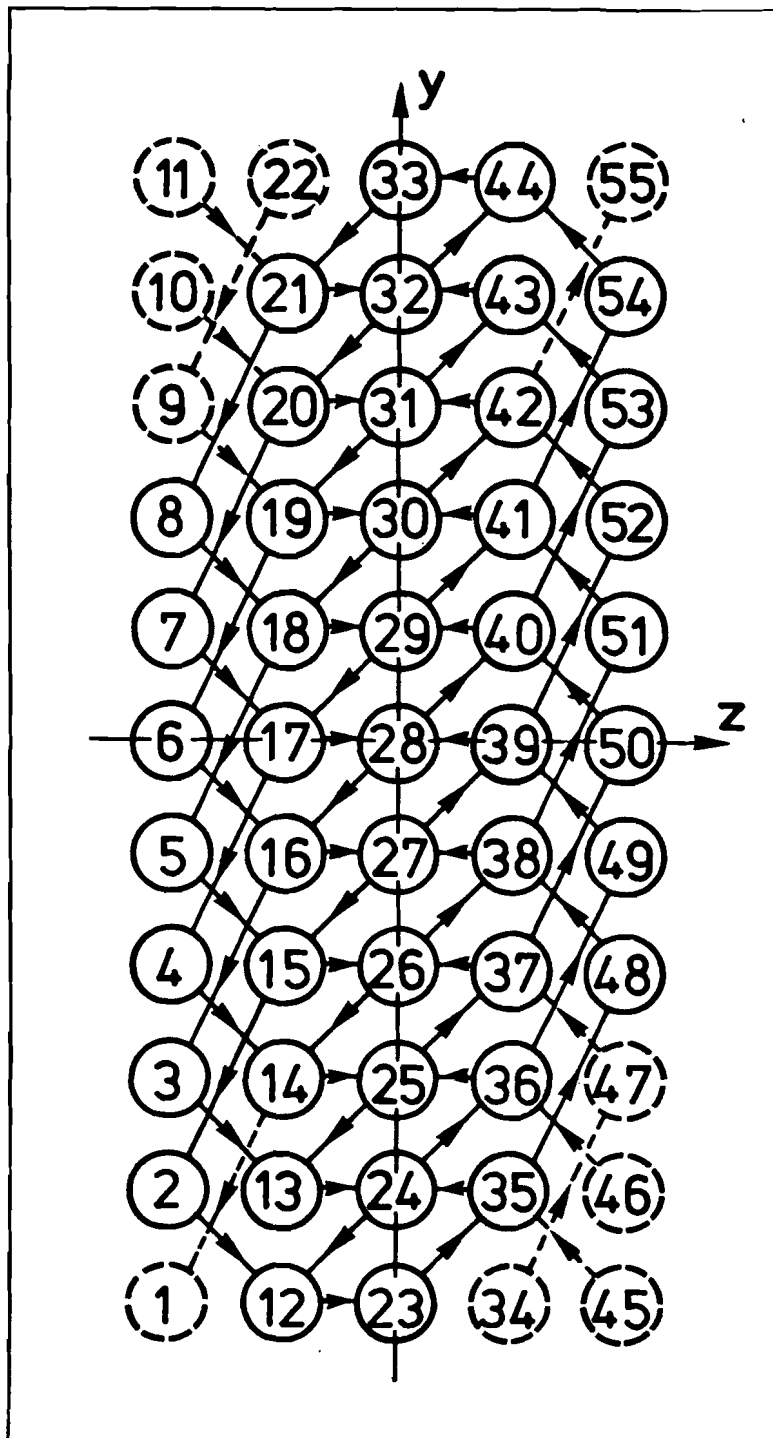


Fig. 6.2. Source with $(N, M) = (5, 11)$.

$$p(M+1+i|i) = \frac{\sin\left(\pi \frac{i+1-M}{M+1}\right)}{2 \cos\left(\frac{\pi}{M+1}\right) \sin\left(\pi \frac{i-M}{M+1}\right)}, \quad M+1 \leq i \leq 2M-1;$$

whereas other transitions have probability one. The corresponding stationary probabilities are:

$$p_s(i) = \frac{1}{M+1} \sin^2 \left(\pi \frac{i-M}{M+1} \right), \quad M+1 \leq i \leq 2M,$$

This result strongly resembles the probabilities from section 2.5.

In Fig. 6.2 a larger model is given with $N = 5$ and $M = 11$. It can be noticed that the states 1, 9, 10, 11, 22, 34, 45, 46, 47 and 55 do not contribute to the entropy and the power spectrum.

Some properties of the sources with a DSV and SDSV constraint are listed below.

- All models have period 4. Consequently, modulation systems with zero disparity codewords have multiple four length (The disparity is the excess of ones over zeros).
- The eigenvalues of the transition matrix Q appear in groups of four elements. According to Perron-Frobenius [13] the complex eigenvalue plane can be rotated over $\pi/2$ without change of configuration. This means that some eigenvalues occur in complex conjugated pairs on the imaginary axis. This results in a fluctuating correlation function.
- Generally speaking: if M tends to infinity, the SDSV constraint is not satisfied and the sequences are the same as the sequences generated by an RDS-bounded Markov source.

6.3 Computational results.

This section presents some results of calculations we have made. For maxentropic sources the entropy and the frequency f_T 0.01% from section 4.3 are given.

It is obvious that the cut-off frequency (and thus sum variance) and the second derivative are unsuitable performance parameters in case of a combination of DSV and

SDSV constraints. The sum variance cannot be used because only one dimension is coincided (RDS). As the second derivative is zero, this parameter also cannot be applied. The mean error rate gives less insight into the case of a two-dimensional source and therefore the frequency $f_T 0.01\%$ is used (Table 6.1) as performance parameter.

N	M	entropy	$f_T 0.01\%$
3	3	.2500	.04396
	5	.3962	.03009
	9	.4638	.02274
	∞	.5000	.01967
5	9	.6271	.02240
	11	.6676	.02001
	13	.6950	.01820
	15	.7141	.01679
	19	.7385	.01476
	∞	.7925	.00959

Table 6.1. Maxentropic (N,M) sources.

Table 6.1 shows the tendency for large M to the maxentropic sources with a DSV constraint only.

In Fig. 6.3 the frequency $f_T 0.01\%$ is plotted as a function of the redundancy $(1-H)$. For maxentropic DSV-constrained sequences the exchange between low-frequency power and redundancy (entropy) can be clearly observed. For sequences with zero second derivative the bound has a steeper slope, indicating that adding some extra redundancy results in a considerable decrease of the low-frequency power.

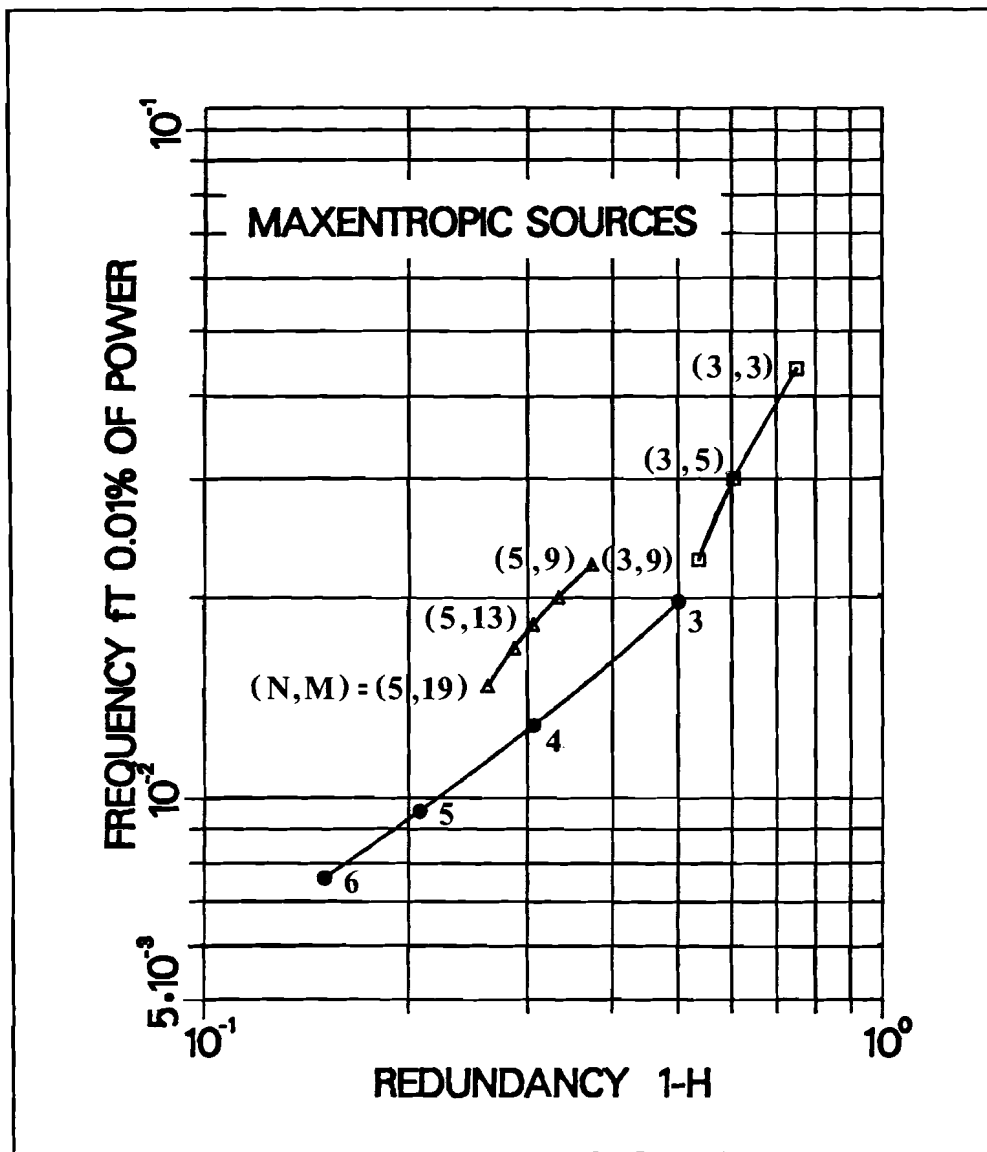


Fig. 6.3. Frequencies f_T 0.01% / redundancy for maxentropic sources.

Interesting is the comparison between the sources with $N = 4$ and $(N,M) = (5,13)$. Both have almost equal entropy (.6942 and .6950, respectively) although the latter has a significant reduction of dc content. This is also shown in Fig. 6.4 where the power spectrum is plotted for several sources. The spectra of DSV-constrained sequences are plotted with broken lines.

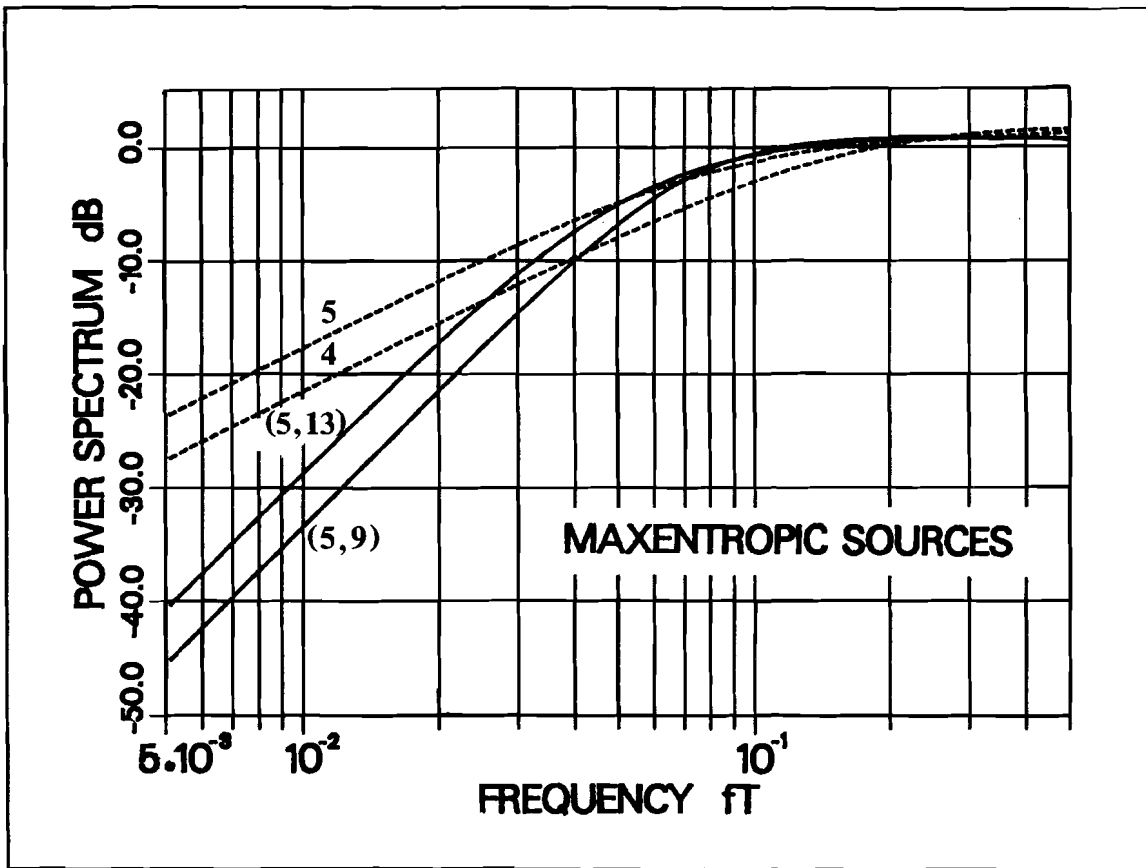


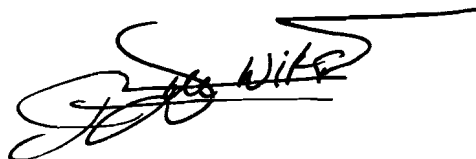
Fig. 6.4. Power spectra of DSV- and SDSV-constrained sequences in comparison with DSV-constrained sequences.

7.0 CONCLUSIONS

We have discussed four performance parameters by which the dc content of a binary spectrum shaping code can be assessed. Besides the description of the sum variance and the second derivative, we have introduced a 'power cut-off' frequency that is related to a certain percentage of the power. This frequency is certainly one of the most objective parameters to compare the dc content of different codes since it is closely related to the low-frequency power. We have also introduced a mean error rate as an important parameter. The mean error rate has a more practical meaning and emphasizes the need for sequences with sufficient alterations of bit polarity.

We have calculated the maximum entropies of RDS-bounded sequences subject to an extra performance constraint. It is shown that the dc content of codes with rates close to the maximum entropy can differ significantly. Analytic computation of the maximum entropy is performed in two cases: the sum variance and the mean error rate criterion. It is shown that the mean error rate yields an example of the general, asymmetric case of the analytic computation of the maximum entropy, whereas the sum variance appears to be a symmetric weighting of the RDS states.

In order to improve the power spectrum for low frequencies, we have introduced RDS- and SRDS-bounded sequences. It is shown that, besides a dc-free power spectrum, these sequences have a zero second derivative at zero frequency, resulting in a substantial decrease of dc content.



Eindhoven, April 1984.
Philips Research Laboratories,
Nederlandse Philips Bedrijven BV.

ACKNOWLEDGMENTS

The author is greatly indebted to Prof.dr.ir. J.P.M. Schalkwijk for his skilful supervision and continuous interest. I want to thank Ir. K.A. Schouhamer Immink for his valuable suggestions. His vast practical insight has been of great importance to this work. In particular I wish to thank Dr.ir. A.J. Vinck for his warm encouragement and interest. I will remember our fruitful discussions, on both scientific and human subjects.

The graduate work described in this Report has been carried out at the Philips Research Laboratories. I want to thank the directors of these laboratories for enabling me to perform the work in this specific research environment.

I wish to thank Dr.ir. A. Huijser for letting me be the first student in electronics performing graduate work in the optical recording group. Furthermore I need to thank J. Schuckard for helping me to review the text.

Met innige dank aan JELLY

REFERENCES

- [1] E. Gorog, 'Redundant alphabets with desirable frequency spectrum properties'; IBM Journal of Research & Development, vol. 12, pp. 234-241, May 1968.

- [2] H. Kobayashi, 'A survey of coding schemes for transmission or recording of digital data'; IEEE Trans. on Comm. Technology, vol. COM-19, no. 6, pp. 1087-1100, December 1971.

- [3] K.W. Cattermole, 'Principles of digital line coding'; Int. Journal of Electronics, vol. 55, no. 1, pp. 3-33, July 1983.

- [4] D.B. Waters, 'Line codes for metallic cable systems'; Int. Journal of Electronics, vol. 55, no. 1, pp. 159-169, July 1983.

- [5] R.M. Brooks and A. Jessop, 'Line coding for optical fibre systems'; Int. Journal of Electronics, vol. 55, no. 1, pp. 81-120, July 1983.

- [6] D.T. Tang and R.L. Bahl, 'Block codes for a class of constrained noiseless channels'; Information & Control, vol. 17, no. 5, pp. 436-461, December 1970.

- [7] G.F.M. Beenker and K.A. Schouhamer Immink, 'A generalized method for encoding and decoding run-length-limited binary sequences'; IEEE Trans. on Information Theory, vol. IT-29, no. 5, pp. 751-754, September 1983.

- [8] Philips Technical Review, vol. 40, no. 6, September 1982.

- [9] J.C. Mallinson and J.W. Miller, 'Optimal

codes for digital magnetic recording';
The Radio and Electronic Engineer, vol.47, no. 4,
pp. 172-176, April 1977.

- [10] H. Yasuda and H. Inose, 'Generalized method for constructing a multi-level balanced code by means of loop-sum-zero transition diagram';
Journ. IECE Japan, vol. 54-A,
pp. 506-513, September 1971.

- [11] J.N. Franklin and J.R. Pierce, 'Spectra and Efficiency of binary codes without DC';
IEEE Trans. on Communications, vol. COM-20,
pp. 1182-1184, December 1972.

- [12] W. Feller, 'An introduction to Probability Theory and its applications';
Wiley & Sons, New York, 1950, vol. 1, Chapter 15.

- [13] F.R. Gantmacher, 'The theory of Matrices';
Chelsea Publishing Company, New York, 1959,
vol. 2, Chapter 13.

- [14] P. Galko and S. Pasupathy, 'The mean power spectral density of Markov chain driven signals';
IEEE Trans. on Information Theory, vol. IT-27, no. 6,
pp. 746-754, November 1981.

- [15] P. Whittle, 'Prediction and Regulation';
The English Universities Press, London, 1963.

- [16] J. Justesen, 'Information rates and power spectra of digital codes';
IEEE Trans. on Information Theory, vol. IT-28, no. 3,
pp. 457-472, May 1982.

- [17] T.M. Chien, 'Upper bound on the efficiency of DC-constrained codes';

- The Bell System Technical Journal, vol. 49,
pp. 2267-2287, November 1970.
- [18] A. Papoulis, 'Probability, Random Variables and
Stochastic Processes';
Mc Graw-Hill, New York, 1965, Chapter 9.
- [19] W.R. Bennet, 'Statistics of regenerative
digital transmission';
The Bell System Technical Journal, vol. 37, no. 6,
pp. 1501-1542, November 1958.
- [20] I.S. Gradshteyn and I.M. Ryzhik, 'Tables of
integrals series and products';
Academic Press NY, London, 1965, fourth edition.
- [21] V.A. DiEuliis and F.P. Preparata, 'Spectrum shaping
with alphabetic codes with finite autocorrelation
sequence'; IEEE Trans. on Commun., vol. COM-26, no.4,
pp. 474-477, April 1978.
- [22] J.M. Wozencraft and I.M. Jacobs, 'Principles of
Communication Engineering';
Wiley & Sons, New York, 1965, Chapter 4.
- [23] R.G. Gallager, 'Information theory and reliable
Communication'; Wiley & Sons, New York, 1968, Chapter 2.
- [24] J. Justesen and T. Hoholdt, 'Maxentropic Markov chains';
Mat. report no. 106, Univ. of Technol. Denmark,
November 1982.
- [25] T. Berger, 'Rate Distortion Theory';
Prentice Hall, New Jersey, 1971, Chapter 2.
- [26] C.E. Shannon, 'A mathematical theory of communication';
The Bell System Technical Journal, vol. 27, no. 3,
pp. 379-423, July 1948.

- [27] D. Slepian, 'On maxentropic discrete stationary processes';
The Bell System Technical Journal, vol. 51, no. 3,
pp. 629-653, March 1972.