Information Theory in the Benelux:
An overview of WIC symposia
1980 – 2003

R.L. Lagendijk, L.M.G.M. Tolhuizen, P.H.N. de With, Eds.

with contributions of

C.P.M.J. Baggen, J. Biemond, G.H.L.M. Heideman,
K.A. Schouhamer Immink, R.L. Lagendijk, B. Macq,
E.C. van der Meulen, A. Nowbakht-Irani, B. Preneel,
J.P.M. Schalkwijk, C.H. Slump, R. Srinivasan,
H.C.A. van Tilborg, Tj.J. Tjalkens, L.M.G.M. Tolhuizen,
P. Vanroose, A.J. Vinck, J.H. Weber,
F.M.J. Willems, P.H.N. de With

sponsored by

Philips Electronics

Information Theory in the Benelux:
An overview of WIC Symposia 1980–2003
R.L. Lagendijk, L.M.G.M. Tolhuizen and P.H.N. de With, editors
Werkgemeenschap voor Informatie- en Communicatietheorie (WIC), Enschede
http:\\www.w-i-c.org
ISBN 90-71048-19-5

Contents

Preface

A symposium on "Information Theory in the Benelux" was organized in Zoetermeer, in 1980. This symposium effectively signifies the informal naissance of the "Werkgemeenschap voor Information en Communicatietheorie" (WIC) – literally translated as "Working Community for Information and Communication Theory". Since 1980, the WIC Information Theory Symposium has become an annual event. The official start of the community originates from February 1984, and the subsequent formal community declaration was established in May 1986. Prof. Boxma (TU Delft), Prof. Gröneveld (Univ. Twente), Prof. Schalkwijk (TU Eindhoven) and Prof. Van der Meulen (K.U. Leuven) are considered the founding fathers of the WIC community, secretarially supported by Dr. Best (Univ. Twente) in the board. Boxma, Gröneveld, and Schalkwijk are honored members of the WIC community; Van der Meulen is still an active member of the WIC board.

The purpose of the WIC – as stated in its Charter, see http:\\www.w-i-c.org – was and still is, first, to coordinate and stimulate the work of professionals in the field of Information and Communication Theory in the Benelux, and second, to further the application of Information and Communication Theory. The community has always stimulated the active involvement of students, for instance by presenting their research results at the WIC's Information Theory Symposia.

Now, 25 years later, in 2004, these principles for WIC symposia still hold and the WIC board is proud to present its 25^{th} symposium to the scientific community. Over the years, the WIC has proven to be relatively small yet active, very much alive and eager to continue communication and exchange of scientific results. The WIC symposium is organized annually as a two-day event and usually takes place at the end of May, and attracts around 50 Information Theory scientists. The symposium is organized without large sponsors and is self-financing, with a relatively low entrance fee to enable students to join the symposia activities.

The 25 WIC symposia reflect the cooperation between the three technical universities in the Netherlands, K.U. Leuven and UCL in Belgium, and Philips Research in Eindhoven. The symposium organization has been rotating between these institutes. Other universities and institutes inside and outside the Benelux have also made significant scientific contributions to the symposia.

The WIC also organizes a midwinter meeting in January. This meeting is a one-day event with tutorials concentrating on a particular theme in information and communication theory and techniques. The event aims at introducing the audience to new developments in specialized fields. This midwinter meeting usually takes place in Eindhoven, because of the large potential technical audience and central location in the Benelux. The meeting attracts between 70 and 150 attendees, and as such has established itself as an important activity of the WIC.

We can safely state that the WIC symposia constitute *the* Benelux forum for the exchange and the in-depth discussion of technical results between Information and Communication Theory specialists. The results presented at the symposia either in oral or poster form, are accompanied by 8-page papers published as the WIC symposia proceedings. This jubilee book summarizes the past 24 WIC symposia and provides an overview of technical results and developments presented at the symposia. The eight chapters have been chosen such that they address particular areas and all cover a reasonable amount of papers. The chapters authors have been invited by the editors to contribute to this jubilee book. In addition to compact reflections on the progress in field, each chapter briefly discusses all published related papers of the past symposia. This jubilee book is therefore not only interesting to read, but we believe that it is also a pleasure to find back the names of the scientists that have contributed to the progress of Information and Communication Theory in the Benelux.

The editors wish to thank all contributors to this book. First, we thank the authors of the chapters, who studied all papers in their category, classified them and provided summaries and related the results to overall developments. Second, we acknowledge Philips Electronics for sponsoring the printing of this book. As the WIC community does not charge a membership fee, only such a sponsorship enables us to carry out a project like this jubilee book. Third, the editors are grateful to Yannick Morvan for the cover design processing, and to Mirjam Nieman for the editorial corrections.

Finally, we would like to say words of appreciation to all authors of the paper published at WIC symposia in the past 25 years. Without any doubt, it were the members of the WIC who have kept the community alive and provided this rich scientific history of Information and Communication Theory and its applications. It was a pleasure to co-author and edit this jubilee book; we hope it will give you the same enjoyment.

Eindhoven, The Netherlands, May 12, 2004.

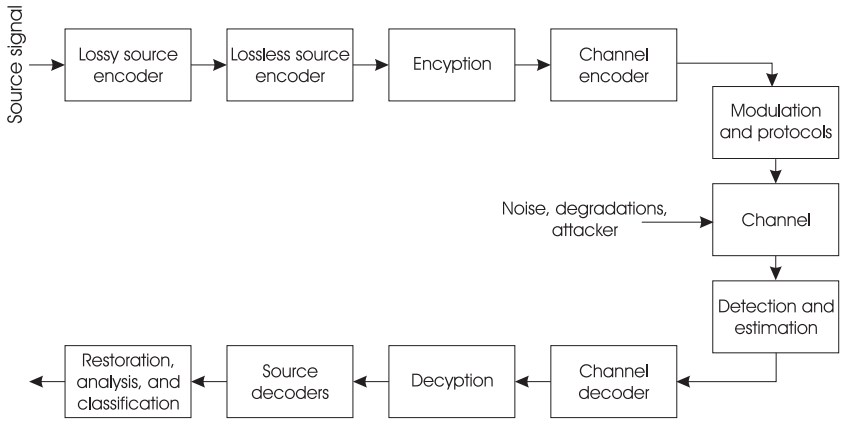
The editors,

Prof.dr.ir. Reginald L. Lagendijk,
Dr.ir. Ludo M.G.M. Tolhuizen, present WIC secretary,
Prof.dr.ir. Peter H.N. de With, present WIC chairman.

Introduction

Information Theory is characterized by a quantitative approach to the notion of information. In 1948, Bell Labs scientist Claude Shannon developed Information Theory [3], and since then the world of communications technology has never been the same. Concepts and theories of Information Theory have found their way to many practical solutions and technologies for communications, consumer electronics, economics, biology, and so on.

At present, Information Theory encompasses not only Shannon's theory of fundamental limits of information representation for reliable transmission and for maximal compression, but also a variety of more design- or engineering-oriented fields. The figure below shows the classical information-theory view on communication systems. This jubilee book on the developments of Information Theory in the Benelux is structured according to this figure.



Information theory view on communication systems.

The first four chapters successively address the building blocks of Information Theory, namely, fundamental Shannon theory of information, lossless (or source)

coding of information, encryption of information, and protection of information by channel codes. The following four chapters increasingly focus on the use of information-theory concepts for solving communication and signal-processing related problems. They address theory and practices of communication and modulation, estimation and detection, signal processing in general and image/video restoration in particular, and finally, compression technology for images and video.

This book discusses all contributions of Information Theory researchers in the Benelux that have appeared in the proceedings of the 24 WIC Symposia between 1980 and 2003. We have categorized the papers into the eight chapters mentioned earlier. Clearly, a substantial number of papers either could have been classified in multiple categories, or fall somewhat outside the eight chapter categories that we selected; we have classified these papers as well as we could. Besides discussing the individual contributions, key references of Information Theory are used for further clarification. In the sequel, we outline the focus and structure of the eight chapters in this book.

In the first chapter, Vanroose, Van der Meulen and Schalkwijk address Shannon Theory and Multi-user Information Theory. The first part of the chapter concentrates on Shannon Theory. After a concise overview of the history of Shannon Theory in the Benelux, the authors address papers on the foundations of Information Theory, including information measures and the relation to statistics, capacity of discrete and AWGN channels, and coding theorems. The second part of the chapter deals with information theory problems in cases with more than one sender and one receiver, i.e. Multi-user Information Theory. The authors summarize theory and papers on five basic multi-user channels (two-way channel, multiple-access channel, broadcast channel, interference channel, and the relay channel), and some other, closely related, communication models.

Willems and Tjalkens discuss source coding in the second chapter. Source coding deals with describing data in the most efficient way, i.e. with the lowest average number of bits per symbol. The chapter starts with the description of the theory and associated papers in the field of non-universal codes. These codes are designed using explicit knowledge about the source behavior. The authors discuss fixed-to-variable and variable-to-fixed codes, as well as several papers addressing applications of these codes. The complementary approach, i.e. designing codes that work for a set of sources with different probabilistic descriptions – called universal codes – is the topic of the second part of the chapter. The main attention in this part of the chapter is paid to the theory of and papers on statistical methods using the Context-Tree Weighting (CTW) method.

In Chapter 3, Van Tilborg, Preneel and Macq address papers on the theory and application of Cryptology. This branch of Information Theory is concerned with the protection of data against malicious parties; in particular, cryptographic primitives try to achieve confidentiality, integrity, and authenticity. The authors start with addressing results on cryptographic primitives, obtained under the assumption that sender and receiver share a common secret. Successively, the authors

focus on private key and public key cryptographic systems. Next, security issues in cryptographic systems are addressed, including policies, key management, and digital signatures. The chapter concludes with a description of results achieved in the fairly recent field of data hiding.

Weber, Tolhuizen and Schouhamer Immink discuss Channel Coding in Chapter 4. Channel coding plays an important role in digital communication and storage systems for combating noise and imperfections of the "channel". The authors first describe the construction and properties of block codes, followed by a discussion on decoding techniques. Subsequently, codes for storage channels are addressed, e.g. run-length-limited (RLL) codes, followed by codes for special channels, such as memories with defects, asymmetric channels, and channels with feedback. The chapter is concluded with the description of papers on applications of channel codes in various areas.

The subject of Communication and Modulation is addressed by Baggen, Vinck and Nowbakht-Irani in Chapter 5 of this jubilee book. The chapter is subdivided into sections dealing with communication and modulation for transmission, for recording, and for networking. The section on transmission discusses papers on coded modulation, single carrier systems, and OFDM. In the section on recording, papers on detection and feedback equalization play a central role. Finally, the section on networking deals with papers on quality of service, routing and queuing, and multiple access.

In Chapter 6, Srinivasan and Heideman discuss research results in the field of Estimation and Detection. Mathematical theories of statistical estimation and detection – in particular Bayesian theories – have laid down guiding principles for processing of signals in a multitude of areas. The chapter starts with a discussion on papers in the field of information-theoretic measures and estimation, including model-order estimation for ARMA processes. The authors continue the chapter with describing papers on detection theory and several applications, like biometrics and the biomedical area. The chapter concludes with papers dealing with statistical classifiers and pattern recognition, including neural networks.

Biemond and Slump address Signal Processing and Restoration in Chapter 7. Digital signal processing concerns the theoretical and practical aspects of presenting, processing, and analysis of information-bearing signals. The first part of the chapter deals with contributions to signal processing problems encountered in the communication between people (audio and speech processing), between people and machines (e.g., biomedical signal analysis), and in the sensing of the environment (e.g., radar and sonar signal processing). In the second part of the chapter, the authors address the numerous papers dealing with image and video restoration, as well as the ensuing processes of image analysis and interpretation.

In the final chapter of this book, De With and Lagendijk address image and video compression. Compression techniques are of prime importance for reducing the amount of data for representing speech, audio, images, and video sequences with-

out loosing too much quality. The authors first give a concise overview of the history of image and video compression theory and technology, and then summarize the WIC Symposia papers in three categories. First, papers on techniques for decorrelating image and video data are described, covering transform and subband coding and motion compensation. Second, papers dealing with scalar and vector quantization theory are summarized. Finally, the authors address papers on advanced topics such as hierarchical, scalable, and embedded compression, as well as alternative compression strategies for particular application domains.

The reference section at the end of the book contains nine parts. The first 114 references are considered key references for Information Theory in general and this book in particular. The following 640 references encompass *all* contributions of Information Theory researchers in the Benelux that have appeared in the 24 WIC Symposia between 1980 and 2003. The WIC references have first been partitioned into categories, corresponding to the eight chapters. Within each category, the WIC references are ordered chronologically.

The October 1998 Commemorative Issue of the *IEEE Transactions on Information Theory* has been a proud testimony of the worldwide accomplishments of five decades of Information Theory. Let this jubilee book be the testimony of the achievements in Information Theory in the Benelux as they were presented at the 1980-2003 WIC Symposia.

CHAPTER 1

Shannon Theory and Multi-user Information Theory

P. Vanroose (K.U. Leuven)
E.C. van der Meulen (K.U. Leuven)
J.P.M. Schalkwijk (TU Eindhoven)

1.1 Shannon Theory

For the research in Shannon theory within the Benelux during the past 25 years, one can distinguish the following clear directions, apart from the research in multi-user information theory.

(i) In the early 80s, when Prof. Y. Boxma was head of the Information Theory Group in the Division of Electrical Engineering at TH Delft, significant research in information theory in Delft focused on the study of information measures, applications of it, and the concept of information in non-probabilistic contexts, resulting in contributions [115, 116, 119, 120]. As these topics are close to the basic question of how to measure information, for which Shannon [3] proposed the fundamental

[1]This chapter covers references [115] – [213]. The work of the second author of this chapter was partially supported by INTAS Project 00-738 and Project GOA/98/06 of Research Fund K.U. Leuven.

quantity

$$H(X) := -\sum_{x \in \mathcal{X}} p(x)\ \log p(x), \tag{1.1}$$

we have grouped these papers in the first section on "Foundations". In that section we have also placed other papers which deal with issues of uncertainty [153], the foundations of probability theory [188], and randomness in connection with typicality [156].

Furthermore, we describe in Section 1.1.1 work by De Bruin and Kamminga on the sum of entropy-type integrals in the time and frequency domain. This research found its origin in Kamminga's Ph.D. thesis (1994), where uncertainty and entropy were addressed in the context of the study of dolphin echo location signals. The study of dolphin sounds was the life-long scientific hobby of Kamminga. We conclude Section 1.1.1 with a description of research regarding the ε-entropy of an ellipsoid in Hamming space carried out by Prelov and Van der Meulen [211].

(ii) In the Department of Mathematics at the K.U. Leuven, significant research was carried out since 1984 by Van der Meulen and Prelov from the Institute of Problems of Information Transmission in Moscow on asymptotic expressions for information-theoretic quantities, such as mutual information and information rate when sending over a stationary channel. Some of this work was done in cooperation with the Russian scientist Pinsker. This research reflects the thinking of the Russian school of information theory, which has built up a great tradition under the influence of Kolmogorov, Dobrushin, Pinsker, Ibragimov and Khas'minskii. The basic concept of information rate $\bar{I}(\boldsymbol{X};\boldsymbol{Y})$ of a pair of sequences of random variables $\boldsymbol{X}, \boldsymbol{Y}$ appears already in the works of McMillan [8] and Khinchin [11], but the main source of reference for the properties of entropy rate, information rate, and conditional information rate is the book by Pinsker [13]. The entropy rate of a stochastic process $\boldsymbol{X} = \{X_i\}$ is defined as

$$H(\boldsymbol{X}) := \lim_{n \to \infty} \frac{1}{n} H(X_1, \ldots, X_n), \tag{1.2}$$

provided that the limit exists. When the sequence $\{X_i\}$ is independent identically distributed (i.i.d.), then $H(\boldsymbol{X}) = H(X_1)$. When $\{X_i\}$ is a stationary Markov chain, the entropy rate can also be easily calculated (cf. Cover and Thomas [84, Chapter 4]). The significance of the entropy rate of a stochastic process arises from the Asymptotic Equipartition Theorem (AEP) for a stationary ergodic process. Although the entropy rate is well-defined for all stationary processes, its calculation into a closed-form expression is, except in a few special cases, not always feasible. Similarly, for the information rate. When a sequence of i.i.d. random variables $\{X_i\}$ is sent over a discrete memoryless channel with transition matrix $\{w(y|x)\}$, then the information rate $I(\boldsymbol{X};\boldsymbol{Y})$ equals the mutual information

$$I(X_1;Y_1) := \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} w(y|x)\ \log \frac{w(y|x)}{p(y)} = H(Y_1) - H(Y_1|X_1) \tag{1.3}$$

between one input and one output of the channel. The study of information rates in various channel and source models is important, as it is connected with other characteristics such as capacity and the rate-distortion function. Therefore, in this line of research, the asymptotic behavior of the information rates is investigated in various models and under various behavior of the parameters specifying these models.

In the continuous case, and for additive noise channels defined by the operation $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}$, if $\boldsymbol{X} = \{X_i\}$ and $\boldsymbol{Z} = \{Z_i\}$ are independent and $\{X_i\}$ and $\{Z_i\}$ are i.i.d. Gaussian sequences with variances $\text{var}(X_1) = P$ and $\text{var}(Z_1) = N$, respectively, the following famous Shannon formula holds

$$I(\boldsymbol{X}; \boldsymbol{X} + \boldsymbol{Z}) = \frac{1}{2} \log(1 + \frac{P}{N}). \tag{1.4}$$

But as soon as one of the sequences $\boldsymbol{X}$ or $\boldsymbol{Z}$ is not i.i.d. Gaussian, no closed form expression exists. Nevertheless, one can search for an asymptotic expression, the first term of which can be easily evaluated and approximates reasonably well the value of $I(\boldsymbol{X}; \boldsymbol{Y})$. At first this led to the investigation of channels with small input signal $\varepsilon \boldsymbol{X}$ ($\varepsilon \to 0$), or equivalently with large noise, as suggested by Dobrushin around 1970, and carried out in the initial work by Prelov (1970) and Ibragimov and Khas'minskii (1972). A good reflection of a great deal of the work which was done in the area of asymptotics of Shannon-theoretic quantities can be found in the papers described in Section 1.1.2 [189, 201, 202, 208, 209, 213].

(iii) In Section 1.1.3 we have grouped together papers addressing problems and situations where the input and output alphabet of the channels and sources under consideration are discrete and where a Shannon-type coding theorem is proved. We begin with a paper by De Bruyn [139] on iterative code construction with a fixed composition list code. Here, advanced concepts and techniques out of the book of Csiszár and Kőrner [55] are used, such as the method of types, a packing lemma, maximum mutual information decoding, and a formulation of the random coding and the sphere packing bound in terms of types.

Rate-distortion theory [24] considers the fundamental problem of data compression under a minimum fidelity criterion, or maximal allowed distortion. There exists a remark by Shannon (1959) on the duality between source coding w.r.t. a fidelity criterion and channel coding subject to a cost constraint. In rate-distortion theory, the problem of successive refinement was investigated by Koshelev [46] and Equitz and Cover [85]. Koshelev and Van der Meulen [203] introduced and analyzed the complementary problem of successive channel coding under increasing cost constraints and obtain sufficient conditions for so-called channel divisibility.

The *multiple description* problem is a rate-distortion theory problem of multi-user information theory, which studies methods for sending different information over the channels, in such a way that if only one channel works, the information received is sufficient to guarantee a minimum fidelity in the reconstruction at the receiver; but should both channels work, the information from both channels can

be combined to yield a higher-fidelity reconstruction. The coding problem was first posed by Gersho, Witsenhausen, Wolf, Wyner, Ziv and Ozarow in 1979 and is still an open problem. A special aspect of the multiple-description problem is minimum breakdown degradation, which is investigated in [155].

(iv) Section 1.1.4 brings together papers dealing with Shannon-type coding theorems for channels with continuous input and output alphabets. Willems [274] investigates the Gaussian side information channel and derives a lower and upper bound for its capacity. In [169], Willems gives a rigorous proof in terms of ε-typical sequences of the result by Shannon [4] that the capacity $\mathcal{C} = \frac{1}{2}\log(1 + P/N)$ can be achieved for an AWGN channel.

Baggen and Wolf [176, 177, 190] introduce and analyze the at that time new concept of a timing jitter channel. Hekstra [178] considers the jitter channel from a different perspective.

Verdú, visiting the 22nd Symposium, introduces the Benelux Information Theory community to new tools for the analysis of power-limited infinite bandwidth channels (also called "very noisy" channels) using the concept of spectral efficiency [210], a topic on which he gave a plenary lecture one year later at the 2002 IEEE International Symposium on Information Theory in Lausanne.

(v) The area of statistical information theory originated with the book of Kullback (1959). In Section 1.1.5 we have grouped together papers which investigate statistical problems involving information-theoretic concepts, such as entropy estimation [126, 166], testing statistical hypotheses using entropy [126, 145], and consistency of statistical estimation procedures as measured by information divergence [192].

Ahlswede and Csiszár [69] introduced the problem of hypothesis testing under communication constraints. Shi [164] continues these investigations. Besides Shannon's information measure, the Fisher information plays an important role in statistical information theory. For a random variable Y with absolutely continuous density $f_Y(y)$, it is defined by

$$\mathcal{J}(Y) := \int_{-\infty}^{\infty} \left[\frac{f_Y'(y)}{f_Y(y)}\right]^2 f_Y(y)\, dy. \tag{1.5}$$

In [193] Prelov and Van der Meulen investigate the Fisher information of the sum of two independent random variables, one of which is small, and obtain an asymptotic generalization of De Bruijn's identity, cf. [84, p. 494].

(vi) Section 1.1.6 is devoted to work in the intriguing area of "ordering". This research domain was originated by Ahlswede, Ye and Zhang (1988). Here the aim is to create order in sequence spaces by information-theoretic methods. In [170], Ye reports on new results in this area.

(vii) We conclude this chapter with a section on applications of Shannon Theory. These concern applications toward human perception, the judged complexity of patterns, economics, system theory, and guidelines for mobile robot design.

1.1.1 Entropy, Foundations, Information Measures, Randomness, and Uncertainty

Shannon [3] and Fisher [1] introduced information measures which gave rise to large research areas. Shannon's information measure finds an important motivation in the source coding theorem, and Fisher's information measure finds application in the Cramér-Rao inequality for the variance of estimators. Later, other information measures were developed which aimed to generalize and extend the properties of the previous two.

In [115], Boekee discusses such new measure, the R-norm information, and its properties. This information measure is pseudo-additive, continuous, symmetric and concave. It yields Shannon's entropy as $R \rightarrow 1$. Boekee [115] also derives a source coding theorem for the R-norm information by a suitable choice of the length-measure of a code satisfying the Kraft-inequality.

Van der Lubbe [120] continues these investigations and compares three different information measures, the Renyi information measure of order α, the information measure of type β due to Daroczy, and the R-norm information. He discusses their properties, and the relationships between their conditional versions with the Bayes error probability. He also derives source coding theorems for the Renyi, Daroczy and Arimoto information measures.

Broekstra [116] addresses the problem of the identification of the structure of a relation between variables in a system. The question here is whether a certain relation R can be decomposed in marginal relations such that R can be reconstructed by a collection of marginal relations with acceptable approximation. The amount of structure in a system of variables is measured by the concept of structural constraint. According to [116], constraint analysis, based on information theory, in particular information measures, can be an effective method for structure identification.

In information theory one usually assumes a stochastic model, where generated symbols are interpreted as realizations of a stochastic process. In a syntactic model, symbol sequences (sentences) are generated without the assumption of an underlying stochastic model. The usual probabilistic approach can therefore not be applied to capture the amount of information in such sentences. Kolmogorov (1965) defined the notion of complexity of a symbol sequence as the length of the shortest binary computer program that describes the sequence. Boekee [119] introduces the concept of *syntactic information* by defining the complexity of a sentence generated by a context-free grammar, and derives from this a measure for syntactic information.

Cover (1975) introduced the concept of ε-typical sequences, cf. [52]. Barbé [156] introduces, as a generalization, the notion of α-typical sequences. It is based on the maximally attainable distance between the actual and expected frequency of successes in a sequence of Bernoulli trials, such that the probability of the set of all sequences satisfying this distance is at most α. Barbé [156] observes that α/ε-typical sequences are not necessarily typically random, but so-called derivative sequences of the basis sequence may be. He develops a theory of higher order derivative sequences, derivative fields, and multi-level α-typical randomness. He shows that the asymptotic equipartition properties remain valid for the α-typical randomness set.

Kamminga [153] discusses the uncertainty principle as applied to the field of signal processing. The classical Heisenberg / Weyl uncertainty relations use the formalism of quantum mechanics. Kamminga presents both Gabor's and Leipnik's form of the uncertainty relation between the time duration for a signal and the frequency width of its Fourier transform. Whereas Gabor (1946) introduced the uncertainty relation in communication theory, Leipnik's uncertainty relation is based on Shannon's information measure.

In [199], De Bruin and Kamminga continue the investigations of [153]. Using the definition of Shannon's entropy, they study the sum of entropy integrals $H_t(s(t)) + H_f(S(f))$ of a Fourier transform pair $(s(t), S(f))$ and show that normalization of the Fourier pair by absolute value integrals in the time and frequency domain leads to a shift and scaling invariant entropy sum. Based on numerical evidence, it is conjectured in [199] that Shannon entropy using absolute normalization is minimal for the Gaussian signal.

Kleima [188] discusses the foundation of probability theory, and argues that this is a question of physics. He gives interesting quotes by Shannon [5] and Kolmogorov (1965) on this foundation, which relate to the theory of secrecy and the theory of information transmission, respectively.

The ellipsoid $E_{\boldsymbol{a}}(r)$ in n-dimensional Hamming space $\{0,1\}^n$ is defined as the set of binary vectors $\boldsymbol{x} = (x_1, \ldots, x_n)$, $x_i \in \{0,1\}$, which satisfy the inequality $\sum_{i=1}^{n} a_i x_i \leq r$, where $\boldsymbol{a} = (a_1, \ldots, a_n)$, $a_i \geq 0$, and $r > 0$. The entropy of the ellipsoid is defined as the logarithm of its cardinality. Pinsker (2000) found an asymptotic representation for it. The ε-entropy H_ε of $E_{\boldsymbol{a}}(r)$ is defined as the logarithm of the minimum number of balls of radius ε which cover the ellipsoid. In [211], Prelov and Van der Meulen investigate the asymptotic behavior of H_ε as $n \to \infty$, when the coefficients a_i take on only two different values. They obtain explicit expressions for the main terms of the asymptotic representation for the ε-entropy of such ellipsoids, under different relations between ε and the parameters defining these ellipsoids.

1.1.2 Asymptotics of Information Rates, Entropy and Mutual Information in Stationary Channels

When the input signal of a continuous alphabet memoryless channel satisfies certain constraints, the evaluation of its capacity requires the optimization of the mutual information function over all probability distributions from a certain class. This is why for most continuous alphabet channels the capacity cannot be calculated explicitly, except for the specific case of an additive white Gaussian noise channel with an energy constrained input. This explains the interest in the investigation of the asymptotic behavior of the capacity of communication channels in situations where certain parameters characterizing the transmission can be designated as small.

Prelov and Van der Meulen [189] derive an asymptotic expression for the Shannon mutual information between the input and output signals of continuous alphabet memoryless channels with weak input signals when the input space is multidimensional. This extends a result by Ibragimov and Khas'minskii (1972) for the one-dimensional case. This asymptotic expression relates the Shannon mutual information and the Fisher information matrix.

Let $\boldsymbol{\xi} = \{\xi_i\}$ and $\boldsymbol{\zeta} = \{\zeta_i\}$ be independent discrete-time second order stationary processes, and consider the stationary channel with an additive noise whose output signal $\boldsymbol{\eta} = \{\eta_i\}$ is equal to the sum $\boldsymbol{\eta} = \varepsilon\boldsymbol{\xi} + \boldsymbol{\zeta}$ where $\varepsilon > 0$ is some constant. The information rate in such a channel is defined as $\bar{I}(\varepsilon\boldsymbol{\xi}; \boldsymbol{\eta})$ where

$$\bar{I}(\boldsymbol{X}; \boldsymbol{Y}) := \lim_{n\to\infty} \frac{1}{n} I(X_1^n; Y_1^n) \tag{1.6}$$

where $I(\cdot;\cdot)$ is the mutual information and $X_1^n := (X_1, \ldots, X_n)$.

In the case where $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are Gaussian, an explicit formula for $\bar{I}(\varepsilon\boldsymbol{\xi}; \boldsymbol{\eta})$ in terms of the spectral densities of the processes $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ is known (cf. Pinsker, 1964). If $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are not Gaussian, the problem of the explicit calculation of $\bar{I}(\varepsilon\boldsymbol{\xi}; \boldsymbol{\eta})$ is rather hard. Therefore, it is of interest to investigate the asymptotic behavior of $\bar{I}(\varepsilon\boldsymbol{\xi}; \boldsymbol{\eta})$ as $\varepsilon \to 0$. This corresponds to a weak signal transmission over the channel in question.

Pinsker, Prelov and Van der Meulen [201] consider the case where $\boldsymbol{\xi}$ and $\boldsymbol{\zeta}$ are obtained by a reversible linear transformation L from a stationary weakly regular process $\boldsymbol{X}$ and a sequence of i.i.d. random variables $\boldsymbol{Z}$, respectively, and obtain an asymptotic expression for the information rate $\bar{I}(\varepsilon\boldsymbol{\xi}; \varepsilon\boldsymbol{\xi} + \boldsymbol{\zeta})$ as $\varepsilon \to 0$ under several assumptions on L and the density function of the noise process.

In [202], Pinsker, Prelov and Van der Meulen consider a general class of stationary channels with a random parameter $\boldsymbol{U} = \{U_i\}$ which is assumed to be a completely singular stationary process independent of the input signal $\boldsymbol{X} = \{X_i\}$. Rather general sufficient conditions are established under which the information rate $\bar{I}(\boldsymbol{X}; \boldsymbol{Y})$ and conditional information rate $\bar{I}(\boldsymbol{X}, \boldsymbol{Y} | \boldsymbol{U})$ coincide, where $\boldsymbol{Y} =$

$\{Y_i\}$ is the output signal. Examples of such channels are provided by channels with additive and/or multiplicative noise ($\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{U}$, $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{X}$, or $\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{X} + \boldsymbol{Z}$ with $\boldsymbol{Z}$ independent of $\boldsymbol{X}$ and $\boldsymbol{U}$).

In [208], Pinsker, Prelov and Van der Meulen consider the problem of calculating the information rate in stationary memoryless channels with additive noise $\boldsymbol{Z}$ and a slowly varying input signal $\boldsymbol{X}$, so that the output is $\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{Z}$. It is not assumed that the power of the input signal goes to zero or that the noise goes to infinity, but rather that $\boldsymbol{X} = \boldsymbol{X}_\varepsilon$ is a finite state stationary Markov chain with transition probabilities tending to zero or one as $\varepsilon \to 0$. Moreover the noise process $\boldsymbol{Z}$ is assumed to be a sequence of i.i.d. random variables, so that the channel is memoryless. Under these assumptions it is shown that the information rate $\bar{I}(\boldsymbol{X}; \boldsymbol{Y})$ is asymptotically equivalent to the entropy $\bar{H}(\boldsymbol{X}_\varepsilon)$ of the Markov chain, and thus that the main term of the asymptotics does not depend on the channel noise.

The problem of investigation of the information rates, capacity and other informative performances of different channels and communication systems, which is of prime importance in information theory, is closely connected with the problem of finding optimal and asymptotically optimal methods of nonlinear filtering and the investigation of their performances in various models of observations. A relationship between information theory and filtering was first observed by Gelfand and Yaglom in 1957.

Let $(\boldsymbol{X}, \boldsymbol{Y})$ be a two-dimensional, discrete-time, second-order stochastic process, where $\boldsymbol{X} = \{X_i\}$ and $\boldsymbol{Y} = \{Y_i\}$ are the unobservable and observable components, respectively. The problem of optimal filtering for the process $\boldsymbol{X}$ consists of constructing, for each time instant n, the optimal (in a certain sense) estimate of X_n from the observations $\{Y_i, i \le n\}$ or from the observations $\{Y_i, -\infty < i < \infty\}$. The implementation of the optimal, nonlinear filter is almost impossible, except for a number of special cases (such as a Gaussian one). Therefore, sub-optimal filters, upper and lower bounds, and asymptotic behavior of the optimal filtering error have been intensively investigated, also by information theoretic methods. In [209] Prelov and Van der Meulen describe some examples of recent results in this direction.

In [213] Prelov and Van der Meulen consider a general class of nonlinear channels with non-Gaussian noise $\boldsymbol{Z}$, defined by the operation $\boldsymbol{Y} = \varepsilon f(\boldsymbol{X}) + \boldsymbol{Z}$, where the transmitted signal $\varepsilon f(\boldsymbol{X})$ is a random function of the input signal $\boldsymbol{X}$. The parameter $\varepsilon > 0$ characterizes the signal-to-noise ratio in the channel. $\boldsymbol{X}$,$f(\boldsymbol{X})$, and $\boldsymbol{Z}$ are assumed to be mutually independent random variables. If $f(\boldsymbol{X}) = \varphi(\boldsymbol{X}, \boldsymbol{U})$ where $\varphi(\cdot, \cdot)$ is a non-random function and $\boldsymbol{U}$ is a random variable independent of $\boldsymbol{X}$ and $\boldsymbol{Z}$, the above model reduces to the model $\boldsymbol{Y} = \varepsilon\varphi(\boldsymbol{X}, \boldsymbol{U}) + \boldsymbol{Z}$ of a channel with a random parameter $\boldsymbol{U}$. For the special cases $\varphi(\boldsymbol{X}, \boldsymbol{U}) = \boldsymbol{U}\boldsymbol{X}$ or $\varphi(\boldsymbol{X}, \boldsymbol{U}) = \boldsymbol{X} + \boldsymbol{U}$ one obtains the models $\boldsymbol{Y} = \varepsilon\boldsymbol{U}\boldsymbol{X} + \boldsymbol{Z}$ or $\boldsymbol{Y} = \varepsilon\boldsymbol{X} + \boldsymbol{Z} + \varepsilon\boldsymbol{U}$ which can be considered as a one-dimensional real-case fading channel and a channel with an additional, contaminating weak noise $\varepsilon\boldsymbol{U}$, respectively. In [213], the higher order asymptotics of the mutual information $I(\boldsymbol{X}; \varepsilon f(\boldsymbol{X}) + \boldsymbol{Z})$ in such

channels is obtained up to terms of order $o(\varepsilon^n)$, as $\varepsilon \to 0$, where n is a given integer ($n \geq 2$), under some conditions on the smoothness and the tails of the probability density function of the noise $\boldsymbol{Z}$.

1.1.3 Shannon-Type Coding Theorems for Discrete Memoryless Channels and Sources

A discrete memoryless one-way channel (DMC) consists of a finite input alphabet $\mathcal{X}$, a finite output alphabet $\mathcal{Y}$, and a transition probability matrix $w(y|x)$, such that

$$w(\boldsymbol{y}|\boldsymbol{x}) = \prod_{i=1}^{n} w(y_i|x_i) \tag{1.7}$$

for all $\boldsymbol{x} \in \mathcal{X}^n$, $\boldsymbol{y} \in \mathcal{Y}^n$. A list code of size L for a set of M codewords has the property that the decoder maps each received sequence $\boldsymbol{y}$ into a list of $1 \leq L \leq M$ messages. A list decoding error occurs if the transmitted message is not on the list of L messages.

In [139], De Bruyn derives a packing lemma for DMCs with fixed composition list codes (FCLCs), i.e., where all M codewords have the same type. Next, De Bruyn derives a random coding bound and a sphere-packing bound for FCLCs, thereby making precise certain statements in Csiszár and Kőrner [55]. Furthermore, De Bruyn [139] gives an iterative code construction of an FCLC used on a DMC such that the corresponding list code (using a maximum mutual information list decoder) satisfies the above-mentioned random coding bound.

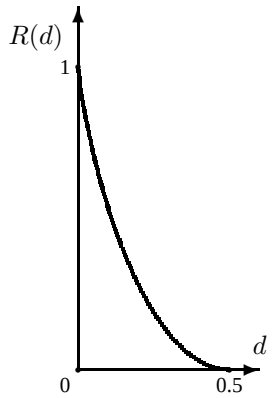


Figure 1.1: *Rate-distortion function for a binary symmetric source.*

For the multiple description problem, consider the situation where two binary channels are used to send information so that even if one channel fails, some data can still be delivered. The rate-distortion function $R(d)$ for a binary symmetric source ($p = 1/2$) and Hamming distortion equals $1 - h(d)$, see Figure 1.1. Remijn [155] considers the problem of minimum breakdown degradation, when

only two binary description channels are available, in the case of no rate excess. The latter means that $R_1 + R_2 = 1 - h(d_0)$, where d_0 is the allowed distortion when both channels are working. The minimum breakdown degradation d_{min} is in this case defined as the smallest achievable distortion when only one of the channels is working. Remijn [155] relates the problem of finding d_{min} to the situation where the decoder must reproduce the source without error if both channels are working. He obtains the value $d_{min} = (\sqrt{2} - 1)/2$, which was also determined by Zhang and Berger [61] using another method.

Koshelev and Van der Meulen [203] explore the duality between source and channel coding, as pointed out by Shannon (1959), from the point of view of successive or hierarchical coding. Multi-level source coding was initiated by Koshelev in 1978 [46], and later investigated by Equitz and Cover [85] under the name of successive refinement of information. Let $R(D)$ denote the rate-distortion function of a source for a given distortion measure. In the problem of multi-level source coding one seeks first an asymptotically optimal description of the source at rate $R_1 \geq R(D_1)$ with distortion not exceeding D_1, followed by an asymptotically optimal refined description at rate ΔR with distortion not exceeding $D_2 < D_1$. The main question is what the minimal value for ΔR is, and whether $\Delta R = R(D_2) - R(D_1)$ can be achieved. Koshelev [46] and Equitz and Cover [85] provided sufficient and necessary conditions for so-called source divisibility.

In [203], Koshelev and Van der Meulen introduce the analogous problem for channel coding, i.e., multi-level channel coding subject to a sequence of increasing cost constraints. Let $C(\tau)$ denote the capacity-cost function, representing the maximum amount of information one can reliably transmit over a DMC at a cost not exceeding τ per channel input, cf. [43, 62]. In multi-level channel coding subject to cost constraints, the goal is to first achieve a coding rate $R_1 \leq C(\tau_1)$ for a code satisfying cost constraint τ_1, and then to send supplementary information at rate ΔR, such that the resulting two-level code satisfies cost constraint $\tau_2 > \tau_1$. The channel is called divisible if $\Delta R = C(\tau_2) - C(\tau_1)$. In [203], Koshelev and Van der Meulen present a coding theorem characterizing the achievable points $(R_1, \Delta R, \tau_1, \tau_2)$, and provide sufficient conditions for channel divisibility.

1.1.4 Gaussian Noise Channels, Jitter Channels, and Power-Limited Infinite Bandwidth Channels

In [274], Willems investigates the Gaussian side information channel, and derives a lower and an upper bound for its capacity. This channel is defined by $Y = X + S + Z$, where S and Z are Gaussian random variables with mean zero and variances N_1 and N_2 respectively. The codewords must satisfy a power constraint P. Shannon [12] found the capacity of the d.m. channel with side information at the transmitter. For the Gaussian channel with side information the capacity $\mathcal{C}$ is

unknown. Willems [274] finds that, using $Q(x) = \frac{1}{2}\ln(1+x)$, it holds that

$$Q\left(\frac{P}{N_1+N_2}\right) \leq \mathcal{C} \leq Q\left(\frac{P}{N_2}\right). \tag{1.8}$$

Shannon [4] proved that the capacity $\mathcal{C} = Q(P/N)$ of the additive white Gaussian noise (AWGN) channel can be achieved, using a geometrical argument. Cover developed the technique of typical sequences to give achievability proofs for discrete multi-user channels. This technique does not work for the Gaussian case, as the cardinality of the typical set in the continuous case cannot be bounded. Willems [169] shows that this difficulty can be overcome and gives a rigorous achievability proof for the single-input, single-output AWGN channel in terms of jointly typical sequences.

In communication theory, one usually assumes that timing is perfect, so the only uncertainty comes from (additive) noise. In 1990, Baggen and Wolf [176] describe a physical situation where timing uncertainty is the limiting factor, resulting in jitter, i.e., wrong timing alignment, at the receiver end. They obtain an upper bound on the capacity of the d.m. *timing jitter channel* (TJC). This work is continued in [177], where a formal proof is given of the capacity of the TJC.

Hekstra [178] proposes a different channel model for timing jitter, the discrete memoryless increments (DMI) TJC, by considering the time shifts as random variables and derives the capacity of this channel model in terms of mutual information. He points out that the capacity of this DMI TJC corresponds to the channel capacity per unit cost, defined by Verdú (1990).

In 1993, Baggen and Wolf [190] consider the combination of additive noise and jitter on the AWGN channel and derive upper bounds on the capacity. They show that in the presence of jitter, capacity is upper bounded, even when signal power is unbounded.

Verdú [210] deals with discrete-time additive noise channels in a general setting (with m complex dimensions) which allow for certain channel impairments such as fading, and investigates the bandwidth/power trade-off for this class of channels in the wide-band regime when the spectral efficiency is small but nonzero. He observes that the trade-off between power and bandwidth is reflected by the trade-off between the information-theoretic quantities spectral efficiency and E_b/N_0 (energy per bit normalized to background noise level), and uses an approach for the wide-band regime to approximate spectral efficiency as an affine function of E_b/N_0.

1.1.5 Information Theory and Statistics

The problem of estimating the entropy of a statistical distribution is well-known in information theory. Shannon [6] already investigated the problem of estimating the entropy of printed English. More generally, one can pose the question how

to estimate the entropy of an unknown distribution, based on a sample of n i.i.d. observations. Here one distinguishes between finitely discrete, finitely denumerable, and absolutely continuous distributions having a probability density function (pdf). The Shannon (or differential) entropy $H(f)$ of a continuous pdf $f(x)$ is defined by

$$H(f) := -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx. \tag{1.9}$$

In [126], Van der Meulen describes an estimate of the entropy of a continuous distribution, based on the order statistics of a sample from the distribution. Exploiting the maximum entropy property of the normal distribution (when the variance is fixed) and of the uniform distribution on the unit interval, one can use this estimate to construct a test for the composite hypothesis of normality and for testing uniformity. In [126] Van der Meulen describes the principle behind these testing procedures and reports Monte Carlo results on the power of the entropy-based test of uniformity, with applications toward the evaluation of random number generators.

In [145] Smit presents a test for the order of a finite-state Markov chain based on the concept of entropy. He assumes a source $Y(a)$, which is modeled by a stationary, aperiodic, irreducible, discrete-time Markov chain of unknown finite order a. The problem is to estimate the order a based on one realization of n symbols of $Y(a)$. Let $X(n)$ be the n-th Markov-approximation of $Y(a)$, and $\hat{X}(n)$ an estimate of $X(n)$ based on the relative frequencies of occurrence of states and transitions in the realization of $Y(a)$. Smit [145] then proposes to use the difference between the entropy of $\hat{X}(n)$ and that of $\hat{X}(n+1)$ as test statistic for the hypothesis that the order equals a for an experimentally determined choice of n.

Ahlswede and Csiszár [69] investigated the problem of testing the hypothesis H_0 : $p(x, y)$ versus the alternative H_1 : $q(x, y)$ for a discrete distribution on a finite set $\mathcal{X} \times \mathcal{Y}$ under communication constraints. They derived an exponent function involving a two-dimensional information divergence based on blocks of length n and the rate of compression, describing the performance of this test. The explicit characterization of this exponent is hard, and Ahlswede and Csiszár provided a lower bound on it. Shi [164] considers the characterization of this exponent for testing the hypothesis H_0 with one-sided data compression, and proposes a characterization of it which he calculates to give larger values than the lower bound of Ahlswede and Csiszár for an example where $\mathcal{X} = \mathcal{Y} = \{0, 1\}$.

In [166], Györfi and Van der Meulen introduce a general class of entropy estimators for estimating the Shannon (or differential) entropy $H(f)$ of a continuous pdf $f(x)$. The general feature of these estimators is that they are based on an L_1-consistent density estimator $\hat{f}_n(x)$. They first consider entropy estimators which involve a histogram-based density estimator $\hat{f}_n(x)$, and state conditions under which these estimators converge a.s. to $H(f)$, with as only condition on f that $H(f)$ is finite. Furthermore, they determine which additional properties

one should impose on an L_1-consistent density estimator $\hat{f}_n(x)$ (not necessarily histogram-based) such that the corresponding empiric entropies are almost sure consistent.

Let $D(f, g)$ denote the information divergence between densities f and g, defined as

$$D(f, g) := \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} \, dx. \tag{1.10}$$

In [192], Györfi and Van der Meulen show that for any sequence $\{\hat{f}_n\}$ of density estimates there is a density f with finite differential entropy $H(f)$ and arbitrarily many derivatives such that $D(f, \hat{f}_n) = \infty$ for all n a.s. This is equivalent to saying that a smooth pdf with finite differential entropy cannot be estimated consistently in information divergence. They also show that, on the other hand, under mild tail and peak conditions on the density functions the almost sure consistency in information divergence can be guaranteed for a suitably defined density estimate.

Prelov and Van der Meulen [193] derive an asymptotic expression for the Fisher information of the sum of two independent random variables X and Z, when Z is small. This asymptotic expression is valid under some regularity conditions on the probability density function of X and conditions on the moments of Z. The first term of the expansion is the Fisher information of X. An asymptotic generalization of De Bruijn's identity is obtained, which provides a relationship between differential entropy and the Fisher information.

1.1.6 Ordering in Sequence Spaces

In [78], an interesting new coding problem is analyzed: how much 'order' can be created in a 'system' when the 'knowledge about the system' and the possible 'manipulations on the system' are restricted? More specifically, the system under consideration consists of binary sequences and the rate or *efficiency* of an ordering algorithm is measured by the logarithm of the total number of different output sequences divided by the sequence length.

Without constraints, the asymptotical rate is 0, since there are only $n + 1$ fully sorted binary sequences of length n. However, for ordering purposes, the algorithm is restricted to operate within a sliding window of size β: only the elements within the window are allowed to be interchanged. If the algorithm has full knowledge of the input sequence, the optimal rate is $1/\beta$. Limitations on the knowledge give higher rates.

In his 1989 contribution, Ye [170] gives a new upper bound for the case of a time-varying algorithm, and proves a conjecture in the case where the incoming order of the elements in the window is exploited.

1.1.7 Applications of Shannon Theory

There are many applications of information theory outside the strict IT domain (i.e., the domains covered by the chapters of this book). The 25 years of information theory in the Benelux have seen several noteworthy applications of the techniques or the results of Shannon theory in other areas. Sometimes, such contributions have even led to new research areas, as can be seen from a glance through this book. Other applications have not (yet) led to fully developed domains of their own, but they witness the broad applicability of information theory.

One area where information theory has been successfully used is psychology. Around 1955, several researchers tried to use selective information theory to understand human perception and especially the judged complexity of patterns. In 1980, Buffart and Collard [117] outline the importance of using information-theoretic complexity measures to quantitatively describe coding efficiency, in order to objectively derive simple representations of a pattern. Collard [125] gives more details on the encoding of structural information in his 1982 contribution.

Another application area is economics. In 1967, Theil published a book on information theory in economics. In his 1983 contribution, Van der Lubbe [132] broadens Theil's approach (based on entropy) to certainty and information and applies this to the concentration index, which measures the uneven distribution of economic goods in a population.

In system theory, De Moor and Vandewalle [157] approach the problem of identifying linear relations from noisy data from an information channel viewpoint: uncertainty in the initial data reflects itself in the uncertainty of the solution set.

In [204], Levendovsky, Kovács, Koller and Van der Meulen propose a new algorithm for adaptive modeling. In order to achieve high performance, the modeling capability of the adaptive system should be of the same degree as of the unknown system. Undermodeling results in loss of performance, whereas overmodeling uses the modeling resources inefficiently. This is typically the case in adaptive noise cancellation when multi-channel cancellation must be performed by a single digital signal processor. As a result, traditional modeling algorithms such as recursive least mean squares must be modified in order to be able to model the degree of the system properly. The methods proposed by Akaike and Rissanen use information-theoretic measures to estimate this degree. These estimation procedures provide rough estimates in practice. The adaptive filter degree algorithm, proposed in [204], not only adapts the weights of an FIR filter, but also adaptively determines the filter degree needed for modeling the system.

Information theory is used by Badreddin [207] to obtain guidelines for mobile robot design, in an abstract setting. Some other research areas have developed more extensive applications of information theory; these areas are covered by subsequent chapters of this book, most notably the chapters on signal processing and on image and video compression.

1.2 Multi-User Information Theory

Multi-user information theory is the part of Shannon theory dealing with communication situations where there is more than just one sender and one receiver. In general, one can think of a channel with several senders and several receivers, where each of the outputs is statistically dependent on each of the inputs. In the discrete memoryless (d.m.) case, there is a transition probability matrix for every receiver, giving the probability of any output symbol, given the input symbols of all channel inputs.

Multi-user information theory originated with Shannon's landmark 1961 paper [14] on the two-way channel (TWC), in which he gave a detailed analysis of this channel. In a TWC two terminals, which are each both sender and receiver, communicate with each other.

In contrast to the one-way communication situation, for most multi-user channels channel capacity has not yet been fully determined. Since there are two or more sources, capacity is multi-dimensional. This leads to the concept of *capacity region* (CR), which is the region containing all rate tuples for which transmission with arbitrarily small error probability is possible. This region is convex, since one can obtain the rate points on the line interval between two points by time-sharing the coding schemes of the end points.

Also in contrast to one-way channels, coding for deterministic multi-user channels is not necessarily trivial, since it involves an interesting trade-off between the transmission rates of the different communication links. Often there exist coding schemes that operate well above the time-sharing line: this means that the terminals can cooperate by using a cleverly designed coding scheme, even without actual mutual communication during transmission (apart from direct use of the channel).

One of the main goals of multi-user information theory is to determine the performance limits of the corresponding channel, i.e., to find an expression for its CR. Shannon [14] found a limiting expression but no single-letter characterization for the capacity region of the general TWC. In fact, the latter is one of the many open problems in multi-user information theory and has resisted a solution for more than 40 years now.

It took several years for the information theory community to assimilate the ideas of Shannon's TWC paper, but at the end of the 1960s and the beginning of the 1970s the field of multi-user information theory gradually emerged.

Apart from the TWC, the following four basic models were defined and investigated:

- The *multiple-access channel* (MAC), where there are two or more senders and just one receiver terminal. The MAC was mentioned as a model by Shannon (1961) but the first investigations on its CR were reported only in

1971 (Ahlswede, Liao).

- The *interference channel* (IFC), with two sender/receiver pairs. The IFC was also mentioned as a possible model by Shannon (1961) but first results on the determination of its CR only appeared in the early 1970s (Ahlswede, Sato, Carleial).
- The *relay channel* (RC), where there is one sender, one receiver and one helper terminal which both receives and sends information. The RC was introduced and first analyzed by van der Meulen (1968).
- The *broadcast channel* (BC), where there is just one sender and two or more receivers. The BC was introduced by Cover (1972).

An interesting, more general communication situation is considered by Salehi and Willems in 1991 [181]: n terminals transmit their message to the others through a 'ring-shaped' network, i.e., there are only one-way connections from terminal i to $i+1$ (modulo n). A single-letter expression is derived for the rate n-tuples for the source coding aspect of this communication situation. For the channel coding aspect, capacity is derived only for the case $n = 2$ and when the channel is deterministic.

In the remaining sections of this chapter we systematically describe the results which were obtained by researchers in the Benelux on the five basic multi-user channel models mentioned above (TWC, MAC, BC, IFC, and RC) and some other, closely related, communication situations.

1.2.1 The Two-Way Channel (TWC)

The TWC has two terminals, each with an input and an output (see Figure 1.2). The output at each terminal is statistically dependent on both inputs. The capacity region, $\mathcal{C}$, of a TWC is the region of achievable rate pairs (R1,R2), i.e. rate pairs that allow essentially error free simultaneous transmission. Shannon [14] derived inner bound and outer bound regions to the capacity region of the TWC in terms of mutual information expressions:

$$\begin{aligned} G_i := \{(R_1,R_2)| \quad & 0 \le R_1 \le I(X_1;Y_2|X_2), \\ & 0 \le R_2 \le I(X_2;Y_1|X_1), \\ & P_{X_1X_2} = P_{X_1} \cdot P_{X_2}\}, \end{aligned} \tag{1.11}$$

$$\begin{aligned} G_o := \{(R_1,R_2)| \quad & 0 \le R_1 \le I(X_1;Y_2|X_2), \\ & 0 \le R_2 \le I(X_2;Y_1|X_1), \\ & \text{arbitrary joint } P_{X_1X_2}\}. \end{aligned} \tag{1.12}$$

These fundamental bounds are generally referred to as "Shannon's inner bound" and "Shannon's outer bound" in the literature. The inner bound region results from independent input probabilities, the outer bound region requires a joint probability.

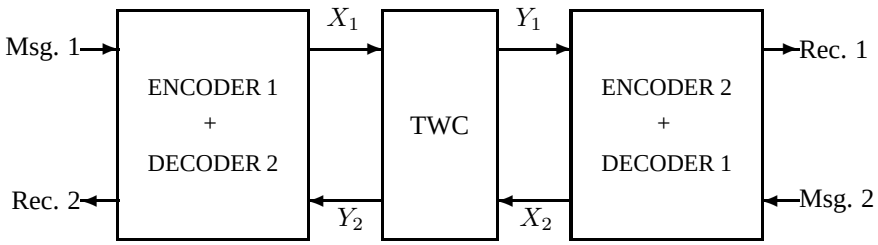


Figure 1.2: *Block diagram of the two-way channel (TWC).*

Sometimes inner and outer bound regions coincide, in which case the capacity region of the particular TWC is known. Of interest are those channels where inner and outer bound differ, i.e., cases where the capacity region is not known.

The prime example of such a TWC is the *binary multiplying channel* (BMC), attributed to Blackwell, where all four alphabets are binary and both outputs are identical and equal to the product of the inputs: $Y_1 = Y_2 = X_1 \cdot X_2$. Note that the BMC is a deterministic channel since both Y_1 and Y_2 are functions of X_1 and X_2. For a deterministic TWC, Shannon's bounds reduce to

$$\begin{aligned} G_i := \{(R_1, R_2) | \quad & 0 \leq R_1 \leq H(Y_2|X_2), \\ & 0 \leq R_2 \leq H(Y_1|X_1), \\ & P_{X_1X_2} = P_{X_1} \cdot P_{X_2}\}, \end{aligned} \tag{1.13}$$

$$\begin{aligned} G_o := \{(R_1, R_2) | \quad & 0 \leq R_1 \leq H(Y_2|X_2), \\ & 0 \leq R_2 \leq H(Y_1|X_1), \\ & \text{arbitrary joint } P_{X_1X_2}\}. \end{aligned} \tag{1.14}$$

Deterministic one-way channels are, in general, not that interesting. However, as in the TWC the information flowing in the direction from terminal 1 to terminal 2 interferes with the information flowing in the opposite direction, one does not need channel noise to make the problem interesting!

Gaal and Schalkwijk [141] classify all 256 binary deterministic TWCs: there are 17 mutually non-equivalent channels, only two of which are non-trivial. Only the BMC has non-coinciding Shannon inner and outer bounds, see Figure 1.3 (a). The other non-trivial channel is $Y_1 = X_1 \cdot X_2$, $Y_2 = X_2$ and it has the capacity region of Figure 1.3 (b).

In [212], Von Haeseler and Barbé generalize the coding problem of the BMC by considering an arbitrary ring $\mathcal{R}$ as the input alphabets, and the channel action as the element-wise product of the two input vectors. If $\mathcal{R}$ consists of the $n \times n$ matrices over F_q, the largest possible uniquely decodable code is the set of all invertible matrices. Also for the ring Z_m, the problem is fully solved.

Figure 1.3 (a)

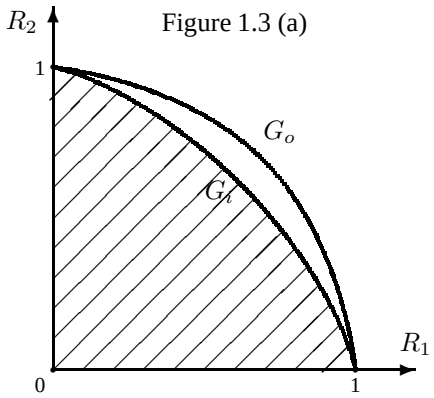
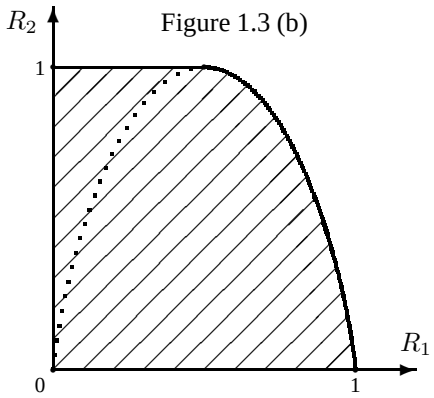


Figure 1.3 (b)



1.2.2 The Binary Multiplying Channel (BMC)

At the end of [14] Shannon remarked that the TWC problem is very difficult. This remark may be part of the reason why between the publication of [14] in 1961 and 1981 very little research on the TWC has been reported. The implicit question of Shannon's 1961 paper was, whether or not there exist coding strategies for the BMC that outperform, for the equal-rate case, the inner bound rate $R_1 = R_2 = 0.61695$. In what follows we will try to sketch in simple terms the steps that eventually led to such a strategy.

The simplest equal-rate coding scheme that operates beyond the timesharing rate $R_1 = R_2 = \frac{1}{2}$ is the following one, attributed to Hagelbarger [14], and it achieves a rate point $(R_1, R_2) = (\frac{4}{7}, \frac{4}{7}) = (0.57142, 0.57142)$ bits per transmission. Both encoders send their binary message bit by bit, where each bit has to be followed by its complement only in the case when the symbol that was received (as a consequence of the other terminal's bit) is a zero. This coding scheme is uniquely decodable. It is schematically represented in the following diagram. The numbers outside of the square represent the information symbols, to the left those of sender 1 and at the bottom those of sender 2, while the numbers inside the square represent the corresponding channel output sequence $\boldsymbol{y}_1$ $(= \boldsymbol{y}_2)$.

1	00	1
0	01	00
	0	1

This coding scheme is a *variable length strategy*, since not all messages require an equally long transmission time. The rate of such a strategy is the reciprocal of the average code word length $\overline{\ell}$ per bit to be transmitted, which in this case is $\overline{\ell} = \frac{7}{4}$, so $R_1 = R_2 = \frac{4}{7}$.

The following coding scheme, described by Schalkwijk and Vinck [128], uses the same idea, but assumes a precoded *ternary* message. This is a simplified version of the strategy presented in [122] (see below), and it enables the authors to clearly demonstrate the essence of the original two-way strategy that was presented earlier.

The encoders send a 0 if the information symbol is a 0, and a 1 otherwise. If they receive a 1, the message pair was (1,1), (1,2), (2,1), or (2,2), which can be resolved as with the Hagelbarger code. If they receive a 0, the message pair was (2,0), (1,0), (0,0), (0,1), or (0,2), which is an L-shape region in the 3×3 square. The encoders work this out further by sending a 0 if the information symbol was a 2, and a 1 otherwise. When a 0 is received, the message pair was (2,0) or (0,2), which is uniquely decodable, and when a 1 is received, one more bit must be sent, namely the information symbol.

2	00	100	11
1	010	101	100
0	011	010	00
	0	1	2

The rate pair of this scheme is $(R_1, R_2) = (\frac{9 \log_2 3}{24}, \frac{9 \log_2 3}{24}) = (0.59436, 0.59436)$. In fact, it can be shown that this is the highest possible rate for the 3×3 message case.

It seems then natural to search for optimum subdivisions (also called *resolutions*) for larger $M \times M$ squares, $M = 4, 5, \cdots$, to thus approach the capacity region. Finding the optimal scheme for alphabet size M thus consists of continuously subdividing the $M \times M$ square, by forcing a channel output 1 in a sub-square of the remaining part. Paper [135] is a first attempt to find such optimal resolutions for $M \times M$ squares of increasing size. This approach has yielded several interesting strategies (see the Table 1.1) but, it is shown later, eventually one gets overwhelmed by the astronomical number of possible resolution strategies on the larger squares.

Table 1.1: *Coding strategy overview table.*

M	$\overline{\ell}$	$R_1 = R_2$
2	$7/4$	0.571428
3	$24/9$	0.594361
5	$98/25$	0.592328
8	$319/64$	0.601881
18	$2216/324$	0.609682
27	$5683/729$	0.609944
58	$32250/3364$	0.611046

Schalkwijk [122] presents the first strategy with a rate point outside Shannon's inner bound region, yielding a common rate $R_1 = R_2 = 0.61914$. The idea underlying this strategy is the following. Consider a binary message sequence, then by putting a decimal point in front of this sequence one obtains a binary expansion of a real number T between 0 (inclusive) and 1 (exclusive), i.e., a message point $T \in [0, 1)$. In the TWC case there are two message points, T_1 for sender 1 and T_2

for sender 2. Thus the combined message pair is represented by a message point $T = (T_1, T_2)$ within the unit square $[0,1) \times [0,1)$. Take two subsets S_1 and S_2 of $[0,1)$. Suppose terminal 1 sends $X_1 = 1$ if $T_1 \in S_1$, and likewise terminal 2 sends $X_2 = 1$ if $T_2 \in S_2$. Then both terminals receive $Y_1 = Y_2 = 1$ whenever $T \in S_1 \times S_2$ and they receive $Y_1 = Y_2 = 0$ if T is in the complement of $S_1 \times S_2$ w.r.t. the unit square. In this fashion the unit square is divided into two sets, i.e. $S_1 \times S_2$ and its complement (see Figure 1.4).

In a similar way each of these two subsets is further divided until the message

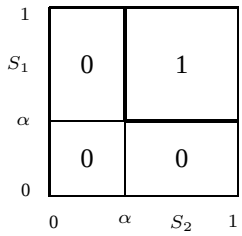


Figure 1.4: *Division of unit square.*

point, T, can be uniquely determined. Schalkwijk calls this type of strategy a *Shannon strategy*.

The original strategy of [122] continually returns to sub-rectangles as resolution products (see Figure 1.5), and only uses resolutions of three different types: an inner bound resolution (for the rectangular regions, see the previous figure), an intermediate one, and a so-called outer bound resolution.

In this way, a first order Markov process with three states is obtained. The rate of

Figure 1.5: *Subdivisions of unit square.*

the complete scheme is an average of the rates of the three strategies, given by the steady state of the Markov process.

By choosing $\alpha = 0.32429$ and $\gamma = 0.52545$ we obtain a rate pair $(0.61914, 0.61914)$. In fact, the intermediate resolution of the original coding strategy can be improved upon, but it was good enough to yield the overall result in excess of Shannon's inner bound rate 0.61695.

After it became apparent that the capacity region of the BMC is strictly larger than its inner bound region, the search for its true capacity region was on.

In his original 1961 paper on TWCs, Shannon showed that the capacity region can be approximated by considering fixed length strategies of increasing length, $n = 1, 2, \cdots$. In fact, the optimum fixed length $n = 3$ strategy [136] comes very close to the variable length strategy by which Schalkwijk obtained the first rate point $R_1 = R_2 = 0.61914$ outside Shannon's inner bound for the BMC. As observed in the paper, the second transmission, the resolution dividing the $y = 0$ region, the $y = 01$ region and the $y = 00$ region can be eliminated using Schalkwijk's bootstrapping technique. One now obtains a very simple equation for a new equal-rate point $R_1 = R_2 = 0.63056$. Because of the simplicity and elegance of the bootstrapped strategy, Schalkwijk initially believed 0.63056 to be on the boundary of the capacity region. However, later much more intricate resolution strategies were found (also using the bootstrapping technique) with slightly higher rates (in the third decimal place). Nevertheless, as János Kőrner says, "the Schalkwijk 1983 strategy *essentially* solves the BMC capacity problem". As of today nobody has been able to determine the capacity region of the BMC.

As observed before, Shannon strategies of increasing length can be seen as resolutions of the unit square. Hence, by somehow upper bounding the efficiency of unit square resolution one could try to tighten Shannon's upper bound to the capacity region. The paper [144] is an effort to upper bound the efficiency of unit square resolutions. However, a mistake right at the beginning of this paper makes the results invalid. Namely, in Figure 2 of that paper, the transition from the α, β thresholds to the α', β' thresholds is, in general, not possible.

The paper [149] is another attempt to construct a converse to unit square resolution. However, in hindsight, it is not possible to get a grip on these resolution strategies that get more and more intricate as the size of the $M \times M$ square increases. A good reference on resolution strategies for larger squares are the Ph.D. theses of Bloemen and Meeuwissen.

Shannon's original paper on TWCs deals with fixed length strategies that have a vanishing probability of error. Schalkwijk's strategy that yielded the rate pair $R_1 = R_2 = 0.61914$ outside the inner bound region is a variable length strategy with zero probability of error. Tolhuizen [150] rigorously proves these fixed and variable length strategies to be equivalent. Tolhuizen shows that Schalkwijk's $R_1 = R_2 = 0.61914$ variable length strategy to be equivalent to a Shannon fixed length strategy. In [158] Van Overveld shows this equivalence to be true for all deterministic T channels, i.e. fixed and variable length strategies yield the same rate if $Y_1 = Y_2$.

In the binary symmetric channel (BSC) we have a simple model that captures the main features of unreliable one-way transmission. Likewise, with the binary two-way echo channel (BTWEC), Schalkwijk [160] tries in a simple model to capture

some essential features of two-way transmission with echoes. Such echoes are, for example, experienced on telephone connections. It is shown that with a simple unit square resolution strategy a rate $R_1 = R_2 = 0.53723$, in excess of the time-sharing rate of 0.5, can be achieved. This echo channel is the first concrete example of a TWC with memory as described in Shannon's 1961 paper. There Shannon shows that TWCs with the recoverable state property do, in fact, have a capacity region. He also says that the concept of a TWC with memory is a very difficult one. Shannon's remark regarding the complexity of TWCs in general, is partly responsible for the fact that very little research on the TWC has been done. Perhaps this remark about the difficulty of the TWC with memory should not hold people back to explore interesting and practically relevant examples of such TWCs.

Shannon showed that strategies of increasing length $n = 1, 2, \cdots$, yield rates approaching the boundary of the capacity region. These fixed length strategies can be represented as unit square resolution strategies. This equivalence allows us to study these Shannon strategies up to say length $n = 8$, as was done by Schalkwijk [167]. Beyond $n = 8$, i.e. Shannon's derived channel K_8, the optimization problem becomes unwieldy. For more on Shannon's derived channels $K_n, n = 1, 2, \cdots, 8$, the reader is referred to the M.Sc. thesis of Smeets. The achievable lower bound $R_1 = R_2 = 0.63056$ of the 1983 bootstrapping scheme is well beyond the rate of K_8. The tightest upper bound, $R_1 = R_2 = 0.64628$, was derived by Hekstra and Willems [148].

Initially, Schalkwijk erroneously thought $R_1 = R_2 = 0.63056$ to be the equal-rate capacity of the BMC. After a long and futile effort to find a converse, [174] he finally succeeded to construct a strategy that improves on the original 1983 bootstrapping scheme in the 8th decimal place! Hence, the problem of the capacity region of the BMC is still open.

The paper [180] is another effort to derive an upper bound on unit square resolution, however, the upper bound found by Hekstra and Willems [148] is sharper. There has been considerable effort to increase the lower bound, 0.63056, on the equal-rate point, see the Ph.D. thesis of Meeuwissen. Several small improvements in rate have been realized, however, 0.630 still stands. The authors suggest to try to lower the upper bound 0.64628. Suggestions made in [183] might be relevant to such an endeavor.

In [184], Bloemen treats the problem of the BMC without feedback, i.e. no strategies but codes are used at both terminals. The ε-error capacity region is known in this case and coincides with the Shannon inner bound region. However, while Shannon considered the case of vanishing probability of error, Bloemen in this paper looks at the stronger requirement of zero probability of error. The simple code found by Benschop yields a rate pair $R_1 = R_2 = 0.52832$, and is hard to improve upon. Later in 1999, Tolhuizen [205] showed that $R_1 = R_2 = 0.58500$ can be achieved. Although $R_1 = R_2 = 0.58500$ is optimal, the full zero error capacity region for the BMC without feedback is unknown.

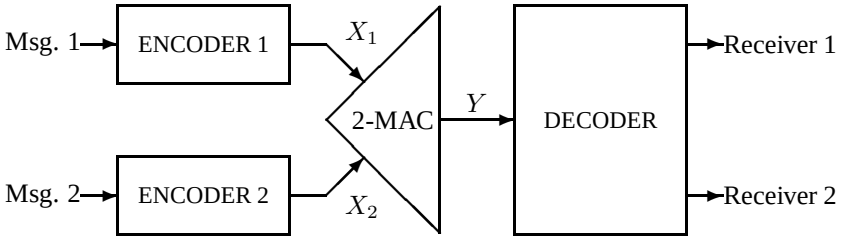


Figure 1.6: *Block diagram of the multiple-access channel.*

Schalkwijk [185] considers an interesting variation on the BMC. Here terminal 1 can use its output Y_1 to construct a code stream that depends on both its message T_1 and on the past Y_1 sequence, i.e. terminal 1 can use a coding strategy. However, the code stream at terminal 2 only depends on its message T_2 and not on Y_2, i.e. terminal 2 is restricted to use a code instead of a strategy. Using the new technique of message percolation Schalkwijk shows that also for the semi-restricted BMC the capacity region is strictly larger than Shannon's inner bound region. It is not known whether the semi-restricted BMC has the same capacity region as the unrestricted BMC.

Bloemen [186] constructs strategies on $M \times M$ squares, $M = 2, 3, \cdots, 25$, using the computer. Meeuwissen [187, 191, 194] extends Bloemen's results to improve Schalkwijk's lower bound 0.63056. Finally, Meeuwissen [198] considers the interesting and realistic result of a TWC with delay. Schalkwijk [197] describes a 2D-weighing technique to find coding strategies.

In conclusion, we can say that considerable progress has been made although the equal-rate capacity of the BMC still eludes us. Between 1981 and 1999 a great effort was made to understand the TWC, i.e. the mathematical dialogue.

1.2.3 Multiple-Access Channel (MAC)

In the communication situation of the T-input multiple-access channel, there are T information sources which are encoded independently by T encoders (see Figure 1.6). The channel thus has T inputs and a single output, which is observed by a single decoder who is to decode all T source messages. The simplest and also the most studied situation is that of the 2-input discrete memoryless MAC (dm-2-MAC). A single-letter expression for the capacity region (CR) of the dm-2-MAC was found in 1971 by Ahlswede and (in a simpler form) in 1972 by Liao:

$$\begin{aligned} \mathcal{C}_{\text{dm-2-MAC}} = \text{co}\,\{(R_1, R_2) | \quad & 0 \le R_1 \le I(X_1; Y|X_2), \\ & 0 \le R_2 \le I(X_2; Y|X_1), \\ & R_1 + R_2 \le I(X_1 X_2; Y), \\ & P_{X_1 X_2} = P_{X_1} \cdot P_{X_2}\}, \end{aligned} \tag{1.15}$$

i.e., a convex hull of the union of pentagon-shaped areas, one for each possible independent assignment of probability distributions to the input alphabets.

In 1981, Van der Meulen [121] gives an overview of recent results for the MAC. He mentions the following five noteworthy facts:

- The discovery of the CR for the dm-2-MAC with uncorrelated sources (Ahlswede, Liao, published in 1974); a strong converse for this coding situation was established 6 years later (Dueck, Ahlswede, 1980).
- The CR of the (power-limited) Gaussian MAC was established shortly thereafter (Cover, Wyner, 1974, 1975). The exact expression is essentially identical to that for the dm-2-MAC.
- For the dm-2-MAC with correlated sources, two (nowadays called "classical") results exist at this moment: when the sources can be decomposed into three independent sources —two private ones and a common one— Slepian and Wolf [32] determined the CR in 1973; in the case of arbitrarily correlated sources, Cover, El Gamal and Salehi (1980) determined an inner bound (or achievable) region.
- Gaarder and Wolf (1975) and Cover and Leung (1976) showed that feedback can increase the capacity of the dm-2-MAC, this in contrast to the one-way channel situation.
- Ozarow (1979) determined the CR of the Gaussian MAC with feedback.

The period between 1982 and 1985 shows a lot of research activity in the Benelux in the area of capacity results for several versions of the MAC communication situation, especially with respect to different amounts of cooperation between the three terminals: either in the form of feedback from channel output to encoders or in the form of cooperation between the encoders.

In 1982, Willems [129] determines the CR of the dm-2-MAC with partially cooperating encoders, in terms of the capacity of the link between the two encoders. Also in 1982, Willems and Van der Meulen [130] determine the CR of the dm-2-MAC with mutually informed (cribbing) encoders, for all five possible cribbing situations, viz., that where one, or both encoders see the full codeword of the other encoder, or only the initial part of it (either including the next symbol to be transmitted or not). Just like for the classical dm-2-MAC, the capacity regions turn out to be the convex hull of the union of pentagons, but in the more 'informed' cases the union must be taken over all *dependent* input distributions $P_{X_1X_2}$, not just $P_{X_1} \cdot P_{X_2}$.

Gaarder and Wolf proved in 1975 that for the binary adder channel ($Y = X_1 + X_2$, see below) with feedback, $R_1 = R_2 = 0.76$ can be achieved, i.e., a rate point outside the non-feedback capacity region ($R_1 + R_2 \leq 1.5$). Willems [123] shows in 1981 that the Cover-Leung region is also achievable with partial feedback (i.e., only feedback to one of the two encoders). He also proves that for a class of MACs,

viz. the ones for which X_1 is a function of Y and X_2, the Cover-Leung region is optimal (i.e., is the CR) in the case of (partial) feedback. The binary adder channel belongs to this class, and the equal-rate point $R_1 = R_2 = 0.79113$, found to be achievable by Van der Meulen (1976), was proved to be optimal by Willems in 1983 [138]. The latter paper also gives an example that shows that the feedback CR of the product of two MACs can be strictly larger than the sum of the CRs of the separate channels, this in contrast to single-user channels.

The dm-2-MACs with feedback are actually equivalent to TWCs for which $Y_1 = Y_2$, the so-called *T channels*. In 1984, Hekstra and Willems [142] prove that the CR of a certain class of T channels equals the Shannon inner bound region. Moreover, when the channels in this class are interpreted as multiple-access channels with feedback, their CR equals the Cover-Leung region. An example is given of a deterministic channel in this class (with ternary alphabets) for which Shannon's outer bound is strictly larger than the inner bound, viz. the channel $Y = |X_1 - X_2|$. In 1985 [148], the same authors give a simpler proof of this result, thereby introducing the concept of dependence increase/decrease of random variables. The result now applies to an even larger class of channels.

In the same time period, there are also four contributions in the area of the so-called "*Slepian-Wolf situation*", i.e., for dm-2-MACs with correlated sources: in 1983, De Bruyn and Van der Meulen [134] give a code construction (based on permutations) for the dm-2-MAC with correlated sources, for the asymmetric situation with just one private source, i.e., encoder 1 sees both its private source and the common source, while encoder 2 only sees the common source. The next year, the same authors prove that in the same situation, feedback cannot increase capacity [140]. In addition, this paper also determines the CR in the general Slepian-Wolf situation for a certain subclass of channels, for the case of feedback to one or both encoders. The authors also prove that for the class of MACs for which X_1 is a function of Y and X_2, with correlated sources, the CR equals the inner bound region of King (1975). This is also the case when partial feedback is available. For the dm-2-MAC with arbitrarily correlated sources in the asymmetric situation, De Bruyn, Prelov and Van der Meulen [147] derive the CR. Actually, they show that the separation principle holds (which is not the case for the general Slepian-Wolf situation), and they also show that feedback does not help in this situation.

For the memoryless additive white Gaussian noise (AWGN) 2-MAC in the Slepian-Wolf situation, Prelov and Van der Meulen [159] determine the capacity region (in terms of the noise power σ^2) in 1987: this region has the expected form, which is similar to the result of Slepian and Wolf for the discrete case, and it generalizes the known results for the classical AWGN 2-MAC (Wyner, 1974) and for the 2-MAC with only one private source (Prelov, 1984).

Finally a word on *strong converses*, which means that a rate point is reachable (asymptotically) for *any* error probability, not just for error probabilities going to zero. In 1980, Dueck provided the first strong converse in multi-user information theory: viz. for the classical dm-2-MAC. In 1987, Verboven and Van der Meulen

[162] use similar techniques to obtain strong converses for the dm-2-MAC in the Slepian-Wolf situation and for the S input MAC with S "hierarchical" sources.

1.2.4 Codes for Deterministic Multiple-Access Channels

A multi-user channel is called *deterministic* if its output(s) is/are unambiguously determined by the channel input(s). Thus all channel transition probabilities are either 1 or 0. Stronger even, transmission with an error probability equal to zero can be obtained, instead of the classical "ε-error" (i.e., an asymptotically decreasing error probability). The corresponding rate region is called the *zero-error capacity region*, and coding schemes operating with zero error are called *uniquely decodable* (UD). In general, the zero-error CR is (strictly) contained within the ordinary (ε-error) CR.

The two-input *binary adder channel* (2-BAC) has two binary inputs and a ternary output $Y = X_1 + X_2$. It was introduced by Van der Meulen in 1971. This channel is sometimes called the binary erasure MAC.

In 1982, Schalkwijk and Vinck [128] give a simple argument to show that all (R_1, R_2) with $R_1 + R_2 = 1.5$ are achievable rate points: Feed one of the channel inputs a binary stream of equiprobable zeros and ones. This transforms the channel from the other input to the output into a binary erasure channel with erasure probability $p = 1/2$. Use this channel at capacity, and after decoding its input sequence recover the equiprobable binary input sequence presented at the first input of the multiple access channel.

For the BAC, Coeberg van den Braak and Van Tilborg [131] continue the work of Kasami and Lin (1976–1983) by explicitly constructing new UD code pairs of relatively small ($n \leq 48$) block sizes, with better rates: their best rate sum is 1.303. Recently it was proved by Urbanke and Li that the zero-error CR of the 2-BAC is strictly smaller than the ε-error CR (where the rate sum is ≤ 1.5, see Figure 1.7 (a)): no UD code pairs can be constructed with sum rates arbitrarily close to 1.5.

As mentioned before, the feedback capacity region of the 2-BAC is strictly larger than the region of Figure 1.7 (a). In the same year, 1983, Vinck [137] uses a technique similar to Schalkwijk's unit square subdivision for the BMC to construct a code for the 2-BAC with feedback, with as rate point $R_1 = R_2 = 0.7909$, i.e., outside the non-feedback capacity region.

In 1984, Vinck, Hoeks and Post [146, 143] numerically evaluate for $R_1 = R_2$ the expression for the full-feedback CR, as found by Willems, for two deterministic 2-MACs with M-ary input alphabets. For both situations, it turned out that this is the total cooperation point. For the 3-user BAC ($Y = X_1 + X_2 + X_3$) with full feedback, for which the CR is not known, [146] also gives two coding strategies with a rate sum above the ARQ bound.

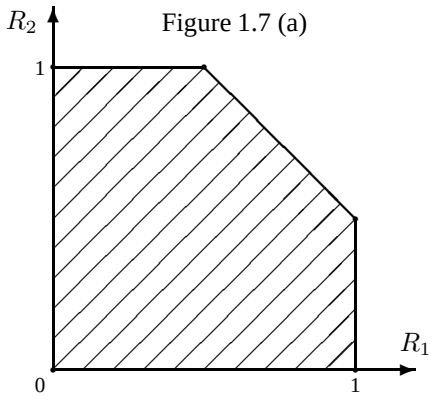


Figure 1.7 (a)

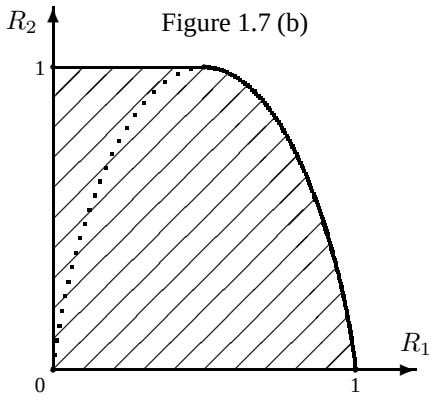


Figure 1.7 (b)

Considering all possible binary input deterministic dm-2-MACs, it turns out that there is only one non-trivial channel besides the 2-BAC. This so-called *binary switching channel* (BS-MAC) was introduced by Vinck in 1984. Its capacity region is shown in Figure 1.7 (b). Code constructions for the noiseless BS-MAC are started in 1986 by Vanroose and Van der Meulen [154] with two classes of UD code pairs, based on MDS codes. Rate pairs are given up to $R_1 + R_2 = 1.33$, still far away from the optimal 1.58496. This work is continued by Vanroose in 1987 [161], who introduces the concept of *tolerated defect patterns* to ease the creation of UD codes. In this paper, Vanroose gives optimal code pairs for block lengths up to 19, but actually he achieves the best rate sum 1.4799 with a relatively simple first-order rate 2/3 convolutional code. Also, the noisy BS-MAC is considered, for which δ-decodable code pairs are to be used. In 1988 [165], Vanroose and Van der Meulen prove that the zero-error CR for this channel coincides with the ε-error region. This means that any rate point, including the only total cooperation point $(R_1, R_2) = (\frac{2}{3}, \log_2(3) - \frac{2}{3})$ of the CR, with rate sum 1.58496, is asymptotically achievable with UD code pairs.

When using UD codes with multiple-access channels, one always assumes codeword (block) synchronization between the two encoders. This is not always a realistic assumption. In [165], also the coding situation is considered where there is no block synchronization between the two encoders. Code pairs for this situation are given; it is not yet clear whether the CR of this *quasi-synchronous* channel is strictly smaller than the classical CR of Figure 1.7 (b).

1.2.5 Broadcast Channel

In the communication situation of the T-user broadcast channel, there are T information sources which are jointly encoded by a single encoder into a single channel input stream (see Figure 1.8). The channel has T separate outputs, each of which is seen by a decoder who is only interested in decoding his source message.

The capacity region for the general broadcast channel is still an open problem.

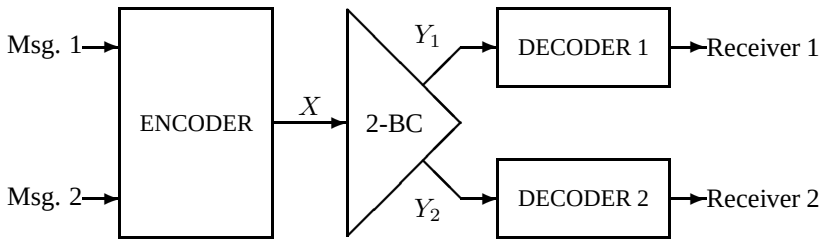


Figure 1.8: *Diagram of the broadcast channel.*

The best inner bound was found by Marton (1979). Van der Meulen [118] gives a simpler proof for this bound at the first Benelux Information Theory Symposium.

The CR has been found for several specific broadcast channel subclasses, especially the "more capable" broadcast channel with common information for both receivers (El Gamal, 1979) which generalizes the "less noisy" broadcast channel which in turn generalizes the broadcast channel with degraded message sets (also called the asymmetric broadcast channel; its CR was determined in 1977 by Kőrner and Marton).

Marton and Gelfand and Pinsker determined the CR of the semi-deterministic broadcast channel (i.e., only Y_1 is a deterministic function of X), which in the fully deterministic case reduces to

$$\begin{aligned} \mathcal{C} = \bigcup_{P_X} \{(R_1, R_2) \quad | \quad & 0 \leq R_1 \leq H(Y_1), \\ & 0 \leq R_2 \leq H(Y_2), \\ & R_1 + R_2 \leq H(Y_1 Y_2)\}. \end{aligned} \tag{1.16}$$

In his 1982 contribution, Van der Meulen [127] gives an overview of the above-mentioned known results for the broadcast channel.

The CR of the Gaussian broadcast channel (with additive white Gaussian noise) is known (Cover, 1972). For the Gaussian broadcast channel with feedback, Ozarow (1979) gave an inner bound. In 1981, Willems and Van der Meulen [124] improve this bound.

The only non-trivial binary output deterministic broadcast channel has ternary input X and outputs $Y_1 = \max(X - 1, 0)$ and $Y_2 = \min(X, 1)$:

	Y_1	Y_2
$X = 0$	0	0
$X = 1$	0	1
$X = 2$	1	1

It was defined by Blackwell (1963), and introduced by Van der Meulen (1975) as the *Blackwell broadcast channel*. The CR (see Fig. 1.9) was found by Gelfand in 1977 to be the convex union of two entropy curves.

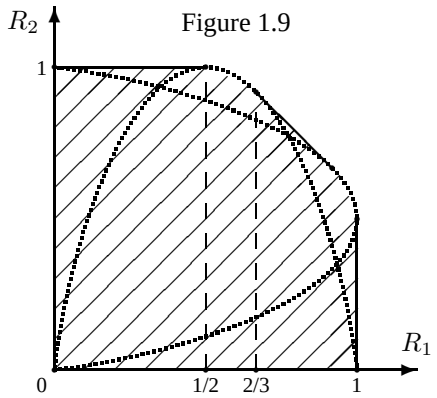


Figure 1.9

The achievability of this CR is outlined by Schalkwijk and Vinck [128] as follows: the $Z \rightarrow X$ information stream is coded as input 0s and input "not-0s". These "not-0s", i.e. 1s or 2s, are used to send the $Z \rightarrow Y$ information. The $Z \rightarrow Y$ channel can now be considered a defect-channel, where the $(Z = 0)$ defects are known to the sender, see Section 1.2.9.

This CR is actually also the *zero-error CR*, as was proved by Vanroose and Van der Meulen in 1989 [172]. Remarkably, their proof makes use of UD code pairs for the BS-MAC.

The Blackwell broadcast channel is a model for a binary "write-once memory" (WOM) that is used twice; it is also a model for write-unidirectional memory (WUM) coding; and also for a binary memory with 1-defects. All three models are described in more detail in Section 1.2.9.

In 1983, De Bruyn [133] describes how one can use permutations of a 'substrate' word as an efficient coding scheme for broadcast channels with degraded message sets. The advantage of this approach is its storage efficiency: adding a single permutation doubles the number of code words. A similar technique is used by De Bruyn in 1984 [139] to construct list codes for the one-way channel.

1.2.6 Identification for Broadcast Channels

In their pioneering paper of 1989, Ahlswede and Dueck [75] introduced a new communication problem, where the receiver's task is not the reconstruction of any transmitted message, but only to decide whether or not one particular message was sent. Their remarkable result is that for a d.m. one-way channel with capacity $\mathcal{C}$, identification at block length n is possible with arbitrarily small error probability for message set sizes up to $2^{2^{nC}}$. Otherwise stated, the capacity for identification equals the transmission capacity, but in a double exponential sense.

In the same year, Verboven and Van der Meulen [168] derive a similar result for

the (general) deterministic broadcast channel. In contrast to the one-way case, here the capacity region for identification is larger than the region for transmission: only the conditions $R_1 \leq H(Y_1)$ and $R_2 \leq H(Y_2)$ remain, the condition $R_1 + R_2 \leq H(Y_1Y_2)$ drops. For instance, for the Blackwell broadcast channel, the CR for identification is the full unit square instead of the region of Fig. 1.9.

1.2.7 Relay Channel and Interference Channel

The relay channel was introduced by Van der Meulen in 1968, see [25] and Figure 1.10. There are very few coding results for the deterministic relay channel. Note that, as opposed to the previous multi-user channels, the relay channel has only a single information stream, but the relay terminal may help the transmission.

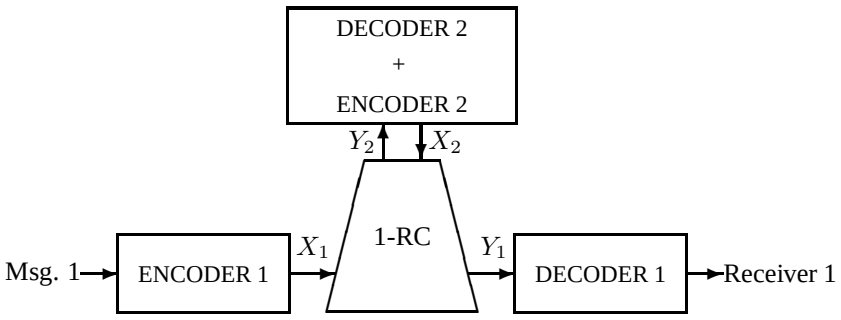


Figure 1.10: *Diagram of relay channel*

In 1990, Vanroose [173] elaborates on coding for three particular relay channels. For the binary channel $Y_1 = X_2$, $Y_2 = X_1 \oplus X_2$, he gives a simple optimal coding scheme which achieves capacity. For two other deterministic relay channels, he presents a suboptimal scheme which makes effective use of the relay terminal.

The *interference channel* was mentioned for the first time by Shannon [14]. A diagram of this channel is depicted in Figure 1.11.

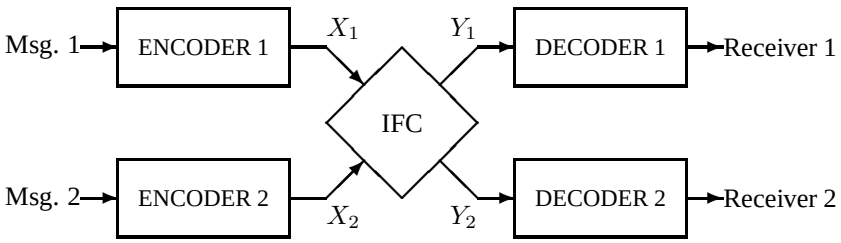


Figure 1.11: *Diagram of the interference channel.*

The general CR is still not known; only for certain special cases a closed expression has been derived. In 1991, Prelov and Van der Meulen [179] derive the CR of the additive almost-Gaussian interference channel.

1.2.8 Non-Cooperative (Jamming) Channels

If one of the transmitters on a multi-user channel actively tries to disturb the communication of the other users, the channel is called a non-cooperative channel, or jamming channel.

In 1995, Vanroose [195] classifies all deterministic jamming channels. It turned out that there are four different possible jamming channel types, one of which is the jamming 2-MAC. The only interesting binary-input jamming 2-MAC is the 2-BAC, with a jamming capacity of 0.5 (as was derived by Ericson in 1986). This 0.5 is actually zero-error capacity, as is outlined in [195]. Vanroose also gives an example of a ternary input jammer 2-MAC for which the capacity differs from the zero-error capacity.

1.2.9 Coding for Memories with Defects or Other Constraints

A memory chip has a high density of memory cells which all can store a single bit, i.e., a 1 or a 0. An unfortunate side effect of the constantly growing storage density is the fact that some (say: a fraction p) of the cells are defective, i.e., they are stuck at either 0 or 1. A defective memory is a noisy communication channel. If both the encoder and the decoder know the location of the defective cells (and hence do not use those), the capacity of the memory trivially is $1-p$. Remarkably, if only the encoder knows the defect locations, the capacity is still $1-p$ and not $1-h(\frac{p}{2})$ (which is the capacity in the case where the encoder is also uninformed about the defect locations).

The capacity was proved in 1974 by Kuznetsov and Tsybakov [35], using message "bins". An outline of this proof can be found in [128]. This coding situation can be seen as a channel with side information at the transmitter, a general setup already considered by Shannon in 1958 [12] and also of interest for data hiding, see Section 3.4. Actually, a memory with 0-defects can also be seen as a noiseless broadcast channel (viz. the Blackwell broadcast channel) since the channel input is ternary (defective 0, stored 0 and stored 1) while for one of the channel outputs, two of these are collapsed into a single 0 read out. Hence a closer look at the derivation of the CR of the Blackwell broadcast channel would reveal that the rate point $(h(p), 1-p)$ is indeed achievable with $P(X=0)=p$.

In the years 1986–1990, there is a renewed interest in coding for defective or otherwise constrained memories. In 1986, Schalkwijk [151] describes a constructive coding scheme for memories with defects known to the encoder only. Schalkwijk observes that, in order to surpass the intuitive $1-h(p/2)$ limit, one has to use knowledge of *all* defects, not just that of the "previous" defect locations. Then he describes how to use Shannon strategies derived from optimal codes of a so-called

derived channel, which in this case is a channel with 4 inputs and 2 outputs and with noiseless feedback.

Willems and Vinck [152] consider a slightly different situation: due to physical limitations, a binary memory can only be overwritten with 0s, not with 1s, during a single pass. In the next pass, only 1s can be written. This so-called *write-unidirectional* memory (WUM) clearly has a capacity between 0.5 and 1 bits per write cycle, since at most two write cycles are necessary to write any possible bit into any memory cell. Similar to the situation of memories with defects, this channel can be seen as a channel with side information at the transmitter, or as an incarnation of the Blackwell broadcast channel, since only the writer knows the 'old' state of a memory cell. Hence it is not a surprise that the capacity for this WUM is strictly larger than 0.5. Actually, the capacity is $0.69424 = \log_2((1+\sqrt{5})/2)$. Willems and Vinck [152] give a coding scheme with rate $\log_2(6)/5 = 0.51699$.

Van Overveld and Schmitt [163] generalize the WUM setup to the situation where the rate of the two passes need not be identical, and they prove that in this case all rate points (R_1, R_2) lying in the Shannon outer bound region of Figure 1.3 (a) are achievable. In 1989, Van Overveld [171] computes the capacity of the q-ary WUM with q alternating cycles, writing q-ary symbols into a q-ary memory. In 1990, Van Overveld and Willems [175] prove that the capacity of the WUM in the situation where both the encoder and the decoder are uninformed of the state of the memory is 0.54588. The achievability part of this result was already stated by Simonyi in 1987, but that proof was not completely satisfactory.

A third type of constrained binary memory is called a *write-once memory* (WOM). Such a memory can be rewritten, but only to change a 0 to a 1. It is assumed that the encoder, but not the decoder, knows the previous state of the memory. This communication situation was introduced by Rivest and Shamir (1982), and the capacity region for T consecutive uses was determined in 1984. In 1997, Fu and Vinck [200] consider the q-ary WOM and derive its zero-error capacity region.

1.2.10 Random-Access Channels

Consider the following communication situation called the (slotted) *multiple-access collision channel* or random-access channel: users are allowed to transmit packets within fixed time slots over a common channel. When two or more users send a packet in the same time slot, these packets "collide" and the packet information is lost. The users obtain information about possible collisions, which allows them to retransmit whenever necessary. This channel was first described by Abramson in [21] and has been widely used in ethernet computer networks. Maximal throughput is $1/e = 0.36788$ effective packets per slot under the assumption of Poisson packet arrivals. This so-called slotted ALOHA system is inherently unstable: once the maximal throughput is surpassed, the system never returns to normal mode.

In 1991, Van der Vleuten [182] proposes a new, low-complexity control algorithm for the slotted collision channel, which automatically adjusts to changes in average traffic intensity and is able to recover from overload situations. So this system is intrinsically stable, in contrast to the ALOHA system.

When there is no feedback present, the only way to avoid conflicts is the use of *protocol sequences*. In 1996, Tsybakov and Weber [196] present a class of conflict-avoiding codes which can be used for this purpose.

In 1999, Vinck [206] considers a slightly different situation, introduced by Chang and Wolf in 1981, called the T user M frequency MAC. This channel model is actually a "classical" multiple-access channel model for the random-access communication situation.

CHAPTER 2

Source Coding

F.M.J. Willems (TU Eindhoven)
Tj.J. Tjalkens (TU Eindhoven)

Introduction

Source coding or data compression deals with the problem of describing data in the most efficient way. By most efficient we usually mean that we want to achieve the *shortest* description.

The source coding problem as it was originally introduced by Shannon [3] considers a probabilistic data source whose output sequence has to be represented in an efficient way, i.e. it has to be as short as possible on average. In this setting it is assumed that all relevant source symbol probabilities are known. Often *blocks* of source symbols are used because the theoretical analysis shows that the best possible compression is achieved for long blocks of data. Methods that devise codes under the condition of a known source are called *non-universal methods*, see Section 2.1. So these methods explicitly use the probabilistic knowledge of the source to design the code. *Universal methods* create coding schemes that will then work for a set of sources with different probabilistic descriptions. Universal methods are the topic of Section 2.2.

The best possible compression that can be achieved for a given source is given

[1] This chapter covers references [214] – [260].

by the source entropy $H(U)$[1]

$$H(U) = -\sum_{u \in \mathcal{U}} p(u) \log p(u) \quad \text{bit per letter (or block).} \tag{2.1}$$

Here $p(u)$ is the source letter (or block) probability.

Sometimes, one would prefer a better compression than the source entropy allows. By Shannon's results we know that this is not possible if one requires a perfect, or error free, reconstruction. For source data such as speech, audio, images, and video, perfect reconstruction is not needed and a better compression can be achieved if some *distortion* is allowed between the source original data and the reproduction. The fundamental limits for this setting, also presented by Shannon, are treated in Chapter 8 of this book.

2.1 Non-Universal Methods

Non-universal methods explicitly use the probabilistic knowledge of the source when designing the code. This knowledge often comes in the form of probabilities of sequences of $n \geq 1$ source letters. If $n > 1$ we often call the sequence a *block*. Although in practice these probabilities are most often unknown they can be estimated from some representative data. e.g. the letter probabilities of English text do not depend very much on the particular text. Therefore, a reasonable performance can be expected from codes using these estimated probabilities. Especially when these codes are used in a larger compression scheme, such as an audio or video compression system, one or a few non-universal codes are used, mainly because their implementation is less complex and so less expensive than a universal method.

2.1.1 Fixed-to-Variable Length Codes

A fixed-to-variable length source code, or *FV-code*, maps sequences of source letters of a fixed length to codewords of variable length. The codewords and especially their length are chosen in such a way as to minimize the expected codeword length.

As an example, consider a ternary memoryless source U with alphabet $\mathcal{U} = \{a, b, c\}$ and probabilities $p(a) = \Pr\{U = a\} = 1/3$, $p(b) = 1/5$, and $p(c) = 7/15$. A FV-code could be the code that maps the source letter 'a' to the binary codeword '00', 'b' to '01', and 'c' to '1'. This code is uniquely decodable since any concatenation of codewords can be decomposed into codewords again in only one possible way and the expected code rate is given by

$$R_1 = \sum_{u \in \mathcal{U}} p(u)\ell(u) = 1.533 \text{ code symbol per letter.} \tag{2.2}$$

[1] We take 2 as base of the logarithm in this chapter.

Here $\ell(u)$ is the length, in code symbols, of the codeword for the source letter u. The entropy of this source is

$$H(U) = -\sum_{u \in \mathcal{U}} p(u) \log p(u) = 1.506 \text{ bit per letter.} \tag{2.3}$$

We see that our code already compresses well.

If we would want to improve our code, we could try a code that assigns codewords to pairs of source letters. Consider the code as described in the next table.

source	code	source	code	source	code
aa	000	ba	1000	ca	011
ab	0100	bb	1001	cb	101
ac	001	bc	0101	cc	11

This is an example of a FV-code with *block length* 2. The expected codeword length of this code is 3.0489 code symbol per pair of source letters, resulting in a code rate of

$$R_2 = \frac{3.0489}{2} = 1.524 \text{ code symbol per letter.} \tag{2.4}$$

Optimal source codes are created with Huffman's algorithm [7]. Since its publication in 1952, the Huffman algorithm has been studied extensively. Not only the compression rate, but also other properties were considered. Members of the WIC community participated in this and we shall report here on their findings.

Complexity Issues

Desmedt, Vandewalle and Govaerts [215] consider the parallel encoding of source symbols by n parallel Huffman encoders. A source symbol is represented by n *parallel* symbols, which are encoded independently and in parallel by the n encoders. However, the resulting parallel sources are usually not independent and some extra redundancy is introduced. One can reduce this so-called *parallel redundancy* by a clever choice of the representation for which the authors derive a heuristic search. The search result can be improved by using the results of their Theorem 1, see [215], which we shall repeat here.

Suppose a letter a_i is represented by the n-tuple $(b_i^1, b_i^2, \cdots, b_i^n)$ and the probability of the parallel symbols b^k is computed by the sum of the probabilities of those original symbols a whose k^{th} component is b^k. Now consider two source symbols a_i and a_j such that for their probabilities p_i resp. p_j holds $p_i < p_j$. In the parallel representation C^1, $(b_i^1, b_i^2, \cdots, b_i^n)$ is the representation for a_i and $(b_j^1, b_j^2, \cdots, b_j^n)$ is the representation for a_j. If

$$\prod_{k=1}^{n} P(b_i^k) \geq \prod_{k=1}^{n} P(b_i^k), \tag{2.5}$$

then interchanging the representations of a_i and a_j reduces the redundancy.

Vanroose and Verbeke [218] also discuss a method to reduce the complexity of the Huffman algorithm. The design of a Huffman code involves a repetition of sorting problems, which are time-consuming operations. On the other hand, it is much simpler to generate the codewords if their length distribution is already known. The authors improve a result of [36] that gives a sufficient condition for a source symbol probability distribution such that the optimal code is (essentially) a block code. Then they consider codes with more than two successive codeword lengths and as an example derive sufficient conditions for the optimality of a code with three successive lengths.

An efficient implementation of a Huffman code is based on the Shannon-Fano code as described by Connell in 1973. Tjalkens [254] considers the actual complexity in terms of the storage cost given a fixed amount of coding time per symbol. In previous discussions on the so-called *Minimum redundancy codes*, usually an ordered (with respect to the probabilities) symbol alphabet was assumed. However, Tjalkens considers the encoding and decoding of blocks of n source symbols from a binary memoryless source and it turns out that the ordering of these blocks, described as the *index computation*, is the most complex operation. Both encoder and decoder have to compute the index of a sequence such that the probabilities are ordered. They then find the codeword using a correctly initialized `base` array. The storage requirements of the `base` array is $\mathcal{O}(n^2)$. The index can be computed in $\mathcal{O}(n)$ operations and $\mathcal{O}(n^3)$ storage cost using pre-computed binomials or in $\mathcal{O}(n^2)$ time and $\mathcal{O}(n^2)$ storage if the coefficients are computed when needed. The latter choice is unacceptable as this would imply a with n increasing amount of time per letter. So the storage cost of the whole method is $\mathcal{O}(n^3)$.

Self-Synchronization

Another topic often considered is recovery from errors. Of course source codes should not contain redundancy, so the goal is not correction of errors in itself, but tackling the more serious problem of error propagation. Because the codewords have varying lengths, errors cause the decoder to lose synchronization, and thereby to continue decoding erroneous words. So the capability of the decoder to regain synchronization is essential.

Already in 1959, the synchronization issue was addressed by Gilbert and Moore, however, without regard to the efficiency, in terms of redundancy, of the code. There the authors defined the notion of a *synchronizing codeword*, which if received, defines a synchronization point of the code stream irrespective of the state of the encoder. The *probability of unconditional synchronization* is equal to the sum of the probabilities of all synchronizing codewords.

A first attempt to find efficient codes containing synchronizing codewords was reported in [65]. In [217], Jansen and Oosterlinck report on the construction of efficient self-synchronizing codes. They devised an algorithm that will produce an efficient code with the highest possible probability of unconditional synchronization, but only in the case where the shortest possible synchronizing codeword has

length $m + 1$, where m is the length of a shortest codeword. Another approach taken in this paper is to consider the expected number of code symbols needed before re-synchronization after an erroneous codeword has been received. A method to calculate this *delay* is given and some experimental results are reported.

One year later, in [70], a more general algorithm for the design of self-synchronizing efficient codes was given. Later, De With [226] reported on an improvement of [70] for special sources that occur in the compression of images. He found that by recursively creating subtrees with synchronization patterns, the number of synchronizing words can be increased significantly.

Special Codes and Applications

The redundancy of a Huffman code is upper-bounded by the source entropy plus one. However, if the source probabilities are of the form 2^{-i} for positive integers i, then the binary Huffman code has no redundancy. This can be generalized such that the r-ary Huffman code has no redundancy if the source probabilities are of the form r^{-i}, again for positive integers i. Stasiński and Ulacha [260] used this basic idea to study the design of more efficient codes. They encode a series of q symbols from an r-ary alphabet together in a binary string of length $b(r^q) = \lceil q \log r \rceil$ bits. If appropriate values for q and r are used, such that r^q is close to an integer power of 2, this representation is efficient, i.e., $b(r^q)/q \to \log r$. We call a device that performs this operation a *combiner*.

The principal approach of Stasiński and Ulacha is to allow codes that use differently sized alphabets for different letters. The letters from non-binary alphabets are processed together in appropriate combiners for each alphabet size. As soon as a combiner has received q symbols, it outputs the $b(r^q)$ bits. The authors claim that the resulting code streams are decodable and that the resulting code is not much more complex than a binary Huffman code.

In [242], Mitrea and De With present the results of a comparative study on the performance and cost of a Huffman coding system versus an arithmetic coding technique in an interesting practical setting. They consider the coding of digital video signals inside video recorders or standard TV applications that are used to reduce the storage cost needed for processing the video data. For small data blocks a so-called Adaptive Dynamic Range Coder determines the minimum and maximum sample value and thus the dynamic range is determined. All samples are quantized adaptively according to the dynamic range. The authors experimentally determined the statistics of the quantizer outputs.

Using a single fixed Huffman code already gives a 10% rate reduction, but using four different codes depending on the dynamic range gives another 10% improvement. The Huffman codes are now simplified by first limiting the codeword length to 16 symbols, which results in a negligible decrease of compression and a fair decrease of complexity. Further reordering of codewords reduces the table size to one-third of the original size. Then an arithmetic code, see Section 2.1.3, is tested

under three conditions. First the same fixed statistical model is used as for the single fixed Huffman code, then the four statistics are used depending on the dynamic range, and finally adaptive codes are used (using symbol counts) separately for each of the four classes.

In terms of compression the arithmetic code outperforms the Huffman codes in all cases, but only with at most 5%. After comparing the results, the authors conclude that the extra complexity of the arithmetic code is not justified given the minor additional compression gain.

In [240], Gerrits, Beuker and Keesman report on the design of a compression system for interactive displays. The display system produces 150 samples of 32 binary symbols data per second, consisting of coordinates and pen pressure information. The channel can transmit 600 bits per second, hence from raw data, a compression of a factor of 8 is required. It turned out that 30-60% of the data is irrelevant and can be ignored without loss of quality. The remaining data is transformed by a second-order differential transformation with limited precision that does not degrade the visual quality of the image. The remaining transformed samples still exhibit dependencies. Several lossless compression methods, Huffman, Lempel-Ziv, and arithmetic coding with a finite order Markov model have been tested. An arithmetic coder with an order-4 model turns out to give the best possible performance. In most cases the required compression can be achieved. If not, the authors suggest to resort to Jelinek's lossy compression method for buffered systems, see [20].

Among the types of data, sampled audio signals have always been difficult to compress losslessly. In [250], Van der Vleuten and Bruekers report on an advanced lossless audio compression scheme. The data is a binary representation of audio signals sampled at 64×44.1 kHz. The data stream is split into frames for $1/75$ seconds worth of samples and these frames are processed independently. First the (sigma-delta) signal is fed into a linear predictor filter $z^{-1}A(z)$, where $A(z)$ is a N-th order filter produced with standard autocorrelation or covariance methods. The predictor coefficients are transmitted to the decoder so it can do the same predictions. When *hard decision* is applied to the prediction, and the resulting error signal is compressed with a run-length code and a well-chosen Huffman code, the compression rate is already impressive. However, it can be improved upon and this is the main contribution of the paper.

The authors found that reliability information can be obtained from the real-valued prediction Z. For a finely quantized absolute value of the predictor, i.e. $|Z|$, a count is kept for the number of successes and failures of the hard decision prediction. This information is used in an arithmetic encoder and the table is also transmitted to the decoder so that the arithmetic decoder can use the same probabilities. This results in a final compression factor of about 2.3. The encoder and decoder are so simple that 2.8 Mbit real-time encoding and decoding is possible in hardware.

2.1.2 Variable-to-Fixed Length Codes

Another type of data compression codes are the variable-to-fixed length codes, *VF-codes*. Here source sequences of varying length, also called *segments*, are encoded into codewords of fixed length. In order to compress the data efficiently, the expected source segment length must be maximized given a fixed number of allowed segments.

As an example, consider again the ternary memoryless source U with $p(a) = 1/3$, $p(b) = 1/5$, and $p(c) = 7/15$. We allow 15 segments and a permissible *segment set* is $\{aa,\ ab,\ aca,\ acb,\ acc,\ ba,\ bb,\ bc,\ caa,\ cab,\ cac,\ cb,\ cca,\ ccb,\ ccc\}$. The resulting expected message length $\bar{L} = 2.5289$ symbols per segment and we need 15 binary codewords, each of length 4, so the resulting code rate is

$$R = \frac{4}{\bar{L}} = 1.5817 \text{ code symbol per source letter.} \tag{2.6}$$

The compression for this example is not very good, but it has been proven that asymptotically, rates arbitrarily close to the source entropy can be achieved. It is also known that VF-codes are almost always a better choice than FV-codes for low-entropy sources, but no clear winner exists. The construction of the optimal code in the sense of the best compression for a given segment set size is known as the Tunstall algorithm, see [19].

The way codewords are assigned to the messages from a message set is rather arbitrary so one can try to assign them such that the result is computationally or storage efficient. A *lexicographical index* is in many cases a way to assign codewords to segments in an efficient manner. To use the lexicographical index, we first have to define an ordering of the segments.: the so-called *lexicographical order*. We define it quickly: let x^n and y^m be two different segments of lengths n and m respectively, where x^n is not a prefix of y^m, nor vice versa. Then using any ordering on the letters of the alphabet, we define

$$x^n < y^m \text{ whenever } \exists i : (\forall j, 1 \leq j < i : x_j = y_j) \text{ and } x_i < y_i. \tag{2.7}$$

Now the lexicographical index $i(x^n)$ (with respect to a message set V) is defined as the number of segments in the set V that are smaller than x^n in lexicographical order.

Complexity Issues

A disadvantage of many implementations of data compression codes is that all possible codewords have to be generated before and stored during the actual encoding and decoding process. Because one would like to encode many source symbols per codeword in order to obtain good compression, the amount of time spent on creating the codewords and the size of the memory needed to store these words is huge, or more practically, one is severely limited in the length of the source sequences.

More efficient methods can be found in the class of *enumerative codes*, see [29, 31]. These codes do not create a list of all possible codewords but *compute* the required codeword when needed using combinatorial computations and often aided by *small* tables. A well-known modern example of this principle is the arithmetic coding technique as mentioned earlier, see also section 2.1.3.

Schalkwijk [214] presents an observation by Petry on their previous enumerative variable-to-fixed length code. Assume that the (memoryless and binary) source probabilities p_0 and p_1 can be represented (or approximated) by r^{s_0} resp. r^{s_1} for a given fixed real valued r and positive integers s_i. Now one can define the set of source messages (of variable length m) as

$$V(n) = \{u_1, u_2, \cdots, u_m | \#(0\text{'s in } u^m) \cdot s_0 + \#(1\text{'s in } u^m) \cdot s_1 \geq n \text{ and } \#(0\text{'s in } u^{m-1}) \cdot s_0 + \#(1\text{'s in } u^{m-1}) \cdot s_1 < n\}. \quad (2.8)$$

This set $V(n)$ for a given n will be the message set for the VF code. The codeword will be the binary representation, in a fixed and sufficiently large number of symbols, of the lexicographical index. So it is important to find the sizes $c(n)$ of these sets $V(n)$. The following holds.

$$c(n) = \begin{cases} 1; & \text{if } n \leq 0, \\ c(n - s_0) + c(n - s_1); & \text{if } n > 0. \end{cases} \quad (2.9)$$

Just as in [31], we can compute the index $i(u^m)$ for any sequence $u^m \in V(n)$ by

$$i(u^m) = \sum_{i=1}^{m} \sum_{y<u_i} c(n - \sum_{j=1}^{i-1} s_{u_j} - s_y). \quad (2.10)$$

This can easily be computed by the following iteration.

```
// inputs are:  symbols u[1], u[2], ..., u[m]
//              prob. parameters s[0], s[1]
//              parameter n
// precomputed: c[1], c[2], ..., c[n]
// output is:  index
index = 0; offset = n; i = 1;
while (offset>0) do
  if (u[i]==1) do
    index = index + c[offset-s[0]];
  endif
  offset = offset - s[u[i]];
  i = i+1;
done
```

Decoding is performed in a similar way, using the same array $c[.]$, and it is easy to extend this method to non-binary sources using a similar linear array $c[.]$.

In [216], Tjalkens and Willems extend the Schalkwijk-Petry result to unifilar Markov

sources. Their algorithm requires one linear array *per* state of the Markov source. In the analysis of the coding scheme, they show that the entropy rate of any unifilar Markov source can be approached arbitrarily close. The authors state finally that this method allows low-redundancy codes for these sources with low storage and computational complexity.

Special VF-Codes

The fact that the optimal variable-to-variable length code is still unknown is one of the reasons why attempts have been made to try and qualify the differences and similarities between FV-codes and VF-codes. In 1996, Keesman [244] presented a unified view on variable length codes by the notion of a *partial code*. Starting with an arbitrary code $\mathcal{C}$ he assigns not a single codeword to a source letter but a whole subset of codewords. If the number of codewords is written as $|\mathcal{C}|$ and the size of the subset for letter i is $|\mathcal{C}_i|$, then it is shown that the *effective rate* is

$$R = \sum_{i=1}^{M} p_i \log \frac{|\mathcal{C}_i|}{|\mathcal{C}|}. \tag{2.11}$$

In fact, this is a re-invention of a method published in 1980 by Guazzo, see [53] and both methods are basically arithmetic codes.

In [223], Willems also creates a FV-code based on a VF-code. His aim is to use the enumerative techniques from [216] to come up with a less complex scheme than the enumerative algorithm of [29]. However, the latter is universal for the class of memoryless sources. The method in [223] can be seen as a combinatorial arithmetic code. The result is indeed a code that achieves a better redundancy for a given complexity than the Pascal triangle method in [29]. Moreover, the storage complexity is linear in the block length, while for the Pascal triangle method, it is quadratic. For non-binary sources, the complexity of the Pascal triangle method increases enormously while the cost of this scheme remains linear.

2.1.3 Arithmetic Coding

Arithmetic codes are based on an observation of Shannon, namely that the *cumulative probability distribution* can serve as the basis of a source code. This code was further improved by Elias, whose result remained unpublished until Jelinek reported on it in his paper [20]. Arithmetic codes in their modern form were first described by Rissanen and independently by Pasco in 1976. Several advantages of arithmetic codes over Huffman codes are worth mentioning. First, arithmetic codes have a rather low complexity. The codewords are only generated when needed, so a costly design phase is not needed and storing the codewords is not necessary. Codewords can be very long, so the code can be very efficient; actually it is limited only by the precision of the computations. Source probabilities that vary from letter to letter are easily accommodated. There is a strong relation between arithmetic codes and enumerative schemes, as was already discussed in e.g. [244, 223].

In 1985, Tjalkens and Willems [219] described the basic structure of arithmetic codes and gave an implementation that uses a finite-precision exponential table to avoid costly multiplications. They derive bounds on the resulting redundancy that clearly show the redundancy cost of limiting the arithmetic precision.

In a sequel in 1986, Tjalkens [221] described *designs*, i.e. choices of coding parameters for a given probability distribution. He shows that designs can be *local*, i.e. a design can depend on the current *position* in the coding interval and such a local design has a lower redundancy. Also an arithmetic code is described that reduces the coding complexity for high-cardinality source alphabet. Finally, a novel technique for carry-blocking is introduced that has the advantage that it fits perfectly in the framework of the coding algorithms discussed.

Because data compression codes can be seen as *probability transformers*, they can be used to produce sequences with special properties. In 1997 Immink and Janssen [246] considered the use of *floating point* representations in enumerative schemes for the generation of $dklr$-sequences or *run-length constrained* sequences. Again the resulting method was an arithmetic code.

2.1.4 More Applications

Trees, a special class of directed graphs, are used in many applications as a convenient way to organize data. In many cases a tree has to be stored or transmitted and this should be done in an efficient way. However, it remains important that the tree structure can be accessed easily.

In two papers, Vanroose [230, 241] discusses efficient tree representations. In the first paper [230], arbitrary rooted trees are considered, so a tree consisting of *nodes* that have an indegree of 1, except for the unique *root*, which has indegree 0. A node is a *leaf* if it has outdegree 0, otherwise it is called an *internal node*. Note that we do not require that every internal node has the same outdegree. Several machine representations of trees are discussed and the cost of storing a tree for each of those representations is evaluated. In [241], Vanroose discusses several complexity measures on trees. He discusses several applications such as variable length source coding, decision trees, and search trees. Different complexity measures are useful for different applications, e.g. the average tree depth is a good measure for FV-codes and the Huffman algorithm minimizes this measure. Another measure is the *average splitting entropy*, which is useful in the construction of classification trees for object recognition.

For a homogeneous tree, where all internal nodes have the same outdegree, a so-called *arrow code* exists that requires only one bit per node. This code is shown to be optimal in the case of outdegree 2, i.e. binary trees. The author prefers a *bracket notation*, which can be seen as the generalization of the arrow code. The cost of representing a tree in this way is $(2n - \ell)\log 3$ bits, where n is the total number of nodes and ℓ is the number of leaves. There are simple algorithms based on the bracket notation that can be used to traverse the tree from the root to a leaf

and also from a leaf to the root. Also modifications of a tree, such as adding, removing, or deleting subtrees, can be done easily. Measures based on the *entropy* of the bracket notation are a good measure of the structural complexity of a tree.

Macq, Marichal and Queluz [237] also consider efficient representations of trees. In their case the tree is the result of a decomposition of a 2-D image into uniform subregions. Their tool is a *truncated run-length code*, where the truncation length is determined by estimated symbol probabilities. This code is applied to the decomposition tree in such a way that the parts of the tree that describe neighboring regions are treated similarly under the assumption that neighbors are correlated. Experimental results support their assumption to the extent that the compression is improved as compared to a level-by-level traversal of the tree. However, the compression obtained is not very high.

In [259], Salden, Aldershoff, Iacob and Otte discuss a method to classify multimedia objects automatically. Classification, as well as prediction and identification, can benefit from a probabilistic problem setting where the object is assumed to be selected from a set of objects with a known or unknown probability. In the case of known probabilities efficient decisions often turn out to be Huffman-like tree structures. When the probabilistic behavior of the underlying selection mechanism is not (completely) known, universal methods (see section 2.2) help in finding the proper model and the efficient decision or classification method.

2.2 Universal Methods

Non-universal codes as described in Section 2.1 can only be designed based on the source statistics. However, it is also possible to construct codes that perform well, i.e. that achieve entropy, for a whole *class* of sources. As an example, we discuss how binary sequences x^N of length N generated by memoryless sources with unknown parameter $\theta \stackrel{\Delta}{=} \Pr\{X = 1\}$ can be *universally* encoded with a prefix-suffix method (see e.g. Schalkwijk [29]). The prefix consists of the number of ones $e(x^N)$ occurring in x^N, therefore the length of the prefix is $\lceil \log(N+1) \rceil$ binary digits. The suffix now specifies the sequence x^N given the number of ones $e(x^N)$ it has, hence the suffix length should be $\lceil \log \binom{N}{e(x^N)} \rceil$ bits. The difference between the average codeword length $\overline{L}(\theta)$ and the sequence entropy $Nh(\theta)$ can now be upper bounded as

$$\begin{aligned}
\overline{L}(\theta) - Nh(\theta) &= \sum_{e=0}^{N} \binom{N}{e} \theta^e (1-\theta)^{N-e} \\
&\quad \cdot \left(\lceil \log(N+1) \rceil + \lceil \log \binom{N}{e} \rceil - \log(\frac{1}{\theta^e (1-\theta)^{N-e}}) \right) \\
&< \log(N+1) - H(E) + 2 \\
&\leq \log(N+1) + 2, \qquad (2.12)
\end{aligned}$$

where $H(E)$ is the entropy of E, the random variable representing the number of ones in x^N. Consequently the code rate $\overline{L}(\theta)/N \leq h(\theta) + (\log(N+1)+2)/N$ bits per source symbol, *for any* $0 \leq \theta \leq 1$. Hence the rate of this simple prefix-suffix code will approach entropy arbitrarily close by increasing N.

As we can see, it is rather easy to construct a code that achieves entropy. However what separates the 'men from the boys' is the redundancy behavior of a code. A good code achieves the Rissanen [67] lower bound; its redundancy is then roughly $\frac{1}{2}\log(N)/N$ per source parameter.

In the next sections, we will first discuss universal codes based on repetition times, and then methods based on statistics (context-tree weighting methods and universal coding based on density estimation). In the third section, we will concentrate on variable-to-fixed length universal codes and in the last section we turn to text compression.

2.2.1 Methods Based on Repetition Times and Dictionary Techniques

There are three papers in this area. In the first paper, published in 1986, Willems [220] proposes and analyzes a noiseless data compression method that encodes each source block by referring to the most recent occurrence of this block. This method should be regarded as a partial explanation for the 1977-Lempel-Ziv data-compression method. In 1986, it was only known that this Lempel-Ziv method achieves entropy in a somewhat superficial manner. Crucial in the analysis in [220] is a result on repetition times that will be stated here. Consider a discrete stationary source with alphabet $\mathcal{X}$ that produces the sequence $\cdots, x_{-1}, x_0, x_1, x_2, \cdots$. First define

$$Q_m(x) \stackrel{\Delta}{=} \Pr\{X_{-m} = x, X_{1-m} \neq x, \cdots, X_{-1} \neq x | X_0 = x\}, \tag{2.13}$$

i.e., the probability that symbol $x \in \mathcal{X}$ with $\Pr\{X_0 = x\} > 0$ has repetition time $m \in \{1, 2, \cdots\}$. If the average repetition time $T(x)$ of this symbol x is defined as

$$T(x) \stackrel{\Delta}{=} \sum_{m=1,2,\cdots} m Q_m(x), \tag{2.14}$$

then $\sum_{m=1,2,\cdots} Q_m(x) = 1$ and

$$\Pr\{X_0 = x\} T(x) = 1 - \lim_{N\to\infty} \Pr\{X_0 \neq x, X_1 \neq x, \cdots, X_N \neq x\}, \tag{2.15}$$

hence the average repetition time $T(x)$ of symbol x is inversely proportional to the probability $\Pr\{X_0 = x\}$ of x for ergodic sources. By encoding a repetition time m with a codeword length roughly equal to $\log m$ binary digits, one achieves entropy for all ergodic sources if the source-block length tends to infinity. Later it turned out that the result in Equation (2.15) became known as Kac's theorem [2]. Consequently in [220], for the first time the connection was made between

universal source coding and Kac's result. This eventually led to the proof that the 1977-Ziv-Lempel algorithm achieves entropy. This proof appeared in Wyner and Ziv [93] in 1994.

In 1990, in [224], Shtarkov and Tjalkens investigated the redundancy of the 1978-Ziv-Lempel data compression method. They focused on the Ma-version of this algorithm. Here the dictionary of strings that can be parsed is always a tree. For this Ma-version they showed that the redundancy of this method decreases not faster than $\mathcal{O}(1/\log(L))$ for memoryless sources. Here L is the codeword length. Actually this is a rather negative result, since we would expect the redundancy to behave as $\mathcal{O}(\log(L)/L)$ according to Rissanen's results [67]. Later the Shtarkov-Tjalkens results were confirmed by Kawabata (1993) for more general sources.

In the third paper in this area [238], Tjalkens and Willems compare the 1977-Ziv-Lempel algorithm to the 1978-Ziv-Lempel method. This very short paper published in 1995 reveals that a weak point of the 1977 method is that the match length has to be specified, while the inefficiency of the 1978 method seems to be related to the limited number of reference points in the past data. Then the authors mention an algorithm that can be seen as an improvement over both the 1977-Ziv-Lempel and the 1978-Ziv-Lempel method in that it does not need to specify the match length in LZ-77 nor that it has a limited number of reference points in LZ-78.

2.2.2 Statistical Methods

Context-Tree Weighting (CTW)

Preliminaries: Context-tree weighting [96] was introduced as a sequential universal source coding method for binary tree sources. Weighting procedures are based on the well-known Elias algorithm (see Section 2.1.3). This method produces for any coding distribution $P_c(x_1 \cdots x_T)$ over all binary sequences of length T a binary prefix code with codeword lengths $L(x_1 \cdots x_T)$ that satisfy

$$L(x_1 \cdots x_T) \leq \log \frac{1}{P_c(x_1 \cdots x_T)} + 2 \text{ for all } x_1 \cdots x_T. \tag{2.16}$$

If the marginals $P_c(x_1 \cdots x_t) = \sum_{x_{t+1} \cdots x_T} P_c(x_1 \cdots x_T), t = 1, \cdots, T$ are sequentially available, arithmetic coding is possible. Accepting a *coding redundancy* of at most 2 bits, we are now left with the problem of finding good coding distributions P_c.

For memoryless binary sources with an unknown parameter θ (the probability of generating a 1), it is reasonable to assign the Krichevsky-Trofimov [57] block probability $P_c(x_1 \cdots x_T) = P_e(a, b)$ to a sequence $x_1 \cdots x_T$ containing a zeros and b ones, where

$$P_e(a, b) \triangleq \frac{1}{2} \cdot \ldots \cdot (a - \frac{1}{2}) \cdot \frac{1}{2} \cdot \ldots \cdot (b - \frac{1}{2})/(a + b)! \quad \text{for } a > 0, b > 0. \tag{2.17}$$

Note that this distribution allows sequential updating. It guarantees uniform convergence of the *parameter redundancy*, i.e., for any sequence $x_1 \cdots x_T$ with actual probability $P_a(x_1 \cdots x_T) = (1-\theta)^a \theta^b$, it can be shown that (see [96])

$$\log \frac{P_a(x_1 \cdots x_T)}{P_c(x_1 \cdots x_T)} \leq \frac{1}{2} \log T + 1 \text{ for all } \theta \in [0,1]. \tag{2.18}$$

In a more general setting the source is not memoryless. We assume that the distribution used by the source to generate the next symbol $X_t, t = 1, \cdots, T$ is determined by the binary sequence $u_t(1) \cdots u_t(D)$, called the *context* of x_t. One can think of sources for which the context consists of the D most recent source outputs, thus $u_t(d) = x_{t-d}, d = 1, \cdots, D$. However, more general context definitions are possible. We assume that the context $u_t(1) \cdots u_t(D)$ is available to the encoder at encoding time and to the decoder at decoding time of symbol x_t.

The mapping M from the context space $\{0,1\}^D$ into the parameter-index set $\mathcal{K}$ is what we call the *model* of the source. To each parameter-index $k \in \mathcal{K}$ there corresponds a parameter $\theta(k) \in [0,1]$. The source generates X_t, with a probability of a 1 equal to $\theta(M(u_t(1) \cdots u_t(D)))$.

If we know the actual model M_a, we can partition the sequence $x_1 \cdots x_T$ in memoryless subsequences and use $P_c(x_1 \cdots x_T | M_a) = \Pi_{k \in \mathcal{K}_a} P_e(a_k, b_k)$ as a coding distribution, where a_k and b_k are the number of instants t for which $x_t = 0$, resp. 1 and $M_a(u_t(1) \cdots u_t(D)) = k$. The image of $\{0,1\}^D$ under M_a is $\mathcal{K}_a$. Again this coding distribution allows sequential updating. For any sequence $x_1 \cdots x_T$, using (2.18) and the convexity of the logarithm, the parameter redundancy can be upper bounded as

$$\log \frac{P_a(x_1 \cdots x_T)}{P_c(x_1 \cdots x_T | M_a)} \leq \frac{|\mathcal{K}_a|}{2} \log \frac{T}{|\mathcal{K}_a|} + |\mathcal{K}_a| \tag{2.19}$$

for all $M_a \in \mathcal{M}$ and $\theta(k) \in [0,1], k \in \mathcal{K}_a$, where $P_a(x_1 \cdots x_T) = \Pi_{k \in \mathcal{K}_a} (1 - \theta(k))^{a_k} \theta^{b_k}(k)$ is the actual probability of $x_1 \cdots x_T$.

If the model is unknown, we can *weight* the coding distributions corresponding to all models M in the *model class* $\mathcal{M}$ and obtain the coding distribution $P_c(x_1 \cdots x_T) = \sum_{M \in \mathcal{M}} P(M) P_c(x_1 \cdots x_T | M)$. Here, $P(M)$ is the a priori probability assigned to the model M in class $\mathcal{M}$. For any sequence $x_1 \cdots x_T$, the *model redundancy* can now be upper bounded as

$$\log \frac{P_c(x_1 \cdots x_T | M_a)}{P_c(x_1 \cdots x_T)} \leq \log \frac{1}{P(M_a)} \text{ for all } M_a \in \mathcal{M}. \tag{2.20}$$

The *total cumulative redundancy* is equal to the sum of the (cumulative) model, parameter and coding redundancies. Using Equations (2.16), (2.19), and (2.20), we can upper bound this total redundancy for any sequence $x_1 \cdots x_T$ in the following

way

$$L(x_1 \cdots x_T) - \log \frac{1}{P_a(x_1 \cdots x_T)} \quad \leq \quad \log \frac{1}{P(M_a)} + \frac{|\mathcal{K}_a|}{2} \log \frac{T}{|\mathcal{K}_a|} + |\mathcal{K}_a| + 2. \tag{2.21}$$

This holds for all models $M_a \in \mathcal{M}$ and parameters $\theta(k) \in [0,1], k \in \mathcal{K}_a$. Rewriting this bound and taking the minimum over all actual source models and parameters, we obtain

$$\begin{aligned} & L(x_1 \cdots x_T) \\ & \leq \min_{M_a \in \mathcal{M}, \theta(k) \in [0,1], k \in \mathcal{K}_a} \left\{ \log \frac{1}{P_a(x_1 \cdots x_T)} \right. \\ & \qquad \left. + \log \frac{1}{P(M_a)} + \frac{|\mathcal{K}_a|}{2} \log \frac{T}{|\mathcal{K}_a|} + |\mathcal{K}_a| + 2 \right\}. \end{aligned} \tag{2.22}$$

Note that Equation (2.22) demonstrates that context weighting methods minimize the *total description length* of a sequence. They exhibit minimum-description-length (MDL) behavior.

In a *tree source*, all contexts that are mapped onto a certain parameter index have a certain prefix in common. We also assume that the context consists of the D most recent source outputs, thus $u_t(d) = x_{t-d}, d = 1, \cdots, D$.

The context-tree weighting method is defined as follows. For each $s \in \{0,1\}^*$ with length $\ell(s)$ not exceeding D, let $a_s(x_1 \cdots x_t)$ and $b_s(x_1 \cdots x_t)$ be the number of times that $x_\tau = 0$, respectively $x_\tau = 1$, in $x_1 \cdots x_t$ for $1 \leq \tau \leq t$ such that $x_{\tau-\ell(s)}^{\tau-1} = s$. The weighted probability corresponding to node s which is denoted by $P_w^s(x_1 \cdots x_t)$ is defined recursively as

$$P_w^s \triangleq \begin{cases} \frac{1}{2} P_e(a_s, b_s) + \frac{1}{2} P_w^{0s} P_w^{1s} & \text{for } 0 \leq l(s) < D, \\ P_e(a_s, b_s) & \text{for } l(s) = D, \end{cases} \tag{2.23}$$

where P_w^s is shorthand for $P_w^s(x_1 \cdots x_t)$ and a_s and b_s for $a_s(x_1 \cdots x_t)$ and $b_s(x_1 \cdots x_t)$, respectively. The weighted coding distribution is now defined as $P_c(x_1 \cdots x_t) \triangleq P_w^\lambda(x_1 \cdots x_t)$, for all $x_1 \cdots x_t \in \{0,1\}^t, t = 0, 1, \cdots, T$, where λ is the empty string. This coding distribution determines the *context-tree weighting method*. It achieves a model redundancy $2|\mathcal{K}_a| - 1$, where $|\mathcal{K}_a|$ is the number of parameters, i.e., the number of leaves in the tree source, or in other words, it holds that $P(M_a) = 2^{1-2|\mathcal{K}_a|}$.

There are 17 papers related to context-tree weighting that appeared in the proceedings of the SITB. In the first one from 1993, Willems, Shtarkov and Tjalkens [231] investigate model classes that extend the tree-model class. Three new recursive weighting methods are specified based on splitting. Crucial is that by making a model class richer we can reduce the parameter redundancy, but then we also need

more bits to specify a model in that class in general. The most general class, class-I, performs "arbitrary splitting". Less general is class-II, which can only "split lexicographically". Class-III refers to "arbitrary-position splitting". The fourth class, the class of tree models, is referred to as "next-position splitting" class.

In [232], Tjalkens, Shtarkov and Willems extend the results of [96] to tree sources with a non-binary alphabet $\mathcal{A}$. Instead of the binary Krichevsky-Trofimov [57] estimator, a Dirichlet estimator is used. To minimize the model redundancy the $(\frac{1}{2}, \frac{1}{2})$-weighting that was used in the binary case is replaced by $(1-1/|\mathcal{A}|, 1/|\mathcal{A}|)$-weighting and each additional leaf costs $h(1/|\mathcal{A}|)/(1 - 1/|\mathcal{A}|)$ bit, where $h(\cdot)$ is the binary entropy function. Then an escape mechanism is used to adjust the Dirichlet estimator to be able to handle sources having symbols that do not occur. Alternatively, sub-alphabet weighting is proposed for such sources. Text-compression simulations show that sub-alphabet weighting is slightly superior to using an escape method. Compression rates as low as 3 bits per ASCII symbol can be achieved.

In 1994, Volf and Willems [234] used an extended version of the CTW method to infer decision trees from classified data. The MDL principle, as in Equation (2.22), should guarantee that the decision tree that is found (using maximizing) is good. The extension consists of using a different model class (a decision tree instead of a context tree) and of applying context maximizing instead of context weighting. The searching complexity is limited using techniques that tell us when splitting a node is certainly useless. Simulations show how the new method compares to techniques proposed by Quinlan and Rivest [77].

In [235] Willems investigates how finite accuracy implementations of the CTW method affect the redundancy. He also studied scaled updating of the Krichevsky-Trofimov estimators and floating-point implementation in the weighted context tree. Better results on the latter topic are presented in Willems (1995).

Tjalkens, Shtarkov and Willems [232], focus on text compression in [236]. Instead of using Dirichlet estimators for non-binary alphabets, they decompose this alphabet into binary components on to which they apply the binary CTW method. Good decompositions can be found using the Huffman [7] method. If many symbols do not occur, the parameter redundancy will be quite high. To avoid this, an adapted version of the Krichevsky-Trofimov estimator from Equation (2.17), called the unary/binary estimator, is proposed in [236]. Later this estimator is referred to as the zero-redundancy estimator.

In [239], Volf and Willems (1995) investigated context-tree maximizing. Maximizing yields the best (MDL) model given a source sequence, whereas weighting averages over all models within the class no matter what the source sequence is. After having determined the MDL model, the encoder encodes the source sequence given this model. An advantage of this method is that the complexity of the decoder can be small compared to that of a decoder for a weighting method. The performance of weighting is better, however. The authors also consider the case

where the decoder complexity is bounded, i.e., where the decoder can only handle tree models having relatively few leaves. For this case they propose the 'Yo-Yo' method. They present model description on-the-fly as a technique to decrease the number of model-specification bits.

In 1996, Volf and Willems [243] studied weighting algorithms for model classes that are more general than the tree-model class, i.e., the class IV. Class III [231] is still more general than the class studied in [243]. Models in class III have the property that they use the "best" context bit for splitting at each point in the context data structure. The model classes that were studied in [243] are tree models extended with 'don't cares'. If at a certain position in the context tree the value of the next context bit is non-informative, it is considered to be a 'don't care'. Two versions are studied; the first one proposed by Suzuki (1995), and a slightly better one presented in [243]. However both methods have a complexity that is comparable to that of class-III methods, but perform poorer.

One year later, Volf and Willems [247] considered branch weighting. In a standard (node) weighting method, the weighted probability of a node is a mix of the estimated probability of that node and the product of the weighted probabilities of its siblings (see Equation (2.23)). Branch weighting produces a product of the mixes of a part of the estimated probability and the weighted probability corresponding to the siblings of the node. Branch weighting can be advantageous for large alphabet sizes.

In 1997, Willems and Tjalkens [248] presented an implementation of the CTW method. Instead of storing both an estimated probability and a weighted probability in each node, they proposed a method that only stores the ratio of two probabilities. This ratio acts as a kind of switch (β) that indicates whether or not further splitting is necessary. The paper also discusses logarithmic representations of (ratios of) probabilities and bit allocations.

An idea of Volf, weighted switching between two source coding algorithms, is studied in [249]. Consider the CTW method and an alternative (companion) algorithm and note that ideally we would like to use *locally* the best of the two. Volf proposes nice weighting techniques to achieve this goal. In [249], Volf and Willems study the performance of several companion algorithms. They achieve a compression that is significantly better than that of standard CTW.

In 1999, Volf, Willems and Tjalkens [252] reported about techniques that can reduce the complexity of implementation for CTW. The number of computations over [248] was reduced by carefully organizing the sequence of operations. Moreover, the binary decompositions that were proposed in [236] were investigated, especially decompositions based on Huffman techniques. Such forward decompositions not only have a positive effect on the redundancy (see [236]), but more importantly, they minimize the number of computations and the number of records that are produced. Within the class of Huffman decompositions, one can search for decompositions that lead to a smaller number of effective parameters and thus

a better compression performance.

In 1999 Vanroose [253] applied the CTW algorithm to language modeling. Language modeling is used in speech recognition. Vanroose studied word-oriented CTW methods. He observed that a perplexity decrease of about 5% was possible relative to classical trigram-based methods. Note that word-based CTW has the (unpleasant) property that each node has many siblings. Applying a context tree of depth 2 is already not straightforward.

Balakirsky and Willems [251] studied a lower bound on the maximal cumulative redundancy of universal coding. The objective of this study was to evaluate the performance of the Krichevsky-Trofimov estimator. Nowbakht, Tjalkens and Willems [255] focused on sources satisfying a permutation property. This property applies to sources whose behavior is determined by the composition of the context and not by its precise value. They first show that the permutation property only applies if all contexts have the same length. Then they present a recursive weighting method resembling the flavor of the class-II method in [231]. Simulations on bi-level images show that this method can outperform classical methods.

In 2001 Stassen and Tjalkens considered parallel implementation of the CTW method at the encoder side. A key result is that a kind of Tunstall [19] procedure yields a well-balanced partitioning of the load over all processors in a two-layer system. A disadvantage of the model is that it requires a pre-scan over the data. Merging the data coming from all the processors is also quite complicated.

Nowbakht and Willems [257] re-investigated the class-I and class-II context weighting methods that were proposed in [231]. They found that models can be realized by different series of splits. By preventing this they could reduce the complexity of these methods. Analysis showed that the improvement was especially significant for class-II methods.

Hekstra [258] studied techniques to reduce the (memory) complexity of context-tree maximizing proposed in [239]. A new pruning method was proposed and the idea (mentioned in [239]) to code the model specification using a Krichevsky-Trofimov estimator is investigated. Hekstra suggested using a short-range Krichevsky-Trofymov estimator to adapt to the fact that nodes that are created initially are more likely to split than nodes that are created during later stages of the compression process. Simulations show that a trade-off is obtained between complexity and performance.

Universal Coding Based on Density Estimation, Infinite Source Alphabets

In 1992, Barron, Györfi and Van der Meulen [227] studied universal coding of finely quantized data. These investigations were based on distribution estimation results that were proposed and analyzed by the authors in [87]. These estimates are consistent in information divergence. Barron *et al.* show in [227] that such distribution estimates lead to universal codes for probability measures that are domi-

nated in I-divergence by a known measure ν.

In an abstract Györfi, Páli and Van der Meulen [228] announce good and bad news for universal noiseless source coding for infinite source alphabets. The bad news is that for any sequence of source coders, there is a memoryless source with finite entropy that produces an infinite average codeword length. However, the good news is that if a fixed coder gives a finite average codeword length for a class of sources, one can construct a universal coder for these sources.

In 1993, Györfi, Páli and Van der Meulen [233] provided proof for their good news result of one year earlier [228]. Their proof was based on distribution estimation techniques of Barron, Gyöfi and Van der Meulen [227, 87].

Closing Remark by the Editors

Despite its relatively small number of contributors, source coding in the Benelux has gained worldwide recognition. The paper introducing the context-tree weighting method by Willems, Shtarkov and Tjalkens was first presented at the 14-th WIC symposium in 1993, see [231]. The full journal paper [96] has received the 1996 IEEE Information Theory Society Paper Award.

2.2.3 Universal Methods for Variable-to-Fixed Length Coding

In 1987, Tjalkens and Willems [222] considered universal variable-to-fixed length codes for binary memoryless sources. They were motivated by a paper of Lawrence [41], who extended the enumerative approach of Schalkwijk [29] to the variable-to-fixed length case. Crucial in the method of Tjalkens and Willems is the probability

$$Q(\underline{x}^*) = \frac{1}{n+e+1}\binom{n+e}{e}^{-1}, \tag{2.24}$$

which is assigned to a sequence $\underline{x}^*$ with n zeros and e ones. Given a design parameter C, the sequence $\underline{x}^*$ is a segment if and only if $Q(\underline{x}^*)^{-1} \geq C$ and $Q(\underline{x}^{*-1})^{-1} < C$. Here, $\underline{x}^{*-1}$ denotes the sequence $\underline{x}^*$ except for the last symbol. If $C \to \infty$, this method achieves entropy, thus $\log(M)/L_{\text{av}}(\theta) \to h(\theta)$ for any source parameter $0 \leq \theta \leq 1$, where $L_{\text{av}}(\theta)$ is the average segment length. Just like Lawrence, the authors proposed using an enumerative approach to do the actual coding. The redundancy behavior of the new method was demonstrated to be superior to that of the Lawrence code.

Three years later, Tjalkens and Willems [225] showed that for any $\delta > 0$, any variable-to-fixed length code with a large enough number M of segments must satisfy

$$\frac{\log(M)}{L_{\text{av}}(\theta)} \geq \left(1 + (1-\delta)\frac{\log\log M}{2\log M}\right) h(\theta) \tag{2.25}$$

for almost all $0 \leq \theta \leq 1$. This result is the variable-to-fixed length memoryless-case counterpart of the famous Rissanen lower bound on the redundancy [67].

Later, Tjalkens and Willems (1992) demonstrated also that their modified Lawrence method proposed in [222] achieves this lower bound on the redundancy.

In 1996, Shtarkov, Tjalkens and Willems [245] studied relative redundancy behavior for binary memoryless sources. Given a code φ, a source segment $\underline{x}$ has a length denoted by $N(\underline{x}|\varphi)$ and an associated codeword length denoted as $L(\underline{x}|\varphi)$. As usual, the *average absolute* redundancy $\overline{r}(\varphi, \theta)$ of code φ, given source parameter θ, is now defined as

$$\overline{r}(\varphi, \theta) \triangleq \frac{\sum_{\underline{x}} P(\underline{x}|\theta) L(\underline{x}|\varphi)}{\sum_{\underline{x}} P(\underline{x}|\theta) N(\underline{x}|\varphi)} - h(\theta). \tag{2.26}$$

The *maximal relative* redundancy $\rho(\varphi)$ is now defined as

$$\rho(\varphi) \triangleq \sup_{\theta} \sup_{\underline{x}} \frac{L(\underline{x}|\varphi)}{-\log(P(\underline{x}|\theta))} - 1, \tag{2.27}$$

hence we compare the codeword length $L(\underline{x}|\varphi)$ to the ideal codeword length $-\log(P(\underline{x}|\theta))$ and search for the worst-case segment $\underline{x}$ and parameter θ. It will be clear that the maximal relative redundancy is unbounded if we do not exclude $\theta = 0$ and $\theta = 1$. In [245], the authors studied both fixed-to-variable length codes as well as variable-to-fixed length codes. They constructed codes based on a probability assignment similar to those in Equation (2.24) and found that the variable-to-fixed length codes outperformed the fixed-to-variable length codes when maximal relative redundancy is the applied criterion.

2.2.4 Text Compression

In 1992, in a one-page paper, Shtarkov and Volkov [229], compared various noiseless techniques for compression of typical computer files. They considered several Ziv-Lempel variants but also string matching techniques (Cleary and Witten (1984)) as well as asymptotically optimal techniques developed by Shtarkov for Markov sources. The best results were obtained by integrating Shtarkov's techniques into partial string matching methods.

CHAPTER 3

Cryptology

H.C.A. van Tilborg (TU Eindhoven)
B. Preneel (K.U. Leuven)
B. Macq (UC Louvain-la-Neuve)

Introduction

Cryptography (see [102] for an excellent handbook) is concerned with the protection of data against malicious parties. In particular, cryptographic primitives try to achieve *confidentiality*, *integrity* and *authenticity*. In Sections 3.1 and 3.2, cryptographic primitives are discussed that assume that sender and receiver, respectively, do not share a common secret key. Section 3.3 discusses the WIC papers on security issues, and Section 3.4 concerns itself with data hiding and related topics.

3.1 Symmetric Systems

In symmetric cryptology, sender and recipient protect the confidentiality and authenticity of the information sent over an insecure channel based on a shared secret key. If one wants to protect the confidentiality of data, one transforms the data (denoted as the plaintext P) under control of a secret key K with the encryption algorithm into the ciphertext C, or $C = E_K(P)$. The recipient can decrypt the ciphertext C with the decryption algorithm to obtain the plaintext or $P = D_K(C)$. It should be infeasible for an opponent who does not know the key K to deduce

[1] This chapter covers references [261] – [345].

information on the plaintext from the ciphertext. One can also assume that the opponent knows part of the plaintext, and tries to deduce the key or additional plaintext; this is called a known plaintext attack. In a chosen plaintext (respectively chosen ciphertext) attack, the attacker can submit plaintexts (respectively ciphertexts) of his choice and try to obtain additional information on plaintexts or on the key.

For data authentication, the sender appends a short string $\mathrm{MAR}_K(P)$ to the plaintext which is a function of the plaintext and the secret key; here MAC is the abbreviation of Message Authentication Code. On receipt of a plaintext P' and its MAC value, the receiver can recompute the value $\mathrm{MAR}_K(P')$; if this equals $\mathrm{MAR}_K(P)$, the receiver can deduce that with high probability $P' = P$, that is, the plaintext is coming from a particular sender and has not been modified. Indeed, an opponent who does not know the key should not be able to predict the correct value of $\mathrm{MAR}_K(P^*)$ for an arbitrary plaintext P^*. Desmedt, Govaerts and Vandewalle study the problem of information authentication from a risk analysis viewpoint [266]: increasing the cryptographic redundancy in the message will increase the security (and hence decrease the expected profit for an active attacker), but it will increase the transmission cost. This results in a simple optimization problem.

This section presents an overview of the state of the art of symmetric cryptography. First secret key systems are treated from an information theoretic standpoint; this is followed by an introduction of the system based and complexity theoretic approach. Next building blocks and designs for practical symmetric systems are discussed. Finally techniques are presented for establishing symmetric keys.

3.1.1 Information-Theoretic Approach

Encryption algorithms are almost as old as writing itself. Until the beginning of the 20th century, most systems were designed for manual operation. The basic operations used are substitutions (permuting the alphabet) and transpositions (permuting the location of letters in a sequence). While none of these systems offer adequate security today, these two operations form essential building blocks for modern symmetric cipher systems. With the development of telegraph and radio communications, encryption techniques gained quickly importance. In this context, the radio engineer Vernam proposed a very simple and elegant system in 1917, known as the one-time pad or the Vernam scheme.

Denote the i-th bit of the plaintext, ciphertext, and key stream with P_i, C_i, and K_i, respectively. The encryption operation can then be written as $C_i = P_i \oplus K_i$, $i = 1, 2, \ldots t$ (here $\oplus$ denotes addition modulo 2 or exor). The decryption operation is identical to the encryption (the cipher is an involution): indeed, $P_i = C_i \oplus K_i$. Vernam proposed to use a perfectly random key sequence, that is, the bit sequence K_i, $i = 1, 2, \ldots$ should consist of a uniformly and identically distributed sequence of bits.

Vernam believed that his cipher was unbreakable, but he did not know how to

prove this. A disadvantage of the Vernam scheme with major practical implications is that the key has the same size as the plaintext. In spite of the long key, the Vernam algorithm is still used by diplomats and spies; it has been used until the late 1980s for the red telephone between Washington and Moscow.

In 1949 – one year after the publication of his landmark paper on information theory [3] – Shannon published his seminal work on cryptology [5]. First he defined what it means for an encryption scheme to be secure against an opponent with unlimited computational capability; a scheme offers *perfect secrecy* if $H(P|C) = H(P)$, or the ciphertext provides the opponent no new information on the plaintext. Shannon proved that the Vernam scheme offers perfect secrecy. Moreover, he showed that the key size of the Vernam scheme is optimal: an encryption scheme can only provide perfect secrecy if $H(K) \geq H(P)$. If one wants to guarantee that the encryption scheme is secure for any plaintext distribution, this implies that the key has to be at least as long as the plaintext.

Most practical systems are imperfect. Shannon proposed the concept of *key equivocation* to study these systems: $H(K|C_1, C_2, C_3, \ldots, C_s)$ measures the uncertainty of the opponent about the key after observing the first s bits of the ciphertext. He defined the *unicity distance* u as the smallest index s^* such that $H(K|C_1, C_2, C_3, \ldots, C_{s^*}) \approx 0$. If an opponent observes u ciphertext bits, he has obtained sufficient information to determine the secret key uniquely (note that the computational power to do this in practice may be beyond reach, but for now it is assumed that this computational power is unlimited). Shannon shows that for a random cipher, the unicity distance is approximately equal to $H(K)/r$ with r the percentage redundancy of the plaintext, or $r = 1 - H(P)/s$, where s is the number of observed ciphertext bits. For a typical English text $r \approx 3/4$, hence if the key is chosen according to a uniform distribution, the unicity distance $u = 4/3$ of the length of the key in bits. It is clear that these information theoretic results can be generalized for an arbitrary alphabet, but in order to simplify the discussion, this section will only consider the binary case.

Van Tilburg and Boekee [269] generalize the unicity distance to the P_e distance of a cipher model (which includes both the properties of the plaintext and the key source): they define this distance as the minimal expected ciphertext length required to "break the cipher" with an average error probability of P_e. This definition can be made concrete by specifying the model in several ways: "breaking the cipher" can mean recovering the plaintext or recovering the key in a ciphertext-only attack. However, one can also consider a known plaintext attack. This contribution studies the variants of the definition and resolves the ambiguity created by the approximation ≈ 0 in the definition of unicity distance, which is not completely satisfactory. Boekee and Van der Lubbe study the security of simple transposition ciphers in this model [273].

In a practical stream cipher, one replaces the random key sequence of the Vernam scheme by a pseudo-random key stream, that is, a key stream that is generated from a short key K but that looks random to an opponent who has limited comput-

ing power. One generates the bit sequence K_i with a finite state machine in which the initial state, the next state function and the output transformation may depend on the key K. Feedback shift registers form an important building block of stream ciphers, since they allow for efficient hardware implementations. The internal state X of such a shift register of length n is denoted with $(X_0, X_1, \ldots, X_{n-1})$, $X_i \in GF(2)$. The next state function is then given by

$$g(X_0, X_1, \ldots, X_{n-1}) = (X_1, X_2, \ldots, X_{n-1}, f(X_0, X_1, \ldots, X_n)) \,. \qquad (3.1)$$

The maximum order complexity of a given sequence is the length of the shortest feedback shift register that can generate this sequence. If the feedback function f is linear over $GF(2)$, that is, $f(X_0, X_1, \ldots, X_n) = \sum_{i=0}^{n-1} a_i X_i$, $a_i \in \mathrm{GAG}(2)$, this is called a Linear Feedback Shift Register (LFSR) over $GF(2)$. The linear complexity of a given sequence is the length of the shortest LFSR that can generate this sequence.

Jansen and Boekee [275] apply information theory to study two classes of stream ciphers. In the first class, the key stream is a sequence $\mathcal{Z} = \{Z_0, Z_1, \ldots, Z_s\}^\infty$, $Z_i \in GF(2)$, which is started in an arbitrary phase j. Hence the key stream sequence equals $\{Z_j, Z_{j+1}, \ldots, Z_{j+s-1}\}^\infty$. They define the Character Uncertainty Profile (CUP) as the sequence of conditional entropies $H(Z_s|Z_1, \ldots, Z_{s-1})$, $s \geq 1$, and the Phase Uncertainty Profile (PUP) as the sequence of conditional entropies $H(j|Z_1, \ldots, Z_{s-1})$, $s \geq 0$. Then they show the following two results: the CUP is monotonically non-increasing and becomes zero after c bits, with c the maximal order complexity of the sequence $\mathcal{Z}$. The PUP is monotonically decreasing and becomes 0 after c bits. From this one can induce that this class of stream ciphers depending on a secret phase is very weak. Next they propose a second class of stream ciphers, for which the user key K selects a sequence from an ensemble and for this stream cipher they study the Sequence Uncertainty Profile $H(K|Z_1, \ldots, Z_{s-1})$.

Several applications (e.g. voting schemes) require not only protection of the data communicated, but also of the identities of the sender and/or receiver. Diaz, Claessens, Seys and Preneel propose an information-theoretic measure to quantify the degree of anonymity and apply this to the concrete problem of targeted advertising with privacy protection [323].

3.1.2 System-Based and Complexity-Theoretic Approach

The information-theoretic approach has as important advantage that the security offered is independent of the computational power or budget of an adversary. Moreover, it also brings fundamental insights into the secure communications. However, Shannon also realized that one needs to use a more pragmatic approach in order to design practical systems. This approach tries to produce practical solutions for basic building blocks such as one-way functions, pseudo-random bit generators (stream ciphers), and pseudo-random permutations. The security estimates are based on the best algorithm known to break the system and on realistic estimates of the necessary computing power or dedicated hardware to carry out

the algorithm. By trial and error procedures, several *cryptanalytic principles* have emerged, and it is the goal of the designer to avoid attacks based on these principles. The second aspect is to design *building blocks with provable properties*, and to assemble such basic building blocks to design cryptographic primitives.

The complexity-theoretic approach, which has been introduced in 1980s develops formal definitions of cryptographic concepts and tries to develop formal reductions and impossibility results in a context where the opponent has limited computing power. For example, one formally proves that if a particular object (e.g., a one-way function) exists, another object exists as well (e.g., a secure stream cipher). While this approach has been very successful, proving lower bounds on concrete problems has remained elusive and cryptology still relies on a large number of primitives that are constructed based on the system-based approach.

An important research problem is how hard it is to invert a specific one-way function. While we cannot prove good lower bounds for any concrete function from n bits to n bits, it is clear that inverting a randomly chosen function on a single element randomly chosen in the range takes on average 2^{n-1} steps. However, in a cryptanalytic context, one often needs to invert the same function multiple times. This is the case if one wants to recover the secret key of a block cipher or a stream cipher or if one wants to recover passwords from their image under a one-way function. Hellman [54] showed that in this case the cost can be reduced to a precomputation of 2^n function evaluations, after which $2^{2n/3}$ $2n$-bit values are stored. Based on this information, a single element can be inverted in $2^{2n/3}$ function evaluations with an average success probability of 1/2. Borst, Preneel and Vandewalle [306] study a variant of this scheme suggested by Rivest. This approach reduces the memory accesses, which significantly reduces the implementation cost of this trade-off.

3.1.3 Building Blocks for Symmetric Cryptography

Following the approach suggested by Shannon [5], block ciphers (cf. Section 3.1.4) consist of a repeated application of two components: small nonlinear building mappings from n to m bits (also known as S-boxes) and linear mappings which diffuse or spread local information. A popular way to construct stream ciphers (cf. Section 3.1.4) is the combination of linear feedback shift registers (cf. Section 3.1.1) and nonlinear Boolean functions or S-boxes. Another approach consists of combining sequences (cf. Section 3.1.1). This section discusses some results on Boolean functions, S-boxes and sequences with cryptographic applications.

Consider a Boolean function $f(\underline{x})$ with domain the vector space $GF(2)^n$ of binary n-tuples $(x_1, x_2, \ldots, x_n)$ that takes the values in $GF(2)$. The *Walsh transform* of $f(\underline{x})$ is the real-valued function over the vector space $GF(2)^n$ defined as

$$\hat{F}(\underline{w}) = \sum_{\underline{x}} (-1)^{f(\underline{x})} \cdot (-1)^{\underline{x} \cdot \underline{w}} . \tag{3.2}$$

This is an orthogonal transform that can be computed in time $n \cdot 2^n$ from the truth table. The minimum distance of the function $f(\underline{x})$ to all affine functions is equal to $2^{n-1} - 1/2 \max_{\underline{w}} \mid \hat{F}(\underline{w}) \mid$. Another useful representation for cryptographic applications is the the *algebraic normal form*:

$$f(\underline{x}) = a_0 \oplus \sum_{1 \leq i \leq n} a_i x_i \ \oplus \sum_{1 \leq i < j \leq n} a_{ij} x_i x_j \ \oplus \ldots \oplus a_{12\ldots n} x_1 x_2 \cdots x_n \,. \quad (3.3)$$

The nonlinear order of a Boolean function is defined as the degree of the highest order term in the algebraic normal form. Jansen and Boekee show in [270] how one can compute the ANF from the truth table of a Boolean function using a fast transform in time $n \cdot 2^n$. Kholosha [343] generalizes these two transforms to the tensor transform, which he studies for functions over $GF(q)$.

A Boolean function is *balanced* if the Hamming weight of its truth table is 2^{n-1}; one can show that this implies $\hat{F}(\underline{0}) = 0$. A Boolean function is called *correlation immune of order* m if $\hat{F}(\underline{w}) = 0$ whenever[1] $1 \leq \mathit{hwt}(\underline{w}) \leq m$. This implies that knowledge of m input bits yields no information on the output. A Boolean function is called *resilient of order* m if it is balanced and correlation immune of order m or $\hat{F}(\underline{w}) = 0$ whenever $0 \leq \mathit{hwt}(\underline{w}) \leq m$. A Boolean function $f(\underline{x})$ satisfies the *propagation criterion of degree* k *(PC of degree* k*)* if $f(\underline{x})$ changes with a probability of one half whenever i ($1 \leq i \leq k$) bits of $\underline{x}$ are complemented. Bent functions are functions that satisfy PC of maximal degree n; the absolute value of their Walsh spectrum is constant and they have maximal distance $2^{n-1} - 2^{\frac{n}{2}-1}$ to all affine functions. Carlet and Klapper [334] provide a new upper bound on the number of bent functions and on the number of resilient functions of order m for m large.

A Boolean function $f(\underline{x})$ of n variables satisfies the *propagation criterion of degree* k *and order* m (PC of degree k and order m) if any function obtained from $f(\underline{x})$ by keeping m input bits constant satisfies PC of degree k. A Boolean function $f(\underline{x})$ of n variables satisfies the *extended propagation criterion of degree* k *and order* m (EPC of degree k and order m) if knowledge of m bits of $\underline{x}$ gives no information on $f(\underline{x}) \oplus f(\underline{x} \oplus \underline{a})$, whenever $1 \leq \mathit{hwt}(\underline{a}) \leq k$. The relation between the propagation criteria and extended propagation criteria was studied by Preneel, Van Leekwijck, Van Linden, Govaerts and Vandewalle in [277].

Daemen, Van Linden, Govaerts and Vandewalle [281] analyze the cryptographic properties of multiplication with a constant modulo $2^n - 1$. They study the properties of the individual output bits, and analyze the correspondence between input and output differences, which is important for the study of differential attacks on ciphers that use this S-box. They also develop an algorithm to find the best multiplication factors for large values of n (the value 32 is of particular interest for software implementations).

[1] Here and in the sequel, $hwt(\underline{w})$ denotes the Hamming weight of the vector $\underline{w}$

Maximum length sequences of length $2^n - 1$ derived from an n-bit Linear Feedback Shift Register (LFSR) are an essential building block for stream ciphers. Another important sequence of length 2^n is a de Bruijn cycle of degree n, i.e., a circular pattern of 2^n bits in which each of the 2^n n-bit patterns occurs exactly once. The number of de Bruijn cycle of degree n equals $2^{2^{n-1}-n}$. Franx, Jansen and Boekee [272] present an efficient algorithm which can construct $O(2^{\alpha_n})$ de Bruijn cycles of degree n with $\alpha_n = 2n/\log_2(2n)$. It is based on the principle of joining cycles of LFSRs. Jansen has proved that the resulting de Bruijn cycles are indeed unique [276]. He also provides an extension by allowing also cycles of nonlinear feedback shift registers with feedback function of nonlinear order r. If $n \geq 2^{r+3}$, this results in a value of α_n equal to $\sum_{i=0}^{r} \binom{n}{r}$; however, the algorithm is no longer efficient.

3.1.4 Practical Constructions of Stream Ciphers, Block Ciphers and Hash Functions

The principles behind the construction of stream ciphers have been introduced in Section 3.1.1. Several concrete designs of stream ciphers have been proposed and analyzed. Daemen, Govaerts and Vandewalle [280] propose a stream cipher using cellular automata; they study the invertibility and the cycle structure of several simple update rules such as $a'_i = a''_{i-1} \oplus a''_i \oplus a''_{i+1}$ with $a''_i = a_{i-1} \oplus \overline{a}_i a_{i+1}$. Here indices are taken modulo N, with N the size of the cellular automaton. They also propose to increase the cryptographic strength by adding a fixed rotation, that is, replacing a'_i in the previous expression by a'_{i+11} for a cellular automaton of length $N = 127$. The same authors study in [289] the resistance of the mapping $a'_i = a_{i-1} \oplus \overline{a}_i a_{i+1}$ with respect to linear and differential cryptanalysis (these attacks are explained below).

Meijer and Jansen [318] construct run-permuted sequences which are obtained by permuting the runs of ones and zeroes of a given sequence. They construct the sequence by combining a set of counters, the bits of which are permuted using a key register; subsequently an S-box maps the resulting sequence to a set of integers, which is then run-length decoded, resulting in a sequence with large period and good uniformity properties. Canteaut and Filiol [333] study the security of filter generators: these are stream ciphers in which a Boolean function is applied to several stages of an LFSR. In a correlation attack, the Boolean function is approximated by an affine function and this approximation is used to deduce information on the secret state of the LFSR. They show that by using all affine functions (rather than just the best ones) the amount of key stream can be reduced at the cost of a higher computational load. This attack however is less dependent on the particular choice of the Boolean function.

A n-bit block cipher with a k-bit key is a set of 2^k permutations on n-bit strings. In contrast to stream ciphers, block ciphers operate on larger blocks; typical values of n are 64 (for DES, the Data Encryption Standard) and 128 (for AES, the Advanced Encryption Standard). Almost all block ciphers are *iterated ciphers*: they consist of an r-fold repetition of a simple key-dependent function. The *key-schedule* al-

gorithm computes from the k-bit user key a number of keys K_i that are used each round. DES (Data Encryption Standard) is the best known example of a Feistel cipher; it was standardized in 1977 by the US government as FIPS 46 (Federal Information Processing Standard 46). The input of a round of a Feistel cipher is divided into two halves denoted with L_i and R_i respectively. The new left half is the old right half, and the new right half is the modulo 2 sum of the old left half and a function of the key and the old right half:

$$L_{i+1} = R_i, \qquad R_{i+1} = L_i \oplus f_{K_i}(R_i)\,. \tag{3.4}$$

The advantage of a Feistel cipher is that decryption is equal to encryption with the round keys in reverse order. Nakahara Jr. , Vandewalle and Preneel [312] study general Feistel networks in which inputs are divided into more than two subblocks. They compare several alternatives with respect to maximal diffusion and present some implications on four contenders in the AES competition. In 1997, the US government launched an open competition to define a 128-bit block cipher which would replace DES. Fifteen candidates were admitted to the competition; in 2000, the Rijndael algorithm was selected as the winner, and the new FIPS 197 standard containing AES was published in 2001. Rijndael was a design of the Belgian cryptographers Daemen and Rijmen.

The most important attacks on block ciphers are linear and differential cryptanalysis. In linear cryptanalysis, one tries to construct an approximation of a cipher of the form $\alpha \cdot P \oplus \beta \cdot C \oplus \gamma \cdot K = 0$ with probability $1/2 + \epsilon$ with positive bias $|\epsilon|$. Here α, β, and $\gamma \in \mathrm{GAG}(2^n)$ and $\cdot$ denotes an inner product. In a differential attack, one tries to find an input difference $P \oplus P'$ that yields a particular output difference $C \oplus C'$ with probability significantly larger than $1/2^n$.

Harpes, Kremer and Massey [287] generalize linear cryptanalysis to threefold sums: $f'(P) \oplus f''(C) \oplus \bigoplus_{i=1}^{r} h_i(K_i) = 0$ with probability $1/2 + \epsilon$ with positive bias $|\epsilon|$. Here f' and f'' are balanced Boolean functions and the h_i's are arbitrary Boolean functions. The authors also analyze carefully which assumptions are required for a general linear attack; one particular element is the piling-up lemma, that is, how can one compute the probability of a threefold sum for a block cipher based on the probabilities of similar sums for each of the round functions.

Standaert, Rouvroy, Piret, Quisquater and Legat [339] use linear approximations over $\mathbb{Z}_4$ of the rounds (as proposed by Parker and Raddum); this results in an approximation of degree 2 over $\mathbb{Z}_2$. If the key addition is performed modulo 2 (which is the case in many block ciphers), one can transform this to linear approximation with a key dependent bias. The authors show that this approach results in an improved attack on the block cipher Q (a candidate submitted to the NESSIE competition; NESSIE was an open European competition for a broad range of cryptographic primitives, started in 2000 and completed in 2003; for more details, see `http://www.cryptonessie.org`).

Ciet, Piret and Quisquater [335] present an overview of attacks on the key schedule of a block cipher. They analyze which key schedules may be vulnerable to related

key attacks, in which an opponent obtains the encryption of two plaintexts under keys with a known relation. They also treat slide attacks, in which two instances of a block cipher are considered with keys and input chosen in such a way that a large number of inner rounds have identical inputs. Two improvements on variants of a slide attack are presented.

A structural attack on a block cipher is an attack which exploits its word-oriented structure, for example by analyzing ciphertexts corresponding to a set of plaintexts which take all values in one input word and are constant in the others. A SQUARE attack is a special case of a structural attack. Nakahara Jr. , Barreto, Preneel, Vandewalle and Kim [328] present a SQUARE attack on reduced versions (2.5 out of 8.5 rounds) of the block cipher IDEA; a novel related key variant of the SQUARE attack is presented as well.

Van Rompay, Preneel and Vandewalle [305] present an overview of the security and performance of the cryptographic hash functions of the MD4-family, which includes MD5, SHA-1 and RIPEMD-160. They evaluate which members offer (second) pre-image resistance and collisions resistance.

Struik proposes two block cipher modes of operation that offer in one pass authenticated encryption [315], which is almost twice as efficient as the encrypt-then-MAC model. In the CBC (Cipher Block Chaining) mode the ith ciphertext block is computed as $C_i = E_K(P_i \oplus C_{i-1})$, where P_i denotes the ith plaintext block ($1 \leq i \leq s$), $E_K()$ denotes encryption with the block cipher E under key K and C_0 is the initial value IV. The redundancy consists of an extra plaintext block P_{s+1} which is computed from $(P_1 \oplus \mathrm{IV}) \oplus \bigoplus_{i=2}^{s+1} A^{i-1}P_i = 0$, with A being a simple linear function in $\mathrm{GAG}(2^n)$. It is shown that this mode offers heuristic security against permuting blocks and known plaintext attacks, but that it may be vulnerable to replay attacks and to certain chosen ciphertext attacks. In the second scheme the linear mapping A is used in the feedback $C_i = E_K(P_i \oplus A(P_{i-1} \oplus C_{i-1}))$ and $P_{s+1} = \mathrm{IV}$.

Van der Lubbe, Spaanderman and Boekee [278] study two transposition systems for image encryption: the first system uses a de Bruijn sequence to define a pseudo-random transposition; the second system swaps pixels of the upper and lower half under control of a pseudo-random sequence. Macq and Quisquater [284] present an algorithm for lossless image encryption which allows for compression after encryption. The main idea is to employ a multi-resolution scheme, in which only the details at higher resolution are encrypted using a permutation of rows or columns.

3.1.5 Symmetric Key Establishment

Symmetric cryptographic mechanisms move the problem of protecting information to the problem of establishing secret keys. Jansen presents a key pre-distribution scheme [267] in which a central entity distributes key material; each of the N parties stores $N-1$ keys to communicate securely with the other parties. He considers the problem of asynchronous updating of these keys and presents a solution

in which every party stores $3N-1$ keys and receives $2N-1$ keys during each key update. In [268] Jansen shows how to generate from a key a simple public identifier at low computational cost. A straightforward solution consists of applying a one-way function to the key, but in 1986 this was too expensive. The proposed alternative selects a random subset of key bits; the information theoretic leakage on the key is analyzed and a practical construction based on LFSRs is presented.

The authenticated key establishment protocol of GSM is described by Van Tilburg [317]; he also presents an overview of the GSM security architecture and discusses its limitations. The shortcomings of the encryption algorithms A5/1 and A5/2 are explained, together with the weakness of COMP-128, a popular choice of the combined entity authentication and key generating algorithm A3/A8 (A3/A8 is operator dependent, while A5/1 and A5/2 are GSM standards). He also offers a perspective on the continued development of GSM standards; he evaluates more in particular the prospects of WAP (Wireless Application Protocol) and STK (SIM Toolkit).

Access of an opponent to the secret key means that the security of a system is compromised completely. In order to mitigate this risk, one can use secret sharing techniques introduced by Shamir [51]: a key is divided into shares, and only an authorized subset of users can recover the secret. In a threshold scheme, an authorized subset consists of t or more out of n users. Nikov, Nikova, Preneel and Vandewalle [329] construct proactive secret sharing schemes, that is, schemes for which the shares are updated regularly; this defeats opponents who can compromise some of the authorized users, but who are never able to subvert all opponents in an authorized set. The construction presented is information theoretically secure and works for all access structures (sets of authorized users) which admit a linear secret sharing scheme.

Hekstra and Van Tilburg [298] propose a solution to the broadcast encryption problem: a message needs to be sent to all users, but only authorized users should be able to decrypt it. The crux of their solution is that all broadcast participants know the decryption algorithms of the other participants, but not of their own. A broadcast message is sent encrypted with all the algorithms of the non-authorized users. Hence, only authorized participants can decrypt and read the message. They show that their scheme is optimal in the Shannon sense.

Bechlagem [330] presents a multi-cast key distribution protocol in with a central entity distributes a common key to n users with the following properties: use of pseudo-random functions rather than block ciphers, mutual entity authentication between the central entity and the parties, guaranteed key freshness and forward secrecy. The ingredients of the protocol are the Chinese Remainder Theorem and Shamir's polynomial secret sharing scheme. The protocol requires that all the n parties are active.

A completely different line of research studies the establishment of a secret key over a noisy channel, as introduced by Wyner [37] in 1975. In Wyner's model,

known as the wire-tap channel, the information of sender (X), receiver (Y) and opponent (Z) form a Markov chain $X \longrightarrow Y \longrightarrow Z$ of random variables. Wyner shows that in sender and receiver can use this channel to agree on a common secret key. He shows that the secrecy capacity of this channel is equal to $C_s(P_{Y,Z|X}) = max_{P_X} I(X;Y|Z)$ for $P_{Y,Z|X} = P_{Y|X} \cdot P_{Z|Y}$. Piret shows that if the channels are binary symmetric channels, the capacity can be achieved using binary linear codes [261]. Wyner's model has been generalized in several ways. Csiszár and Körner study the secrecy capacity of the Broadcast Channel with Confidential messages (BCC), which has discrete memoryless channels between sender and recipient and between sender and opponent, or $X \longrightarrow (Y, Z)$.

Maurer [91] and Alshwede and Csiszár analyze the secrecy capacity if a noiseless authenticated public channel is added to the BCC. Maurer's protocol can be divided into three phases: a coding gain phase, in which sender and receiver exchange coded information and make a reliability decision; a reconciliation phase, in which sender and receiver exchange redundant information and apply error correction techniques to generate a shared secret string; and a privacy amplification phase, in which sender and receiver distill a shorter string on which the opponent has only non-negligible information. Van Dijk [294] generalizes the reliability estimation technique for the coding gain phase and shows that the coding gain can be improved.

A BCC can also be realized using quantum channels; information can then be transmitted through e.g., the polarization of photons. The security of the resulting protocol is then based on the assumption that quantum physics offers an accurate model of our physical world, and more in particular on the validity of Heisenberg's uncertainty principle. Van Dijk and Koppelaar [300] study protocols for the BCC with public channel in which the opponent can intercept and resend photons. They compute a probabilistic upper bound on the amount of information leaked to the opponent as a function as the number of errors observed between the strings of sender and receiver.

Balakirsky [310] studies the secrecy capacity of the binary multiplying channel, where the opponent can only observe the logical AND of the input of sender and receiver; sender and receiver observer this result together with their own input. It is shown that the asymptotic secrecy capacity equals 0.292893... keys bits/communicated bit, and a construction is provided that achieves this bound.

Sometimes the goal of an interaction is not the transmission of a particular message, but it is sufficient for the receiver to know whether or not a particular message has been sent; this is known as the identification problem. Verboven studies this problem for a stochastically varying channel [279].

3.2 Asymmetric Systems

In 1976, Diffie and Hellman [40] introduce the novel idea of public key cryptosystems. In such systems, each user will have two matching algorithms at his disposal: a public one and a matching second one that has to remain secret. How these systems work will become clear from Section 3.2.2.

3.2.1 The Discrete Logarithm System

In the same publication [40], Diffie and Hellman describe a public key agreement scheme which is based on the difficulty of computing logarithms over a finite field. Let α be a primitive element of a finite field $GF(q)$. This means that each nonzero element c in $GF(q)$ can be written as a power of α, so $c = \alpha^m$ for some $0 \leq m < q - 1$.

For a given value of m, one can compute c very efficiently by means of repeated squaring and/or multiplication by α in a way that is indicated by the binary representation of m. For instance, the binary representation 10101011 of $m = 171$ leads to the following exponentiation:

$$\alpha^{171} = (((((((\alpha)^2)^2\alpha)^2)^2\alpha)^2)^2\alpha)^2\alpha. \tag{3.5}$$

The opposite problem of finding m given c is assumed to be difficult in general. It is called the *discrete logarithm problem* (see e.g. [102]). This discrepancy in computing time can be used to make a public key distribution system. This system makes it possible to agree on a common secret over a public channel. Later, the same principle has been used to design cryptosystems and digital signature schemes. So, here we assume that A and B want to communicate with each other using a conventional cryptosystem, but have no secure channel to exchange a key. They proceed as follows.

Diffie-Hellman Key Exchange
Preliminary work: Each user U chooses a secret exponent m_U, $1 \leq m_U < q-1$, at random, computes $\alpha^{m_U} = c_U$ and makes c_U public.

Key Determination: Users A and B can easily agree on the secret key $k_{A,B} = \alpha^{m_A m_B}$. Indeed, A can compute $k_{A,B}$ by raising the publicly known c_B to the power m_A, which only A himself knows. This follows from

$$c_B^{m_A} = (\alpha^{m_B})^{m_A} = \alpha^{m_A m_B} = k_{A,B}. \tag{3.6}$$

Similarly, B finds $k_{A,B}$ by computing $c_A^{m_B}$.

If somebody else is able to compute m_A from c_A (or m_B from c_B), she can compute $k_{A,B}$ just like A or B did. By taking q sufficiently large, one can make the computation time of solving this logarithm problem prohibitively large. Diffie and Hellman suggest to let q be a prime of about 100 digits long. Now we would rather suggest to take 300 to 600 digits long numbers. A different way of finding $k_{A,B}$

from c_A and c_B does not seem to exist.

Already at an early stage people, realized that other group structures could be used for a secure key exchange. The most notable example described years later was the elliptic curve addition group (see [102]).

Massey explains in [264] a method to take discrete logarithms in arbitrary groups that is known under the name of the "Baby-step Giant-step" method. The method allows a complete trade-off between running time and required memory: q^u time complexity versus q^{1-u} memory, for any $0 \leq u \leq 1$.

A further method that he explains is the Pohlig-Hellman technique to reduce the original discrete logarithm problem into several smaller ones (and for the cryptanalyst preferably much smaller ones) by making use of the factorization of $q-1$ and the Chinese Remainder Theorem.

3.2.2 The RSA Cryptosystem

In 1978, R.L. Rivest, A. Shamir and L. Adleman [49] proposed a public key cryptosystem that has become known as the RSA system. It makes use of the following theorem.

Euler's Theorem
Let $\phi(n) = |\{1 \leq i \leq n \mid \gcd(i,n) = 1\}|$ be Euler's ϕ-function. Then for all integers a and n with $\gcd(a,n) = 1$, one has $a^{\phi(n)} \equiv 1 \pmod{n}$.

The RSA cryptosystem
Preliminary work: *Each user U chooses two large, different prime numbers, say p_U and q_U. Let $n_U = p_U \cdot q_U$ (so $\phi(n_U) = (p_U - 1)(q_U - 1)$). Secondly, U chooses a public exponent $1 < e_U < \phi(n_U)$ such that $\gcd(e_U, \phi(n_U)) = 1$. Then user U computes (e.g. with the extended version of Euclid's Algorithm) the secret exponent d_U from $e_U \cdot d_U \equiv 1 \pmod{\phi(n_U)}$. User U publishes e_U and n_U, but keeps d_U secret.*

Encryption: If user A wants to send a secret message to user B, he represents his message by a number m, $0 < m < n_B$. User A looks up e_B and n_B and sends the ciphertext

$$c \equiv m^{e_B} \pmod{n_B}. \tag{3.7}$$

Decryption: User B can recover m from c by computing $c^{d_B} \pmod{n_B}$. Indeed, for some integer l one has that $c^{d_B} \equiv m^{e_B d_B} \equiv m^{1+l\phi(n_B)} \equiv m.(m^{\phi(n_B)})^l \equiv m \pmod{n_B}$.

A cryptanalyst can compute m from c in exactly the same way as B, once he knows the secret d_B. Just like B, he is able to compute d_B from the publicly known e_B if he knows $\phi(n_B)$. To find $\phi(n_B)$ from the publicly known n_B, a cryptanalyst has

to find the factorization of n_B. However, factoring is infeasible if the primes are chosen large enough.

With the RSA cryptosystem, one can also digitally sign electronic files.

In [263], Lenstra discusses the problem of primality tests and factorization algorithms. Probabilistic primality tests are very fast. If they declare a number to be non-prime, it is non-prime, but if a number is not declared non-prime, no such conclusion can be drawn. Rigorous primality tests are much slower. Also factorization algorithms have a probabilistic character, but of a different nature: the final result is unambiguous, but the running time is probabilistic.

Because the RSA cryptosystem, just like the Diffie-Hellman key exchange involves computations with very large numbers, it is often tempting to relax some of the conditions in applications, especially when they involve smart cards with their limited computing facilities. For instance, when the secret exponent d is stored on a smart card, one may want to restrict the size of the secret. Of course, a cryptanalyst should not be able to guess d. Wiener [82] shows that it is not safe to let d be less than $n^{1/4}$. Note that a 200-digit modulus n still makes a 50 digit d possible and that 10^{50} possibilities are impossible to check. He shows that the continued fraction approximations of e/n, where e is the public exponent, will include one in which the secret d appears as a factor of the denominator!

In [303], Verheul and Van Tilborg show that Wiener's method is not worthless when d is a little bit bigger than $n^{1/4}$. Their analysis shows that when the binary representation of d is l bits longer than that of $n^{1/4}$, the work factor for finding the secret d grows with factor 2^l. Boneh and Durfee improve on Wiener's attack by defining a particular lattice and by finding a short basis of this lattice by means of the L^3 algorithm. De Weger improved on this by adding a bound on the difference of the prime divisors of the modulus. Laguillaumie and Vergnaud [338] adapt these results to apply them to RSA-like systems, like LUC, KMOV, Demytko, and the HMT scheme.

3.2.3 The McEliece Cryptosystem

The McEliece cryptosystem [47] is based on the inherent difficulty of decoding arbitrary linear codes (see [45]). McEliece suggests to make use of Goppa codes but to hide their structure by means of random linear transformations. We recall the following facts.

Goppa code
Each irreducible polynomial of degree t over $GF(2^m)$ defines a binary, irreducible Goppa code of length $n = 2^m$, dimension $k \geq n - tm$ and minimum distance $d \geq 2t + 1$. A decoding algorithm with running time nt exists. There are about $2^{mt}/t$ irreducible polynomials of degree t over $GF(2^m)$.

The McEliece Cryptosystem

Preliminary work: *A typical user, say U, chooses a suitable $n_U = 2^{m_U}$ and t_U. User U selects a random, irreducible polynomial $p_U(x)$ of degree t_U over $GF(2^{m_U})$ and chooses a generator matrix G_U of the corresponding Goppa code. The size of G_U is $k_U \times n_U$. Next, user U chooses a random, dense $k_U \times k_U$ non-singular matrix S_U and a random $n_U \times n_U$ permutation matrix P_U and computes $G_U^* = S_U G_U P_U$. User U makes G_U^* and t_U public, but keeps G_U, S_U and P_U secret.*

Encryption: Suppose that user A wants to send a message to user B. He represents his message by a binary vector $\underline{m}$ of length k_B, and sends to B the ciphertext

$$\underline{c} = \underline{m} G_B^* + \underline{e}, \tag{3.8}$$

where $\underline{e}$ is a randomly chosen vector (error pattern) of length n_B and weight $t \leq t_B$.

Decryption: Upon receiving $\underline{c}$, B uses his secret permutation matrix P_B to compute

$$\underline{c} P_B^{-1} = \underline{m} S_B G_B P_B P_B^{-1} + \underline{e} P_B^{-1} = (\underline{m} S_B) G_B + \underline{e'}, \tag{3.9}$$

where $\underline{e'} = \underline{e} P_B^{-1}$ also has weight t. With the decoding algorithm of the Goppa code, B can now retrieve $\underline{m} S_B$. Multiplying this on the right with S_B^{-1} (only known to B) results in the original message $\underline{m}$.

The reason why an error pattern is added in the computation of the ciphertext is of course to make it difficult for the cryptanalyst to retrieve $\underline{m}$ from $\underline{c}$. Indeed, to the cryptanalyst, matrix G_B^* looks like a huge random matrix (note that without $\underline{e}$, it would be simple linear algebra to determine $\underline{m}$ from $\underline{c}$). Parameters suggested by McEliece are $t = 50$ and $m = 10$.

The encryption function maps binary k-tuples to binary n-tuples. This mapping is clearly not a surjection and so it follows that the McEliece system cannot be used directly for digital signatures.

In [282] Preneel, Bosselaers, Govaerts and Vandewalle summarize two types of attacks on the McEliece cryptosystem. The first category tries to recover the original G or an equivalent G. This approach is much more time consuming than the second approach, in which the cryptanalyst tries to find k error-free coordinates on which G^* has full rank and to recover $\underline{m}$ directly with Gaussian elimination. They quote that in view of this attack a choice of $t = 39$ for $m = 10$ is much more appropriate. The authors describe a specific software implementation of the encryption and decryption algorithm (including a decoding algorithm).

Several people, in particular Rao and Nam, have tried to make the McEliece cryptosystem more practical by considering much shorter codes at the price of turning the system into a secret key cryptosystem.

Rao-Nam Secret Key Cryptosystem
Secret Key: A $k \times n$ generator matrix G, a dense $k \times k$ non-singular matrix S, a permutation matrix P of order n, and a set Z of binary vectors of length n and average weight $n/2$ no two of which are in the same coset of the code C spanned by G.

Encryption: A message $\underline{m}$ is encrypted by selecting a random $\underline{z}$ from Z and computing the ciphertext $\underline{c} = (\underline{m}SG + \underline{z})P$.

Decryption: First calculate $\underline{c}' = \underline{c}P^{-1} = \underline{m}SG + \underline{z}$. Compute the syndrome of $\underline{c}'$. This determines $\underline{z}$ uniquely, since $\underline{m}SG$ is a codeword. Determine $\underline{c}'' = \underline{c}' - \underline{z}$, which is $\underline{m}SG$. Now compute $\underline{m} = \underline{c}''(SG)^{-R}$, where $(SG)^{-R}$ is a right inverse of SG.

Struik, Van Tilburg and Boly [271] describe a chosen-plaintext attack on this scheme and extend it to a ciphertext-only attack. The pre-computation of the chosen-plaintext attack involves $kN \log N$ encryptions, where $N = 2^{n-k}$, and nN bits of memory. Breaking the system (i.e. finding the encryption matrix) takes $kn|Aut(\Gamma)|^k$ operations, where $Aut(\Gamma)$ denotes the automorphism group of a graph Γ that is defined by the code C (it has N points).

Also digital signature schemes have been proposed that are based on the difficulty of decoding linear codes. One of them is the *Alabbadi-Wicker Public Key Signature Scheme*. Its description can be found in [89].

Van Tilburg [286] shows that this scheme is not secure if one is able to verify n signatures with linear independent vectors. In general, a few more signatures are needed to get n linear independent error vectors. The same author shows in [296] that all signature schemes (like that by Alabbadi-Wicker) that are based on the Bounded Hard-Decision Decoding problem can only be secure if a signature cannot be verified in polynomial time! In [311], Xu and Doumen go one step further. They demonstrate a universal forgery attack on the Alabbadi-Wicker scheme, meaning that an attacker can put the right signature over any message message $\underline{m}$. To this end, they first recover the parity check matrix H, which can be done if n signatures with independent error vectors can be obtained.

3.2.4 The Knapsack Problem

Two years after the introduction of the notion of public key cryptography, Merkle and Hellman [48] proposed a public key encryption method that is based on the knapsack problem.

Knapsack Problem
Let $a_1, a_2, \ldots, a_n$ be a sequence of n positive integers. Let also S be an integer. Question: does the equation

$$x_1a_1 + x_2a_2 + \cdots + x_na_n = S \tag{3.10}$$

have a solution with $x_i \in \{0, 1\},\ 1 \leq i \leq n$*?*

Although the knapsack problem is known to be NP-complete (see [50]), for some $\{a_i\}_{1\leq i\leq n}$ sequences it is easy to find an explicit solution! For example, given the sequence $a_i = 2^{i-1},\ 1 \leq i \leq n$, there will be a solution if and only if $0 \leq S \leq 2^n - 1$ and finding the solution is very easy. A much more general class of sequences $\{a_i\}_{i=1}^n$ exists, for which this equation is easily solvable. This is the class of *superincreasing* sequences. A sequence $\{a_i\}_{i=1}^n$ is called superincreasing if $\sum_{i=1}^{k-1} a_i < a_k$, for all $1 \leq k \leq n$.

It is easy to determine the solution in this case. Working backwards, one has $x_n = 1$ if and only if $S \geq a_n$, followed by $a_{n-1} = 1$ if and only if $S - x_na_n \geq a_{n-1}$, etc., and ending with "a solution exists" if and only if $S - \sum_{i=1}^n x_ia_i = 0$. Based on the apparent difficulty of solving the knapsack problem and the ease to solve this problem for superincreasing sequences, the following cryptosystem has been proposed [48].

Knapsack Cryptosystem
Preliminary work: *Each user U selects a superincreasing sequence $\{u_i\}_{i=1}^{n_U}$ of length n_U, and selects a modulus m_U and constant w_U, such that $m_U > \sum_{i=1}^{n_U} u_i$ and $\gcd(w_U, m_U) = 1$. Finally, user U computes the numbers $u_i' \equiv w_U \cdot u_i \pmod{m_U},\ 1 \leq i \leq n_U$. User U makes the sequence $\{u_i'\}_{i=1}^{n_U}$ known as his public key, but keeps m_U, w_U and the original superincreasing sequence $\{u_i\}_{i=1}^{n_U}$ secret.*

Encryption: *If A wants to send a message to B, he looks up the public encryption key $\{b_i'\}_{i=1}^{n_B}$ of B. User A represents his message by a binary sequence $\{m_i\}_{i=1}^{n_B}$ of length n_B and sends to B the ciphertext $C = \sum_{i-1}^{n_B} m_i \cdot b_i'$.*

Decryption: *User B computes $w_B^{-1} \cdot C \equiv w_B^{-1} \cdot \sum_{i=1}^{n_B} m_i \cdot b_i' \equiv \sum_{i=1}^{n_B} m_i \cdot b_i \pmod{m_B}$. Since $\sum_{i=1}^{n_B} m_i \cdot b_i < m_B$, this can be rewritten as $\sum_{i=1}^{n_B} m_i \cdot b_i = (w_B^{-1} \cdot C \pmod{m_B})$. The solution $\{m_i\}_{i=1}^{n_B}$ is now easily found since the sequence $\{b_i\}_{i=1}^{n_B}$ is superincreasing.*

Although the knapsack cryptosystem can not be used to digitally sign documents, it was enormously popular for a while, basically for the simplicity to implement it. It is a good idea for each user U to publish a permuted version of his public knapsack. A further recommendation of [48] is to iterate the modular multiplication of the knapsack.

Example *Consider the knapsack $(u_1, u_2, u_3) = (5, 10, 20)$. Multiply this with the multiplier $w = 17$ modulo $m = 47$ to get $(u_1', u_2', u_3') = (38, 29, 11)$ and multiply this in turn with $w' = 3$ modulo $m' = 89$ to get $(u_1'', u_2'', u_3'') = (25, 87, 33)$.* It is an easy exercise to show that it is impossible to find integers w'' and m'' that map

(u_1, u_2, u_3) directly into (u_1'', u_2'', u_3'') by means of $u_i'' \equiv w'' u_i \pmod{m''}$.

Desmedt, Vandewalle and Govaerts in [262] warn against exaggerating the security of the knapsack cryptosystem:

i. The cryptanalyst does not need the original superincreasing sequence to break the system. (The above example shows this. The final sequence is itself superincreasing!)

ii. In fact, infinitely many deciphering keys exist.

iii. Not all x_i's of the original message have to be found in general, because of the redundancy in the plaintext.

A year later, the same three authors [265] attempted a more positive approach, most likely tempted by the ease of implementation of the knapsack cryptosystem and the resulting achievable transmission speed. They describe how transformations by means of linear equations can be used to provide a trapdoor for the knapsack problem. Their method generalizes all known ways (at that time) to construct public enciphering keys and shows new ways to make them. The effect of iterations is better understood. They repeat that to break a cryptosystem one does not need to deal with all the original transformations.

In 1982, Shamir [58] did break the single multiplication version of the system (demonstrating (i) and (ii)). A year later, Lagarias and Odlyzko [64] showed that the knapsack cryptosystem is not safe in general.

3.2.5 Implementation Issues

Given the fact that all public key cryptosystems work with very large numbers or very big matrices and tables, it does not come as a surprise that great attention needs to be paid to their implementation, especially if part of the calculations take place on a smart card that typically has limited computing power and storage facilities.

Béguin and Quisquater address in [291] the situation that a smart card wants to make use of a powerful auxiliary unit (server) to do its calculations. The server may be under the influence of an opponent, so calculations by the server must be verified and the card must protect its secrets. A practical protocol is described that computes a RSA signature in this way. The protocol is secure against active attacks, i.e., the server may send false information to the card to get some secret information. The authors point out that one part of the protocol seems to be vulnerable to passive attacks.

Bosselaers, Govaerts and Vandewalle [288] describe an extensive software library, written in ANSI C, and discuss the design criteria in particular. The functionality of the library is grouped into the following categories:

i. conversion between types and I/O;

ii. low-level arithmetic (like bit operations, addition, multiplication, etc. , but also gcd and modular inverse);

iii. high-level arithmetic (like modular exponentiation or prime-number generation).

The authors also pay attention to number representation, error handling and memory management.

Multiplications in $GF(2^n)$ play an important role in public key cryptosystems, especially in elliptic curve cryptography. An efficient multiplication is essential for their performance. For scalable hardware implementations, one cannot rely on special properties of the irreducible polynomial that defines the field. For this reason, a *normal* basis is not suitable. Batina, Jansen, Muurling and Xu in [325] describe a scalable multiplier architecture that combines the classical bit-serial method with Montgomery's modular multiplication algorithm. In the same volume, Potgieter, Van Dyk and Tjalkens [326], with the same application in mind, come to the same conclusion with regard to the choice of the polynomial and propose a similar, flexible multiplier that is twice as fast as previous methods at the expense of 50% more chip area.

For better performance of calculations over a finite field, it is often advantageous to use a trinomial as defining polynomial for the finite field. In [332], Ciet, Quisquater and Francesco prove that for $p \equiv 13$ or $19 \pmod{24}$, irreducible trinomials of prime degree p do not exist.

It is well known [99] that the variations in power consumption during the calculation of an exponentiation on a smart card may leak information about the secret exponent. Normally, it is assumed that a multiplication consumes more time and energy than a squaring. In [331], Batina and Jansen assume a scenario in which information only leaks on the total number of these operations. They conclude that for practical bit lengths, the information obtained in this way (in an information theoretic sense) is far from exploitable. For instance, when $n = 1024$, the leakage amounts to 6.06 bits. In [337], the same authors make their analysis more precise. In their first paper, the assumption was that the secret exponent was a random odd number. Here, the assumption is (as it should be) that the secret exponent is coprime with the Euler ϕ function of the modulus. The results differ only marginally from [331], also for the case where the prime numbers involved are *strong primes* (see [102]): the leakage is at most 3.6 bits for $n = 1024$.

3.3 Security Issues

3.3.1 Internet Security Standards

Vandenwauver, Govaerts and Vandewalle [302] give an overview of the existing Internet security standards. The following services need to be present:

i. *Data authentication*: both the integrity of the data as well as their origin need to be authenticated;

ii. *Non-repudiation*: a sender of a message should not be able to deny having sent it; a receiver cannot deny having received the message (nor change its contents);

iii. *Data confidentiality*: unauthorized disclosure of the message should not be possible.

The basic approach consists of the following ingredients. The data are encrypted with a symmetric cryptosystem (for reasons of performance) with a key that is exchanged with a public key cryptosystem. A digital signature of the sender is added to the message. Most of the standards do not incorporate all services, in particular non-repudiation of delivery is often missing.

The public keys of the different parties involved are distributed or guaranteed by a Certification Authority by means of a certificate. Guidelines for these certificates are given by X.509. An important standard is Secure Socket Layer (SSL). New Internet standards that are briefly discussed in [302] are S/MIME, PGP/MIME, MOSS and MSP.

As noted above, the issue of non-repudiation of receipt is often not addressed. Kremer and Markovitch in [319] describe two protocols proposed by J. Zhou. The first one involves a *Trusted Third Party* that acts as notary. Since this solution may create a communication bottleneck, Zhou's second protocol avoids such a TTP, but assumes that sender and receiver are honest. The authors demonstrate some weaknesses of this model and present a solution that involves an active, offline TTP and a resilient channel (i.e., data may be delayed but always arrive eventually). This new protocol guarantees fairness and timeliness.

3.3.2 Security Policies and Key Management

The security of a system (or a network of systems) that performs computations or operations is obviously of utmost importance. Any unauthorized action, such as altering the system files, may cause loss of valuable data or even complete system failures. For this reason, a proper security model is an important tool in the design of a system. One of the earliest models for this purpose [30] is the Bell-LaPadula model, which describes four levels of security clearance (unclassified, confidential, secret, and top-secret) and access rights that amount to: someone with a lower security level cannot read the information that belongs to a higher security level; such a person should, however, be able to write to the higher level. There are some problems with this model, for instance, such a linear system does not always reflect reality. Also, the system should be flexible (it should be possible to change permission rights).

Verschuren, Govaerts and Vandewalle in [283] concentrate on the model above in a distributed environment. They consider the situation where Application Processes (APs) are running on different end systems which are connected by a public

communication channel. It is assumed that communicating end systems make use of the Reference Model for Open Systems Interconnections (OSI-RM).

To minimize the number of keys involved and taking the OSI-RM protocol into account, the authors arrive at the following optimal scheme. Without loss of generality, we assume that the APs are numbered according to their clearance.

Key Distribution Proposal

1. AP_1 (with lowest clearance) chooses a key that can handle all the data that it is allowed to handle.
2. AP_i is equipped with all the keys of the APs with lower clearance and one key that can handle the data classes that are unique to its clearance.

Note that AP_i has i keys.

Radu, Vandenwauver, Govaerts and Vandewalle [292] consider the access of a personal database by different organizations. The database is located on the non-volatile memory of a multi application smartcard. The paper outlines a subject view mechanism that guarantees that only eligible organizations can execute the actions they are entitled to. The authors propose to substantiate the information necessary for authentication and authorize the access as tickets to be release and signed by a trusted authority. The tickets are supposedly stored in the computer system of the eligible organizations. During an access transaction no on-line communication takes place with the trusted authority.

Verschuren [297] lies the foundation for an evaluation method of the security aspects of a computer network. He represents the communication subsystems of the various users (APs), by means of finite-state machines (FSM). Each FSM in turn can be described by a table. The table consists of rules "input, old state $\rightarrow$ output, new state". For APs with different clearances, different parts of the table apply. The evaluation method checks whether requests and indications at an AP are in accordance with its security policy.

Seys and Preneel [345] discuss the setting of an ad-hoc network that has no fixed infrastructure. A new connection is created as soon as a mobile device (node) enters the vicinity of one or more other nodes. These nodes may have to rely on other nodes to forward their messages. The wireless nodes are allowed to move around and will typically have limited power and limited communication means. The authors wanted to realize two objectives:

i. distributed trust to ensure robustness, and

ii. strong authentication.

In such a network, some nodes may be there to control the network and to help realize the objectives. A distributed and hierarchical public key infrastructure is proposed that depends on a protocol that securely establishes and manages cryptographic keys.

3.3.3 Side Channel Attacks and Biometrics

In the 1990s, the cryptographic community has broadened its view from studying the security of mathematical models only to evaluating the security of physical implementations. Even if a cryptographic algorithm is mathematically secure, its implementations may be vulnerable to attacks exploiting physical side channels (timing information [99], power consumption, electromagnetic emanation, ...) and attacks inducing deliberate faults in the computations.

Timing attacks are studied by Hachez, Koeune and Quisquater [307]; they present improved attacks on Montgomery modular exponentiations with a secret exponent. Borst, Preneel and Vandewalle [316] compare countermeasures at the hardware, software, algorithm and protocol level. Ciet, Piret and Quisquater [342] propose a new block cipher with a built-in error-correcting code to increase resistance against fault attacks.

Verbitskiy, Tuyls, Denteneer and Linnartz [341] study the problem of verifying biometric templates that uniquely determine human beings. Problems that have to be addressed are:

i. Robustness to noise (since measurements will differ slightly each time);

ii. Security;

iii. Privacy protection (centrally stored data on the biometrics of people should be protected).

It is pointed out that a universal authentication scheme satisfying these three requirements does not exist. The authors propose a scheme that makes use of side information and evaluate its performance. They do not make use of error-correcting codes to tackle the problem of noise in the data measurements; their

3.3.4 Signature and Identification Schemes

There are several methods to digitally sign documents. They are based on the RSA system or on the difficulty to take discrete logarithms. An example of the first one is the Guillou-Quisquater (GQ) signature scheme [73].

Guillou-Quisquater Signature Scheme
Preliminary work by the Signer: The signer selects two large prime numbers, say p and q, computes $n = p \times q$, selects an exponent e that is prime and computes the corresponding exponent d from $e \times d \equiv 1 \pmod{(p-1)(q-1)}$ (see also Section 3.2.2). The signer selects a number $I,\ 1 < I < n$, which serves as his identifier (it may contain his name, date of birth, etc.) and also computes the solution D to $I \times D^e \equiv 1 \pmod{n}$ (called *authentication number*). Let $h : \{0,1\} \rightarrow Z_n$ be a hash function.

Signature generation: To sign a message M, the signer selects a random $r,\ 1 < r < n$, and computes $R \equiv r^e \pmod{n}$. He computes the hash value $T =$

$h(M, R)$, called *question*, and then he determines the so-called *witness* $S \equiv rD^T \pmod{n}$. The signature on M is given by the pair (S, T).

Signature verification: The verifier should obtain an authentic public key (n, e, I) of the presumed signer. He computes $U \equiv S^e I^T \pmod{n}$ and $T' = h(M, U)$. He accepts the signature if and only if $T = T'$. The reason why this works is: $U \equiv S^e I^T \equiv r^e D^{Te} I^T \equiv r^e (D^e I)^T \equiv r^e \equiv R \pmod{n}$.

Delos and Quisquater in [285] address the problem of signature schemes in which several signers interact. One can think of a situation where the power to sign has to be shared (maybe even all have to sign more or less at the same time). In their proposal, also an intermediate entity plays a role. Imagine two smart cards, each securely storing the authentication number D_i corresponding to its identity I_i, related by $D_i^{e_i} I_i \equiv 1 \pmod{n},\ i = 1, 2$. The intermediate can simulate an identity $I \equiv I_1 I_2 \pmod{n}$, with $e = 2e_1 e_2$, and authentication number D following from $I \times D^e \equiv 1 \pmod{n}$. The signature of the intermediate on behalf of the two signers consists of the signing identities, the global witness, the global question (computed from the initials questions), and the global challenge.

In an identification scheme, a person called Prover can convince another person called Verifier of its identity, without having to reveal a secret. In a *group identification scheme* (GIS), the Prover can convince the Verifier that he belongs to a certain group of people. A GIS should have the following properties: correctness, soundness, anonymity, unlinkability and traceability. Gaddach [324] proposes a GIS that is based on the composite discrete logarithm problem: given two elements a and b in Z_p^* and a generator g of Z_p^*, are there x and y such that $a^x b^y \equiv g \pmod{p}$? The proposed GIS has the advantages that only one initialization phase is needed in order to create several groups and that a coalition of dishonest members can be traced.

So-called *designated verifier schemes* only provide authentication of a message to an intended receiver, so nobody else can be convinced of its validity. Such schemes do not provide non-repudiation (cf. Section 3.3.1). As a matter of fact, the intended receiver could have made the signature himself in an indistinguishable way. These schemes may be needed in situations, where the receiver should not be able to show the document to others with a signature of the sender that can be verified by others. A third person could still try to intercept the sent message before it is received and then identify the sender. In [344], Saeednia, Kremer and Markowitch give a solution to this problem. Such a scheme is said to have the strong designated verifier property. The proposed method is based on Schnorr's signature scheme and is very efficient.

Delos and Quisquater in [290] announce a signature scheme in which the ability of a signer to sign messages is limited to a fixed number of signatures.

3.3.5 Electronic Payment Systems

To make electronic payment systems more acceptable, some degree of integrity has to be offered. Basically, this means that it should not be possible to forge or copy money. Radu, Vandenwauver, Govaerts and Vandewalle [295] point out disadvantages of a coin-based solution (too elaborate) and suggest a counter-based solution: a tamper-resistant device (smart card) that contains a counter representing money. However, customers, of course, want a certain degree of anonymity (intractability). In this proposal, the above is realized by two cryptographic primitives. One is a blind signature scheme, the other is a double-spending detection mechanism. The authors present the design of an efficient off-line traceable counter-based cryptosystem based on the intractability of taking RSA roots (see Section 3.2.2), in particular also on the Guillou-Quisquater identification scheme.

Clearly, anonymity offered to the customers can easily be misused by criminals, e.g. for money laundering or illegal purchases. This means that mechanisms to revoke the offered anonymity have to be present. Claessens, Preneel and Vandewalle in [313] discuss this aspect for a number of current electronic payment systems. The SET protocol does not provide privacy nor does Proton. ECash, which basically works online, uses blind signatures and does offer privacy. The CAFE payment system uses restrictive blind signatures; the identity of the user can be determined if the same money is spent twice. There are two common types of tracing mechanisms:

i. those that trace the owner of a coin, and

ii. (ii) those that trace the coin itself.

The authors observe that anonymous communication between the various parties in an electronic payment system is necessary to have real anonymous cash. Mix networks and Anonymizers may solve this problem. Several proposals in this direction are discussed.

3.3.6 Time Stamping

Time is an important ingredient for documents having a long lifetime. For instance, when a key pair in a public key cryptosystem is compromised and revoked, and one wants to check whether that document has been signed within the period when the secret key was valid. As another example, think of the date on a patent. *Time stamping* is a solution to these problems. It should meet the following two requirements.

i. It must be infeasible to timestamp a document with an incorrect date or time.

ii. It must be infeasible to change even a single bit of a timestamped document without the change being apparent.

The basic solution for timestamping relies on a trusted third party, the Time Stamping Authority (TSA). The TSA appends the current time and date to the document

and digitally signs the result to produce the timestamp. Compressing the document first by means of a cryptographically secure hash function (meaning that it is collision-resistant and one-way) can improve the efficiency greatly.

Of course, each TSA needs to have a time that differs minimally from a chosen standard, which is for instance the *Network Time Protocol*. Van Rompay, Preneel and Vandewalle in [308] address the problem of minimizing the trust that one needs to have in the TSA. A basic solution is that all timestamps issued by a TSA are linked: each new timestamp includes information from the previous timestamp. For this, another collision resistant, one-way hash function is needed. This approach results in relative temporal information. Timestamping additional documents (e.g. random numbers) may further narrow the time window down. Another possibility is a periodic publication in an authentic medium like a newspaper.

TSAs can, of course, be incorporated in public key infrastructures. A TSA which also authenticates the client and verifies the contents of the submitted documents is called a *Notary Authority*.

Linking all timestamps in a linear way poses a high demand on cooperation and may also impose a long computation time before a trusted timestamp is encountered on the chain. A solution to this would be to divide the timestamping procedures in rounds. At the end of each round, a timestamp is calculated that depends on all requests during that round and on the timestamp of the previous round. If the mutual order of the timestamps does not matter, one can compute the time stamp of a particular round from the hash values y_i of the documents presented during that round by means of a binary authentication tree or a function for y_i like $g^{\prod_{i \neq j} y_j} \pmod N$.

Massias, Serret Avila and Quisquater in [309] present a design and implementation of a timestamping system for the Belgian project TIMESEC. They also prefer to minimize the trust in the TSA. As an example of their method, let there be 8 documents to be signed in a particular round and let $y_i,\ 1 \leq i \leq 8$, be their hash values. The concatenation of y_1 and y_2 is hashed to produce $H_{1,2}$, similarly, $H_{3,4}$ $H_{5,6}$, and $H_{7,8}$ are computed. Then the concatenation of $H_{1,2}$ and $H_{3,4}$ is hashed to produce $H_{1,4}$, etc. Finally, the top value (here $H_{1,8}$) is concatenated with the hash value of the previous round, say RH_{i-1}, and then hashed to produce the new round value RH_i. Periodically, some of these round values are published in a newspaper or in another widespread medium. To check the timestamp of y_1 one needs $y_2, H_{3,4}, H_{5,8}$, and RH_{i-1}.

3.4 Data Hiding

In the last decade there has been a considerable increase in the interest in the Information Theory community devoted to data hiding. As a matter of fact, the rapid growth of broadband Internet has brought many concerns related to the protection of multimedia contents. In the digital world, security and privacy are implemented

through the use of cryptographic algorithms and protocols. In the case of multimedia intangibles, the digital contents have to be provided at the end point in an analogue form: the digital image is transformed into light through a screen; the digital sound is transformed into acoustic waves. Capturing and re-digitizing these analogue signals for illegal redistribution is always possible. This is a first and main goal for data hiding: providing secret, robust and invisible marks for copyright protection and usage tracing. Other applications may be related to copy control (as has been proposed for the DVD-RW) or to the authentication of multimedia data. Data hiding in the particular context of protection of multimedia contents is generally called watermarking.

Data hiding is not only a concern for Information Theory but also for signal processing, game theory and risk analysis. The goals for Information Theory are to ensure secrecy of the communication (cryptographic coding) and to maximize the capacity of the hidden channel (channel coding). Signal processing is useful for the design of imperceptible channels in different media, while game theory is able to model the global compromise between the actors of the chain, namely, the content owner, the opponent and the receiver.

The WIC community has been very active, and comprises some of the main pioneering contributors in the field. The works related to watermarking and data hiding has followed two veins: some of the researchers have tried to design effective systems dedicated to particular applications, while others have developed theoretical frameworks for determining bounds and expected performances, which is only possible for simple enough situations. This second vein has mainly enhanced and developed further the initial framework set by Shannon [12] and Costa [63] describing transmission over channels with side information. In information hiding, the transmission channel is the media content itself. If it is considered as noise, no advantage is taken of the fact that the content is completely known to the watermark embedder (and detector, if the original unwatermarked content is available as part of the detection process). Most of the authors view watermarking as an example of communication with side information described by Shannon.

A practical approach to the problem of transmitting a message through an AWGN channel with side information where only the current and past channel states are considered is presented by Willems [274]. His encoding scheme utilizes a regular lattice, but does not follow Costa's approach to adapt to the known interference in an optimum way and thus suffers from capacity loss. Later he generalizes his work in [320], where he proposed a framework for computing the capacity of such channels.

Boucqueau, Bruyndonckx, Lacroix, Mertès, Macq and Quisquater [293] describe in 1995 the use of watermarking for the protection of broadcasted digital TV signals in contribution (inter-studio) and distribution (to the consumer) links. The watermarking is used at two levels: one for copyright claims and the other one for traitor tracing. Langelaar, Van der Lubbe and Biemond [299] describe a very efficient way to individualize MPEG streams by data hiding for video-on-demand

applications. Each copy accessed from a video server is marked imperceptibly by information allowing retrieval of the transaction.

In [304], Kalker shows that all correlation based watermark methods are not secure if the detector is publicly available such as is the case for DVDs. In [314], Kalker, Oostveen and Linnartz study the optimal detection of optimal watermarks. Multiplicative watermarks are of great interest due to Weber's law applied to image distortions: modification in the luminance profile are less visible in the white areas of the images than in the dark ones. The optimal detector of such watermarks is no longer a linear correlator, but the signal should be squared before applying the correlation detector. Under a limited set of assumptions, the authors demonstrate the optimality of such a detection structure.

In 2000, Van Dijk and Willems [321] propose codes for embedding data in grayscale images. These are in fact codes for channels with side information. Ingredients of these codes are Hamming and Golay codes. The codes proposed are optimal, i.e. alternative codes with the same block length must either have a smaller or an equal embedding rate, or a larger or an equal distortion. As described above, watermarking is closely related to the Costa side-information problem [63]. Costa showed that the capacity of a Gaussian channel depends only on the transmitter power and the variance of the noise that is not known to the transmitter. At the time, the watermarking community was trying to design coding techniques that approach the Costa limit as closely as possible. New codes that operate both as quantizer and as a channel code are described in 2002 by Van den Borne, Kalker and Willems in [322].

Some particular applications, like medical-image distribution, require a reversible process for the data hiding. The message is there for copyright protection or authentication but has to be removed when it is used in a secure reader for fine diagnosis purposes. WIC authors have addressed this challenge as pioneers. In [340], Maas, Kalker and Willems propose bounds for such a particular situation, and also address the case of watermarked images with a small distortion.

The research of Moulin describing a complete game theoretic model was presented in a tutorial paper in [336]. This paper reviews recent research on information-theoretic aspects of information hiding. Emphasis is put on applications requiring a high payload (e.g., covert communications). Information hiding may be viewed as a game between two teams (embedder/decoder vs. attacker), and optimal information-embedding and attack strategies may be developed in this context. This paper focuses on several of such strategies, including a framework for developing near-optimal codes and universal decoders. The suboptimality of spread-spectrum strategies follows from the analysis. The theory is applied to image watermarking examples.

Finally, alternative methods to watermarking of images can rely on the visual hashing of images. This is an extension of the audio fingerprints of Kalker. In [327], Lefebvre, Macq and Legat develop a visual hash strategy based on the Radon

transform, which exhibits good properties for resistance against affine transforms (zooming and rotation). The hash can either be used for image retrieval or for the resynchronization of a watermarking algorithm.

3.5 Conclusions

The cryptographic community in the Benelux can be considered very active. Their activities cover more or less the whole scope of modern cryptography and related security issues. It is an amazing coincidence that the WIC was founded more or less at the time when several university groups in the region became interested in cryptographic research.

CHAPTER 4

Channel Coding

J.H. Weber (TU Delft)
L.M.G.M. Tolhuizen (Philips Research Eindhoven)
K.A. Schouhamer Immink (University of Essen/Turing Machines)

Introduction

Channel coding plays a fundamental role in digital communication and in digital storage systems. The position of channel coding in such a system is depicted in Figure 4.1 overleaf. The *channel encoder* adds redundancy to the (possibly source encoded and encrypted) messages generated by the information source, in order to make them more resistant to noise and other disturbances affecting the modulated signals during transmission over the channel. The *channel decoder* exploits the redundancy when trying to retrieve the original information based on the demodulator output. The choice of a channel coding scheme for a particular application is a trade-off between various factors, such as the *rate* (the ratio between the number of information symbols and the number of code symbols), the *reliability* (the bit or message error probability), and the *complexity* (the number of calculations required to perform the encoding and decoding operations).

[1]This chapter covers references [346] – [450].

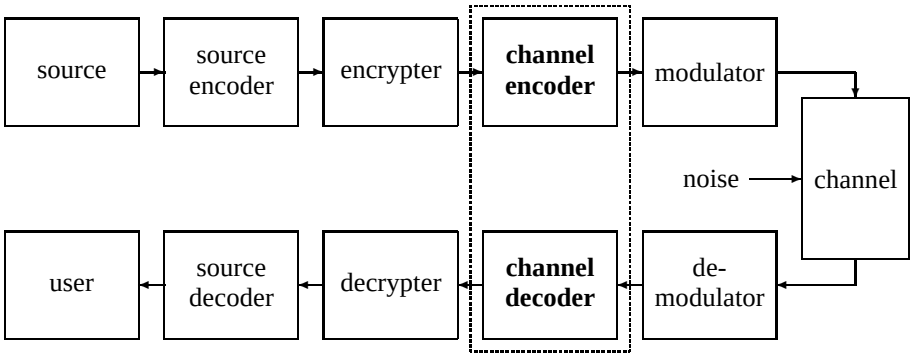


Figure 4.1: *Channel coding as a component in a communication or storage system.*

In his landmark paper [3], Shannon showed that virtually error-free communication is possible at any rate below the channel capacity. However, his result did not include explicit constructions and allowed for infinite bandwidth and complexity. Hence, ever since 1948, scientists and engineers have been working to further develop coding theory and to find practically implementable coding schemes. The paper of Costello, Hagenauer, Imai and Wicker gives a good overview of applications of error-control coding. Some of the codes emerging from coding theory as developed in the 1950s and 1960s have been applied in mass consumer products like the CD (developed jointly by Philips and Sony in the 1970s and 1980s) and GSM (1990s). Classical reference works are the book of MacWilliams and Sloane [42] and that of Blahut [62]. A more recent reference is the 2-volume work [108]. The above books focuses mainly on block codes; the book of Johannesson and Zigangirov [110] deals exclusively (and expertedly!) with convolutional codes. In [109], a comprehensive overview is given of (modulation) codes particularly designed for data storage systems, such as optical and magnetical recording products.

The introduction of turbo codes in 1993 [90] caused a true revolution in error-control coding. These codes allow transmission rates that closely approach channel capacity. Also, the re-discovery of Gallager's low-density parity-check (LDPC) codes [15] contributed to the large present interest in iterative decoding, both theoretically and practically (iterative decoders are being applied in UMTS).

Sessions on channel coding have been part of the Symposia on Information Theory held in the Benelux since 1980. On average, about four channel coding papers were presented per symposium. Among the highlights of the many Benelux contributions to this field are the celebrated Roos bound on the minimum distance of cyclic codes [352], Best's work on the performance evaluation of convolutional codes on the binary symmetric channel [368], and the comprehensive survey papers by Delsarte on the association schemes in the context of coding theory [400], [444].

In this chapter, we briefly describe the over one hundred papers on channel coding presented at the Symposia on Information Theory in the Benelux. Some structure has been pursued by classifying each paper into one of the following categories: *constructions and properties of (block) codes* (Section 4.1), *decoding techniques* (Section 4.2), *codes for data storage systems* (Section 4.3), *codes for special channels* (Section 4.4), and, finally, *applications* (Section 4.5). Some categories have been divided further into subcategories. The classification is not always unambiguous, since many papers deal with more than one aspect (e.g., a paper presenting a code construction together with an accompanying decoding method). The final choice represents the main contribution of the paper in the opinion of the authors of this chapter.

4.1 Block Codes

4.1.1 Constructions

In this section, we discuss papers that deal with the construction of block codes. Some papers in this section might just as well have been discussed in the next section, as they aim at constructing codes with special properties, e.g. a large minimum distance.

The well-known Griesmer bound states that the length n of a binary $[n, k, d]$ code satisfies the following inequality:

$$n \geq g(k, d) := \sum_{i=0}^{k-1} \left\lceil \frac{d}{2^i} \right\rceil . \tag{4.1}$$

In [349], Van Tilborg and Helleseth explicitly construct, for each $k \geq 4$, a binary $[2^k + 2^{k-2} - 15, k, 2^{k-1} + 2^{k-3} - 8]$ code that is readily be seen to meet the Griesmer bound with equality. It is claimed that for $k \geq 8$, up to equivalence, the constructed codes are the only ones with these parameters. In [380], Kapralov and Tonchev construct self-dual binary codes from the known 2-(21,10,3) designs without ovals, and study the automorphism groups of these codes.

In [401], Ericson and Zinoviev give three methods for constructing spherical codes (i.e., sets of unit norm vectors in $\mathbb{R}^n$) from binary constant weight codes. Bounds are given on the dimensionality, the minimum squared Euclidean distance, and the cardinality of the resulting spherical codes, and numerical examples are given.

In [403], Peirani studies a class of codes obtained by application of the well-known $(u, u + v)$ construction to a simplex code U and a code V from a class of codes with normal asymptotic weight distribution. It is shown that the resulting codes have an asymptotically normal weight distribution as well, by using properties of the dual of the $(u, u + v)$ code, the MacWilliams identity, and the central limit theorem.

According to the Singleton bound, the cardinality of a code C of length n and minimum distance d over a q-ary alphabet Q is at most q^{n-d+1}. In case of equality, C is called an MDS code. Examples of MDS codes are Reed–Solomon codes, which are defined if Q is endowed with the structure of a finite field (and hence q is a power of a prime). In [404], Vanroose studies MDS codes over the alphabet $\mathbb{Z}_m$. His main results are the following. Let $N_m^L(k)$ denote the largest length of a linear MDS code over $\mathbb{Z}_m$ with m^k words. Then $N_m^L(2) = p+1$, and $N_m^L(k) \leq p+k-1$, where p is the largest prime factor of m. Note that the demand that the code is linear over $\mathbb{Z}_m$ is quite restrictive: if m is the power of a prime, doubly extended Reed-Solomon codes are $[m+1, k, m+2-k]$ codes for each $k \in \{1, 2, \ldots, m+1\}$.

In [422], Van Dijk and Keuning describe a construction of binary quasi-cyclic codes from quaternary BCH codes. The length and dimension of the binary code is determined by the generator polynomial of its originating quaternary code; its minimum distance is at least the minimum distance of the quaternary code. For some example codes obtained with this construction, the true minimum distance (found by computer search) equals the best known minimum distance for binary linear codes of the given length and dimension.

An (n, w, λ) *optical orthogonal code* is a set of binary sequences of length n and weight w such that for each $x \in C$ and integer $\tau \in \{1, 2, \ldots, n-1\}$,

$$\sum_{t=0}^{n-1} x_t x_{t+\tau} \leq \lambda, \tag{4.2}$$

and for any two distinct $x, y \in C$ and each integer $\tau \in \{0, 1, \ldots, n-1\}$,

$$\sum_{t=0}^{n-1} x_t y_{t+\tau} \leq \lambda. \tag{4.3}$$

The subscripts are to be taken modulo n. Optical orthogonal codes can be used to allow multi-user optical communication.

In [426], Stam and Vinck give a good overview of the known results in this area. They also introduce a property they call "super cross-correlation": for all distinct x, y and z in C, and integer $\tau \in \{0, 1, \ldots, n-1\}$, it is demanded that

$$\sum_{t=0}^{n-\tau-1} x_t y_{t+\tau} + \sum_{t=n-\tau}^{n-1} x_t z_{t+\tau} \leq \lambda. \tag{4.4}$$

Codes satisfying this extra property could be used in applications with partial synchronization between different codewords and where the mutually synchronized words typically are not sent simultaneously. In [436], Martirosyan and Vinck describe a construction of optical orthogonal codes with $\lambda = 1$. If a certain parameter in their construction is small enough, their code contains, in a first-order approximation, as many words as possible. Specific examples of good codes resulting from the construction are tabulated.

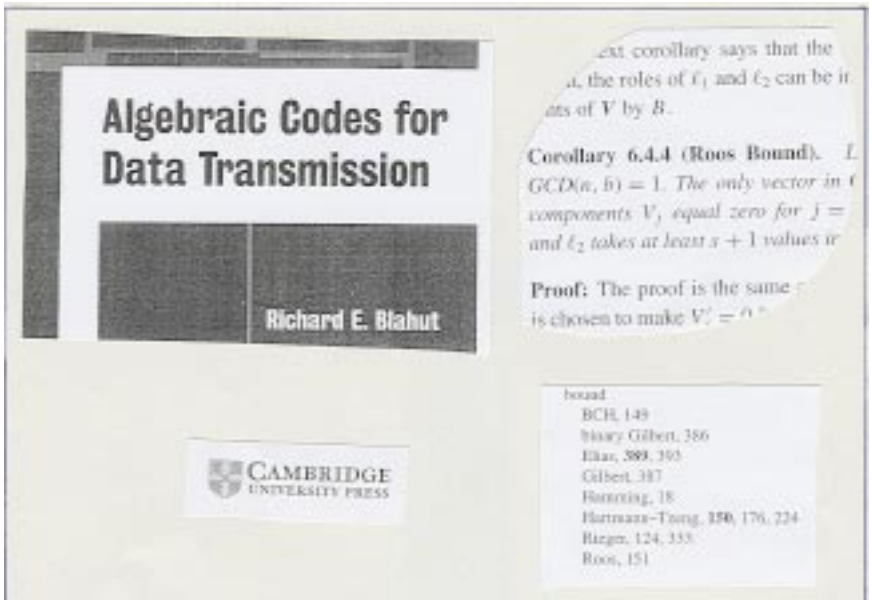


Figure 4.2: *Citation of the Roos bound in a textbook from 2003.*

4.1.2 Properties

Over the years, properties like the length, cardinality, minimum distance, or weight distribution of codes belonging to a particular family have been studied extensively. In this section we review miscellaneous results in this area as presented at the various Benelux Information Theory symposia.

In [352], Roos states and proves what in present textbooks (see Figure 4.2) is referred to as the "Roos bound" for the minimum distance of cyclic codes. The bound reads as follows. Let α be an n-th primitive root of unity in GF(q). Let b, c_1, c_2, δ and s be integers such that $\delta \geq 2$, $(n, c_1) = 1$, and $(n, c_2) < \delta$, and let

$$N := \{\alpha^{b+i_1 c_1 + i_2 c_2} \mid 0 \leq i_1 \leq \delta - 2, 0 \leq i_2 \leq s\}. \tag{4.5}$$

Let C be a cyclic code over GF(q) such that each element of N is a zero of C. That is, for each word $\mathbf{c} = (c_0, c_1, \ldots, c_{n-1}) \in C$ and each $\beta \in N$, we have that $\sum_{i=0}^{n-1} c_i \beta^i = 0$. Then the minimum distance of C is at least $\delta + s$. The Roos bound is often applied to prove a lower bound on the minimum distance of a subfield subcode of C. For example, let α be a 51^{st} root of unity in GF(2^8). Let B be the binary cyclic code with zeroes α, α^5 and α^9. The conjugacy constraints imply that all elements of $N = \{\alpha^i \mid i \in \{7, 8, 9, 10, 13, 14, 15, 16\}\}$ are zeroes of B. It follows from the Roos bound, with $b = 7, c_1 = 1, c_2 = 6, \delta = 5$, and $s = 1$, that the code $C := \{(c_0, c_1, \ldots c_{50}) \in (GF(2^8))^{51} \mid \sum_{i=0}^{50} c_i \beta^i = 0 \text{ for all } \beta \in N\}$ has minimum distance at least six. As B is a subcode of C, its minimum distance

is surely at least six.

In [354], De Vroedt considers formally self-dual codes. For such codes, with the property that all weights are multiples of some constant $t > 1$, he derives the weight enumerator through computation of the eigenvalues and eigenvectors of the so-called Krawtchouk matrix, rather than by using the traditional method based on invariant theory.

In [357], Bussbach, Gerretzen and Van Tilborg study properties of $[g(k,d),k,d]$ codes, i.e., codes that meet the Griesmer bound from Equation (4.1) with equality. It is shown that the maximum number of times a coordinate in C is repeated equals $s := \lceil \frac{d}{2^{k-1}} \rceil$. Moreover, it is shown that the covering radius ρ of such codes is at most $d - \lceil \frac{s}{2} \rceil$, with equality if and only if a $[g(k+1,d),k+1,d]$ code exists. For $s \leq 2$, all $[g(k,d),k,d]$ codes with $\rho = d - \lceil \frac{s}{2} \rceil$ are described; for fixed k and sufficiently large d, there exist $[g(k,d),k,d]$ codes with $\rho = d - \lceil \frac{s}{2} \rceil$.

In [400], Delsarte gives a comprehensive survey of some of the main applications and generalizations of the MacWilliams transform relevant to coding theory. The author, one of the world's most respected contributors to this area, considers in this paper both the generalized MacWilliams identities for inner distributions of dual codes and the generalized MacWilliams inequalities for the inner distributions of unrestricted codes. The latter leads to the linear programming bound in general coding theory. The paper also contains an introduction to association scheme theory, which is an appropriate framework for non-constructive coding theory. In [444], again a survey paper by Delsarte, the Hamming space, particularly important to coding theory, is viewed as an association scheme. The paper provides an extensive overview of those parts of association scheme theory that are especially relevant to coding problems. Special emphasis is put on several forms of duality inherent in the theory. The Hamming space is also considered by Canogar in [424]. The author studies an example of a non-trivial partition design of the 10-dimensional Hamming space. He shows that this partition can be reconstructed from its adjacency matrix.

Gillot derives in [402] bounds on the codeword weights of cyclic codes by using bounds on exponential sums. In particular, the author pays attention to a family of codes defined by Wolfmann, for which the parameters can be expressed in terms of numbers of solutions of trace equations.

Maximum-likelihood decoding of a linear block code can be efficiently performed with a trellis. An important parameter for judging the complexity of trellis decoding is the state complexity of the code. In [419], Tolhuizen shows that a binary linear code of length dimension k, Hamming distance d and state complexity at most $k-3$ has length $n \geq 2d + 2\lceil d/2 \rceil - 1$, and constructs a [15,7,5] code attaining this bound with equality.

A superimposed code in n-dimensional Euclidean space is a subset of vectors with the property that all possible sums of any m or fewer of these vectors form a set of

points which are separated by a certain minimum distance d. Since known bounds on the rate of such a code are not so useful for small values of m, Vangheluwe [425] studies experimentally the case $m = 2$ using visualization software packages, leading to plots for both the random-coding bound and the sphere-packing bound.

4.1.3 Cooperating Codes

Two (or more) error-correcting codes can be combined into a new code, which has good error correction capabilities for combinations of random and burst errors. The new (long) code can make use of the encoding and decoding algorithms of the (short) constituent codes, so the encoding and decoding complexity can be kept rather low. *Product Codes* and *Concatenated Codes* are two important classes of such cooperating codes. In the product coding concept, two (or more) codes over the same alphabet are combined. In the concatenated coding concept, a hierarchical coding structure is established by combining an inner code over a low-order (mostly binary) alphabet with an outer code over a high-order alphabet.

The product coding concept was introduced by Elias in 1954 [9]. In the two-dimensional case, the codewords are arrays in which the rows are codewords from a code $\mathcal{C}_1$, while the columns are codewords from a code $\mathcal{C}_2$. After (row-wise) transmission, the received symbols are collected in a similar array, in which first the rows are decoded according to $\mathcal{C}_1$ and next the columns according to $\mathcal{C}_2$. In this way, random errors are likely to be corrected by the row decoder, while remaining burst errors, which have been distributed over various columns due to interleaving, are to be corrected by the column decoder.

In [370], Blaum, Farrell and Van Tilborg consider simple product codes using even-weight codes (requiring only a single parity-check bit) as constituent codes. They propose a diagonal read-out structure (instead of the traditional row-wise procedure) together with an efficient decoding algorithm, which enables the correction of relatively long burst errors.

In [385], Tolhuizen and Baggen show that a product code is much more powerful than commonly expected. Product codes generally have a poor minimum distance, i.e, there may exist codes of the same length and dimension with a higher minimum distance. Nevertheless, they may still offer good performance, since many error patterns of a weight exceeding half the minimum distance can be decoded correctly, even with relatively simple algorithms. The authors derive upper bounds on the number of error patterns of low weight that a nearest neighbor decoder does not necessarily decode correctly. Further, they also present a class of error patterns which are decoded correctly by a nearest neighbor decoder. This class suggests possibilities beyond those already known in 1989 for the simultaneous correction of burst and random errors.

Concatenated codes were introduced by Forney [18] in 1966. The classical concatenated coding scheme consists of a binary *inner code* with 2^k words and an

outer code over GF(2^k), typically a Reed-Solomon code. Information is first encoded using the outer encoder. Next, each of the generated symbols is considered as a binary vector of length k, which is encoded using the inner code. After transmission, the received bits are decoded by the inner decoder, leading to symbols which are decoded using the outer decoder. In order to further increase the burst error correction capabilities, one can insert an interleaver between the outer and inner encoder, and a corresponding de-interleaver between the inner and outer decoder. A popular concatenated coding scheme (e.g., for deep space missions) uses a rate 1/2 convolutional inner code of constraint length $k = 7$, and a Reed-Solomon outer code over GF(256) of length 255 and dimension 223.

In [373], Van der Moolen proposes a decoding scheme for a concatenated coding system with a convolutional inner code and a Reed-Solomon (RS) outer code, with block interleaving. For bursty channels, if a symbol error occurs in an RS word, the symbols at the corresponding positions in the previous and next codewords are suspicious. Based on this observation, Van der Moolen develops a "decoding with memory" strategy. The basic idea is that if the RS decoder succeeds, then at all the locations of the (corrected) symbol errors, the Viterbi decoder is (re-)started to decode the corresponding symbols of the subsequent codewords with the new initial states. Furthermore, the author gives a 12-state Markov model describing the process of decoding with memory for the concatenated coding system.

In the same year, Tolhuizen [375] considered the generalized concatenation construction proposed by Blokh and Zyablov in 1974. The BZ construction uses a code $\mathcal{A}_1$ over GF(q) of dimension k and r (outer) codes $\mathcal{B}_i$ over GF(q^{a_i}), where $\sum_{i=1}^{r} a_i = k$. The author indicates how these ingredients should be chosen to obtain a good code, i.e., a code with high minimum distance given its length and dimension.

At the 1989 symposium in Houthalen, prof. T. Ericson from Linköping University in Sweden gave an invited lecture on recent developments in concatenated coding [386]. In particular, he discussed decoding principles, the construction of optimal codes via concatenation (e.g., a construction of the Golay code using a Reed-Solomon outer code and a trivial distance-1 inner code), and asymptotic bounds.

In the late 1990s, Weber and Abdel-Ghaffar studied decoder optimization issues for concatenated coding schemes. Instead of exploiting the full error correction capability of the inner decoder with Hamming distance d, they use this capability only partly, thus leaving more erasures but less errors for the outer decoder. Since it is easier to correct an erasure than an error, there is a trade-off problem to be solved in order to determine the optimal choice. In [420], the inner code error-correction radius t is optimized over all possible values $0 \leq t \leq \lfloor (d-1)/2 \rfloor$, either by maximizing the number of correctable errors or by minimizing the unsuccessful decoding probability. For small channel error probabilities, a strategy that is optimal in the latter respect is also optimal in the former respect. However, for large channel error probabilities, a strategy that is optimal in one respect may

be suboptimal in the other. In [430], the erasing strategy is not determined by the inner code error-correction radius, but it is made adaptive to the actual reliability values of the inner decoder outputs. The authors also determine the maximum number of channel errors for which correction is guaranteed under such an optimized erasing strategy.

In 1995, Baggen and Tolhuizen [409] introduced a new class of cooperating codes: *Diamond codes*. The two constituent codes, C_1 and C_2, have the same length n and are defined over the same alphabet. As illustrated in Figure 4.3, the Diamond code consists of the bi-infinite strips of height n, where each column is in C_1 and each slant line with a given slope is in C_2. In contrast to CIRC (Cross Interleaved

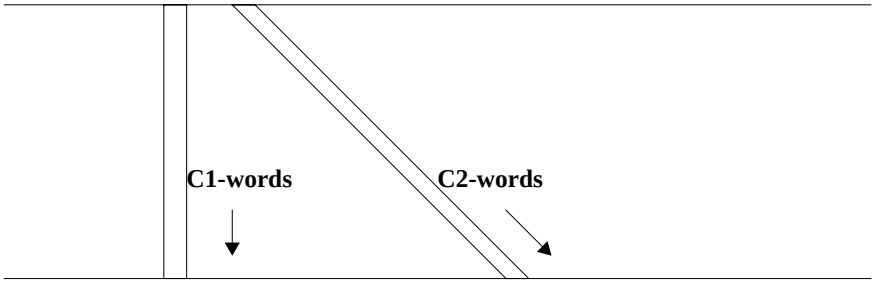


Figure 4.3: *The format of Diamond codes.*

Reed-Solomon Code, used in the CD system), all symbols of the Diamond code are checked by both codes. In the area of optical recording, the application of Diamond codes can enhance storage densities significantly. In the accompanying paper [410], Tolhuizen and Baggen consider block variations of Diamond codes in order to make these more suited for rewritable, block-oriented applications.

4.2 Decoding Techniques

In the previous sections, we considered papers dealing with properties of codes and constructions of codes. In the present section, we review papers on the decoding of error-correcting codes, both block codes and convolutional codes. Various contributions to the decoding of convolutional codes are described.

4.2.1 Hard-Decision Decoding

Hard-decision decoders operate on the symbol estimates delivered by the demodulator. A hard-decision decoder may decode up to a pre-specified number of errors and declare a decoding failure otherwise; in that case, we speak of a *bounded-distance* decoder.

In [359], Simons and Roefs describe algorithms for the encoding and decoding of $[255, 255-2T, 2T+1]$ Reed-Solomon codes over GF(256) that allow an efficient

implementation in digital signal processors. The decoding algorithms contain the following conventional steps: syndrome computation, solving the key equation, and error location and evaluation. Significant savings in the number of computations are reported for Fast Fourier Transform techniques (strongly advocated in the then recent book of Blahut [62]) used for encoding, syndrome computations and for determining the error values.

In [379], Stevens shows that the BCH algorithm can be used to decode up to a particular instance of the Hartmann-Tzeng bound. By applying this result while trying all values of a set of judiciously chosen syndromes, he obtains an algorithm for decoding cyclic codes up to half their minimum distance. For various code parameters, the cardinality of this set of syndrome values to be tried is minimized, and thus efficient decoding algorithms are obtained.

Van Tilburg describes [387] a probabilistic algorithm for decoding an arbitrary linear $[n, k]$ code. It refines the following well-known method. A set of k of the n received bits is selected at random. It is hoped that these k bits are error free. If the positions corresponding to these k bits form an information set, the unique codeword corresponding to these k bits is determined, and it is checked whether the codeword so obtained is sufficiently close to the received word. If not, another group of k bits is selected. The method proposed by Van Tilburg features a systematic way of checking, and a random bit swapping procedure.

In [415], Heijnen considers binary $[mk, k]$ codes that are quasi-cyclic. That is, if

$$(c_1, c_2, \ldots, c_k \mid c_{k+1}, \ldots, c_{2k} \mid \ldots \mid c_{(m-1)k+1}, \ldots, c_{mk}) \tag{4.6}$$

is a codeword, then the vector obtained by simultaneously applying a cyclic shift on each of the m blocks

$$(c_k, c_1, \ldots, c_{k-1} \mid c_{2k}, c_{k+1}, \ldots, c_{2k-1} \ldots \mid c_{mk}, c_{(m-1)k+1}, \ldots, c_{mk-1}) \tag{4.7}$$

is a codeword as well. Three general decoding methods are compared: comparison to all codewords, syndrome decoding (where the quasi-cyclic property allows reduction of the number of coset leaders to be stored), and "error division". The latter method is based on the observation that an error vector of weight t has a weight of at most $s = \lfloor \frac{t}{m} \rfloor$ in at least one of its m blocks. For each i, $1 \leq i \leq m$, and each vector $\mathbf{e}$ of length k and weight at most s, the codeword is computed that in the i-th block equals to the sum of $\mathbf{e}$ and the i-th block of the received word. The Hamming distance of the codeword so obtained and the received vector is used to select the codeword to decode to.

4.2.2 Soft-Decision Decoding

While *hard-decision decoders* do their job solely based on the symbol estimates delivered by the demodulator, *soft-decision decoders* also take into consideration the reliability of those estimates. This leads to better performance, at the expense of higher complexity. Over the years, many soft-decision decoding techniques

have been proposed. Although a *maximum-likelihood* (ML) decoding algorithm minimizes the decoding error probability, other algorithms are of interest as well, due to the (prohibitively) high computational complexity of ML decoding for long codes.

Generalized Minimum Distance (GMD) decoding, as introduced by Forney [17] in 1966, permits flexible use of reliability information in algebraic decoding algorithms for error correction. In subsequent trials, an increasing number of the most unreliable symbols in the received sequence is erased, and the resulting sequence is supplied to an algebraic error-erasure decoder, until the decoding result and the received sequence satisfy a certain distance criterion. In Forney's original algorithm, the unique codeword (if one exists) satisfying the generalized minimum distance criterion is found in at most $\lceil d/2 \rceil$ trials, where d is the Hamming distance of the code. In 1972, Chase [28] presented a similar class of decoding algorithms for binary block codes, in which unreliable symbols are inverted (instead of erased) in various decoding trials. From the list of generated codewords the most likely one is chosen as the decoding result. Although the Forney and Chase decoding approaches are rather old, they are still highly relevant. The resulting decoders are not only used as stand-alone decoders, but also as constituent components in modern techniques like iterative decoding of product codes.

In [391], Hollmann and Tolhuizen present a new condition on GMD decoding to guarantee correct decoding. They apply their weakened condition on the decoding of product codes, and describe a class of error patterns that is corrected by a slightly adapted version of the GMD-based Wainberg algorithm for decoding product codes is described. This class of error patterns equals the class that Tolhuizen and Baggen [385] showed to be correctable by a nearest neighbor decoder two years before, cf. Section 4.2.1.

In the early 2000s, Weber and Abdel-Ghaffar considered reduced GMD decoders. They studied the degradation in performance resulting from limiting the number of decoding trials and/or restricting (e.g., quantizing) the set of reliability values. In [431], they focus on single-trial methods with fixed erasing strategies, threshold erasing strategies, and optimized erasing strategies. The ratios between the realizable distances and the code's Hamming distance for these strategies are about 2/3, 2/3, and 3/4, respectively. A particular class of reliability values is emphasized, allowing a link to the field of concatenated coding. In [437], asymptotic results on the error-correction radius of reduced GMD decoders are derived.

Recently, limited-trial versions of the Chase algorithm were introduced as well. The least complex version of the original Chase algorithms ("Chase 3") [28] uses roughly $d/2$ trials, where d is the code's Hamming distance. In [442], Kossen and Weber show that decoders exist with lower complexity and better performance than the Chase 3 decoder. It also turns out that optimization of the settings of the trials depends on the nature of the channel, i.e., AWGN and Rayleigh fading channels may require different arrangements. In [449], Weber considers Chase-like algorithms achieving bounded-distance (BD) decoding, i.e., decoders for which

the error-correction radius (in Euclidean space) is equal to that of a decoder that maps every point in Euclidean space to the nearest codeword. He proposes two Chase-like BD decoders: a static method requiring about $d/6$ trials, and a dynamic method requiring only about $d/12$ trials. Hence, the complexity is reduced by factors of three and six, respectively, compared to the Chase-3 algorithm.

4.2.3 Decoding of Convolutional Codes

The Viterbi algorithm [110, Ch. 4] is a well-known method for decoding convolutional codes that minimizes the sequence-error probability. It is the most popular decoding algorithm for decoding convolutional codes with a short constraint length. In literature, quite some attention has been paid to implementation aspects of the algorithm. Also some contributions to the WIC symposia dealt with implementation aspects of the Viterbi algorithm.

In [369], Nouwens and Verlijsdonk discuss (in Dutch) soft-decision Viterbi decoding of a rate $R = 1/2$, $K = 3$ convolutional code with generator polynomials $1 + D + D^2$ and $1 + D^2$ that is used on an AWGN channel. The effect of quantization of the bit reliabilities that serve as input to the Viterbi decoder is studied. An equally-spaced quantizer is assumed, and the level spacing is determined to optimize the union bound on the error probability after decoding.

Baggen, Egner and Vanderwiele [448] discuss quantization for a Viterbi decoder used on a Rayleigh fading channel. Also here, an equally-spaced quantizer is considered. The level spacing is now computed in such a way that the cut-off rate of the discrete channel resulting from this quantization is optimized. The optimal spacing depends only weakly on the average SNR, and it is better choose one that is too large than one that is too small. Simulation results suggest that the spacing that maximizes the cut-off rate is optimal for Viterbi decoding as well.

Quantization of the bit reliabilities is not the only important practical aspect of Viterbi decoding; one also has to determine which numerical range suffices for performing the required computations. In [393] and [397], Hekstra gives results on the maximum difference between path metrics in Viterbi decoders. From this maximum difference, he derives consequences for reduction of the required numerical range.

The Viterbi algorithm operates on a trellis that has a number of states that is exponential in the encoder constraint length. Consequently, the implementation of the Viterbi algorithm is impractical for convolutional codes with a large constraint length. In this case, *sequential decoding* [110, Ch. 6], which can be seen as a backtracking decoding method, can be applied. In the basic stack algorithm, a search is performed in a tree, while a list is maintained of paths of different lengths ordered according to their metrics. The path with the highest metric is extended and subsequently removed, while the new paths are placed within the ordered list (stack). The stack algorithm suffers incomplete decoding because the stack is full ("stack overflow"). Its number of required computations depends on the actual noise se-

quence. In [351], Schalkwijk describes several ways of reducing the complexity of sequential decoders, using the syndrome of the received vector. One of the observations is that extension of a noise sequence with a "zero" digit is much more likely than extension with a "one" digit, and that one has to consider more noise digits at each decoding step to obtain two *a-priori* equally likely extensions. Simulations results are given.

The m-algorithm is a list decoding algorithm [110, Ch. 5]. It is a non-backtracking method and, in contrast to sequential decoding, its decoding complexity does not depend on the actually received sequence. The idea of the algorithm is that at each time instant, a list of the m most promising initial (equal length) parts of the codewords is extended. In [383], Van der Vleuten and Vinck describe an implementation of the m-algorithm. Paths for which the metric is below the median are extended; the others paths are not. As finding the median of m numbers is linear in m, the time complexity of the algorithm is linear in m. Their ingenious trace-back method allows use of a small trace-back memory.

Assume that we generate the list of the m most likely transmitted words from a convolutional code, given the received sequence. If messages include a CRC check sum, the most likely codeword in the list that has a correct CRC checksum can be selected as final decoding result. In this way, a significant decoding gain over conventional Viterbi decoding ($m = 1$) can be obtained. In [447], Hekstra proposes to generate an unordered list of *all* words for which the path metric exceeds that of the most likely path with at most B. In this way, sorting of paths according to their path metrics is avoided. An algorithm for generating this list is given. The length of the list is a random variable. A strategy is described for choosing B in such a way that the list size remains reasonable. Simulation results are presented, showing a decoding gain of about 1.5 dB for the coding scheme employed in GSM/GPRS on a static AWGN channel.

In 1983, Best [353] describes a convolutional decoder that outputs reliability information. This decoder seems to be a re-discovery of the BCJR algorithm or forward-backward algorithm described by Bahl, Cooke, Jelinek and Raviv in 1974 [34] and well forgotten until its usage in the decoding of Turbo codes in the 1990s. Best considers such a decoder "not useful for practical purposes because of speed limitations", but he does find it useful for theoretical insight in what happens in decoding. He mentions that the likelihood of a state in a most likely path is almost always equal to one, until the decoder is forced to choose between two paths with almost the same metric. In that case, the probability drops to about one half, and remains on that value until paths merge. As a result, Best was led to modify a Viterbi decoder so that it outputs both alternative paths in case of a close decision. In a concatenated code system, the outer code then can decide which path is the correct one.

The Viterbi algorithm minimizes the *sequence* error probabilities, while the BCJR algorithm [34] minimizes the *bit* error probability. In concatenated coding schemes, it seems more important to minimize the error probability of the *symbols* entering

the outer decoder. Willems and Pašić [413] describe an implementation of such a decoder with a complexity much lower than that achieved before, but still significantly larger than that of a Viterbi decoder. Simulations with a specific convolutional code show that the symbol error output rate of the proposed decoder is only negligibly lower than with Viterbi decoding. The proposed decoder has the advantage of generating soft-output information about the symbols, which can possibly be used by the outer decoder.

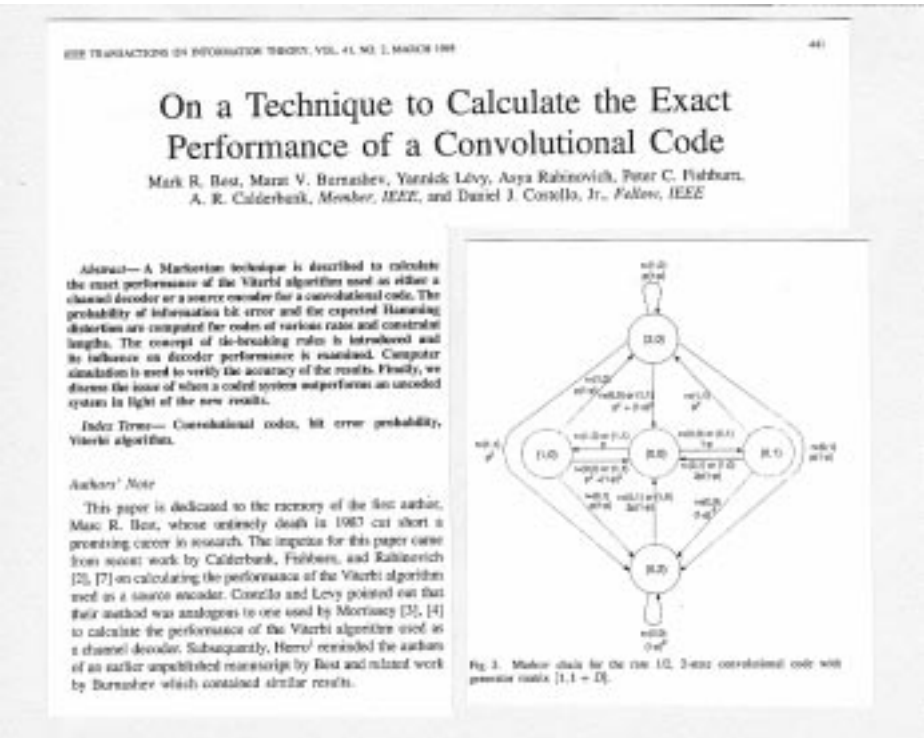
We finalize this section by discussing papers dealing with the performance of Maximum-Likelihood (ML) decoded convolutional codes employed on a binary symmetric channel with error probability p.

Post [346] describes an upper bound for the first error event probability of ML decoding. First, with the aid of the codeword enumerator of the code, he derives lower bounds on the weights of error patterns of a given length that a ML decoder does not decode correctly. Next, by analyzing a related random walk, he determines the probability of occurrence of error patterns satisfying these lower bounds. For small p, the well-known union bound is sharper, but for larger p, Post's bound is sharper.

Schalkwijk [348] describes a syndrome decoder for ML decoding of convolutional codes with the aim of analyzing the first error event probability. A diagram incorporating metrics and states is studied, and a Markov chain technique is applied for estimating the error event probability. This approach was continued and extended by Best, who shows in [368] that a convolutional coding scheme with ML decoding over a discrete memoryless channel can be modeled as a Markov chain. This model allows exact analysis of the statistical behavior of the errors. The method is illustrated with a $R = 1/2$ code with constraint length 1, used over a binary symmetric channel. Unfortunately, the amount of computation grows rapidly with the constraint length of the code. For example, according to the author, for the "standard code" with constraint length 3 and generator polynomials $1 + D$ and $1 + D + D^2$ used on a binary symmetric channel, the Markov model has as many as 104 states. In 1995, this work was reported on in [94], dedicated to the memory of Mark Best – see Figure 4.4.

4.2.4 Iterative Decoding

The introduction of *turbo codes* [90] in 1993 caused a true revolution in the field of error control coding. In their original form, turbo codes combine two recursive convolutional codes along with a pseudo-random interleaver in a parallel concatenated coding scheme. Through a maximum a posteriori (MAP) iterative decoding process, performances very close to the Shannon limit are achieved. As mentioned by Wicker in [108, Ch. 25, Sect. 11], turbo codes initially met with some skepticism, but already four years after their introduction, a turbo code experimental package was launched into space aboard the Cassini spacecraft. Further research on iteratively decodable codes resulted in the rediscovery of Gallager's



On a Technique to Calculate the Exact Performance of a Convolutional Code

Mark R. Best, Marat V. Burnashev, Yannick Lévy, Asya Rabinovich, Peter C. Fishburn, A. R. Calderbank, *Member, IEEE*, and Daniel J. Costello, Jr., *Fellow, IEEE*

Abstract— A Markovian technique is described to calculate the exact performance of the Viterbi algorithm used as either a channel decoder or a source encoder for a convolutional code. The probability of information bit error and the expected Hamming distortion are computed for codes of various rates and constraint lengths. The concept of tie-breaking rules is introduced and its influence on decoder performance is examined. Computer simulation is used to verify the accuracy of the results. Finally, we discuss the issue of when a coded system outperforms an uncoded system in light of the new results.

Index Terms— Convolutional codes, bit error probability, Viterbi algorithm.

Authors' Note

This paper is dedicated to the memory of the first author, Marc R. Best, whose untimely death in 1987 cut short a promising career in research. The impetus for this paper came from recent work by Calderbank, Fishburn, and Rabinovich [2], [7] on calculating the performance of the Viterbi algorithm used as a source encoder. Costello and Lévy pointed out that their method was analogous to one used by Morrissey [3], [4] to calculate the performance of the Viterbi algorithm used as a channel decoder. Subsequently, Herro[1] reminded the authors of an earlier unpublished manuscript by Best and related work by Burnashev which contained similar results.

Figure 4.4: *Paper in IEEE Transactions on Information Theory based on [368].*

low-density parity-check (LDPC) codes [15], dating from the 1960s. Currently, both turbo codes and LDPC codes are studied extensively and are considered as the most promising candidate codes for many application areas. For example, turbo codes have been implemented in UMTS, the third-generation mobile communication standard.

In [421], Tolhuizen and Hekstra-Nowacka consider turbo coding schemes employing serial (instead of parallel) concatenation. They focus on the word error rate after decoding, for which they give the average union bound. In order to compute this bound, one needs the input-output weight enumerator of the inner decoder. The authors provide an explicit formula for this enumerator, and apply it to some specific examples.

Dielissen en Huisken [432] explain four implementation techniques for the soft-input soft-output (SISO) decoding module of a third-generation mobile communication turbo decoder. They compare the performance and implementation costs (in terms of silicon area and power dissipation). The final choice is not trivial, but a trade-off between different aspects.

The inputs and outputs of an a-posteriori probability (APP) decoder as used in turbo decoding can be represented as log-likelihood ratios (LLRs). Hagenauer's box function $\log((1+e^{x+y})/(e^x+e^y))$ can be used to establish an explicit input-output relation of an APP decoder. Janssen and Koppelaar [433] consider turbo codes with BPSK modulation over an AWGN channel. They show that the random variable $\mathbf{z}$ that is the output of the box function exhibits the LLR property, that is, for each z,

$$\log \frac{p_{\mathbf{z}}(z \mid b=0)}{p_{\mathbf{z}}(-z \mid b=0)} = z. \tag{4.8}$$

They study the effect of mismatched inputs to the box function, and give upper and lower bounds on the LLR at the output of the box function as a function of mismatch.

Le Bars, Le Dantec and Piret [443] focus on the design of the interleavers in a turbo coding scheme. The authors present an algebraic interleaver construction method leading to codes with a high minimum distance. The performance of these codes are very good at high signal-to-noise ratios.

In [435], Balakirsky describes a realization of the Maximum-Likelihood (ML) decoding algorithm for messages encoded by an LDPC code and transmitted over a binary symmetric channel. The algorithm is based on the introduction of a tree structure in a space consisting of all possible noise vectors and principles of sequential decoding with the use of a special metric function. The author derives an upper bound on the exponent of the expected number of computations in the ensemble of low-density codes and shows that it is much smaller than the exponent for the exhaustive search. It should be noted that this work is based on a (Russian) paper by the author dating from 1991, i.e., from well before the world-wide rediscovery of LDPC codes!

Steendam and Moeneclaey [441] derive the ML performance of LDPC codes, considering BPSK and QPSK transmission over a Gaussian channel. They compare the theoretical ML performance with that of the iterative decoding algorithm. It turns out that the performance of the iterative decoding algorithm is close to the ML performance when the girth of the code is sufficiently high.

4.3 Codes for Data Storage Systems

Given the continuing demand for increased data storage capacity, it is not surprising that interest in coding techniques for mass data storage systems, such as optical and magnetic recording products, has continued unabated ever since the day when the first mechanical computer memories were introduced in the 1950s. Evidently, technological advances such as improved materials, heads, mechanics, and so on have been the driving force behind the "ever" increasing data storage capacity, but state-of-the-art storage densities are also a function of improvements in channel coding, the topic addressed in this section. The book by Immink [109] and the survey article by Immink, Siegel and Wolf [107] offer a comprehensive description

of the literature on this topic.

Optical recording, developed in the late 1960s and early 1970s, is the enabling technology of a series of very successful products for digital consumer electronics systems such as Compact Disc (CD), CD-ROM, CD-R, and Digital Video Disc (DVD). The design of codes for optical recording systems is essentially the design of combined *dc-free, run-length limited* (DC-RLL) codes.

An encoder accepts a series of information words as an input and transforms them into a series of output words, called codewords. Binary sequences generated by a (d, k) RLL encoder have, by definition, at least d and at most k 0s between consecutive 1s. Let the integers m and n denote the information word length and codeword length, respectively. The *code rate*, $R = m/n$, is a measure of the code's efficiency. The maximum rate of an RLL code, given values of d and k, is called the *Shannon capacity*, and it is denoted by $C(d, k)$ [3].

Early examples of RLL codes have been given by Berkoff [16] some forty years ago, and since then the chase of various code designers in the world has been the creation of "practical" RLL codes whose rate approaches Shannon's theoretical rate limit. Hundreds of examples of RLL codes have been published and/or patented over the years. Dc-free codes, as their name suggests, have no spectral components at the zero frequency and suppressed spectral content near the zero frequency.

4.3.1 RLL Block Codes

One approach that has proved very successful for the conversion of source information into constrained sequences is the one constituted by block codes. The source sequence is partitioned into blocks of length m, called *source words*, and under the code rules such blocks are mapped onto words of n channel symbols, called *codewords*. In order to clarify the concept of block-decodable codes, we have written down a simple illustrative case of a rate 3/5, $(1, \infty)$ block code. The codeword assignment of Table 1 provides a simple block code that converts source words of bit length $m = 3$ into codewords of length $n = 5$. The two left-most columns tabulate the eight possible source words along with their decimal representation. We have enumerated all eight words of length four that comply with the $d = 1$ constraint. The eight codewords, tabulated in the right-hand column, are found by adding one leading zero to the eight 4-bit words, so that the codewords can be freely cascaded without violating the $d = 1$ constraint.

The code rate is m/n = 3/5 $< C(1, \infty) \simeq 0.69..$, where $C(1, \infty)$ denotes the maximum rate possible for any $d = 1$ code irrespective of the complexity of such an encoder. The *code efficiency,* expressed as the quotient of code rate and Shannon capacity of the (d, k)-constrained channel having the same run length constraints, is $R/C(d, k) \simeq 0.6/0.69 \simeq 0.86$. Thus the very simple block code considered is sufficient to attain 86% of the rate that is maximally possible.

Table 4.1: *Simple* $(d = 1)$ *block code.*

source		*output*
0	000	00000
1	001	00001
2	010	00010
3	011	00100
4	100	00101
5	101	01000
6	110	01001
7	111	01010

It is straightforward to generalize the preceding implementation example to encoder constructions that generate sequences with an arbitrary value of the minimum run length d. To that end, choose some appropriate codeword length n. Write down all d-constrained words that start with d zeros. The number of codewords that meet the given run length condition is $N_d\,(n - d)$, which can be computed with generating functions or recursive relations [23].

A maximum run length constraint, k, can be incorporated in the code rules in a straightforward manner. For instance, in the $(d = 1)$ code previously described, the first codeword symbol is at all times preset to zero. If, however, the last symbol of the preceding codeword and the second symbol of the actual codeword to be conveyed are both zero, then the first codeword symbol can be set to one without violating the $d = 1$ channel constraint. This extra rule, which governs the selection of the first symbol, the *merging rule*, can be implemented quite smoothly with some extra hardware. It is readily conceded that with this additional 'merging' rule the $(1, \infty)$ code turns into a (1,6) code. The process of decoding is exactly the same as that for the simple $(1, \infty)$ code, since the first bit, the "merging" bit, is redundant, and in decoding it is skipped anyway. The (1,6) code is a good illustration of a code that uses state-dependent encoding (the actual codeword transmitted depends on the previous codeword) and state-independent decoding (the source word can be retrieved by observing just a single codeword, that is, without knowledge of previous or upcoming codewords or the channel state).

The first article describing RLL block codes was written by Tang and Bahl [23] in 1970. It describes a method where (d, k) constrained info blocks of length n' are cascaded with merging blocks of length $d + 2$. Twelve years later, it was shown by Beenker and Immink [60] that their method can be made more efficient by constraining the maximum number of zeros at the beginning and start of the (d, k) constrained info blocks to $k - d$. Then merging blocks of length d are sufficient to cascade (glue) the info blocks. The authors presented two constructions. In the first construction, the merging block is the all-zero word (as in Table 1), while in the second (more efficient) construction, the merging blocks depend on the two neighboring info words.

The methods described by Weber and Abdel-Ghaffar [392] [395] offer a more flexible and efficient method for cascading RLL blocks than that described in the early literature, specifically for the case where k is rather small. The method presented by Tjalkens [394] does not use 'merging bits' to cascade the RLL info blocks, but Tjalkens, alternatively, shows that with the set of (d, k) constrained codewords that start with at least d zeros and end with at most $k - 1$ zeros one may construct a RLL block of maximum size. Later constructions showed that merging blocks of length less than d can be used, where the merging algorithm can alter both the merging block and (small) parts of the info word.

The article by Hollmann and Immink [390] addresses the problem of generating RLL sequences, where we have the additional demand that a certain, prescribed, sequence of run lengths is not allowed to be generated. Said specific sequence of run lengths that should be avoided is called a *prefix*, which is normally used in recording practice as a synchronization pattern.

In essence all articles mentioned above discuss block codes. The article by Hollmann [398] uses a completely different approach, as codes generated by his constructions must be decoded by *sliding-block* decoders. A sliding-block decoder observes the n-bit codeword plus r preceding n-bit codewords plus q trailing n-bit codewords. Such a sliding-block concept leads to codes having a high efficiency, involving small hardware, and that usually do not have too many significant drawbacks. A drawback of codes that are decoded by a sliding-block decoder is *error propagation*, as the decoding operation depends on $r + q + 1$ consecutive codewords. In practice, the increased efficiency and reduced hardware of a sliding-block decoder outweigh the extra load on the error correction unit. There are various coding formats and design methods with which we can construct such codes. Immink [114] has recently shown that very efficient sliding block codes can be designed. For example, a rate 9/13, (1,18) 5-state encoder has a redundancy of 0.2%, while a rate 6/11, (2,15) 9-state encoder has a redundancy of 0.84%.

The article by Abdel-Ghaffar and Weber [412] addresses run-length-constrained channels, where there is, as in the prior art, a maximum run length constraint, and additionally a maximum run length constraint on both the odd and the even positions of the encoded sequence. These codes are often called $(0, G/I)$ constrained, where G denotes the maximum run length constraint on the sequence, and I denotes the maximum run length imposed on the symbols at the odd and even positions. Abdel-Ghaffar and Weber study block codes, where they show results on the maximal size of a set of $(0, G/I)$ constrained codewords of length n that can be freely concatenated without violating the specified $(0, G/I)$ constraint.

Closing Remark by the Editors

The work described in several WIC papers of Schouhamer Immink *et al.* summarized in this subsection on RLL codes has found its way in consumer electronics products, such as CD and DVD. His contributions to these products have gained

him acknowledgment from several international institutions and societies.

4.3.2 Dc-Free Codes

Dc-balanced or *dc-free* codes, as they are often called, have a long history and their application is certainly not confined to recording practice. Since the early days of digital communication over cable, dc-balanced codes have been employed to counter the effects of low-frequency cut-off due to coupling components, isolating transformers, and so on. In optical recording, dc-balanced codes are employed to circumvent or reduce interaction between the data written on the disc and the servo systems that follow the track. Low-frequency disturbances, for example due to fingerprints, may cause completely wrong read-out if the signal falls below the decision level. Errors of this type are avoided by high-pass filtering, which is only permissible provided that the encoded sequence itself does not contain low-frequency components, or, in other words, provided that it is dc-balanced.

Rejection of LF components is usually achieved by bounding the accumulated sum of the transmitted symbols. Common sense tells us that a certain rate has to be sacrificed in order to convert arbitrary user data into a dc-balanced sequence. The quantification of the maximum rate, the capacity, of a sequence given the fact that it contains no low-frequency components has been reported by Chien [22]. The articles by Immink [358] and De With [360] provide a description of key characteristics of dc-free sequences generated by a Markov information source having maximum entropy. Given the fact that a Markov source, which describes a dc-balanced sequence, is maxentropic, we can substitute the maxentropic transition probabilities. Then computation of the spectrum is straightforward. Knowledge of ideal, "maxentropic" sequences with a spectral null at dc is essential for understanding the basic trade-offs between the rate of a code and the amount of suppression of low-frequency components. The results obtained in [358] and [360] allow us to derive a figure of merit of implemented dc-balanced codes that takes into account both the redundancy and the emergent frequency range with suppressed components (notch width).

Beenker and Immink [367] present a category of dc-free codes called dc^2*-free codes*. This type of codes offers a larger rejection of low-frequency components than is possible with the traditional codes discussed in the prior art. Besides the trivial fact that they are dc-balanced, an additional property of dc^2-free codes is that the second (and even higher) derivative of the code spectrum also vanishes at zero frequency (note that the odd derivatives of the spectrum at zero frequency are zero because the spectrum is an even function of the frequency). The imposition of this additional channel constraint results in a substantial decrease of the power at the very low frequencies for a fixed code redundancy as compared with the designs based on the conventional 'bounded accumulated sum' concept. The drawback of this new scheme is the implementation of the codes, as it demands significantly more hardware and large codewords at high coding rates.

4.3.3 Error-Detecting Constrained Codes

The paper by Immink [374] offers coding techniques for simple partial-response channels. He showed that the simple bi-phase code can be used as an inner code of an outer code designed for maximum (free) Hamming distance. The paper by Weber and Abdel-Ghaffar [389] discloses a class of run-length-limited codes that can detect asymmetric errors made during transmission. Baggen and Balakirsky [450] consider data transmission over so-called bit shift channels with $(2, \infty)$ RLL constraints, and obtain bounds on the entropy of the output sequences.

4.4 Codes for Special Channels

4.4.1 Coding for Memories with Defects

In 1974, Kusnetsov and Tsybakov introduced [35] the following model for *coding for memories with stuck-at defects.* In some memory cells, known to the encoder, only one particular symbol (known to the encoder) can be written. The decoder does not know in which positions stuck-at errors occur. The question is how much information can be stored in such a memory with stuck-at defects. Kusnetsov and Tsybakov [35] gave upper bounds on the rate that can be obtained if a fraction p of the positions contain stuck-at errors. With a random coding argument, they obtained the surprising result that the capacity of a stuck-at channel with stuck-at probability p equals 1-p.

Some ten years later, coding for stuck-at defects was a popular subject at various WIC symposia. In 1985, Van Pul [361] described an *explicit* construction for obtaining the capacity of the stuck-at channel with stuck-at probability p. In the same year, Baggen [362] showed that MDS codes achieve the upper bound on the information rate, given the number of stuck-at errors combined with random errors. Vinck [363] varies on the theme by using convolutional codes for correcting *bursts* of defect errors, separated by guard spaces. In [382], Peek and Vinck give an explicit algorithm for the binary stuck-at channel. Bounds for the bit error rate and the decoding complexity are also obtained. Schalkwijk and Post [381] take an information-theoretic approach to coding for stuck-at errors. Indeed, suppose that information is stored in elementary blocks of n bits. The memory with known defects is then equivalent to a noisy channels with input and output alphabets of size 2^n. This "superchannel" can be described by a strategy in which an n-bits input block is to be used for a particular input message and defect pattern. In a memory with known defects, the bit values that are eventually read out become available at the moment of storing. In other words, the equivalent super channel has perfect feedback, and repetition feedback strategies can be used [26] – see also Section 4.4.5. Strategies for small n are described.

Vinck and Post [376] discuss the following combined test and error-correction procedure. A message m of even length is initially written in memory as $x(m) = (0, m, P)$, where P is the parity of m. Upon reading a word z from memory, we check if it has an even number of ones. If so, we leave it unchanged; if not, we

invert all its bits and obtain z′. If z originates from $x(m)$ by a single stuck-at error, then all bits of z except for the stuck-at bit are actually inverted; the stuck-at bit keeps its value that is incorrect for $x(m)$. Consequently, z' is the complement of $x(m)$. We see that m can be represented by two messages, namely $x(m)$ and its complement, as long as at most one stuck-at error occurs in the bits of word. Note that both $x(m)$ and its complement have an even number of ones. We keep applying the same procedure. A next single stuck-at error that occurs in the course of time is detected, as inversion of the word leads to a 0 in the leftmost bit. Upper and lower bounds on the mean time before a memory fails with this procedure are given, and an extension of the procedure for combination with coding for random (non-permanent) errors is indicated.

In 1989, Bassalygo, Gelfand and Pinsker [76] introduced the model of *localized errors*. In this model, the encoder knows a set of E of codeword positions in which an error *may* occur; outside E, no errors occur. The decoder does *not* know E. Coding for this model received quite some attention in the early nineties, as indicated by Bratatjandra and Weber in their paper from 1997 [417]. In this paper, the authors take for E a set of multiple burst errors, that is, E is the union of a collection of disjoint sets of consecutive positions. In literature, the main attention is on the sets E consisting of all set of positions up to a certain cardinality. Bratatjandra and Weber assume that both encoder and decoder know an upper bound m on the number of bursts, and an upper bound b on the length of each burst. They give a "fixed-rate" scheme for this situation. They also give a "variable-rate" scheme that allows the transmitter to send more information information if the actual number of burst errors is below m, or one or more of the burst lengths is below b.

4.4.2 Asymmetric/Unidirectional Error Control Codes

Most classes of error control codes have been designed for use on binary symmetric channels, on which $0 \rightarrow 1$ cross-overs and $1 \rightarrow 0$ cross-overs occur with equal probability (*symmetric errors*). However, in certain applications, such as optical communications, the error probability from 1 to 0 may be significantly higher than the error probability from 0 to 1. These applications can be modeled by an asymmetric channel, on which only $1 \rightarrow 0$ transitions can occur (*asymmetric errors*). Further, some memory systems behave like a unidirectional channel, on which both $1 \rightarrow 0$ and $0 \rightarrow 1$ errors are possible, but per transmission, all errors are of the same type (*unidirectional errors*).

Codes that detect and/or correct symmetric errors have been studied extensively since the 1940s. Of course, these codes can also be used to detect and/or correct asymmetric or unidirectional errors. However, it seemed likely that it should be possible to design codes that detect and/or correct asymmetric or unidirectional errors which need less redundancy than a comparable symmetric error correcting code. Pioneering work in this area was done by Varshamov [33] in the 1960s and 1970s. In the Benelux, the topic was further explored by Weber and various coauthors in the late 1980s and early 1990s.

In [377], Weber, De Vroedt and Boekee propose a method to construct codes correcting up to t asymmetric errors by expurgating and puncturing codes of Hamming distance $2t+1$. The resulting codes are often of higher cardinality than their symmetric error-correcting counterparts, but are mostly nonlinear. The same group of authors derived bounds on the sizes of codes that correct unidirectional errors [378], and they determined necessary and sufficient conditions for a block code to be capable of correcting/detecting any combination of symmetric, unidirectional, and asymmetric errors [384].

For practical purposes it is highly desirable that a code is *systematic*, i.e., that the message is to be found unchanged in the codeword. In [399], Weber and Kaag present a construction method for systematic codes which are able to correct up to t asymmetric errors and detect from $t+1$ up to d asymmetric errors.

Finally, in [405], Weber studies the asymptotic behavior of the rates of optimal codes correcting and/or detecting combinations of symmetric, unidirectional, and/or asymmetric errors. The main conclusion is that, without loosing rate asymptotically, one can upgrade any error control combination to simultaneous symmetric error correction/detection and *all* unidirectional error detection.

4.4.3 Codes for Combined Bit and Symbol Error Correction

In 1983, Piret introduced [355] binary codes for compound channels where both bit errors and symbol errors occur, where a symbol is a fixed group of bit positions. He introduces a distance profile to measure the error control capabilities and gives some examples of codes for combined bit and symbol error control.

Two years later, Van Gils published the first of a series of 3 papers dealing with the construction of codes for combined bit and symbol error correction. In the application that Van Gils has in mind, a symbol corresponds to a module in a processor. An erased symbol thus corresponds to a module that is detected to be in error, while an erroneous symbol corresponds to a malfunctioning module that is not detected to be in error. In [366], Van Gils announces binary $[3k, k]$ codes for $k = 4, 8, 16$ that can correct one single symbol error (i.e., one of the three groups of k bits is in error), up to $k/4$+1 bit errors, and one single symbol erasure plus up to $k/4$ bit errors (for $k = 4, 8$) or 3 bit errors (for $k = 8$). In addition, for $k = 8$ and $k = 16$, $k/4$+2 bit errors can be detected. In [371], he describes a binary [27,16] code, with symbol size 9, that can correct single bit errors, detect single (9-bit) symbol errors and detect up to four bit errors. Finally, in [372], Boly and Van Gils suggest to construct codes for controlling bit and symbol errors by representing the symbols from a symbol-error correcting code with respect to a judiciously chosen basis.

4.4.4 Coding for Informed Decoders

In 2001, Van Dijk, Baggen and Tolhuizen introduced informed decoding [438]. This concept was inspired by the following practical application. The address of

a sector of an optical disc is part of a header that is protected by its own error-correcting code. In many circumstances, the location of the reading/writing head is approximately known. The question is whether it is somehow possible to use this information on the actual sector address for retrieving the header more reliably. With informed decoding, it assumed that the decoder is informed about the value of some information symbols of the transmitted codeword. The authors show that with judicious encoding, the decoder can employ such information to effectively decode to a subcode with a larger minimum distance. Three ways to encode well-known codes that lead to favorable decoding capabilities are presented.

In [440], Tolhuizen, Hekstra, Cai and Baggen discuss two aspects of coding for informed decoding. Firstly, they propose to use a certain Gray code for addressing sectors in such a way that all addresses of sectors close to a target sector have many coordinates in common. In this manner, it is ensured that whenever the reading/writing head lands close to the target sector, many coordinates of the address of the sector in which the head actually lands are known. It is claimed that the proposed method yields the maximum number of common coordinates for each maximum deviation of the target sector. The other aspect aims to improve decoding for data encoded using a formed informed decoding, but where no information about known information symbols is supplied to the encoder. This is done by combining the codewords of several consecutive sectors, which usually have many information symbols in common.

4.4.5 Coding for Channels with Feedback

Already in 1956, Shannon proved [10] the surprising fact that feedback does not increase the capacity of a discrete memoryless channel. Feedback may, however, significantly reduce the complexity that is required to obtain reliable communication. In 1971, Schalkwijk presented simple fixed-length feedback strategies for the binary symmetric channel with error probability p [26]. It is assumed that the feedback is error-free and instantaneous, that is, immediately after the transmission of a bit, the transmitter knows which bit value has been received. Schalkwijk's strategies achieve an upper bound on the rate below which reliable communication is possible and can be described as follows. A message index s is pre-coded to an n-bits message m that does not contain a run of k equal symbols. The transmitter consecutively transmits the bits of m until the feedback reports the occurrence of an error. In such a case, the bit that was meant to be transmitted is repeated k times and transmission continues until the next error occurs. If all bits of m have been transmitted successfully, a tail is added until n bits have been transmitted. The receiver decodes as follows. Working its way back from the last received bit, it replaces subsequences 01^k by 1 and 10^k by 0, respectively, and afterwards, it removes the tail.

In the 1990s, Veugen and Bargh, two Ph.D. students of Schalkwijk, build further on his research on channels with feedback. The remainder of this section describes their work as presented at various WIC symposia.

A possible choice for the tails in Schalkwijk's strategy is the alternating sequence $0101\ldots$. In [407], Veugen studies conditions on the tails that are sufficient for correct operation of Schalkwijk's strategies. In [396], he introduces the following generalization of Schalkwijk's scheme. Each bit of the message m is transmitted c times in c consecutive transmissions. If not all c received bits are equal, the receiver neglects them, and the transmitter again transmits the intended message bit c times, until c equal bits are received. If the receiver decodes incorrectly, which happens if the channel produces c consecutive errors, the transmitter acts like that in Schalkwijk's scheme: it inserts the last message bit k times in the message m. This scheme reduces to Schalkwijk's scheme if $c = 1$. For $c > 1$, it introduces large redundancies, so it is not suitable for small p. For each $p < 1/2$, a strategy can be found that has a positive rate. The schemes need less than 1 bit feedback per transmitted bit, as for each c bits, the encoder only needs to know if they were all zero, all one, or not all equal.

In [406], Veugen considers the following extension of Schalkwijk's scheme to non-binary channels. If the transmitter observes that symbol j was received, although it sent symbol i, it immediately repeats symbol i k_{ij} times. A pre-coder takes care that in the data stream to be transmitted, subsequences of the form $ji^{k_{ij}}$ (with $i \neq j$) do not occur. Veugen considers decoding with a fixed delay D. That is, suppose the sequence $(x_n)_{n\geq 0}$ is transmitted, and the sequence $(y_n)_{n\geq 0}$ is received. Symbol y_n will be decoded as follows. The sequence $y_n, y_{n+1}, \ldots, y_{n+D}$ is scanned from right to left, and each subsequence $ji^{k_{ij}}$ is replaced by i. The leftmost symbol of the resulting sequence is the estimate $\hat{x}_n$. By comparing $\hat{x}$ and y, the pre-coder inverse can locate the errors and eliminate the error correction symbols. Veugen studies the error probabilities for these schemes. Combining calculations on random walks and a plausible conjecture, he computes the error exponent of the strategy.

In [414], Schalkwijk and Bargh consider the situation where the feedback link is without delay and noiseless, but operates at a smaller rate than the forward channel. They combine Ungerboeck's set partitioning technique and feedback schemes for full-rate feedback. The feedback scheme is used to see if the received signal was in the correct subset of signal points. If so, convolutional decoding is expected to retrieve the remaining information correctly. If not, the label of the subset of signal points is repeated. An example with feedback rate $1/2$ and a $\nu = 2$ convolutional code shows a much better performance than a much more complicated $\nu = 6$ convolutional code.

In [423], Bargh and Schalkwijk compare the block coding strategies discussed above with a recursive scheme. In the latter case, decoding takes place after a fixed delay D. A new strategy is discussed, and results on the rate and error exponent are obtained. In [428], Bargh and Schalkwijk introduce Soft-Repetition Feedback Coding and its recursive decoding method for binary input, soft-output symmetrical Discrete Memoryless Channels. The method is explained with a binary-input, quaternary output channel.

In [429], Bargh and Schalkwijk give an overview of error correction schemes in DMCs and AWGN channels with noiseless, instantaneous and full-rate feedback. They distinguish between two classes. In the first class, which they call "repeat to resolve uncertainty", the transmitter conceptually reconstructs the list of candidate codewords for the decoder, and aims to reduce this list size with every transmission. The second class of schemes, called "repeat to correct erroneous reception", the transmitter repeats a message segment if it is received incorrectly. In such schemes, a mechanism is required to signal to the receiver whether transmission is repeated, or a new segment is transmitted.

4.5 Applications

Channel coding theory is applied in a wide range of areas: deep space communication, satellite communication, data transmission, data storage, mobile communication, file transfer, digital audio/video transmission, etc. For an overview of applications in the first fifty years following Shannon's 1948 "noisy channel coding theorem", we refer to [105]. One of the most notable success stories for the Benelux in this respect is the development of the compact disc (CD) in the late 1970s and early 1980s [109]. In this section we provide an overview of various applications reported at the symposia on Information Theory in the Benelux.

In [347], Roefs discusses candidate concatenated coding schemes (cf. Section 4.1.3) for European Space Agency (ESA) telemetry applications in the early 1980s. The inner code is fixed as the standard rate 1/2 convolutional code of constraint length 7, but several candidates for the outer code are considered: Reed-Solomon codes with interleaving, Gallager's burst-correcting scheme, and Tong's burst-trapping scheme. Their performances are compared for dense burst channels with widely varying burst and guard space lengths. This work is continued in [350]. In this paper, Best and Roefs again take as inner code the conventional rate 1/2 convolutional code of constraint length 7. As outer code, they use a [256,224] Reed-Solomon code C over GF(257). To be more precise, they propose to encode 224 non-zero symbols (in GF(257)) systematically into a word from C. If a generated parity symbol happens to be zero, it is replaced by the element 1 (in GF(257)). The authors argue that the encoding error probability introduced by this replacement is negligible compared to the symbol error probability of the Viterbi decoder. The choice for GF(257) instead of GF(256) is motivated by the resulting possibility to employ the Fermat Number Transform for more efficient encoding and decoding.

Van Gils [364] describes dot codes for product identification (as an alternative to the well-known bar codes). As a product carrying a dot code word can have several orientations with respect to the read-out device, the same product is identified by several dot code words. It is indicated that for certain error-correcting codes, this ambiguity can be efficiently resolved.

At the time when telephony, telegraphy, and postal services were still all carried out by the PTT, Haemers considered the protection of a binary representation of the

postal code, as printed on envelopes, against read-out errors. In [365] he proposes the use of an (extended) Hamming codes for this purpose, with a small modification in order to increase the burst error detection capability.

Belgian bank account numbers consist of 12 digits, $a_9a_8 \ldots a_1a_0c_1c_0$, where c_0 and c_1 are such that $\sum_{i=0}^{9} a_i(10)^i \equiv 10c_1 + c_0 \pmod{97}$. The check digits c_0 and c_1 serve to detect the most common errors made by humans when processing digit strings (single errors, transpositions of consecutive symbols). Stevens [388] shows that replacing the modulus 97 by 93 slightly increases the error detection probability. Another slight increase is obtained if it is stipulated that the bank account number be divisible by 93, *i.e.*, that $\sum_{i=0}^{9} a_i(10)^{i+2} + 10c_1 + c_0 \equiv 0 \pmod{93}$.

Offermans, Breeuwer, Weber and Van Willigen [408] consider error-correction strategies for Eurofix, an integrated radio navigation system that combines terrestrial Loran-C and the satellite-based Global Positioning System (GPS). Differential GPS messages are transported via the Loran-C data link, which is disturbed by continuous wave interference, cross-rate interference, atmospheric noise, etc. In order to combat these phenomena, the authors propose a coding scheme based on the concatenation of a Reed-Solomon code and a parity check code.

In [411], Hekstra considers the following synchronization problem. Suppose that when a bit string $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is written down, then either $\mathbf{x}$ or one of its cyclic shifts, i.e., a string of the form $(x_{1+i}, x_{2+i}, \ldots, x_n, x_1, \ldots, x_i)$, could be read out. The problem is how to efficiently encode much information into strings such that all cyclic shifts of two distinct information strings are different. The author proposes the following method for efficient encoding of nearly the maximum amount of information. Suppose that $n = 2^m - 1$. Then encode $k = n - m$ information bits systematically to a cyclic Hamming code of length n, and subsequently invert the leftmost parity symbol. Synchronization is re-established by single-error correction, followed by shifting the received sequence until the error position corresponds to the leftmost parity bit.

In [418], De Bart shows that the channel coding scheme of the Digital Video Broadcasting (DVB) satellite system, based on the concatenation of a Reed-Solomon code and a convolutional code, has to deal with ambiguities that cannot be solved by the Viterbi decoder. The channel and the QPSK demodulator may cause transformations (rotations, shifts, etc.) yielding an incorrect sequence that resembles a codeword of the original convolutional code. Joined synchronization of the Viterbi and Reed-Solomon decoders should solve the problem.

A method for error correction in IC implementations of Boolean functions is proposed by Muurling, Kleihorst, Benschop, Van der Vleuten and Simonis [434]. The methods corrects both manufacturing *hard* errors and temporary *soft* errors during circuit operation. A systematic Hamming code is used, which can be implemented through additional logic or even through software tools.

Desset [439] considers error control coding for Wireless Personal Area Networks

(WPAN) in 2002. In a Wireless Personal Area Network, power consumption plays a very important role. High-performance channel coding strategies can be used to obtain coding gain and thus reduce transmit power. The average energy required per bit in a typical situation is about 15 nJ/bit. In addition, power consumption due to the complexity of encoding and decoding has to be considered. The complexity of Hamming codes, Reed-Muller codes, Reed-Solomon codes and Convolutional and Turbo codes has been analyzed. The two constraints are in contradiction and an optimum solution has to be found. The paper proposes a strategy to select error correcting codes for WPANs. For applications with different average bit energies ranging from 100 pJ/bit to 10 nJ/bit, the authors recommend Hamming codes, short constraint-length convolutional codes, and turbo coding, respectively.

CHAPTER 5

Communication and Modulation

C.P.M.J. Baggen (Philips Research Eindhoven)
A.J. Vinck (University of Essen)
A. Nowbakht-Irani (TU Eindhoven)

Introduction

Surprisingly, the earliest paper in this chapter originates from the seventh WIC symposium, testifying that the "transmission and modulation community" within the Benelux at first did not identify itself with the WIC. Actually, the advent of coded modulation and the interest in modulation issues of people having a background in coding and information theory led to a growing stream of WIC papers in this field. Also upcoming industrial applications like digital storage and transmission in the eighties (e.g., CD, GSM and DAB) stimulated research and publications within the WIC.

The chapter on Communication and Modulation is subdivided into the sections Transmission, Recording and Networking. The papers in each section are clustered according to their subject. Background information and extensive bibliographies can be found in standard texts like [71, 74, 88, 101, 112].

[1]This chapter encompasses references [451] – [510].

5.1 Transmission

The section Transmission is subdivided into the subjects Coded Modulation, Single Carrier Systems and OFDM (multi-carrier or multi-tone systems). Coded modulation [59, 83] found and finds its main applications in transmission systems, where the channel is known (due to soundings) relatively well to both the transmitter and receiver, and which need to have a high spectral efficiency, e.g., the by now classical modems (19.6 kbit/s) and other cable transmission systems such as ADSL and DVB-C. Within the Benelux, research in this particular field was mainly of an academic nature. On the other hand, communication-theoretic aspects of single carrier systems (among which we also count digital optical communication), channel estimation, equalization and synchronization issues were and are of interest to a widespread community within the Benelux, which began to see the WIC as a forum where they could present the more theoretical results. OFDM [80, 95] was studied because of its applications, first in DAB (Digital Audio Broadcast) and later in DVB-T (Terrestrial Digital Video Broadcast), where these types of modulation systems, in combination with appropriate channel coding systems, are used for efficiently transmitting digital information via a frequency-selective (broadcast) channel. Also for cable transmission, (trellis-coded) OFDM is used, but this did not lead to a WIC paper. By the end of the nineties, we saw that OFDM was also being used in WLAN systems such as the IEEE802.11a and upcoming MIMO systems.

5.1.1 Coded Modulation

In 1988, Dekker and Smit [455] first explain that a hexagonal packing of signal points achieves asymptotically a 0.58 dB gain with respect to a rectangular signal set because of the denser packing of signal points in D2. Next, they consider trellis-coded modulation (TCM) using a 4-dimensional lattice D4. As in [59], they find that doubling the number of signal points, combined with a set-partitioning approach, where the last 2 bits are encoded using a convolutional encoder, leads to a coding gain of approximately 3 dB on an AWGN channel.

In 1990, a low-complexity approach is taken by De Bot and Vinck [458], achieving basically also an asymptotic coding gain of 3dB on an AWGN channel. An example explaining their idea applied to 4-PSK works as follows. First, double the number of signal points by taking 8-PSK. Next, partition a block of m 8-PSK symbols into an even and an odd set of m 4-PSK symbols each, where the odd set differs from the even set by a rotation of $\pi/4$ for each symbol. A total of $2m$ user bits is transmitted using these m symbols, where the coding is done as follows: the even set is chosen if the parity of the $2m$ user bits is even, and otherwise the odd set. Note that the intra-set Euclidean distance in each set is $\sqrt{2}$ larger than for 4-PSK because the parity in each set is prescribed. It turns out that the Euclidean distance between the sets is at least as large as the intra-set distance for $m \geq 8$.

In 1995, De Bart and Willems [480] introduce enumerative techniques for obtaining shaping gain and to simultaneously combat intersymbol interference in a

PAM signaling scheme. As trellises are being used, this shaping technique can be combined with error correcting codes, thus providing both coding and shaping gain. The computational complexity is rather high.

In 1997, Bargh and Schalkwijk [482] present an extension of low-rate noiseless feedback coding strategies (cf. Section 4.4.5) for the BSC to AWGN channels, in order to achieve coding gain as in coded modulation. They consider sequences of transmitted QAM symbols, using a set partitioning along each of the transmitted dimensions. In traditional coded modulation, the "weakest" bits may be protected by a distance providing code, while the "strong" bits remain uncoded. Similarly, the authors propose to apply a temporal binary feedback coding strategy on the weakest bit in each dimension in order to ensure a reliable decision for these weak bits, while the remaining bits are uncoded, thus aiming at a coding gain of 6 dB. The main advantage claimed is an enormous complexity reduction compared to traditional coded modulation for a comparable performance. Of course, the existence of a virtually error-free feedback channel is required.

In 1999, Peek [486] introduces multirate block codes which may simultaneously provide spectral shaping, Hamming distance, and change of sampling frequency. The input x to the coding system is assumed to be a binary string $x_i \in \{-1, +1\}$, which is partitioned into blocks of equal size L. Each such block is multiplied by a $K \times L$ matrix A, which is $\{-1, +1\}$-nonsingular, to obtain a coded output block of L symbols, where the output alphabet depends on A. Depending on the column properties of A, one can enforce spectral nulls, e.g., at zero frequency or the Nyquist frequency. It turns out that such spectral nulls may lead to an increased minimum Hamming distance between the possible output sequences of a given block.

In 2001, Gorokhov and Van Dijk [495] consider the effect of choosing different bit labelings for a bit-interleaved (convolutionally) coded modulation scheme, while using iterative demodulation. In this setup, the combination of convolutional code, bit interleaver and (QAM or PSK) mapper is considered as a serial concatenated coding system, where the mapper acts as an inner code. The bit labeling defines the code properties of the inner code. For non-iterative decoding, a Gray mapping is known to be good as it minimizes the number of bit errors of the demapper for the SNR region of interest. For iterative decoding, however, it turns out that it is beneficial to choose the mapping such that it maximizes the minimum Euclidean distance between signal points that have labels with Hamming distance 1. In this way, the inner decoder is better capable of improving the LLRs after the first iteration, where it is mostly faced with single errors for interesting SNRs.

5.1.2 Single-Carrier Systems

In 1991, De Bot [462] presents a simple phase-recovery algorithm for the detection of M-PSK. In particular, he is interested in the detection of differentially encoded PSK (DPSK). It is known that coherent detection of DPSK asymptotically performs 3 dB better than incoherent detection (i.e., than looking only at phase

differences between two successive symbols). Let ϕ be the unknown common phase deviation of a sequence of received signal values. For each received signal $r_i = |r_i|e^{j\vartheta_i}$ with $\vartheta_i = \frac{2k_i\pi}{M} + \phi + \theta_i$, where θ_i is the phase deviation caused by the AWGN, De Bot considers r_i^{ϕ}, which is obtained from r_i by rotating it by a suitable multiple of $2\pi/M$ such that $\arg r_i^{\phi} \in (\phi - \pi/M, \phi + \pi/m)$. By simple operations using r_i^{ϕ}, he obtains estimates of ϕ that are ML-like for a series of consecutive observations i, thus leading to almost coherent detection. He also introduces an adaptive variant for time-varying channels or channels having frequency offsets.

In 1993, Van Linden, De Bot and Baggen [464] present an analytical derivation of the error rate performance of 2-DPSK using non-coherent detection on a Ricean fading channel. It is shown that, both for 2-PSK (coherent detection) and for 2-DPSK (incoherent detection), the performance on a Ricean channel resembles the performance on a Gaussian channel for low SNR, while it is more like the performance on a Rayleigh fading channel for large SNR. The transition point depends on the K-factor of the channel. An intuitive physical explanation for this phenomenon is given.

Krapels and Jansen [478] expand on previous work of Jansen in 1995. This work considers a dual signal receiver using successive interference cancellation, for simultaneous reception of two BPSK modulated co-channels. The authors investigate various alternative detection schemes, among which a joint ML detection scheme, for improving the performance in the notoriously difficult situation where the two co-channels have about equal strength at the joint receiver. They find that even joint ML detection gives little improvement over conventional successive interference cancellation for uncoded BPSK.

In 1999, Gerrits, Koppelaar, Taori, Sluijter, Baggen and Hekstra-Nowacka [485] present the Philips proposal for an adaptive multi-rate (AMR) GSM system. The AMR system comprises a set of speech and channel coders where, for a fixed given channel bit rate and depending on the channel quality, the combination of speech and channel coder giving the best speech quality is selected. A solution for a fast and seamless adaptation to a time-varying channel quality is explained and demonstrated. Although the system did not end up in the standard, several of its ideas can be found in the current GSM-AMR.

In 2000, Jansen and Slimana [490] consider the BER performance of successive interference cancellation (SIC) or "onion peeling" of a received signal being a sum of N independently modulated M-PSK signals (using the same carrier frequency) and AWGN. Assuming that the amplitude and phase of each signal is known at the receiver, the performance of a coherent SIC system is approximated analytically and simulated. Assuming that the signal amplitudes A_i are geometrically related, $A_k = \alpha^{k-1}A_1$, $k = 2, \ldots, N$, they find that such a system can work reliably for all N if α and A_1 are sufficiently large, depending on M. They also consider the extra margin in α that is required if the amplitude and phase of the received signals are not perfectly known at the receiver.

In 2001, Meijerink, Heideman and Van Etten [493] consider an optical communication system using Optical Code Division Multiple Access (OCDMA). In this set-up, the phase noise of each transmit laser (assumed to be independent between M different transmitters) is effectively used as its signature. Such a system is known to suffer from so-called beat noise, of which the power is proportional to M^2. The authors replace the delay elements traditionally used in OCDMA by a bank of filters and delay elements, both at the sender and the receiver, in such a way that the arrangement at the receiver forms a matched filter for the arrangement at the (wanted) transmitter. In this way they can make the beat noise proportional to M. The same authors consider optical communication using OCDMA again in 2002. They note that, e.g., because of temperature drift, it is difficult to accurately match the delays of the transmitter and receiver, which is required for coherent detection using BPSK. They analyze as a function of the number of users M, the performance of OOK and DPSK, which are less sensitive to drifts in phase. They find that DPSK using phase diversity detection performs almost as well as BPSK using balanced detection, while OOK has several disadvantages leading to a performance degradation with respect to that of DPSK.

In 2002, Levendovszky, Kovács and Van der Meulen [501] analyze the performance of a blind adaptive equalizer (DFMMSE) compared to an equalizer using a training set (MMSE). Both equalizers use the Robbins-Monroe stochastic approximation for adapting the equalizer coefficients, where the blind equalizer replaces the assumed known transmitted symbols (in case of the presence of a training sequence) by the hard decisions made at the output of the equalizer for the blind case. They confirm, both from computations and simulations, that the DFMMSE equalizer converges to the same performance as the MMSE equalizer, provided that the initial error rate is less than 10%.

In 2003, Janssen [509] presents a method to increase spectral efficiency in the downlink of a cellular system by simultaneously addressing multiple users with a single compound QAM signal. The technique is based on stacking a number of M-PSK modulated signals, each intended for a different user. The signal amplitudes and phases are optimized for given link gains and interference levels, in order to obtain a required symbol error probability performance at each of the user locations with minimum transmit power. The QAM compound signal and a successive cancellation detection structure are described. Comparisons with alternative signaling methods show the power gain of the presented scheme, especially in the situation where system capacity is basically interference limited. The scheme is very similar to the hierarchical modulation scheme suggested for DVB, and to the degraded broadcast channel [27].

Also in 2003, Levendovszky, Kovács, Olah, Varga and Van der Meulen [510] consider a bit detector for an ISI channel, where the bit detector consists of a FIR equalizer followed by a threshold detector. Classical equalizers use ZF or MMSE algorithms for optimizing the tap weights of the equalizer. The authors propose an algorithm that chooses the tap weights such that the resulting BER is minimized.

The algorithm considers all binary sequences of length L, where L has to be sufficiently large given the memory length of the channel and equalizer. Therefore, the algorithm is exponentially complex in L. They also propose a simplified (suboptimal) algorithm which only considers those binary sequences of length L that are most influential in determining the BER. Although they are much more complex than ZF or MMSE algorithms, the new algorithms are shown to have a better performance on two examples of two-tap channels for equalizer lengths from 2 to 10.

5.1.3 OFDM

In 1993, De Bot [470] considers (spatial) antenna diversity for OFDM systems. He first discusses various antenna combining techniques for a flat Rayleigh fading channel. Next, he observes that in the context of DVB-T, the channel is severely frequency selective, which is the reason why OFDM is used. He also observes that all of the considered wide-band combining techniques give little improvement for the frequency selective channel using OFDM. This is because different OFDM subchannels have their own (independent) fading parameters for each antenna, and hence need to be combined in a different manner. The solution for OFDM is to apply a baseband combining technique for each of the subchannels separately, giving large performance improvements for the frequency selective channel.

Also in 1993, Koppelaar [469] considers an OFDM system in the situation that the channel impulse response is larger than the guard interval, or even an OFDM system without a guard interval. In such cases, successive OFDM symbols suffer from intersymbol interference. He develops a formalism based on a vector channel (a vector corresponding to an OFDM symbol), using it to describe a (vector) DFE equalizer, the (LMS-type) algorithms that are required to compute the equalizer coefficients and to compute their performances. It turns out that to reduce the complexity, one can use band-matrices. In an example, excellent results are obtained by using only 2 tri-diagonal matrices for the OFDM DFE.

Van Linden [468] presents an attempt to analytically derive the performance of a coded OFDM system on a frequency-selective Raleigh fading channel in 1993. Because of the limited delay spread, the signal quality of different subcarriers of the OFDM system are correlated, leading to burst errors in the frequency domain. Comparing computations with simulations, Van Linden shows that a generalization of the Gilbert-Elliott burst-noise model can be used to fairly predict the performance of an interleaved algebraic code for SNRs up to 30 dB. It also turns out that an interleave depth of about twice the coherence bandwidth is required for approximating the performance on an infinitely interleaved Rayleigh fading channel. For high SNRs, the behavior of the error rates is not correctly described by the theoretical model, for which an explanation is given.

In 1994, Van de Wiel and Vandendorpe [473] consider a combination of OFDM and DS/SS, where the spreading is applied to the composite OFDM signal. Furthermore, because of spectral efficiency, the guard interval is removed, which leads

to inter symbol interference (between successive OFDM symbols) and inter channel interference (between different subcarriers). At the receiver, these interferences can be mitigated using 2-dimensional (time-frequency) equalizers. Modeling this problem as a MIMO equalization problem, the authors consider 2-D MMSE equalization leading to the LMS algorithm, and they also consider an RLS-type of equalization leading to a Kalman filter. They find that the RLS-type of equalizer performs much better than the LMS-type, in particular for large search spaces.

In 2000, Bakker and Schoute [487] describe the design and partial implementation of an experimental wireless platform that operates in the 2.4 GHz ISM band. They focus on the baseband digital signal processing module, which is a kind of software radio having a CPU board using the Linux operating system. The module is capable of performing 16 carrier OFDM demodulation (inclusive the corresponding synchronization algorithms), and error correction using a BCH code, at data rates over 1 Mbit/s. The aim of the platform is to provide the flexibility for real-time experiments using different types of baseband signal processing algorithms.

In 2002, Taubock [499] considers an equivalent baseband transmission system, where the complex additive (Gaussian) noise is not circular complex (i.e., it does not have a uniform phase distribution), which they call rotationally variant complex noise. First the author shows that, for a given noise power, the entropy is maximal if it is circular. Next, he shows that the capacity of an additive noise channel having an average input power constraint (and an average noise power) is increased if the noise is rotationally variant. However, this capacity increase can only be found and used if one considers the "pseudo-covariance" matrix of the noise. Essentially, one has to exploit the rotationally invariance of the noise by using a proper loading of the real and imaginary components of the channel ("water filling"). An application would be OFDM transmission, where the presence of non-white noise at the input of the FFT leads to rotationally variant additive noise at the subcarriers.

In 2003, Cendrillon, Rousseaux, Moonen, Van den Boogaert and Verlinden [508] consider a MIMO channel with channel state information available at the transmitter. They explain that an optimal transmitter and receiver structure can be found by considering the eigen-decomposition of the channel. The corresponding eigenvectors are used to decompose the MIMO channel into a set of parallel channels for which "water filling" can be applied and for which the capacity is easily found. Furthermore, they show that when the spread of the eigenvalues of the channel is large, a power constraint per transmitter is more detrimental to the capacity than a power constraint on the total transmitted power, as the latter leaves more freedom to the power allocation.

In 2003, Van Houtum [504] first explains the physical layer of the IEEE802.11a system. Next, he compares the performance obtained from simulations of this system on an AWGN channel with information theoretic bounds and union bounds. Finally, he gives plausible reasons for the differences (13 dB) between theoretical obtainable curves and simulated performances.

5.2 Recording

Within the Benelux, research in the area of recording is mainly related to Philips activities in the area of optical and magnetic recording [68, 97, 86]. This typically concerns the application of runlength-limited (RLL) modulation codes (cf. Section 4.3), initially both in optical and magnetic recording. In high-density magnetic recording, one has abandoned the use of (d, k)-constrained codes because of the application of PRML detection. In optical recording, (d, k)-constrained codes are still being used because the combination of removable media with simple detectors requires much greater robustness.

In 1986, Bergmans [451] studies the optimum performance of the decision feedback equalizer (DFE) for partial response (PR) channels with D-transform in the form $g(D) = (1 - D)^n(1 + D)^m$. He derives a closed-form expression for the minimum mean-square error (MMSE) at the bit detector input. From the expression we see that the MMSE depends on g_0. Since $g_0 = 1$ for all PR channels of the above mentioned form, as well as for the non-partial response channel ($g(D) = 1$), he concludes that unlike for the linear equalizer, the optimum performance of the DFE is independent of the PR channel used.

In 1987, Bergmans and Jansen [452] derive the DFE with an optimum mean-square performance in the presence of a mixture of intersymbol interference (ISI), noise and channel parameter variations. They use a transform that J. Zak introduced in 1967 in the field of quantum mechanics. The Zak transform of a continuous-time signal is the discrete Fourier transform of a version of the signal that has been sampled with a specified sampling phase. The Zak transform therefore is a natural tool to introduce the timing errors into the optimization of the DFE, and is used by the authors to find a closed-form solution. The superior performance of the DFE with an optimum resistance to uniformly distributed timing errors with respect to the conventional MMSE DFE is demonstrated by means of computer simulations.

In 1988, Schouhamer Immink [454] proposes to code digitized audio samples **s** with a rate $(n - 1)/n$ binary code, where n is a power of 2. The coding has several interesting properties. First, decoding is simple: **s** can be recovered from the binary codeword **x** by performing a Hadamard transform $\mathbf{y} = H_n\mathbf{x}$ followed by a slicer. The Hadamard transform has low complexity since H_n is a binary matrix. Second, the code is error resilient: it is constructed in such a way that the MSB of **s** is placed in the most reliable frequency band of **y**, and so on, until the LSB which is placed in the most unreliable frequency band. As a result, an increase in additive noise or a reduction of bandwidth results in a graceful degradation of the audio SNR.

In 1989, Van der Vleuten and Schouhamer Immink [456] describe the implementation and performance of a class IV $(1 - D^2)$ PR magnetic recording system. The authors build two detectors: the classical threshold detector and the maximum likelihood (ML) Viterbi detector (VD). Experiments were performed in order to assess if the VD indeed has better performance as predicted by theoretical anal-

ysis (3 dB improvement with respect to the threshold detector for AWGN). The $(1 - D^2)$ VD consists of two independent $(1 - D)$ VD used in *ping-pong*. Two experiments were performed: in the first, the system was optimally adjusted to achieve the smallest possible bit error rate (BER). The VD achieved a reduction of the BER by a factor of 2.9 with respect to the threshold detector. In the second experiment, a tracking error was introduced which increased the BER. The VD showed to be more robust than the threshold detector and reduced the BER by a factor of 9.3.

In 1990, Bergmans [459] shows that run-length-limited (RLL) codes lead to poorer pre-detection SNRs than uncoded recording for a high-density recording system with optimum mean-square DFE. More specifically, he shows that the merit factor introduced by the use of RLL codes through spectral shaping is not enough to compensate for the loss in minimum mean-square error that results from the fact that the RLL codes have a rate $R < 1$. Losses are lower bounded for a number of practical codes as well as for maxentropic (d, k) sequences.

In 1991, Bergmans [461] revisits the implications of binary modulation codes on PR channels. He considers a continuous-time transmission system with ISI and noise in which signaling occurs by means of non-overlapping rectangular pulses and binary modulation codes with rate $R = 1/N$ (N is a positive integer). He shows that the common assumption that the effect of coding on the channel is a SNR loss by a factor of R does not necessarily apply to PR channels. He computes the actual loss for most PR channels and shows that it differs from R. Furthermore, he shows that coding implies more ISI for some PR channels.

In 1993, Ribeiro [467] considers the robustness of frame synchronization for a digital magnetic tape recorder (S-DAT). Each frame starts with a sync pattern, which does not appear elsewhere in the frame. Experimental error analysis shows that the main source of synchronization errors are deletions and insertions. Burst and random errors are rarely found. His synchronization strategy uses a flying wheel, a search window, and a number of sync levels. The flying wheel memorizes the position where the next sync pattern is expected. The search window defines how many bits around the expected position are checked for the sync pattern. At sync level 0, the search window is always open. When the pattern is found the sync level jumps to 1. If the sync level $L \neq 1$ and the sync pattern is found at the expected position, the synchronizer jumps to level $L + 1$; otherwise it jumps back to $L - 1$. Simulation results show that this strategy improves robustness against false alarms (due to the search window) and that the optimum number of levels to be considered is $L = 1$.

In 1994, Siala and Kawas Kaleh [472] derive bounds on the total SNR loss due to equalization and coding. Furthermore they derive the cut-off rate for the normalized information density $\delta = \tau/T$, where $1/T$ is the user bit rate and τ represents the impulse width of the Lorentzian channel model. Both bounds are depend on m, which defines the PR channel $g_m(D) = (1-D)(1+D)^m$. They conclude that for magnetic recording, the channel requires little equalization to match the class-4 PR

channel ($m = 1$). At higher recording densities, $m = 2$ represents a better choice. From the plot of the cut-off rate, they conclude that for high SNRs, it is more interesting to work with large values of m (neglecting the non-linearities). They also conclude that for a large interval of SNRs, the system equalized to $m = 1$ outperforms the one equalized to $m = 0$. They therefore recommend to equalize to $m = 1$ since it offers a good compromise between efficiency and complexity, and presents low nonlinearity effects compared to $m > 1$.

In 2003, Riani, Bergmans, Van Beneden, Coene and Immink [505] derive the MMSE linear equalizer for a Two-Dimensional Optical Storage (Two-DOS) system. Data is stored in a hexagonal two-dimensional lattice. They also consider the design of an optimum 2D target response. They derive an expression for the BER of the 2D PRML system. By means of numerical simulations they are able to find the optimal 2D target response in the sense of minimizing the resulting BER.

5.3 Networking

In this section, we consider quality of service (QoS), routing and queuing problems, and multiple access (MA). Currently, most networking issues typically are found in the higher layers of the OSI stack [71]. On the other hand, CDMA, although it is a Multiple Access technique, is mostly considered part of the physical layer of the OSI stack.

Multi-user information theory (cf. Section 1.2) seems at this date to have little influence on actually implemented multi-terminal networks. In fact, practical networking systems use a lot of bandwidth (or capacity) in executing their algorithms for getting a network up and running, thus wasting the hard-won capacity on the "PHY" layer in protocol overhead. An example is the IEEE 802.11a system, where the actual user throughput is only about half the data rate realized on the PHY layer. A future unified approach might lead to better insights and performances of practical multi-user systems.

5.3.1 Packet Transmission

In 1993, Prasad, Jansen and Van Deursen [466] propose to enhance the throughput of slotted ALOHA by using more than one transmitting frequency (channel). Transmitted packets are distributed at random over a number of frequencies. It is assumed that a packet is received correctly if its power exceeds the total interfering power by the capture ratio. An expression for the total network throughput is derived and evaluated for different channel conditions, like uncorrelated log-normal shadowing, Rician and Rayleigh fading.

The ALOHA collision resolution scheme is based on using feedback at the end of each time slot to signal that a collision occurred. One of the several forms of feedback is multiplicity feedback, where all users are informed of the multiplicity of the collision. The capacity of the multiplicity feedback scheme is 1 (proved

by Pippenger in 1981), and can be obtained by random coding. In 1994, Ruzinko and Vanroose [474] describe a constructive protocol that has throughput arbitrarily close to 1. The protocol is based on earlier work by Györfi and Vajda using protocol sequences.

Vvedenskaya and Linnartz [479] consider a wireless network with two base stations and many mobile users transmitting packets in 1995. The users in a particular cell compete for random access, using the stack algorithm with feedback from the respective base station. Two different cases are considered: one where both base stations share the same channel and thus interference may occur, and one where both base stations use different channels and thus no interference is assumed. To avoid interference in the second situation requires two different channels, each with half the bandwidth. The first situation is modeled with a 2-state Markov channel model with a "good" (no interference) and a "bad" (interference) state. Performance of this two-cell system is analyzed. Simulations show that splitting bandwidth into two separate channels yields worse results than using one single-channel system for both base stations handling all traffic. The results suggest that it might be advantageous to allow nearby cells to use the same channel in lightly loaded wireless networks with bursty traffic.

In 2002, Levendovszky, David and Van der Meulen [502] remark that a major bottleneck in multicast communications is the number of NACKs generated by the receivers for a sender's packet that is received erroneously. If the network is flooded with these signaling packets, the throughput will decrease considerably. To circumvent this effect, a suppression mechanism of NACKS is introduced by sampling a stochastic timer. The authors design optimal stochastic timers for feedback mechanisms in multicast communication. The sender is assumed to include a timer probability density function in the message to a receiver. When sending feedbacks, the receiver samples the timer probability density function and waits accordingly. If no feedback from other nodes arrive during the waiting period, then a feedback is generated; otherwise the feedback is suppressed. The challenge is to prevent that the network is flooded with NACKs but, at the same time to ensure secure feedback to the sender. The goal of the paper is to develop optimal timer distributions that lead to specified properties of the distribution of the aggregated NACKs. Results are given in the case of uniform distances between the sender and receiver and among the receivers themselves. For nonuniform distances, the central limit theorem is used to derive the results. An optimal feedback mechanism is presented that uses a Markovian control scheme.

5.3.2 Routing and Queuing

In 1998, Boxma [483] gives a performance analysis of communication networks in a tutorial presentation. He focuses in particular on congestion problems that are not likely to disappear with the introduction of fast networking. The distributed structure of modern computer-communication networks, as well as the nature of traffic arrival processes and service request offered to those networks, pose new challenges to queuing theory. Queuing models also lead to accurate predictions of the

behavior of complex computer systems. As an example, the performance analysis of ATM networks gives rise to stochastic networks that still comprise traditional single- or multiple-server queues, but also often have complicating features like intricate priority structures. In order to take full advantage of the available network bandwidth, one should make use of statistical multiplexing effects. LAN, Internet, WAN, VBR video are examples of networks with traffic that is self-similar or has a long-range dependence. The occurrence of heavy-tailed active (and/or silent) periods of sources seems to provide the most natural explanation of long-range dependence and self-similarity in aggregated packet traffic. The changing traffic distributions forces one to consider novel non-exponential stochastic networks. An example is the investigation of the effect of non-exponential service time distributions in ordinary single-server queues.

Vvedenskaya [475] investigates in the distribution of message delay in a network with many multiple routes in 1995. As a network model, a single input node is connected to N server nodes. An arriving packet is transferred to the least busy server out of a randomly selected set of m servers. This means that the node is informed about the server queues. The probability distribution for the message delay is computed for the case where N goes to infinity, making queues independent. Simulation results are presented that suggest the existence of a stationary probability distribution of the queue length at a server.

One year later, Vvedenskaya [481] gives another example of optimal message routing in a complete graph network model with N nodes. The model forwards a message of length m from node I to node J with probability p, or it divides the message into unit-length packets and forwards the packets individually on one of the two-link connections for the path from node I to node J. Each two-link path is selected with probability $1/(N-2)$. The end-to-end delay of a message is the delivery time of its last packet. The asymptotic performance is defined as the mean end-to-end delay as N goes to infinity. For a given message length distribution and flow intensity, the optimal value of p that minimizes the mean end-to-end delay is investigated. The optimum value for p is shown to be $p = 0$ or $p = 1$, depending on system parameters. Simulations support numerical results.

In 1989, Giannakouros and Laloux [457] describe a system of multiple queues served by a single server under the exhaustive service discipline. They first analyze priority polling systems and give explicit approximations for the mean waiting times at individual stations for a given group of polling sequences. Then, they propose an elegant definition of a special group of polling sequences, which enable both performance and system optimization. In particular, they find that consecutive polls of the priority station increases its average waiting time if all normal stations are symmetric. In 1990, the same authors consider a similar problem and present an expression for the optimum relative frequency, with which different stations should be visited during a polling cycle for minimizing the average waiting time [460].

In 1998, Levendovszky, Elek and Van der Meulen [484] argue that efficient traf-

fic control is imperative in ATM networks when statistical multiplexing results in bursty aggregate traffic. ATM cell loss occurs when there is a buffer overflow. To maintain a previously negotiated level of Quality of Service (QoS), a Call Admission Control (CAC) function must be performed. They model an ATM switch as a buffer connected to a single server with deterministic service time. They seek to develop a fast algorithm that evaluates the tail of the stationary distribution of the underlying queuing system. The algorithm is expected to support real-time operation. Based on the outcome of the algorithm, user calls are admitted or rejected.

Vitale, Stassen, Colak and Pronk [496] present a new diffuse data routing concept based on multi-path signal propagation aided with adaptive beam-forming methods in 2002. The multi-path data flow incorporates redundancy and therefore increases resilience. The beam-forming method allows the multi-path channel to be used in an energy-efficient manner. To increase the energy efficiency further for low-power operation, multi-path channels are bounded within a diffusive data flow region determined by the strength of the signals. The operation of the multi-path diffuse routing algorithm is demonstrated with a simple example network topology. The multi-path diffuse routing has the potential to provide low-power and resilient communications in dense networks of low-cost devices in changing and noisy environments.

In 2001, Levendovszky, Fancsali, Vegso and Van der Meulen [492] investigate the problem of ensuring QoS in packet communication networking. That is, the selected route has to satisfy given end-to-end delay or bandwidth requirements. In this contribution a path is selected which guarantees the end-to-end QoS criteria with maximum probability. This type of selecting is called Maximum Likely Path Selection (MLPS) procedure. If link parameters are random variables, the problem becomes an NP-hard problem. The MPLS is reduced to a quadratic optimization that can be carried out by a Cellular Neural Network. As a result, the QoS requirements are met, even in the case of incomplete information.

In 2002, Bargh, Van Eijk and Salden [498] study the role and of a service broker in a Personal Service Environment (PSE), and define its functionality. The PSE has to integrate complex and distributed heterogeneous entities such as wireless and fixed networks, terminals, services users and organizations. In the PSE two planes deliver personalized mobile services: a data or service plane, and a brokerage or control plane. The data plane contains service components, governed by the brokerage plane, that store, forward, and adapt the data units and logic in mobile services. A broker is in charge of the control plane and handles all issues of mobility. If all involved agents, and hence the actors they represent (end-users, end-devices, network operators, service providers and policy makers) are pleased with the proposed settings of mobile services in the service plane, the PSE has reached an acceptable QoS level. The paper studies the role and functionality of a service broker in the PSE by investigating the basic mechanisms from a privacy perspective, and from the perspective of distributed QoS management.

5.3.3 Multiple Access

In 1993, Prasad [465] reviews CDMA systems for future universal personal communication systems. One of the important topics considered is the choice of a multiple access technique. Performance results are presented for a DS CDMA network in macro-, micro-, and pico-cellular systems that use DPSK and BPSK modulation and perfect power control, in terms of throughput and delay for fast and slow Rician fading channels. The paper further summarizes the research carried out in the Traffic Control Systems Group of TU Delft.

The papers of Rodrigues, Vandendorpe and Albuquerque [471] and Jacquemin, Rodrigues and Vandendorpe [477] combine multi-h continuous-phase modulation (CPM) with DS-CDMA in order to exploit the benefits of both principles. These benefits include low-cost receivers, interference rejection and multiple-access capabilities. As a result, a finite state description for the signal structure permits to define a periodic trellis and thus enables maximum likelihood sequence detection by means of the Viterbi Algorithm. In [471], simulation results are presented for the AWGN channel and several types of indoor channels. In [477], the authors develop an analytical model for the performance evaluation in a multipath Rayleigh fading indoor channel corrupted by multiple user interference. Previously, results were obtained for the AWGN channel. The evaluation is based on the constructed trellis and its transfer function, see also [471]. Simulations validate the model.

In 1992, Çamkarten [463] studies the design of an optimum CDMA receiver for a fixed number of fixed or mobile terminals. An accurate statistical model of a multiple-access Rayleigh fading channel and of the received signal is developed to optimize the use of the allocated channel bandwidth and to maximize the throughput of a packet radio network. Single-user coherent and partially coherent multiuser base station receiver structures are designed for uncoded BPSK packet transmissions over uncorrelated Rayleigh fading linear channels using CDMA. The corresponding exact bit error rates are evaluated, and the feasibility and robustness of the new systems developed are discussed.

In 2002, Vanhaverbeke and Moeneclay [497] investigate CDMA for the situation where the users are divided into two groups. This is called OCDMA/OCDMA (O/O). Set-1 contains as many users as the spreading factor of the CDMA system. The rest of the users are supposed to be in set-2. The perfectly synchronized users of the two orthogonal signature sets are allowed to have a different average input-energy constraint. The sum capacity of the O/O system can be made arbitrarily close to the upper bound imposed by the Gaussian Multiple-Access Channel if the set-1 users are assigned a higher power than the set-2 users. Making the power of the set-2 users higher than that of the set-1 users drastically reduces the sum capacity of the O/O system.

In 2000, Vinck [488] considers Frequency Hopping (FH) as an alternative to DS CDMA. He generalizes a binary FH scheme to M-ary symbols and calculates the maximum throughput that can be obtained. He shows that uncoordinated M-ary

Frequency Hopping gives rise to an efficiency of about 70%. The same paper discusses transmission of signatures in a multi-user environment where the set of active users is small compared to the total amount of users. Two classes of signatures are described: uniquely decipherable signatures, where the individual signatures are detected uniquely from the composite signature; and uniquely distinguishable signatures, where the presence of a particular signature can be detected uniquely. Upper and lower bounds on the length of these signatures are given.

In 2003, De Lathauwer, De Baynast, Vandewalle and De Moor [506, 507] discuss an algebraic technique for blind signal separation of constant modulus (CM) signals, received on multiple antennas. They apply this technique for estimating (blindly) a MIMO equalizer that separates a convolutive mixture of multiple CM signals. Another application is the separation of a mixture of DS-CDMA signals (also of the CM-type), received on multiple antennas. Their approach consists of using a matrix formulation of the MIMO channel model, where the CM property is used to infer that a solution for the separation problem can be found by looking for dominant singular values and a simultaneous diagonalization of a set of matrices.

Tang, Deneire and Engels [494] consider Link Adaptation (LA) to maximize the spectral efficiency in high-speed wireless networks in 2001. To approach the instantaneous channel capacities, the adaptation of the system parameters needs a general optimal LA switching scheme. Using a block-by-block adaptation mode instead of a symbol-by-symbol approach, Tang *et al.* determine channel quality thresholds obtaining a target bit error rate and spectrum efficiency. These parameters lead to the optimization problem that maximizes throughput for a given average power budget, or minimizes power under an average throughput constraint. The paper also presents numerical calculations verified by simulations. For a study case, the presented scheme could provide 18 dB gain, using adaptive modulation as an example.

CHAPTER 6

Estimation and Detection

R. Srinivasan (University of Twente)
G.H.L.M. Heideman (University of Twente)

Introduction

The early part of the last century saw the development of the mathematical theories of statistical estimation and detection. Since then, these theories have played an important role in many areas of engineering. They have laid down guiding principles for processing of signals in the areas of communications, radar, sonar, radio astronomy, seismic processing, meteorology, underwater and deep space exploration, and biomedical research. These principles have given rise to powerful algorithms in numerous applications, as evidenced by the highly reliable and sophisticated processing systems that are in use today. The applications are too many to list here. However, a common conceptual thread that links them all is the extraction of information from signals that are inherently stochastic in nature.

Bayesian reasoning and the *principle of maximum likelihood* (ML) are the classic paradigms of statistical estimation and decision theory. The development of optimal signal detection techniques and the associated processing algorithms has its roots firmly embedded in statistical decision theory and the testing of hypotheses. In digital communications, for example, optimum statistical signal processing is crucial in order to achieve, or at least to come close to achieving, the benefits of reliable information transfer as promised by the fundamental limit theorems of

[1]This chapter covers references [511] – [561].

information theory. Whereas some of the coding theorems of information theory are predicated on the assumption of maximum likelihood decoding, the ML principle and Bayesian approach have guided the development of optimum estimation and detection structures that achieve minimum probability of error performances in a variety of realistic environments. Another landmark that occurred more than half a century ago is the use of likelihoods (by Woodward, Kotelnikov, and others) in devising optimum methods for target detection in radar systems. At the other end of the applications spectrum these same principles, together with measures of information inspired by Shannon's work, have resulted in estimation and detection techniques for the processing of signals arising from biological phenomena. This has led to the development of powerful systems for the detection and diagnosis of medical anomalies in humans and animals.

Despite the existence of an immense literature on estimation and detection as distinct areas of research, their roles are usually hard to delineate in the operation of any real processing system. Nevertheless, in this chapter, we have attempted to categorize papers on the two topics in separate sections, notwithstanding the close interrelationships that exist in some cases. An attempt has also been made, as far as possible, to provide a commentary on these WIC contributions while keeping information theoretic considerations in mind. The papers have roughly been grouped into three categories: estimation, detection, and pattern recognition and classification. The few papers that fall outside this categorization but nevertheless fall within the general purview of the aim of this chapter have been treated separately at the end.

6.1 Information Theoretic Measures in Estimation

Several theoretical and application oriented papers on estimation are described in this section.

6.1.1 Time Delay Estimation

The use of entropy and mutual information measures have produced several results in estimation applications. An important application has been the analysis of electroencephalogram (EEG) signals in animal and human brains for understanding the mechanisms that cause epileptic seizures. Several results in this area, which are due to Moddemeijer, are described herein. Estimation of time delays between recordings of EEG signals from different channels is a principal approach for analysis of these signals.

Several methods are in use for time-delay estimation. The cross-correlation and mutual information methods search for the maximum correspondence of pairs of samples $(\underline{\mathrm{X}}(t), \underline{\mathrm{Y}}(t+\tau))$ as a function of the time shift τ, disregarding the dependence of subsequent sample pairs. Other well-known methods are maximum likelihood delay estimation, see Knapp and Carter [38], and those that employ autoregressive moving average (ARMA) modeling (cf. Section 6.1.2). In addition

to these, there is a large number of phase measurement methods defined in the frequency domain which use the same signal model as that in [38].

The connection between time-delay estimation and mutual information and entropies (and therefore probability density functions) is relatively easy to illustrate. The time shift τ that maximizes the mutual information between the X and Y signals is considered to be a good estimate of the delay between the two signals. As is well known, mutual information can be expressed as a function of individual and joint entropies. Estimation of these information measures therefore requires knowledge (or at least estimates) of underlying density functions. Consequently, estimation of joint density functions has been the subject of many research efforts, and several methods have been developed.

In [524], a histogram method is presented for estimating a two-dimensional continuous probability distribution, from which estimates of entropy and mutual information are obtained. Using bias correction and variance estimation, results at least as good as those reported for other estimation techniques have been obtained.

In [529], an attempt at developing a unifying concept underlying the different methods of time-delay estimation mentioned above is discussed. It resulted in the proposed maximum average log-likelihood (MALL) method. The concept is based on (a generalization of) defining an average log-likelihood function and using it as an estimate of the mean log-likelihood (MLL). Then a search is carried out for a parameter vector which maximizes this average. The maximum thus obtained, or MALL, is then considered to be an estimate of the negative entropy, where the latter is well approximated by the maximum of the MLL. This leads to an estimate for an unknown probability density function that can be used in time-delay estimation. The different biases of this procedure are related to the histogram-based estimators proposed in [524]. Jumping ahead to [556], Moddemeijer studies the probability distribution of the MALL statistic. He shows that, under certain conditions, the distribution of the MALL is a sum of independent contributions. In particular, in the asymptotic situation of a large number of observations, it is obtained as the sum of a normal distributed component and a χ^2 distributed component. These findings indeed provide theoretical justification for the assumptions made by the author in his earlier results ([552] and [554]) on AR order estimation based on hypotheses testing. The latter are described in the sequel.

An interesting further result due to Moddemeijer is an information theoretic time-delay estimator [531]. The proposed method is model-free and non-parametric, and sets up a measure of mutual information between processes to define time delay. Two stochastic processes are considered, where one process is a sample sequence shifted j samples in the future. Each process is partitioned into two parts: an infinite sample sequence representing the past and one representing the future. The past vectors of both processes are concatenated into one past vector, and the same is done for the future vectors. A mutual information measure is set up between the joint past and joint future by considering both original processes to be of length $2M$ and then allowing $M \rightarrow \infty$. It is shown that for station-

ary processes and under certain convergence conditions, this mutual information possesses a unique minimum with respect to the time shift j. This minimizing value of j is then defined as the information theoretic time delay between the two processes. The interpretation is that for this specific time shift, there exists a joint process with a minimum transport of information between the past and future. The minimum mutual information method proposed herein is discussed in comparison with other methods. It is shown for example that this method is, to an approximation, a generalization of the maximum likelihood method. For exposition of this estimator, normally distributed sequences are considered. It is demonstrated that the mutual information can be calculated by operations on the determinants of estimated covariance matrices of the processes. Numerical results are promising.

6.1.2 Autoregressive Processes

The modeling of time series data using autoregressive (AR), moving average (MA), or mixed ARMA processes has long been a powerful approach for characterizing various kinds of signals arising in practice. These are signal models which are driven, usually, by stationary uncorrelated Gaussian sequences of known or unknown variance. Such models lend themselves well to estimation activities, especially for methods based on Kalman and least- squares filtering and prediction. Multichannel ARMA processes are closely related to the state-space models arising in Kalman-Bucy filtering. This is a reason for their importance in the statistical analysis of speech, biomedical signals, weather data, and a host of other applications. We remind the reader that a scalar (single-channel) stationary ARMA process $\{x_n\}$ has a model that can be written as

$$x_n = \varepsilon_n - \sum_{i=1}^{m} a_i x_{n-i} + \sum_{i=1}^{p} b_i \varepsilon_{n-i}. \tag{6.1}$$

It is a model driven by the stationary white Gaussian noise sequence $\{\varepsilon_n\}$ with variance σ^2 and the model may include initial conditions. The parameters a_i and b_i denote the AR and MA parameters, respectively. Together with σ^2, they represent the model parameters in an application. It is usual to refer to the process as an ARMA(m, p) sequence with AR order m and MA order p.

In practice, choosing a model, determining model order, and estimating parameters within the model are real problems to be solved. The decision to model a process by ARMA, AR, or MA models usually depends on some prior information regarding the physics of the phenomenon under study. The second two estimation tasks are handled by well-known powerful methods. For example, the model order can be determined using Akaike's information criterion (AIC), final prediction error (FPE), or the minimum description length (MDL) information theoretic criterion, with parameter estimation based on ML or on least squares methods.

In [523], Liefhebber describes the *minimum information* approach for model selection and order determination. It is in fact an application of the *principle of maximum entropy*, a formalism based on statistical estimation and information theoretic

considerations that arose almost 40 years ago. The minimum information approach to model identification involves the use of a normalized power spectrum (as a spectral density function) to define a spectral entropy and then maximizing this entropy subject to a set of constraints on the correlation coefficients estimated from a finite realization of a discrete random process with continuous power spectrum obtained as observed data. Such a procedure is considered to provide a process model which is least presumptive or minimally prejudiced to the observations. The result is a parametric model for the power spectrum as a representation of the observed data. By means of spectral factorization, an equivalent time-domain model is obtained. It is finally shown that an *a priori* choice for an AR, MA, or ARMA model for the observed data is violated if the minimum information principle is imposed on the data. In the first two cases, applying the principle leads to increased-order *a posteriori* representations for the data, whereas the ARMA case leads to a nonparametric representation. The author recommends further investigations into this problem.

Using the ARMA model approach, Moddemeijer presents in [527] a slightly different method for order determination than conventional ARMA estimation. EEG signal models typically involve a large number of parameters. While the Akaike criterion is used to select the optimal model, the parameter space of the ARMA model signal is split into two parts, containing active and inactive parameters. Optimization of an appropriate cost function is then carried out with respect to the active parameters. Application of this approach using numerical examples indicates somewhat better results when compared with the conventional method.

Continuing this line of research in [554], Moddemeijer uses a distinction between the correct or true AR model and an optimal model to present an algorithm for model identification. These two models differ in the following way. If in the correct AR model a parameter is small, then it is neglected or set equal to zero in the optimal model. This is carried out for all the parameters. Such a procedure sacrifices flexibility but reduces the variance by allowing some bias to enter into the estimation. In practice, neither the AR order nor the number of negligible parameters is known *a priori*. An algorithm to estimate the configuration of significant parameters is proposed based on the ARMA estimation algorithm studied in the preceding paragraph combined with an AR order estimation procedure using a *modified information criterion* suggested by the same author. An AR model order and values of the nonzero coefficients of the model are first estimated. This model has a parameter vector consisting of independently adjustable parameters. Fixing some of these parameters to zero leads to a reduced dimension for the parameter vector. Models with different configurations (or parameter vectors) are treated as multiple hypotheses. Then the optimal configuration is selected via hypotheses testing based on an *a-priori* specified value of false alarm probability of selecting an excessively high order. The hypotheses testing aspects ([552]) are dealt with in Section 6.2.4 for papers written by Moddemeijer. Using examples, the author shows that the method performs satisfactorily.

6.1.3 Miscellany

In [512], Boel addresses the question of estimating the intensity of a Poisson process. An explicit, recursive, optimal estimator is sought. Boel shows that the solution is a stochastic linear partial differential equation with the observed Poisson process as input. In an example, it is assumed that the intensity is the square of an Ornstein-Uhlenbeck process, which is related to models for optical communications and communication networks.

In [514], Kwakernaak proposes an algorithm for the fundamentally important problem of estimating arrival times and heights of pulses of known shape in the presence of additive white noise. In the realistic situation of an unknown number of pulses, maximum likelihood procedures encounter the same difficulties as for order estimation of an unknown system. He proposes a solution for this based on Rissanen's *shortest data description* criterion (equivalent to the MDL mentioned in Section 6.1.2) and establishes consistency of the estimation algorithm. An example from seismic data processing serves to illustrate the algorithm.

The mathematical paper by Berlinet, Györfi and Van der Meulen [548] concerns the ever important problem of estimating the quality of density estimators. In particular, the Kullback-Leibler number or information divergence of two densities is used. They study a histogram-based density estimator proposed by Barron in [72] and a related distribution estimator proposed by Barron, Györfi and Van der Meulen in [87]. In the latter paper, the authors established sufficient conditions for consistency, based on information divergence, of the histogram density estimator. In the present paper ([548]), a limit law is derived for the centered information divergence of the same estimator. The centered divergence is defined as the random part of the information divergence. It is shown that a suitably normalized form of the centered information divergence is asymptotically normal with asymptotic variance less than or equal to unity. They show that the centered divergence is smaller (asymptotically) than the non-random part of the information divergence, the latter representing the expected global error in estimation. The result therefore strengthens the proposed density estimation procedure.

6.2 Detection Theory and Applications

In this section we attempt to describe the work carried out in detection. The topics dealt with are diverse, ranging from abstract concepts through typical signal detection problems in communications to biomedical applications.

6.2.1 Change Detection

Jump or change detection (also called the change-point problem) has been studied by several researchers because of its importance in many applications. A rather large body of literature exists on various aspects of this problem. Applications of jump detection are in image processing, oil exploration, underwater signal processing, radar tracking of maneuvering targets, and in many more areas. The basic

problem is one of detecting a sudden jump in a noisy signal. The size of the jump may be known or unknown. The so-called "quickest detection" problem can also be considered as a case of change detection. It is one of detecting the change in the shortest time possible.

Much is known about optimal methods for detecting jumps in random signals when the size of the jump is known. Relatively less is known about how to deal with the general case of unknown jump size. In the latter case, the problem naturally becomes one of simultaneous detection and estimation. This is the subject of the paper by Vellekoop [558]. A brief background on this problem is useful. The setting is one wherein the noise is additive and white Gaussian. It has been established that for a known jump size in the stochastic signal, the optimum detection rule produces an alarm whenever the conditional probability that a jump has occurred exceeds a certain threshold. This conditional probability can be determined in terms of a likelihood ratio. This is referred to as a Shirayev detector [44]. On the other hand, when the time of occurrence of the jump is known, the solution to the estimation problem is just the Kalman filter. The Kalman filter of course is optimal if the signal has a Gaussian distribution. The general case where both jump size and time of occurrence are unknown is much harder. In the present paper, Vellekoop proposes an algorithm which projects the nonlinear filtering Zakai equation on a statistical manifold using the Kullback-Leibler information criterion. This results in a structure which is a mixture of the Shirayev detector and the Kalman filter. The equations provide estimates of the conditional probability that a jump has occurred and size of the jump. The paper then establishes convergence properties of the filtering algorithm.

In the two papers [547] and [550], written before the one by Vellekoop discussed just above, Hupkens studies the problem of quickest detection of changes in random fields. The classical quickest detection problem, as solved by Shirayev, is defined for unidirectional stochastic processes, i.e. those that evolve in time. The solution is specified in terms of a stopping rule given by a generalized sequential probability ratio test. If the signal under study is a random field, this causality is no longer available. The change may be present at any arbitrary site of the field from which measurements are taken. Examples of such a situation arise in several spatial search applications. In his first paper [547], Hupkens develops a mathematical formulation of this problem. He demonstrates that in its full generality, the change detection problem for random fields is difficult to solve. Assuming that the prior distribution of changes is known and making some simple assumptions on a cost function, he approaches the problem from a Bayesian viewpoint in his second paper [550]. Thus a Bayes cost is set up, and a Bayes stopping strategy that minimizes the cost is the required solution. Even here it is shown that the problem cannot be solved explicitly without making further restrictions. For cases where change detection can be modeled as a simple hypotheses testing problem, the author obtains an approximate solution, and he provides numerical results which match well with the exact solutions for some simple cases.

6.2.2 Biomedical Applications

An early paper on transient detection in EEG signals is the one by Kemp [518]. A simple model describes the EEG signal as observations of a known amplitude modulated signal in additive white Gaussian noise. The author makes use of Ito's differentiation rule and a filter result of Wonham. Using a martingale representation of the amplitude modulated transient, he derives an optimal estimator-detector structure for sleep states. The relationship between the estimation and detection operations is examined.

The detection of brain state during sleep using EEG observations is the subject of the paper by Kemp and Jaspers [521]. Here, brain state is modeled as a 4-state Markov process. Using a feedback loop driven by white noise with the Markov process as a modulating signal, they adopt a generator model for the EEG signal. Then martingale theory is used to derive filtered estimates of the state. Optimal state decisions are then obtained by minimizing the average cost in the usual Bayes cost formulation employing uniform costs. It is shown that the resulting detection rule is easy to implement and that extension to a larger number of states is straightforward.

In a further attempt toward developing automated sleep stage monitoring systems, Kemp in [528] proposes a model for the occurrence of bursts of rapid eye movements (REMs). Various stages of human sleep produce different eye and body movements. REMs occur irregularly, but exclusively during waking or during a sleep stage called REM-sleep. In this paper, REM bursts are modeled as stochastic processes simulated by a Poisson counting process with a rate that depends on a binary Markov sleep state. Using this model, a stochastic differential equation driven by a martingale process results, and this describes the REM burst counting process. The likelihood ratio for the problem of testing whether or not the observations belong to a REM state is set up. The detection problem is then investigated using a Bayes optimal threshold, the latter being obtained by simplifying the Poisson rate to be one of two constant values. The rates are the reciprocal of the average sojourn times in each state (REM and non-REM), experimentally observed, and their ratio forms the test threshold. The structure of this minimum probability of error detector is derived, and the required processing is revealed. Although performance results have not been presented, the author feels that better detectors can be obtained using these methods.

More recent research on the analysis of EEG recordings is contained in the paper by Cremer and Veelenturf [549]. The problem investigated is that of spike-wave detection, an application somewhat different from the one mentioned in the preceding paragraph. Spike waves are randomly occurring waveforms sometimes present in EEG signals, and they usually mark the start of an epileptic seizure. They are difficult to characterize mathematically, as they have very different shapes and durations. Detection of such phenomena is therefore only possible by learning from examples. This is the motivation for the authors to use neural networks, in particular Kohonen's neural network. Using single-channel EEG data, they implement

and compare 6 different detection methods. Three of these use a variant of the Kohonen network. The conventional detection methods used are correlation detection, parametric, and non-parametric density estimation for determining likelihood functions. The neural based methods (combined with statistical signal detection) are non-parametric and semi-parametric density estimation, and parametric signal detection. The conclusion is that parametric signal detection combined with a neural network gives the best trade-off between the number of calculations required and the occurrence of false alarms.

6.2.3 Communications

Bergmans [525] presents a clear and concise description of the principal operations of equalization, detection, and channel coding in a digital transmission system. This is done with the motivation of comparing the three operations with respect to their respective abilities to combat intersymbol interference (ISI), noise, and channel fluctuations. A comparison is made between the signal-to-noise ratio improvements, implementation complexities, and adaptivity. Equalizer types discussed are the linear, decision feedback, and ISI cancelers using feedback and feedforward filters. As an alternative for combatting ISI, Viterbi detection is considered. Finally, he considers channel coding for protection against noise and burst errors. As is well known now, the study concludes that channel coding has the highest complexity, but also is most effective in dealing with channel variations. Based on complexity, the ISI canceller is found to be preferable to the Viterbi detector.

In [557], Levendovszky, Kovács, Jeney and Van der Meulen address the well-known problem of developing low-complexity alternatives to maximum likelihood multiuser detection (MUD) for direct sequence code division multiple access signals. In this work, the authors employ a neural network to perform blind MUD, where channel characteristics are not known and no training sequences are used. The network used is a stochastic Hopfield net. A decorrelating algorithm is suggested that performs inverse channel identification and which can combat multiuser and intersymbol interference. Mean-square convergence of the algorithm is established and performance evaluation of the system by simulation demonstrates "near optimal" MUD detection performance.

6.2.4 Autoregressive Processes

Moddemeijer and Gröneveld address a composite hypotheses testing problem in [537]. Although not directly on AR processes, the problem discussed here has a close bearing on AR order estimation, as described in a following paper. It deals with estimation of parameters of the density function of an observed random vector. The problem is posed as one of hypotheses testing wherein one probability density function is to be selected from a set of hypothesized density functions. In this paper the set is restricted to two density functions, each containing a vector of parameters that are unknown. Thus it constitutes a composite hypotheses testing problem. Consequently, a generalized likelihood ratio test is proposed as a solution. As in [529] discussed in Section 6.1.1, the average log-likelihood is used

as an estimate of the mean or expected log-likelihood and a maximization of the former is sought with respect to the unknown parameter vector. A test is derived and an improved test is suggested that compensates for the bias introduced by the approximation of the MLL.

In [552], Moddemeijer provides a solution to the problem of AR model order estimation based on composite hypotheses testing. The AIC is used as a test statistic, with the maximum of the MLL replaced by the MALL. Convergence properties of the MALL are analyzed. A modification of the test in the framework of the Neyman-Pearson criterion is suggested. Simulations carried out by the author indicate excellent match with theory.

6.2.5 Biometrics

There are two interesting papers on this subject in these proceedings: [560] and [561], which address problems in biometrics using concepts of optimal hypotheses testing. Briefly, biometric verification attempts to confirm the identity of a user based on a biometric signature data (or feature vector) provided by the user. The process typically uses stored templates obtained from a large number of users. Quite akin to signal detection, such problems are modeled well in the framework of hypotheses testing. In [560], Veldhuis, Bazen and Boersma formulate a certain multi-user verification problem. It is assumed that each of the (uncountable) multiple users can be characterized by a feature vector possessing a probability density function. A likelihood ratio test is set up for a user and its performance, in terms of a threshold and false-acceptance and false-rejection rates. By averaging over the distribution of the feature vector, an optimization problem is solved to determine optimal threshold settings. They show that the overall false-rejection rate is minimized if thresholds for all users are set to the same value. Using, as they say, an exotic example, the authors proceed to illustrate their formulation by obtaining performance curves. The example involves using signals resulting from tapped rhythms as biometric features.

In [561], Goseling, Akkermans and Baggen look at the verification problem using a somewhat different hypotheses testing formulation. A noisy version of the biometric feature of a user is available. A noisy version of another biometric feature is presented, and it has to be decided whether this new feature belongs to the first user or to a new one. Employing Gaussian distribution models for the underlying processes, the authors set up a likelihood ratio test solution. The structure of the test is examined in detail and compared with standard solutions available in the detection theory literature. A conclusion from the analysis is that the optimal decision rule is not equivalent to a situation where the reference feature can be assumed to be noiseless and adding an extra noise source to the new measurement.

6.2.6 Miscellany

The paper by Van Schuppen [515], addresses some problems in estimation and detection. It was published as a short abstract in the WIC proceedings. The topics

covered here include Markov processes, stochastic filtering, Kalman-Bucy filters, detection algorithms, false alarm probabilities, and Chernoff bounds.

Gröneveld and Kleima examine m-fold detection in a general setting in [519]. They show that each optimal detector uses a partition of the $(m-1)$-dimensional simplex of the likelihood ratios in convex regions. The proof is based on optimality criteria that do not use prior distributions and loss functions. A converse is also shown wherein every partition represents an optimal detector. It turns out that selecting an optimum detector implies always selecting a Bayes detector which in turn implies certain priors and loss function.

In [540], Vanroose addresses the well-known *NP*-complete problem of constructing optimal binary decision trees and test algorithms for the identification of objects. With a simple example, he points out the deficiencies of various heuristically proposed cost functions that have been used for designing test algorithms. The author then introduces the aspect of reliability by assigning probability distributions to the important features of the objects to be identified. This is incorporated into the cost function and an unreliability measure is set up and interpreted as a conditional entropy. A test procedure based on evaluation of such a measure is then proposed as a more reliable method.

6.3 Pattern Recognition

In this section we describe papers that deal with the subjects of classification and pattern recognition, including the use of neural networks in applications.

6.3.1 Neural Networks

The brain is the most advanced information processing machine, and therefore it should be of much interest to information theorists to know how neural networks can mimic some properties of the brain. At least there is some hope that neural networks do so. It is somewhat surprising that neural networks received so little attention in the WIC community. From the ten papers that are devoted to neural networks in the past 25 years, half of them appeared in the proceedings of 1989. The other half is distributed over the next ten years.

In 1989, a lot was known about different types of neural networks: multi-layer networks, Kohonen networks, Hopfield networks, and so on. Therefore, most of the papers are concerned with learning algorithms, i.e., Hebbian rules, stability and convergence problems, and applications of neural networks in different classification and estimation applications.

A popular learning algorithm is the back-propagation learning algorithm for multilayer feedforward networks. In order to effect learning, one has to determine the weights of the connections between neurons of different layers. To do so, we need an error function. This may be a nonlinear function of the state of the output lay-

ers. Usually the gradient descent method is used.

One problem with the back-propagation algorithm is the slow convergence in some cases. De Wilde suggests in [532] to use the Marquardt algorithm. This method is a hybrid between the gradient descent and the Gauss-Newton methods. He shows that the Marquardt algorithm can be used for online learning in a similar way as gradient descent.

The article of Piret [534] is devoted to the analysis of a class of Hopfield associative memories. It analyzes a modification of the common Hebbian rule. An application of a neural network with Hebbian learning and with transmission delays can be found in the paper of Coolen and Kuijk [533]. They show that such a system will automatically perform variant pattern recognition for a one-parameter transformation group. Such a network needs a learning phase in which static objects are presented as well as objects that continuously undergo small transformations. The system does not need any *a-priori* knowledge of the transformation group itself. It learns from the information contained in the "moving" input and creates its internal representation of the transformation.

In [536] Vandenberghe and Vandewalle also mention the central problem in the use of neural networks for pattern recognition and image and signal processing, i.e., the development of training and learning algorithms. The authors discuss a number of dynamic properties of neural networks and indicate how these considerations can lead to improvements. They realize that specifications on the behavior of neural networks can generally be written as linear equations with unknown coefficients. They suggest that a systematic approach to derive adaptive training algorithms should consist of applying classical relaxation methods of solving sets of linear inequalities. They demonstrate their ideas with a design of a neural network that should recognize characters $(0, 1, ..., 9)$ as images of 15×20 pixel size and for edge detection and noise removal.

An important property of neural network design and analysis is the robustness of the construction in the presence of possible weight errors. The paper of Levendovszky, Mommaerts and Van der Meulen [544] determines some basic properties of neural networks, i.e., the convergence speed and tolerated level of inaccuracy in the implementation of the weight matrix. This kind of network qualification is suitable for engineering design in terms of computing these properties in advance. Tolerance analysis is of particular interest for both feedforward and Hopfield neural networks. The authors compute the basic properties of the nets from the weight matrix and assess the minimum tolerated weight error. In carrying out the tolerance analysis on Hopfield nets, a statistical evaluation of the network can be performed, providing statistical bounds for the convergence speed and the tolerated level of inaccuracy.

It is often said that neural networks, specifically multilayer feedforward networks, can outperform other statistical techniques because they do not estimate parameters of the classes to be distinguished, but directly "learn" the class-separating hy-

perplanes. Multilayer feedforward networks can approximate any class-separating function arbitrarily well, provided that enough neurons are available. In [543], De Bruin raises the question: how do multilayer feedforward networks perform the mapping? Therefore he carries out an experiment with a feedforward neural net with one hidden layer, containing 5 neurons. He concludes that the neural classifier does not simply make decisions on features in the first layer which are then combined in the second layer. His conclusion is that the idea that neural network class-separating hyperplanes are built up from parts of hyperplanes defined by hidden-layer neurons may not be correct.

Most of the results on neural networks are obtained by simulations on conventional computers. However, some advantages of neural networks are lost during simulation: speed, parallelism, fault tolerance. Dedicated VLSI processors can make networks more interesting than conventional computers. In [535], Verleysen, Martin and Jespers present a VLSI architecture for a Hopfield-like fully interconnected network with capacitors as synaptic interconnections instead of resistors or current sources. However, the connection weights are restricted to some discrete values. This type of architecture offers several advantages: the accuracy that can be reached with capacitors is increased, and the number of synapses that can be connected to the same neuron is greater. Also only the relative values of the capacitors are important; their size can be reduced to very small values. An 8-neuron network with discrete components has been realized.

A specific application of a two-layer network is proposed in [546] by Levendovszky, Van der Meulen and Poszyai for estimating the tail of aggregate traffic emitted by users of ATM networks for Call Admission Control (CAC). The authors interpret CAC as a set-separation problem. A traffic configuration is admitted or not. Learning can be regarded as a search in the parameter space to find the best point which minimizes the number of lost calls. They also compare the results. The neural network yields best approximation of the original admitted region (the number of lost calls is much lower than obtained by the Chernoff bound and also much lower than that obtained by the Hoeffding inequality).

In further work on the same application, Levendovsky, Meszaros and Van der Meulen [553] propose and evaluate various neural based learning algorithms for classification. This is done with the aim of implementing fast CAC in multi-access systems. Using non-uniform costs for the two kinds of errors, the authors study directed gradient and penalty function methods for performing classification. Based on comparisons made via numerical simulation, it is concluded that penalty function classifiers have a higher learning speed at the cost of a slight decrease in performance.

6.3.2 Classification and Expert Systems

The papers that appear here are diverse. They treat classification with and without teachers, data analysis, expert systems, and so on. We have dealt with them in a chronological order.

The first paper [511], by Backer, written in Dutch, is about minimal distortion relations in classification without a teacher. In this paper, special attention is given to the treatment of the minimal distortion criterion. The special attention to this important consideration provides insights into fuzzy relations that can lead to more sophisticated models. The author shows that decomposition of fuzzy relations can lead to new essentials in hierarchical classification.

In [513], Duin provides a discussion of the need for using *a-priori* knowledge in developing a pattern recognition system. Various possibilities and difficulties are treated. Special attention is given to a comparison of statistical and structural approaches. Also, the use of fuzzy concepts is discussed in various ways; fuzzy labeling, fuzzy relations, fuzzy classification, etc. It appears that the use of a fuzzy labeled learning set puts higher demands on the teacher and the features used than a hard labeled set does. The use of a fuzzy intermediate classifier improves the possibilities of a multistage classifier.

From the same author there is the paper [517] about small sample size considerations in discriminant analysis. This paper discusses a practical rule for avoiding the peaking phenomenon in discriminant analysis. This phenomenon is: the classification error made by a discriminant function based on a finite set of learning objects increases if the number of features used for representing the objects has been increased far enough. The conclusion is that the addition of new features should be stopped before the number of learning objects per point are in the order of one.

After a period of silence, the paper [526] by Backer and Eijlers was published. It describes an attempt to develop a knowledge base (CLUSAN1) for the expert system DELFI2. It should help the user to obtain validated results of an explorative data analysis. The resulting system appears to be particularly suitable for potential users which are non-experts but familiar with the subject matter. The art of knowledge engineering and the resulting structure of the knowledge base are reviewed.

Backer, Van der Lubbe and Krijgsman treat the modeling of uncertainty and inexactness in expert systems in [530]. The problem is that it is very difficult to represent uncertainty, inexactness, and belief that may be attached to expert opinions, judgments and solutions in a rigorous mathematical way. A proposition may be uncertain or inexact or may have a degree of belief, the degree of which can be represented by probabilities, possibilities, fuzzy sets and belief functions which when used in a particular calculus will yield an inexact reasoning. This paper attempts to put the major calculi into perspective as far as their functioning and

performance related to mathematical assumptions are concerned.

The article [538] by Kleihorst and Hoeks is concerned with optical pattern recognition. The subject is identification of machine-printed characters in the electronic representation of an image, acquired by a camera or a scanner. The idea is that parts of characters can be detected with template matching. Detection of a part may be indicated by a connected cluster of pixels, called blobs. An automatic learning system constructs a list of "best" blobs, which were detected when the templates were applied to the example character images. The quality measure for blobs is based on techniques from fuzzy set theory. It involves reliability, support, and fuzziness (fuzzy entropy) of the detection blobs and the discriminative power of the template. For a limited set of input characters, the proposed system can recognize characters at high speed with a false recognition rate of 3.5%. An improvement may be reached with a larger description, though such modifications may cause some missed characters.

Design principles and some features of EDAPLUS (Exploratory Data Analysis) are presented by Backer in [539]. Exploratory data analysis is characterized by multiple statistical testing, validations, and complex reasoning. Quite a number of statistical procedures have to be applied in order to understand the peculiarities of the data at hand. Such a reasoning process is associated with knowledge-based systems. There is a need for more intelligence in statistical software packages. As such, EDAPLUS is designed as a knowledge-based software package for cluster analysis. The author describes the decision network in terms of clustering tendency based upon low-level, intermediate-level, and high-level rules. An application in the domain of signal analysis is included.

Hierarchical cluster analysis is a widely used method to represent a finite number of objects in the form of a tree or dendrogram. The paper by Lankhorst and Moddemeijer [542], presents a novel approach to the automatic categorization of words from raw data. The authors count occurrences of word pairs in text and use a hierarchical clustering technique on the frequency data to obtain a classification of words into linguistic categories. The loss of mutual information, caused by combining two clusters in a single new cluster, is used as a criterion in the clustering process. Using this method, words are not only classified on the basis of their syntactic categories, but also with respect to aspects that are related to their meaning. They suggest that this method can form the basis of a system that uses a much finer categorization of words than is feasible using traditional grammar-based approaches.

Another contribution to pattern classification is treated in [545] by Vanroose, Van Gool and Oosterlinck. In this paper, the authors propose BUCA (a bottom up classification algorithm) as a general-purpose supervised learning algorithm based on the average splitting entropy concept. The classification tree is built starting from the leaves, as opposed to other classical methods. BUCA can be applied to any training set which includes class information. BUCA differs from top-down classification systems in two aspects. It recursively joins two training data subsets into

a new set in a way similar to the well-known Huffman source coding algorithm, maximizing the joint dissimilarity of the two subsets with respect to the rest of the training set. Dissimilarity of the two classes is defined to be the average splitting entropy, i.e., the average log-probability of a feature value belonging to one subclass, which will be classified into another subclass erroneously. It sometimes outperforms classical classifiers, both in terms of correct classification rate and in execution time.

6.4 Miscellaneous Topics

The paper [516] of Veelenturf belongs to the subject of automata theory. He considers the adaptive identification of sequential machines. It is known that an n-state discrete-time sequential machine can be identified if the set of all input-output sequences of length $2n-2$ is given. Algorithms that do this are complex. Performing identification using a smaller set is difficult. The author suggests an adaptive procedure which constructs a sequential machine stage by stage. The steps are described in detail and the algorithm is shown to be of reduced complexity.

In [520], written in Dutch, Schripsema and Veelenturf study Petri-networks as a representation of learning behavior. They conclude that Petri networks can be used to simulate learning behavior, but are inefficient for specific applications of learning behavior.

In [551], Slump describes applications in optics from an information theoretic viewpoint, mainly using Gabor's interpretation of information as degrees of freedom of phenomena. With optical image formation as a starting point, it is shown how the wave function characterizing an object can be expanded in terms of the Whittaker-Shannon interpolation (sampling) equation. This is used to determine the number of degrees of freedom. Then radiological imaging is described. For the case where light levels are low, the author shows that noise analysis and detection theory are required. The covariance function of the stochastic image is computed for the example of an X-ray imaging detector. The author states that a spatial information capacity can be defined and computed for such applications.

In [555], Van Someren, Wessels and Reinders tackle the important problem of information extraction from genetic data consisting of high-dimensional signal sets measured at relatively few time points. This task, of inferring gene interactions, is approached by modeling them with a linear genetic network. Advantages of the simplified model adopted include the use of a few network parameters that are easily interpretable, and the possibility of applying constraints without introducing errors in fitting the measured data. Their approach is based on empirical observations that show that genetic networks tend to be sparsely connected. The authors provide a description of the general linear model followed by a procedure to optimize it from the point of view of alleviating the dimensionality problem. In experiments conducted on real data sets, they find computational complexity to be a major obstacle. A clustering procedure is suggested to partially address this is-

sue. In related work, Reinders [559] is concerned with the analysis of genetic data that comprise DNA microarrays. By studying gene expressions (in the enormous amounts of data produced by numerous genome projects worldwide), one can gain a better understanding of gene function, regulation, and interaction in fundamental biological phenomena. The article describes various computational tools used in microarray analysis.

CHAPTER 7

Signal Processing and Restoration

J. Biemond (TU Delft)
C.H. Slump (University of Twente)

Introduction

Digital Signal Processing (DSP) concerns the theoretical and practical aspects of representing information-bearing signals in digital form and the use of processors or special purpose hardware to extract that information or to transform the signals in useful ways. Areas where digital signal processing has made significant impact include telecommunications, man-machine communications, computer engineering, multimedia applications, medical technology, radar and sonar, seismic data analysis, and remote sensing, to name a few.

Boaz Porat starts his book "A Course In Digital Signal Processing" (Wiley 1997), by quoting Thomas P. Barnwell (1974):

> Digital Signal Processing: *That discipline which has allowed us to replace a circuit previously composed of a capacitor and a resistor with two anti-aliasing filters, an A-to-D and a D-to-A converter, and a general purpose computer (or array processor) so long as the signal we are interested in does not vary too quickly.*

[1] This chapter covers references [562] – [664].

This "definition" relates signals and systems with (digital) signal processing, as illustrated in Figure 7.1.

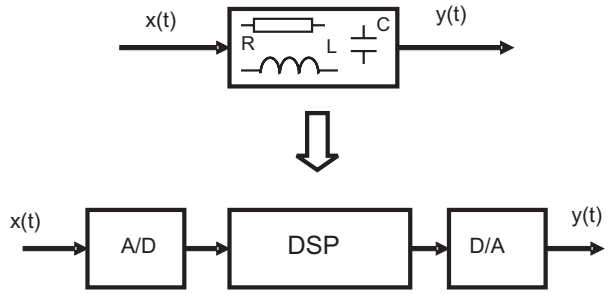


Figure 7.1: *The relation of the signals $x(t)$ and $y(t)$ with circuits and systems.*

A signal refers to a physical quantity that varies with time, frequency, space or any other independent variable or variables. Examples are electromagnetic waves such as the visible light (reflections) in human vision, the sound waves we perceive in a music hall, the electrocardiogram (ECG) that shows the differences in electric potential on the human body due to the activity of the heart. We assume that a sensor has transformed the signal into the electrical domain; the situation shown in the upper half of Figure 7.1. Digital signal processing, shown in the lower part of Figure 7.1, has developed rapidly over the past 30 years.

Figure 7.1 also applies to two-dimensional signals, usually called images, and image sequences. In general, light is reflected by a scene and picked up by a sensor at the input of an imaging system. The imaging system converts the sensor signal into a digital matrix of picture elements ready for display, storage or further processing steps. Digital image processing is based upon two main application areas: improvement of pictorial information for human interpretation; and processing of image data for storage and transmission. In this chapter we highlight key developments in the broad area of one- and multi-dimensional signal processing in the past 25 years, and summarize the contributions of Information Theory researchers in the Benelux. We have chosen the following subdivision.

- *Signal Processing*: We characterize the contributions in this category based upon the consideration that signals are carriers of information that are used to communicate between people, between people and machines, and are used to sense the environment. We start with audio and speech processing after which we pay attention to sampling, biomedical signals and signal analysis before we turn via radar and sonar to signal processing for telecommunications. Finally, we address signal processing hardware.

- *Image Restoration* (Image Processing and Analysis): We have chosen for the image restoration paradigm to classify and describe the papers in this

area as this takes into account the overall impact of the different papers. We first concentrate on still image restoration, followed by image sequence restoration and the notion of object motion. Next, the focus will be on the consecutive analysis and interpretation steps within the image processing chain.

7.1 Signal Processing

This section addresses the (one-dimensional) digital signal processing topics as presented in the past decades at the WIC Symposia. Signals are carriers of information that are used to communicate between people, and between people and machines. Signals are also used to sense the macroscopic world around us, by radar (electromagnetic waves) and sonar (acoustical waves), and the microscopic world by optical and electron optical techniques. The papers of the symposia contribute to these various aspects of signals in the digital signal-processing field of research. We have grouped the papers in the following sections: (1) Audio and Speech Processing, (2) Sampling, (3) Biomedical Signals and Applications, (4) Signal Analysis and Modeling, Parameter Estimation, (5) Radar and Sonar, (6) Signal Processing for Communications, (7) Signal Processing Hardware and a final (8) Miscellaneous.

We remark that signal processing research is not exclusively algorithm oriented. New algorithms often result from the need to improve efficiency and to reduce the cost of implementation. But also computation-intensive algorithms stimulate new design and implementation techniques. Over the last decades, the growth of signal processing in consumer products, computing, communications and networking has been tremendous. This spectacular expansion in applications and capabilities is due to the exponential development in the microelectronics and semiconductor industry, well-known as Moore's law. Over the last three decades, the integration density of integrated circuits has been increased at a rate of 50% every year. At the same time, the clock frequency of circuits has doubled every three years, resulting in more performance and computational power. Signal processing applications implemented on a rack full of printed circuit boards in the early years of the symposia are now implemented in a single chip. Several papers at the WIC Symposia pay attention to the implementation aspects of signal processing algorithms.

7.1.1 Audio and Speech Processing

Speech is one of the most important forms of human communication; it therefore has attracted much attention in the last decades. Speech coding has made voice communication (viz. mobile phones) and storage effective and efficient. Together with speech synthesis technology, speech recognition has created interactive information systems that, if faster processing power becomes available, may evolve to transparent human-computer interaction.

The human speech production system is illustrated in Figure 7.2. The main vo-

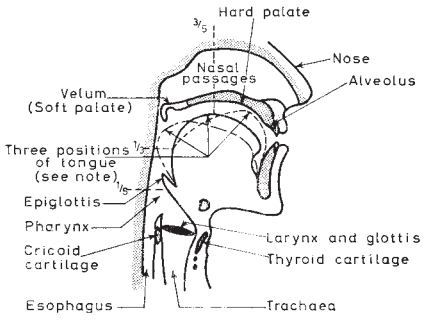


Figure 7.2: *Elements influencing the vocal tract.*

cal tract extends from the larynx to the lips, by lowering the velum for certain sounds, the nasal cavity is coupled to the main vocal tract. During speech production, the passage from pharynx to esophagus is closed. The width of the larynx is variable and is called the glottis. Speech sounds are classified into three classes corresponding with the excitation. Voiced sounds are produced by the vocal cords which vibrate open and closed, thus interrupting the flow of air forced through the glottis in a rapid sequence of pulses. The pulse rate is also known as the pitch frequency. Unvoiced sounds result from noise like turbulence excitations produced with open glottis if air is forced at high velocities through a constriction in the vocal tract. Plosive sounds result from releasing the air pressure built up in a complete closure in the vocal tract.

Speech generation systems model the human speech production, see for example Figure 7.3, where the vocal tract is transformed into a mechanical model of an acoustical speech production system. The system in Figure 7.3 is transformed to the signal processing domain in Figure 7.4. This scheme is the basis for several analysis-by-synthesis type of speech coders. The widely applied speech coder in mobile telephony (GSM) is also of this type.

Speech Recognition

In [628], Hermus, Wambacq and Van Compernolle consider the degradation of speaker recognition due to the presence of noise, e.g. disturbing sounds from the surroundings. The paper proposes a method based on Singular Value Decomposition (SVD) to improve the robustness against the influence of additional noise at moderate SNR ratios. The noise reduction is obtained by suppressing low-energy singular value components in the Hankel matrix, while the formant structure of the speech is preserved.

Vanroose [663] considers the problem of improving automatic speech recognition from audio fragments containing background music. The problem is put into the

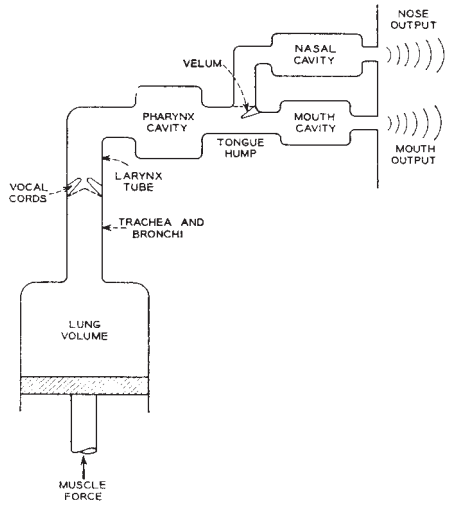


Figure 7.3: *The vocal tract as acoustical speech production system.*

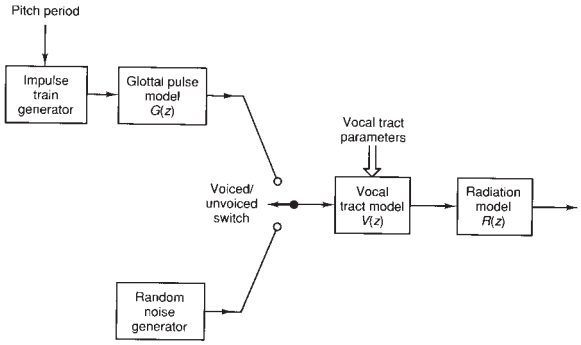


Figure 7.4: *Signal processing model of speech production.*

framework of linear source separation, where the music component is subtracted from the signal, thereby aiming at better speech recognition, but not necessarily at a better subjective audio quality. A pattern classifier depends on the input features that have to be both highly discriminative and compact.

In [630], Demuynck and Wambacq describe an alternative to the commonly used Linear Discrimant Analysis (LDA) for finding linear transformations that map large feature vectors onto smaller ones while maintaining most of the discriminative power. The new proposed set of methods is based upon the mutual information error or the minimal classification error. The new methods, called Minimal Mutual Information (MMI) and Minimum Classification Error (MCE), take all information on the individual class distributions into account while searching in an

optimal subspace. An example of classification of a speech segment is presented.

Speech Coding, Modeling with Sinusoids

Traditionally, speech and audio coding have been two separated research areas. Vos and Heusdens [636] present a method for coding both speech and audio signals. Speech coders obtain a low bit rate by heavily exploiting *a priori* knowledge of the speech signal. This does not apply to audio. In video coding applications such as MPEG-4, there is a need for coding of speech signals within the context of audio coding. Both speech and audio signals are modeled with complex exponentials. The presented method can efficiently represent the "attacks" in the audio signal. Jensen, Heusdens and Veenman [650] propose an algorithm for encoding the model parameters for sinusoidal coding of audio and speech signals. Sets of amplitudes, frequencies and phases of the sinusoidal components are estimated for consecutive signal segments. The differential encoding with respect to values of components in the previous segment achieves a bit rate reduction up to 39% compared to non-differential encoding schemes.

In [651], Hermus, Verhelst and Warnbacq present a scheme for perceptual speech and audio coding. The Total Least Squares (TLS) approach is a flexible tool for modeling short signal segments approximately by a finite sum of damped sinusoids. Close fits with transitional segments in natural speech are obtained. The paper proposes dividing the speech signal into a number of subband signals, which turns the TLS approach in a feasible optimization problem.

Burazerovic, Gerrits, Taori and Ritzerfeld [643] report on the use of time-scale modification (TSM) for speech coding. The time scale of a speech signal is compressed prior to coding, which leads to a lower bit rate representation. After decoding, the original time scale is restored. The paper compares the Synchronous OverLap Add (SOLA) method including a special inverse time scaling of unvoiced segments with other speech coders.

Speech Synthesis

Vanroose [644] discusses part-of-speech tagging in the field of natural language processing. This technique assigns to each word in a sentence its morphosyntactic category. Annotating a text with part-of-speech tags is a standard low-level text-preprocessing step. The new approach in the paper is the modeling of the language as an information source followed by a channel. The Shannon capacity is a bound for the percentage of correct tagging by any tagging algorithm.

Speech Transmission

In [627], Slump presents a signal recovery approach to the problem of speech transmission over a non-ideal channel. The *a priori* knowledge about the speech generation process that is usually well applied in the speech coding area is used in the receiver's signal detection. In this way the transmission capacity is exploited effectively. In [629], Slump, De Bont, Mertens and Verwey address the speech

quality to be expected from the new TErrestrial Trunked RAdio (TETRA) digital mobile communication system for public order and safety. The TETRA standard was developed for this application field by the European Telecommunication Standards Institute (ETSI). With the Perceptive Speech Quality Measure (PSQM), the resulting speech quality is evaluated by simulation of different channel conditions.

7.1.2 Sampling

Digital signal processing starts with the conversion of the signal from the continuous-amplitude continuous-time domain into the discrete-amplitude discrete-time domain. This process is called sampling. The roots of sampling theory are in the work of Shannon and that of mathematicians before him. The design and realization of the devices doing the conversions, the Analog-to-Digital converters, is a research field of its own in solid-state circuits and systems. The progress in this field of microelectronics does not follow Moore's law.

Wiersma [579] argues that although the classical sampling theorem for bandlimited signals is well-known, it is often of no practical importance. The paper defines bandwidth and time-duration based upon the second moments of the signal and the spectral power. These definitions allow the definition of a class of finite energy signals that have both a finite time-duration and a finite bandwidth. The paper shows that the total number of degrees of freedom of a signal is bounded by the product of time-duration and bandwidth.

In [607], Moddemeijer argues that sampling is an application of linear algebra. The paper shows that sampling and reconstruction of signals with a minimum mean square error corresponds to the computation of inner products of basic functions with the time signal to be sampled, followed by an orthogonalization step and reconstruction by a coefficient weighted sum of basic functions. The linear algebra approach leads to alternative sampling procedures with basic functions other than sinc-functions.

Van der Laan [614] extends this approach. A geometrical representation of sampling is generalized to an approximation in a subspace of the signal space. Two sampling operators in spline spaces are presented and properties are discussed. The abstract [649] points out the use of bandpass sampling in telecommunications, e.g., for software radio. Bandpass sampling holds the promise of a much lower sampling rate than twice the maximum frequency, which is useful for mobile terminals as it implies an AD converter with lower power consumption.

Sampling theory also applies to the conversion of optical scenes into video signals. This conversion is implemented using appropriate color filters. In [616], Hoeksema describes two methods for selecting a 3×3 matrix to be used in color video imaging for correcting the color signal for non-ideal transmission filters in the video camera.

7.1.3 Biomedical Signals and Applications

Biomedical signals and applications have always inspired the creativity of the algorithm researcher. Sometimes the bio-system itself is copied in part; examples are neural networks and the human visual system. In [567], Heideman proposes to use a model of the human visual system for image coding purposes. The method comprises an image characterization, a model of the human observer, and a coding part. Heideman and Veldhuis [568] continue this line of research by using a model of the visual cortex in order to be able to decide what image details are not relevant and can therefore be discarded in image coding. Circular bounded functions are described for this purpose as a superposition of orthogonal basics functions, see also Veldhuis and Heideman [572].

Rompelman [570] studies the behavior over time of a biological signal source, namely the human heart. From the Electro Cardio Gram (ECG) signal, the heart rate is determined and the variability is analyzed. This research result from 1982 was applied much later in determining the real-time digital filtering requirements for baseline drift removal in ECG monitoring during physical exercise. In [593], Rompelman indicates that many processes in nature can be described as a series of repeatedly occurring identical events, which leads to a characterization by a stochastic point process. This enables the use of simple algorithms for filtering, spectral analysis and correlation analysis.

Mars [573] discusses the biological signal source of epilepsy: the almost simultaneous firing of neurons in the brain. In some cases there appears to be a “focus location” in the brain from where the firing starts. In order to locate the focus, Mars determines the time delays between simultaneously recorded Electro Encephalo Graphic (EEG) signals during epileptic seizures. Cross-correlation between the EEG signals is an often-used technique. However, its success is limited due to non-linearities. The paper presents a new method based upon mutual information, which results in more robust estimates of delay time and thus of the location of the focus, which is highly relevant for cases where the focus must be removed surgically.

In [574], Rompelman analyzes repetitively occurring waveforms such as neural spike trains and electrocardiographic signals. The shape of the signals is often very similar; the information contained in the signal is represented by the Waveform Occurrence Time (WOT). The analysis requires two steps: detection of the waveform, and estimation of the WOT, respectively. Because the signal needs to be sampled in order to enable digital signal processing, also the maximum signal frequency present in the waveform is of importance. The paper presents a method for obtaining this frequency, exploiting also phase spectrum information.

In [580], Koenderink discusses the human visual system. The analysis of the human visual system by Fourier-based concepts from optical systems theory such as the Modulation Transfer Function (MTF) by Schade in the 1950s has led to the television system. Koenderink also points out that in case of a “lazy eye”, the im-

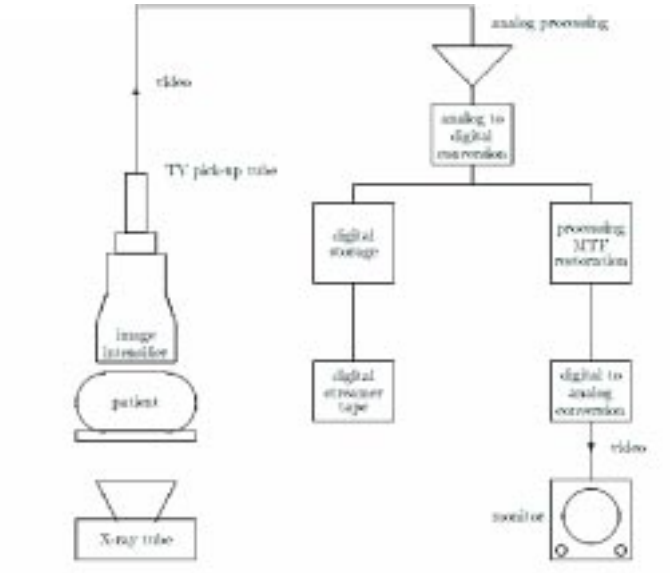


Figure 7.5: *Block diagram of a digital diagnostic X-ray system in the early nineties.*

age on the retina is about the same for both eyes, but that the visual acuity is also determined by the way the neurons in the brain detect the simultaneous order in the visual stimulus.

Slump [604] describes a way to avoid subtraction artifacts in Digital Subtraction Angiography (DSA). DSA is a less invasive imaging technique of blood vessels by intravenous injection of contrast material and subsequent X-ray exposures, see Figure 7.5. By subtracting images from a pre-contrast mask image, the blood vessels are visualized. Subtraction artifacts deteriorate the image quality, however, which are due to the periodic motion of the arteries by the contraction of the heart and the propagation of the blood pressure. By triggering the X-ray exposures with respect to the ECG signal, the motion artifacts are reduced. In cardiology, coronary angiography is the *de facto* standard imaging modality used to visualize the condition of blood vessels. Usually the percentage area and percentage diameter of a stenosed vessel segment are determined. This measure does not provide information about the blood flow. In [615], Lubbers, Slump and Storm report about an approach in which the relative flow distribution between the two main branches in the left coronary artery are determined from acquired digital angiograms. This method may reveal the functional clinical relevance of a stenosis in one of the branches.

Lerouge and Van Huffel [631] discuss the preoperative discrimination between benign and malignant ovarian tumors. A reliable classifier assists clinicians in selecting patients for whom minimally invasive surgery or conservative manage-

ment suffices versus referral to oncology. The paper reports a first approach to use a neural network for this classification task. To validate the performance of the different classifiers, the Receiver Operating Characteristic (ROC) test criterion is used. The ROC curve plots the percentage of correctly classified malignant tumors (sensitivity) versus the percentage of false positives (specificity). Different neural-network-based classifiers are compared by computing the area under the ROC curve.

7.1.4 Signal Analysis and Modeling, Parameter Estimation

Signal analysis, the study of random signals and power spectrum estimation, is one of the core competencies of signal processing. The level of mathematics necessary for a rigorous characterization of stochastic processes is high, but tools have become available which enable relatively easy implementation of various algorithms. Simulation greatly helps in understanding the algorithms and their performance. In practical situations such as radar tracking in air traffic control, linear filtering may just not be sufficient. Non-linear filtering may give a significant improvement.

Many of the widely applied signal processing algorithms are based on second order statistics. Among these most well-known algorithms we find Principal Component Analysis (PCA) and Independent Component Analysis (ICA). In [657] an alternative approach to the Canonical Component Analysis (CCA) is proposed, based on the observation that CCA does not work for certain classes of data. As a side result an efficient algorithm is proposed for computing ICA under certain circumstances.

In [565], Blom discusses the implementation of the representation result of a differential equation for the conditional density of the state of a Markov process subject to additive white Gaussian noise. Direct application to finite-state Markov processes is possible. In many optimization problems, e.g. parameter estimation, one has to find the global extremum of a function of several variables. In most cases, iterative techniques must be used, and the problem becomes one of finding the global optimum instead of a local extremum near the starting point of the iterative search procedure. Slump, Hoenders and Ferwerda [575] discuss a method that provides the total number of extrema in the area of interest. This information is useful for tracking the location of the extrema to ensure that the true global optimum was found.

In [578], Boekee and Van Helden discuss the relation between distance and distortion measures. Statistical distance measures are widely applied in e.g. pattern recognition. In speech recognition on the other hand, distortion measures are used based upon power spectral densities. Paper [578] investigates the relation between the two types of measures.

Veldhuis, Jansen and Vries [584] discuss algorithms for the restoration of unknown samples embedded in a neighborhood of known samples. For signals that can be modeled as autoregressive processes, an adaptive iterative solution to the restoration problem is given that produces good and stable results.

Chen and Vandewalle [602] present a comparative study of the adaptive IIR filter with the adaptive FIR filter. The adaptive IIR filter is composed of two tapped delay lines; one is fed with the input of the filter and the other is the feedback path fed from the output or the residual error signal. A comparison is made based on convergence properties and applications in adaptive noise cancellation, adaptive line enhancement and spectral estimation. The IIR filter outperforms the FIR filter. However, it is potentially unstable. The contribution [603] of Callaerts and Vandewalle shows that the Singular Value Decomposition (SVD) provides a unifying framework and a numerically robust approach for use in signal separation problems. Two applications are presented; extraction of the foetal electrocardiogram (fECG) from ECG recordings of the mother, and signal-to-noise enhancement in speech disturbed by noise.

Beck [610] presents an algorithm that estimates the parameters of multiple sinusoids from a finite number of noisy discrete-time observations. The method is essentially a statistically efficient variant of Prony's method. The linear prediction equations are solved using Total Least Squares. The Toeplitz structure of the resulting error matrix leads to a computationally efficient procedure.

Van der Wurf describes in [611] the generation of synchronous random pulse trains by linear pulse modulation. The input signal of linear pulse modulation is a discrete-time signal and the output is continuous in time. Van der Wurf calls this type of system a hybrid system. The paper describes the analysis of these systems; expressions are given for impulse response, frequency response, convolution, autocorrelation and power spectral density.

Albu and Fagan [656] treat the problem of unwanted echoes produced by a microphone if it picks up the reflections of the speech via different delay paths. If the reverberation time is in the order of a few hundred milliseconds, an adaptive Echo Cancellation Filter (ECF) with a long impulse response is required. The well-known normalized LMS (NLMS) algorithm has been used for this purpose, but the convergence is slow. The affine projection algorithm is a generalization of the NLMS algorithm. However, its implementation as the Fast Recursive Least Squares algorithm is not numerically stable. In this paper, the implementation of several Fast Affine Projection (FAP) algorithms using the Logarithmic Number System (LNS) is investigated. Successive Over-Relaxation (SORFAP) proves to be marginally more complex than NLMS but a better alternative in different voice applications.

De Lathauwer, De Moor and Vandewalle [660] consider the problem of signal separation. For example, when a microphone picks up the signals from several sources, the problem is to find the source signal. Many source separation algorithms are based on an approximate diagonalization by means of a simultaneous unitary similarity transformation. In this paper, the authors derive a new algorithm for the approximate diagonalization of a set of matrices by means of a simultaneous non-unitary congruence transformation.

In [658], De Lathauwer, Fevotte, De Moor and Vandewalle generalize the well-known SOBI technique (Second Order Blind Identification) for blind source separation to convoluted mixtures. The algorithm is based upon joint block diagonalization of a set of covariance matrices by means of a unitary similarity Jacobi transformation. In [661], De Lathauwer, De Moor and Vandewalle link the blind identification of a MIMO FIR filter to the calculation of the Canonical Decomposition (CANDECOMP) in multi-linear algebra. This allows blind identification of systems that have more inputs than outputs.

7.1.5 Radar and Sonar

Radar

Not many papers have been presented in the WIC Symposia on the topic of radar signal processing. One likely reason is that the radio frequencies applied for radar are very high and that the digital processing technology was just not fast enough in the past in order to process the signals. In 1980, Van der Spek [563] presented the design of a radar system based upon a phased-array. Conventional radar systems employ a rotating antenna. Therefore it is not possible to allocate radar energy and observation time in a flexible way. Phased-array antennas overcome this problem. The pencil beam of the phased-array antenna can be positioned very fast in any desired direction within a field of view. The surveillance application with the new phased-array-based radar concept is discussed. The 1986 abstract [589] by Van der Spek introduces the Inverse Synthetic Aperture Radar (ISAR). With this technique, an aircraft is tracked by radar with a coherent pencil beam and echoes are obtained during several seconds at a sufficiently high repetition rate, as a one-dimensional "image" from the object of interest.

Sonar

Digital signal processing techniques have developed more rapidly for active and passive sonar systems in comparison with radar systems because of the lower sampling frequencies. Time-delay estimation in the observation process has been an area of significant practical importance in underwater acoustics. The understanding of how the biosonar of dolphins works has been the research topic of Kamminga and co-researchers. In [595], Braadbaart and Kamminga compare four current definitions of time resolution for biosonar, the echolocation waveforms of two bottlenose dolphins, Tursiops truncates. In the abstract [597], Kamminga considers the structural information theory of biosonar, the Odontocete echolocation signal of dolphins. The "uncertainty product" of the time duration and bandwidth of the echolocation waveforms of these dolphins is low; therefore, the analytic Gabor elementary signal description seems appropriate.

In [609], Kamminga describes an echolocation experiment carried out with a captive born Tursiops truncates to obtain the threshold figure for time difference perception in echo structures. The blindfolded animal was able to differentiate to almost 8 mm in range, which corresponds to a time difference of 10.6 μs. The

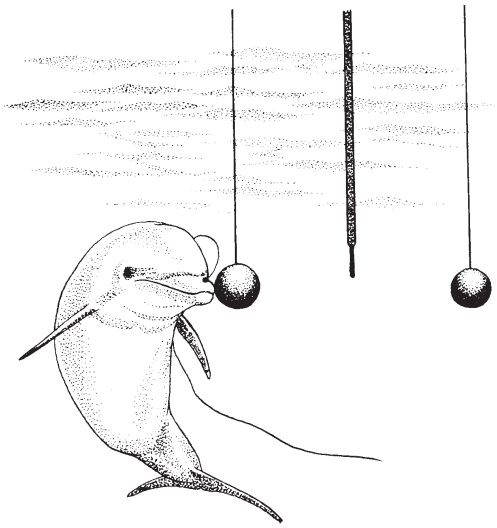


Figure 7.6: *Dolphin Doris approaches the echolocation targets.*

theoretical definition of the time resolution of sonar clicks corresponds to these experimental results. Decreasing the range differences to 4 mm and ultimately 2 mm lowered the success rate to 50%. This seems to suggest that the animal is capable of breaking through a theoretical resolution bound derived from a Gaussian wave shape.

Cohen Stuart [618] describes a method to investigate the similarity and discrepancy of waveforms of dolphin echolocation signals from dolphins that belong to the same species but have different dominant frequencies. A re-sampling technique is applied in order to normalize the dominant frequency and to get the same number of data points per cycle in both signals to be compared. This improves the correlation and shows that the waveforms of two different animals are similar. In [619], Cohen Stuart and Kamminga model the polycyclic sonar waveform of the Phocoena phocoena using Gabor's elementary signal. They show that the sonar click consists of a primary click and the first reverberation; both contributions are described with Gabor's model.

The paper [622] by Kamminga and De Bruin deals with the additive entropy measure of uncertainty applied to echolocation signals of dolphins. A modified minimum principle for the sum of entropies in the time domain and the frequency domain is applied to the analytic form of limited time-duration-frequency bandwidth signals. The minimum is obtained for a Gaussian pair under the constraint of limited variance of the signal. The presented formulation reveals the additive nature of entropy, other than Gabor's uncertainty relation, which is based on the variances in both time and frequency. An application is presented for echolocation

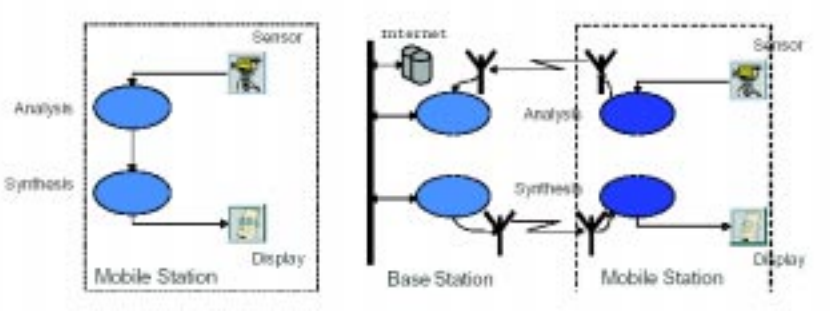


Figure 7.7: *Analysis and synthesis of information in UbiCom [633].*

signals of dolphins in a perceptual context to establish whether there would be a preference for the time domain or the frequency domain or that there is equilibrium between the two.

In [642], De Bruin and Kamminga minimize the uncertainty product of composite signals, in particular signals composed of pure signal waveforms, to which a time-delayed replica has been added. If the pure signal is Gabor's elementary wave packet, then the uncertainty product shows local maxima and minima as a function of the time delay. This effect is of importance for the interpretation of the reverberation phenomenon in the echolocation signals of dolphins.

7.1.6 Signal Processing for Communications

In [633], Lagendijk describes the TU Delft research program Ubiquitous Communications (UbiCom). UbiCom was a multidisciplinary research program at Delft University of Technology. The program aimed to develop wearable systems for mobile multimedia communications, i.e., (i) visual information processing such as context-aware augmented reality in real time, (ii) high bit-rate communication at 17 GHz, (iii) architecture and design optimization. The paper discusses the views on UbiCom, and motivates the research objectives of the program: low power, negotiated quality of service, system level approach (see Figure 7.7).

Communication systems are hard to characterize analytically with respect to performance evaluation. Fast stochastic simulation methods based upon Importance Sampling (IS) have been successfully applied to a large number of situations that involve non-linearities, memory effects and non-Gaussian stochastic processes. Examples are coded modulation systems with Viterbi decoding, CDMA systems with fading channels.

Srinivasan [655] describes the concepts of fast simulation by IS applied on communication systems and signal processing detection. The paper provides an introduction to adaptive IS theory and techniques, describing various biasing schemes

that can be used to estimate probabilities of rare events. An IS technique for estimating density functions of sums of random variables is also provided. The article goes on to describe various applications. The two main applications presented are the estimation of probabilities of error in some digital communication systems and false alarm in constant false alarm rate detection algorithms. Several numerical results are presented to demonstrate the huge savings in computational effort obtained relative to conventional Monte Carlo simulation.

7.1.7 Signal Processing Hardware

Only few papers at the WIC Symposia were devoted to the design of signal processing hardware. In [585], Lohman discusses digital optical computing. Photons as well as electrons can be used as carriers of information. Electrons have strong interaction, whereas photons normally do not interact. In non-linear optical materials, photon interaction and therefore logical functions can be realized. The paper points to areas where optical computing and optical processors could play a role.

In [599], Verbakel describes the high-level description language SILAGE that is used in the silicon compiler Cathedral II for digital signal processing, developed by IMEC and Philips Research. The paper describes an overview of the language and presents a simulation of an adaptive echo canceler. The synthesis of combinatorial logic is important for the design of integrated circuits for all kinds of signal processing systems.

In [647], Benschop describes the decomposition of any Boolean function of a number of binary inputs into an optimal inverter coupled network of symmetric Boolean functions. Threshold logic cells can implement these functions. The cells can be mapped onto silicon with a proper CAD tool.

7.1.8 Miscellaneous

In [606], Van der Vlugt describes a system that enables accurate registration of behavior. The researcher registers behavior by means of pushing preprogrammed keyboard keys effecting the tagging of labels onto the video registration of the behavior to be analyzed.

7.2 Image Restoration

This section deals with image processing and analysis papers. The major part of the papers is devoted to the topic of image and video restoration. A much smaller part deals with image processing steps such as analysis and interpretation. Therefore, to describe the papers in a consistent way, we have chosen for the restoration paradigm to classify and describe them. We will first concentrate on still image restoration followed by image sequence restoration and the notion of object motion. Next, we will focus on the consecutive steps in the image (sequence) processing chain.

Figure 7.8: *Restoration of an old photograph; (left) noisy defocused image, (right) restored image with visible ringing due to inadequate boundary conditions.*

7.2.1 Still Image Restoration

Images are produced to record or display visual information. Because of imperfections in the imaging and capturing process, however, the recorded images invariably represent a degraded version of the original scene. Although the degradations may have many causes, two types of degradations are usually dominant: blurring and noise. The field of image identification and restoration is concerned with the problem of restoring these imperfections. Identification and restoration is crucial to many of the subsequent image processing tasks, such as compression, analysis and interpretation.

Since the introduction of restoration in digital image processing in the sixties, a variety of image restoration methods have been developed, with applications in astronomy, satellite imagery, electron microscopy, medical imaging, forensic sciences, and cultural heritage. Figure 7.8 shows a restoration example of an old photograph with out-of-focus blur. Although the restored version is clearly an improvement on the originally blurred version, some ringing artifacts at the image boundaries are visible, showing the difficulty of the restoration problem.

The research on still image restoration as represented in the proceedings of the WIC symposia can be best described by the general scheme in Figure 7.9. The combined identification and restoration of images is sometimes referred to as the *a posteriori* restoration scheme [81]. It shows the complete restoration problem, in which prior to the restoration filtering, the characteristics of the blur Point-Spread Function (PSF) must be estimated, as well as the statistical properties of the origi-

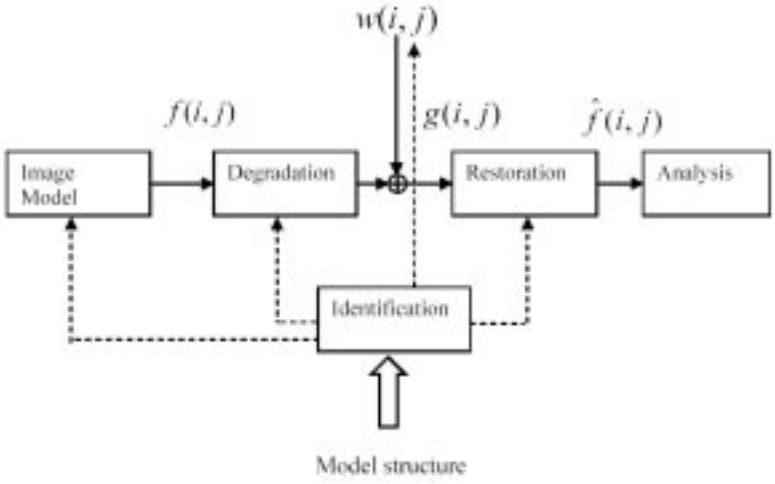


Figure 7.9: *A posteriori identification and restoration scheme.*

nal image and the noise. Here, the recorded or observed image is given by

$$g(i,j) = d(i,j) \otimes f(i,j) + w(i,j), \tag{7.1}$$

while the restored image is given by $\hat{f}(i,j)$.

The first papers from the early eighties, however, concentrated on *a priori* restoration, in that they assume that the PSF of the degradation process and the image and noise characteristics are known *a priori*. The focus in these days was on image modeling and stochastic linear least-squares filtering methods (Wiener, Kalman), and especially the extension of the recursive Kalman filter for the restoration of noisy, degraded images. Kalman filtering theory was well established in one dimension (for time signals), and an intriguing question at that time was how to extend the 1-D causal filter concept to two (spatial) dimensions with applications to image restoration.

Image Formation and Recording

Accurate models for image formation and recording (sampling) are prerequisite for good consecutive restoration. Biemond [562] introduces a state-space representation for a scanned digital image which gives a recursive description of the relations between intensities of pixels in the original image and those in the noise-corrupted observations.

Slump, Hoenders and Ferwerda [571, 583] study image formation and recording for low-dose electron microscopy. The electron dose is a compromise between the requirements of minimal radiation damage and a sufficient signal-to-noise ratio for

subsequent image interpretation. The images become a realization of a stochastic process due to the low electron dose. Both papers discuss the stochastic process that governs the low-dose image formation and present some aspects of the evaluation of the information about the object's structure contained in the noisy images. In [625], De Bruijn, Schrijver and Slump observed that cardiac X-ray images tend to be relatively noisy due to the low exposure. The assumption is made that if noise is not correlated with the signal, it does not contain any diagnostic information. A compression scheme is proposed exploiting the different spectral distributions of signal and noise.

Veldhuis and Heideman [572] introduce a sampling model for space-limited two-dimensional signals, followed by an implicit sampling model for images [577]. Here, implicit sampling means that samples are not taken at predetermined locations, but at locations where the signal fulfills some specified conditions. The way the samples are taken is consistent with a model for a part of the human visual system. In [596], Heideman, Hoeksema and Tattje discuss multi-channel sampling of sequences. Simon [621] introduces a particular class of multi-resolution transforms, the smooth non-symmetrical interpolation functions for a quad-tree representation of images, which are aimed at representing an image in a visually acceptable way.

Inverse Filtering and Least-Squares Filtering

An inverse filter is a linear restoration filter whose known Point-Spread Function (PSF) is the inverse of the blurring function $d(i, j)$. There are two problems associated with the inverse filter. First, the inverse filter may not exist because the PSF has zeros at certain spectral frequencies. Second, the inversely filtered noise may be magnified enormously because the PSF has near-zero values at certain frequencies.

To overcome the noise sensitivity of the inverse filter, a number of restoration filters have been developed; they are collectively called least-squares filters (Wiener filter, constrained least-squares filter, Kalman filter). In [564], Biemond derives an optimal line-by-line recursive Kalman filter for restoring images degraded by linear spatially-invariant degradation phenomena (motion, defocusing) in the presence of additive white noise.

Woods [588] observed that linear shift-invariant (noise) filtering is of limited utility in many image processing problems, such as restoration. The main difficulty is that the constraint of shift-invariance leads to blurring of the edges in the images. This effect has motivated the introduction of many adaptive procedures to track the apparent spatial inhomogeneity (non-stationarity) in images. Woods [588] introduced the doubly stochastic random field model for image restoration, which has apparent inhomogeneity on a local scale as well as homogeneity on a global scale using the reduced-update Kalman filter.

De Haan and Slump [608] report about a study to reduce folding distortion of digitized analog medical images without anti-alias pre-filtering. The approach followed is to consider the folding distortion as noise which can be partly filtered out by a Wiener filter. In a consecutive paper, Slump [613] reports on the development of image restoration algorithms (inverse Fourier filtering) to reduce the Moiré interference patterns arising from anti-scatter grids in the application area of medical diagnostic X-ray imaging.

Iterative Restoration Techniques

It is not easy to integrate the prior knowledge that image intensities are always positive in the linear filtering techniques described above [79]. The Kalman and Wiener filters may produce negative intensities, simply because negative values are not explicitly prohibited in the design of the restoration filter. For reasons like these, iterative procedures for image restoration have been introduced: they allow one to incorporate physical constraints on the data, to deal with nonlinear or shift varying blurs, they allow man-machine interaction and make it unnecessary to determine the inverse distortion operator.

Biemond and Katsaggelos [581] introduce an iterative procedure whose iteration equation consists of a prediction part that is based on a noncausal image model description, and an innovation part that is weighted by a gain factor. This procedure can be interpreted as an iterative procedure with a statistical constraint on the image data.

In [592], Lagendijk and Biemond extend this work. They use three kinds of *a priori* knowledge to solve the ill-posed restoration problem. The first type imposes an upper bound on the residual signal, the second type restricts the high-frequency content of the restored image, and the third kind of *a priori* knowledge is a deterministic constraint, representing a closed convex set in the solution space. Further, the concept of weighted norms is introduced in order to incorporate fundamentally spatially varying image statistics.

Lagendijk, Biemond and Boekee [598] extend the iterative restoration procedure with a nonlinear model for the image formation and recording process. This model incorporates the blurring of an image, and a nonlinear transformation to account for the response of the recording device.

Identification of Model and Blur Parameters

In the use of the image restoration filters so far, it was assumed that the degradation an image has suffered (the blur model), the image model, and the variance of the noise are known *a priori*. Since these parameters are unknown for practical images of interest, they have to be estimated from the noisy blurred images themselves. In [601], Lagendijk and Biemond propose a maximum-likelihood-based estimator to simultaneously identify the unknown image and blur parameters and to restore the image by employing an iterative procedure called the expectation-maximization

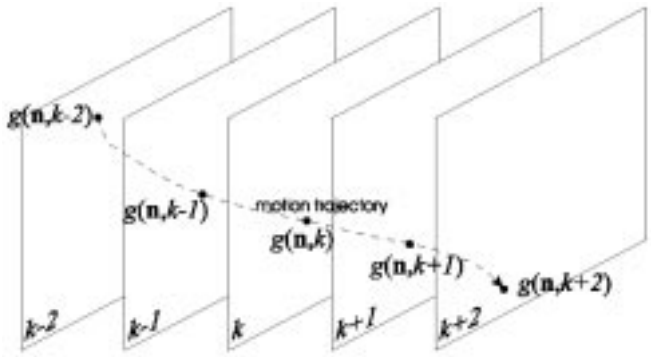


Figure 7.10: *Noise filter operating along the motion trajectory of the picture element* $(\mathbf{n}, k)$*, where* $\mathbf{n} = (i, j)$.

(EM) algorithm. The advances of this method are reported in [605]: its ability to solved the problem of estimating the coefficients of relatively large PSFs, and the estimation of the support size of PSFs in general. Hereto a hierarchical blur identification approach based on the EM algorithm is proposed.

7.2.2 Moving Picture Restoration

A video source is a much richer source of visual information than a still image. This is primarily due to the capture of *motion*; while a single image provides a snapshot of a scene, a sequence of images registers the dynamics in it. The registered motion is a very strong cue for human vision; we can easily recognize objects as soon as they move, even if they are inconspicuous when still. Motion is equally important for image sequence processing (filtering, restoration, interpolation) and compression for two reasons. First, motion carries a lot of information about spatio-temporal relationships between image objects. Second, image properties such as intensity and color have a very high correlation in the direction of the motion, i.e., they do not change significantly when tracked in a picture sequence. This can be used for example to remove temporal video redundancy (compression); in an ideal situation, only the first picture and the subsequent motion (vectors) have to be transmitted. It can also be used for general temporal filtering of a noisy picture sequence. In this case, the spatial detail in the picture is not affected by one-dimensional temporal filtering along a motion trajectory (Figure 7.10).

Motion Estimation

The goal of motion estimation is to estimate the motion of image points, i.e., the *2-D motion* or *apparent motion*. Such a motion is a combination of the motion of objects in a 3-D scene and that of a 3-D camera. Since motion in an image sequence is estimated (and observed by the human eye) based on variations of intensity, color, or both, the assumed relationship between motion parameters and

image intensity plays a very important role. The usual, and reasonable assumption made is that image intensity remains constant along a motion trajectory, i.e., that the brightness and color of objects does not change when they move. In order to develop a motion estimation algorithm, one has to consider three important elements: *motion models*, *estimation criteria*, and *search strategies*.

Block matching is the simplest algorithm for the estimation of local motion. It uses a spatially constant and temporally linear motion model over a rectangular region of support. Although this is a very restrictive model assumption, when applied locally to small blocks of pixels it is quite accurate for a large variety of 3-D motions. An average error criterion is usually used, although other measures are possible, such as a maximum error (min-max estimation). An exhaustive search gives the lowest matching error, but is computationally costly and does not *a priori* provide a smooth motion vector field. De Haan developed a "3-D recursive search block-matcher", as reported in Kleihorst, De Haan, Lagendijk and Biemond [617], which allows an extremely fast implementation, and a smooth reliable motion (vector) field. The "bi-directional convergence" of the algorithm overcomes the inherent slow convergence of the (block) recursive algorithm.

Noise Filtering

In [612], Kleihorst, Lagendijk and Biemond propose an image sequence noise filtering scheme that operates in the temporal direction. Due to the movements in the scene, the noisy signal

$$g(i,j,k) = f(i,j,k) + n(i,j,k) \tag{7.2}$$

cannot be modeled as a stationary signal. Thus, one way of dealing with the non-stationarities in the temporal signal is to use motion estimation of objects and to filter along the motion trajectories. In this way, motion estimation is used to find the path of maximal correlation in the temporal direction and indirectly creates a more stationary signal. Motion estimation is a very time-consuming operation in general, and does not successfully work in for example occluded areas.

Kleihorst, Lagendijk and Biemond [612] therefore investigate a noise filtering approach for image sequences that removes the non-stationarities in the temporal signal in a different way, namely by *trend removal and normalization*. The decomposition is done by estimating the local statistics of the signal with the aid of ordered statistics estimators. After the decomposition, the stationary part of the signal can be filtered by a regular noise filter, the result of which is combined with the non-stationary part to produce the final filtered sequence.

In [617], Kleihorst, De Haan, Lagendijk and Biemond extend their previous filter with an additional motion-compensation step. This will remove additional non-stationarities due to the filtering along the motion trajectory. However, because of the incompleteness of the motion model, a compensated signal still contains a lot of non-stationarities. Therefore the signal is additionally decomposed into a stationary and non-stationary part, resulting in a noise filtering scheme with a double

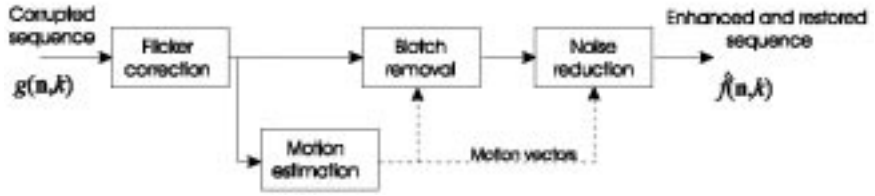


Figure 7.11: *Some processing steps in the removal of noise, blotches and intensity flicker from video.*

compensation for motion. For the motion-compensation step, the *3-D recursive search block-matcher* is used. Excellent filter results are reported for moderate amounts of noise. For low signal-to-noise ratios, the uncompensated results are better. This is because the motion estimator tends to match the noise.

Restoration of Archived Film and Video

Another important application of image sequence filtering and restoration is for preservation of motion pictures and video tapes recorded over the last century. These unique records of historic, artistic, and cultural developments are deteriorating rapidly because of aging of the physical reels of film and magnetic tapes that carry the information. The preservation of these fragile archives is of interest not only to professional archivists, but also to broadcasters as the archives themselves form a cheap alternative to fill the many television channels that have come available with digital broadcasting and the Internet.

However, it only makes sense to reuse old film and video material in a digital format if the visual quality can meet the standards of today. For that reason, the archived film or video is first transferred from the original reel or magnetic tape to digital media. Second, all kinds of degradations (noise, flicker, blotches) are removed from the digitized picture sequence to increase the visual quality and commercial value. Intensity flicker refers to variations in intensity over time, caused by aging of the film, by copying or format conversion (for instance from film to video), and in case of earlier film, by variations in shutter time. Blotches are the dark and bright spots that are often visible in damaged film. The removal of blotches is essentially a temporal detection and interpolation problem. Where blotches are spatially highly localized artifacts in video frames, intensity flicker is usually a spatially global, but not stationary, artifact. In practice, picture sequences may be degraded by multiple artifacts. Therefore, a sequential procedure is usually followed where artifacts are removed one by one. Figure 7.11 illustrates the order in which flicker, blotches, and noise are removed.

The reasons for the modular approach described above are the necessity to judge the success of the individual steps (for instance for an operator), and the algorithmic and implementation complexity. Blotch removal and noise reduction (see Figure 7.12) use motion-compensated interpolation and filtering based on a mo-

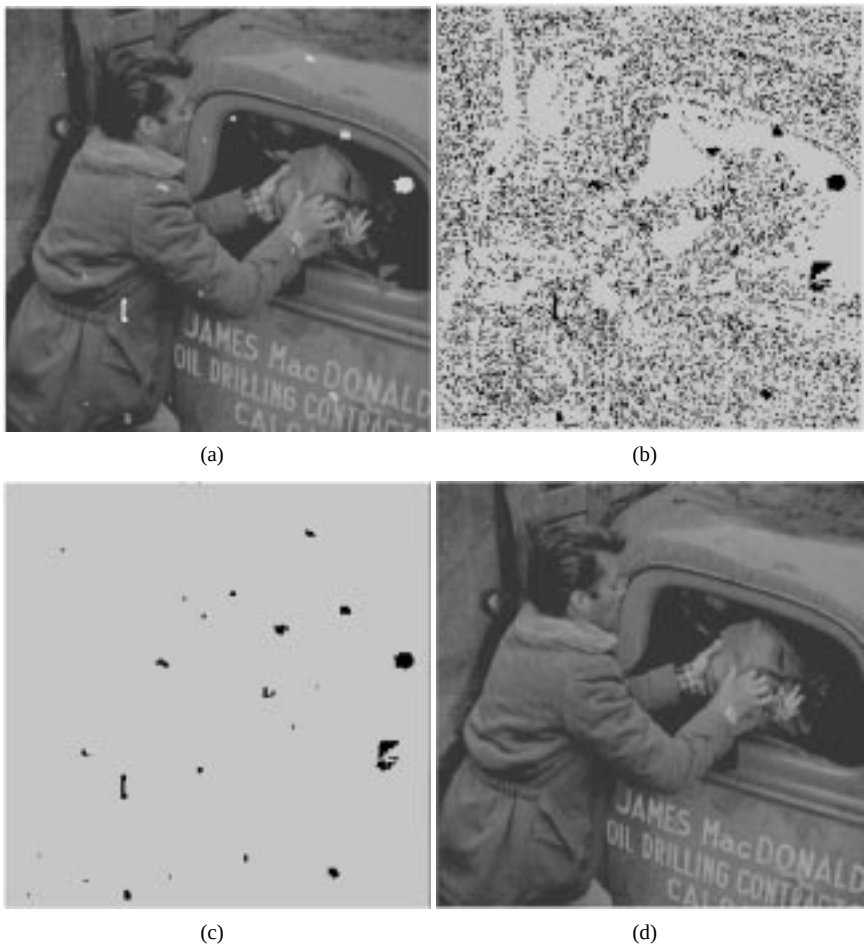
tion estimator on the flicker-corrected data, respectively. It is important to mention that the estimation of motion from degraded sequences is problematic in general. This particularly holds for picture sequences that contain flicker, because virtually all motion estimators are based on the constant luminance constraint. Therefore, motion estimation is performed on the flicker-corrected data. Further, the focus is on robust motion estimators to the different artifacts, with the possibility to repair incorrect motion vectors.

Because the objective of restoration is to remove irrelevant information such as noise, it restores the original spatial and temporal correlation structure of digital picture sequences. Consequently, restoration may also improve the efficiency of the subsequent MPEG compression of image sequences. However, there are situations where current restoration/filtering techniques are still failing. In some of these cases, the quality of parts of the restored sequence is even worse. For instance, in sequences where objects or persons perform complex motion, called *pathological motion*. Rares, Reinders and Biemond [654] extend and improve the restoration scheme in Figure 7.11 by taking into account these complex motion events.

Motion-Compensated Picture Rate Conversion and De-interlacing

In an early paper, Van Otterloo, Rohra and Veldhuis [587] identify two main drawbacks of conventional television systems (625 lines per frame, 50 fields per second, 2:1 interlace) as large area flicker and line flicker. The paper gives a theoretical analysis describing the effects of increased field rate on moving objects in an observed sequence of pictures, where the increased field rate is obtained by temporal interpolation without and with motion-compensated interpolation. It was concluded that to prevent the interpolated sequence from artifacts (blurring) one needs motion compensation. However, fast and reliable motion estimation was not yet possible at that time.

De Haan [638] gives an overview of the progress in spatial scaling, picture rate conversion, de-interlacing, and motion estimation as important tools for video format conversion, which has become a key technology for multimedia systems. By the end of the twentieth century, there was a strong convergence between PC and TV, due to the fact that video entered the personal computer through DVD, CD, and the Internet. This convergence led to an explosion of video formats, as an addition to the two main broadcast formats (interlaced 50 and 60 Hz formats with 625 and 525 scanning lines, respectively), PC monitors with picture rates between 60 Hz and 120 Hz, and spatial resolutions in a broad range (VGA, SVGA, XVGA, etc.). Also television receivers profited from these techniques and decoupled their display format from the historically determined transmission format to eliminate flicker artifacts (as discussed above), and/or to adapt to new display principles, which resulted in new flicker-free (100Hz), non-interlaced (Proscan), and/or widescreen (16:9) formats on cathode ray tubes, plasma panel displays and liquid crystal screens. Currently, also video telephony, video from the Internet, and graphics are being merged with broadcast signals.



(a) (b)

(c) (d)

Figure 7.12: *(a) Video frame with blotches, (b) blotch detection mask (incl. noise) (c) Blotch detection mask after post processing; (d) blotch-corrected frame.*

7.2.3 Image and Video Analysis

After restoration, one of the possible goals of processing an image (sequence) digitally is to analyze the image content, in order to extract information about the phenomena which are represented by the image. Image analysis can thus be described as an *image-to-data* transformation, the output data being, e.g., a set of measurement values, a set of labeled objects, or even a description of the imaged phenomena. One of the crucial steps in the analysis process is the segmentation of an image, i.e., the partitioning of the image plane into regions which are homogeneous according to some predefined criteria. The result of the segmentation stage is thus a map of the various regions, which is intended to be meaningful with

respect to the imaged phenomena.

Two major approaches exist to image segmentation: *region-based*, and *edge-based* methods. In region-based methods, areas of images with homogeneous properties are found, which in turn give the boundaries. In edge-based methods, the local discontinuities are detected first and then connected to form larger, hopefully complete, boundaries. The two methods are complementary and can also be combined to a certain extent. The segmentation results combined with for example motion information can be used for object tracking, object recognition and scene modeling. It should be noted that color information is a highly important feature in image analysis and recognition tasks. Finally, image analysis plays an important role in the searching and accessing of stored visual information.

Region-based Segmentation

Gerbrands [566] describes image segmentation as a pixel labeling or classification problem, because the ultimate goal of segmentation is to assign a label to each and every pixel. The label indicates to which one of the various image components or regions the pixel belongs. He introduces a probabilistic procedure, which is an iterative procedure to use contextual information to reduce local inconsistencies in label assignment.

Kruisbrink [569] applies syntactic pattern recognition for image segmentation. In certain patterns (image components) to be classified simpler sub-patterns (pattern primitives) are first searched and these are applied to segment muscle cell pictures.

Gerbrands and Backer [582] introduce a split-and-merge method for the segmentation of side-looking airborne radar (SLAR) imagery, i.e. the detection of boundaries of agricultural fields. Based on an image formation model, the agricultural fields are represented by regions in the image that differ in mean value (depending on crop type, crop coverage, moisture, etc.). A region is examined as a candidate for splitting or for merging based on some predefined criteria. In [600], Gerbrands, Backer, Hoogeboom and Kleijweg improve their proposed split-and-merge segmentation algorithm for SLAR imagery by using *a priori* knowledge about the agricultural scene in the form of topographical maps, remote sensing data from other sources or from previous occasions, and, eventually, geo-information systems.

Gerbrands, Backer and Cheng [591] introduce a multi-resolution segmentation algorithm based on split-and-merge procedure generating variable-sized multi-resolution data units, which are used in a clustering procedure to extract regional features followed by a nonlinear probabilistic relaxation procedure to conduct the final labeling of the blocks. It is shown that a large reduction in data processing is attained by using processing blocks rather than pixels (as in a previous method) and still the result reasonably approximates the true segmentation.

Gonzalez, Katartzis, Sahli and Cornelis [646] discuss the identification of man-

made objects like land mines from polarimetric infrared (IR) images. The performance of IR systems for the detection of shallowly buried land mines is limited due to the background clutter. For this reason, IR polarization filters were introduced for improving the low target-to-clutter ratio in infrared scenes. The paper proposes a pixel-fusion approach for combining the polarization information with image analysis techniques such as image enhancement and segmentation.

Farin and De With [637] describe a fast and flexible implementation of region merging as a spatial segmentation algorithm using different merging criteria including region sizes and quad-tree decomposition as a preprocessing step to be applied in object-oriented video coding, such as MPEG-4.

Finally, Brox, Farin and De With [653] develop a multistage generalization of conventional region merging for image segmentation again with applications in MPEG-4. A sequence of different criteria is used to achieve a semantically and subjectively superior segmentation result. Instead of starting the algorithm with single-pixel regions, a pre-segmentation with the watershed algorithm for edge detection is performed on a gradient map of the input image.

Edge-based Segmentation

Gerbrands, Backer and Van der Hoeven [586] discuss a sequential method of edge detection which uses dynamic programming to detect the optimal edge in a specific region of interest. The problem of finding the optimal edge can be formulated as the problem of searching for the optimal path from the bottom to the top through a matrix of cost coefficients. This method is developed for the detection of the left ventricular contour in cardiac scintigrams.

Vanroose [644] describes the implementation of a complete recognition system for flat objects in a picture taken by a camera with unknown parameters and position. As a consequence, the objects, as seen in the picture, can be distorted by an arbitrary projective transformation with respect to their counterparts in a sample database. Contours in an image are then found by standard edge detection followed by spline fitting, contour segment transformation, and they are identified with respect to a training database.

In [626], Vanroose reflects on the information flow and spatial locality of image processing operators, such as thresholding, histogram, convolution and edge detection. Special attention is paid to the *edge following* step of the edge detection operation. The potential quality improvement resulting from the use of a less local algorithm is studied.

Object Detection, Tracking, and Recognition

Detection and tracking of (moving) objects is important for robot control, human face recognition in for example video surveillance applications, augmented reality, motion-compensated prediction/interpolation/restoration and object-based coding.

Backer and Gerbrands [594] design a flexible and intelligent system for fast measurements in binary images to enable object tracking for in-line robot control. Rares and Reinders [641] introduce an object tracking system for film archive restoration based on statistical models. An object selected in a frame by a user is tracked throughout the sequence by using a blob-like description of its features, statistically represented by a mixture of Gaussians. To deal with the initial incomplete data about the object's appearance, as well as to integrate the acquired knowledge about these appearances and to cope with changes in them, the object models are updated statistically by an on-line version of the expectation-maximization (EM) algorithm.

Persa and Jonker [635] describe a real-time system for human computer interaction through gesture recognition and 3-D hand tracking. One camera is used to focus on a user's hand to which a small rigid dark square is attached.

Ravyse, Sahli and Cornelis [645] present an approach for automatically segmenting and tracking faces in color image sequences. The goal is to analyze a moving person's head in front of a static camera, relevant for applications in video telephony, animation, and virtual conferences. Segmentation of faces is based on skin color and shape verification. The tracking is realized using a 3-D ellipsoidal model and optical flow. Here the optical flow is interpreted in terms of rigid motion of the 3-D ellipsoid.

Zuo and De With [659, 664] concentrate on exploiting human face information for surveillance applications in a consumer home environment. Their system features robust, real-time human face detection and facial feature identification to be inserted in a video-security system architecture, where MPEG-4 coding techniques enable low bit-rate video transmission over a home network environment.

3-D Scene Modeling

Scene modeling aims to reconstruct, as accurately as possible, the exact shape of 3-D objects which are (partly) visible in several (2-D) views. This shape can be used to the recognize 3-D objects as well as to determine the object's position and orientation in 3-D world coordinates.

Mieghem, Gerbrands and Backer [590] follow a stereo vision approach, where two images are obtained from calibrated camera positions. Three-dimensional object features are then computed and used as attributes in an inexact graph matching recognition stage to recognize trihedral objects. Lei and Hendriks [640] focus on the extraction of 3-D shape information. The necessary low-level feature extraction is approached in a unifying way, employing phase information which is robust to noise, shading and contrast variations in an image.

Vanroose [639] rephrases the 3-D scene modeling process in information theoretic terms using a source-channel model. An optimal 3-D model is obtained by maxi-

mizing the mutual information as a measure of the goodness-of-fit of a 3-D model to the imaging data. Pasman and Jansen [634] deal with virtual reality for mobile use, where virtual objects can be projected in overlay with the real world for applications such as remote maintenance. A latency-layered system is proposed that combines fast position tracking and rendering using approximate geometric models, with slower but more accurate techniques.

Vanroose, Kalberer, Wambacq and Van Gool [662] present a method to animate the face of a speaking avatar, i.e., a synthetic 3-D human face, such that it realistically pronounces any given text, based on the audio only. Special attention is given to the lip movements, which must be rendered carefully and perfectly synchronized with the audio in order to look realistic, from which it should in principle be possible to understand the pronounced sentence by lip reading.

On the Use of Color Information

Color information has proven to be very useful in image analysis and recognition tasks. For example, for the viewpoint- and illumination-independent recognition of planar color patterns such as labels, postcards, pictograms, which typically have a high pictorial content. Mindru, Moons and Van Gool [632] present new invariant features which are based on the moments of powers of the intensities in the individual color bands and combinations thereof and test the discriminant power and classification performance on a data set of images of real, outdoor advertising panels. In [648], Mindru, Moons and Van Gool concentrate on a model for the photometric changes of planar surfaces under internal and external illumination changes between two different color (R,G,B) images of a same object or scene.

Video Content Analysis

Since digital libraries for storing large amounts of textual, audio and visual information are becoming widespread, there is a need for efficient methods for searching and accessing these libraries for example through the Internet. Hanjalic, Lagendijk and Biemond [624] discuss the achievements and the challenges in the visual search of video, especially for consumer home-digital libraries, such as automation of shot-change detection and optimization of key-frame extraction by taking into account users' specifications.

In [652], Hanjalic and Xu address the problem of extracting the affective content of video, defined as the amount of feeling or emotion contained in and mediated by a video toward a viewer. A method is developed to extract this type of video content based on the *dimensional approach to affect* known from psychophysiology, where the affective content can be represented as a set of points in the "3-D emotion space". The availability of methodologies for automatically extracting affective video content should lead to a high level of personalization and a way of efficiently handling and presenting the data to various categories of viewers.

7.3 Discussion and Conclusions

The growth and maturity of the signal processing field can be measured by the many text books on signal processing, the many patent applications in this area, and the major signal processing conferences. We mention the annual IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), and Eurasip's EUropean SIgnal Processing COnference (EUSIPCO). The growing importance of image processing has led to the prestigious IEEE International Conference on Image Processing (ICIP) in 1994.

Researchers of Information Theory Groups at Delft, Eindhoven, Leuven and Twente and of Philips Research have contributed substantially to the field of signal processing and in particular the image processing. Signal processing has been and remains to be an exciting and economically vital area. The past decades have been particularly exciting as each new wave of faster computing hardware has opened the door to new applications. Most likely, this trend will continue in the near future.

CHAPTER 8

Image and Video Compression

P.H.N. de With (TU Eindhoven/LogicaCMG)
R.L. Lagendijk (TU Delft)

Introduction

Compression techniques are of prime importance for reducing the large amount of data needed for the representation of speech, audio, images and video sequences without losing much of its quality, judged by human viewers. Of the previously mentioned areas, digital video compression is the one most recently established and has gained strong interest and popularity. Many different compression – or lossy source coding – methods, all firmly based on rate-distortion principles, can be found in a variety of Internet applications, television broadcasting, music distribution, and consumer digital video applications, such as DVDs and DV camcording.

An abundance of standards for image and video compression has been put forward since the beginning of digital compression technology in the late 1970s, each reflecting the state-of-the-art when released. The performance of these standards – from the H.120 DPCM-based video compression standard and DCT-based JPEG image compression standard, to the most recent JPEG2000 wavelet-based image compression standard and H.264 video compression standard – have been improved upon time after time. In fact, at the time of writing, new initiatives

[1]This chapter covers references [665] – [755].

emerge for yet another improved video compression standard (H.265). Video standards have greatly influenced compression technology, because they focused the research and development leading to interoperable products and they also contributed to concentrated VLSI realizations and architectural innovations.

In the first part of this chapter we review the development of compression theory and technology. We will consistently use the word *compression* to distinguish between the *lossy* source coding discussed in this chapter and the *lossless* source coding discussed in Chapter 2. In the second part of this chapter, we highlight key developments in compression in the past 25 years, and summarize the contributions of Information Theory researchers in the Benelux. We have chosen to subdivide this part into three interrelated areas, namely:

- fundamental techniques to decorrelate image and video data prior to quantization. Papers will be discussed that deal with image transforms such as DCT and subband decompositions, as well as papers that discuss the problem of motion estimation and compensation for video.
- quantization theory, covering rate-distortion theory, vector quantization, bit allocation, and perceptual optimization of image and video compression.
- hierarchical, scalable and embedded compression, and other extended or alternative compression strategies for particular application domains.

8.1 History of Compression Theory and Technology

A lossy source coding or compression method is one where compressing a signal (image, video, but music and speech as well) denoted by $x(n, m)$ with image coordinates (n, m), and then decompressing it, retrieves a signal $\hat{x}(n, m)$ that may well be different from the original, but is "close enough" to be useful in some way. The difference between the original signal and its reconstructed version can be expressed in two performance measures.

- *Distortion* D between the signal amplitudes, often called compression or quantization error. The most straightforward way to express this difference is in terms of variance of the quantization error:

$$D = \quad \sigma_q^2 = \quad \mathrm{E}\left[(x(n,m) - \hat{x}(n,m))^2\right] \tag{8.1}$$

  Although this measure has the significant drawback that it does not reflect the human perception of compression errors in images and video very well, it is still the *de facto* performance number for comparing systems.

- *Average number of bits* R used per signal sample, yielding a bit-per-pixel (bit/pixel) measure. For video, sometimes the average number of bits per second is used, yielding the bit rate in kilo- or Megabit per second (kbit/s or Mbit/s). The ratio between the average number of bits per sample or bit rate of the original (uncompressed) signal and the compressed signal is called the compression factor.

The fundamental problem of compression is the optimal trade-off between the distortion σ_q^2 and the required bit rate (information) R that needs to be communicated from sender to receiver. This optimality problem, known as *rate-distortion theory*, was first addressed by Shannon [3] and later on by Berger [24]. In a theoretical setting, the rate-distortion problem can be formulated as the minimization of the mutual information $I(X;\hat{X}) = H(\hat{X}) - H(\hat{X}|X)$ between the source X and the received signal $\hat{X}$ as a function of the behavior of the communication channel, given a maximal distortion D^*, or

$$\min_{Q_{\hat{X}|X}(\hat{x}|x)} I(X;\hat{X}) \quad \text{subject to:} \quad D \leq D^* \tag{8.2}$$

Here $Q_{\hat{X}|X}(\hat{x}|x)$ is the conditional PDF of the communication channel, which in practical systems reflects the behavior of the compression algorithm in probabilistic terms. Solving the rate-distortion problem yields expressions for the smallest bit rate needed to compress a signal with a distortion no larger than D^*. Unfortunately, the rate-distortion relation can only be calculated for relatively simple signal models. A well-known and important example is when the signal X can be modeled as a Gaussian iid process with variance σ_x^2, and D is the mean-squared error σ_q^2 between the original and compressed signal, as expressed by Equation (8.1). In this case, we find

$$\sigma_q^2 \quad = \quad \sigma_x^2 2^{-2R} \tag{8.3}$$

Practical image and video signals often do not follow such simple stochastic models; in fact complete stochastic modeling of image and video signals is utterly infeasible. For that reason image and video compression theory has always been complemented by the art of designing video systems and by making the theorems practically feasible.

The heart of any compression method is the *quantizer*, which rounds continuous-valued signal amplitudes to a set of suitably chosen discrete values (called representation levels). The discrete values are then represented by bit patterns, which are communicated to the decoder. The mapping of the quantizer representation levels to binary code words is an entropy coding problem, for which techniques can be used as described in Chapter 2, such as runlength, Huffman, and arithmetic coding. It is the rounding process of the quantizer that causes the decompressed signal values to be different from the original ones, hence the quantizer is the primary element that is responsible for achieving the trade-off between bit/information rate and distortion. Theory and optimal design of scalar quantizers under different constraints has been widely studied [106], resulting in different categories of quantizers such as uniform, Lloyd-Max, and Uniform Threshold quantizers.

The multidimensional extension of scalar quantization, called *vector quantization (VQ)* [66], was a major step toward reaching the rate-distortion bounds for dependent sources. However, this requires the processing of infinitely long sequences. For image and video compression very long series of pixels are indeed available, as was first realized by Gersho [56] in 1982. However, the single most important

explanation for the impediment of the widespread usage of VQ is the computational complexity of the codebook search process.

A more successful attempt to exploit dependencies in signals was the use of *predictive or differential compression* strategies. In predictive compression, signal amplitudes are predicted on the basis of neighboring signal amplitudes. In order for the decoder to be able to reproduce the prediction made by the encoder, the prediction mechanism operates on already quantized signal amplitudes. This leads to the basic scheme for any predictive compression technique, usually called Differential PCM (DPCM), which is illustrated in Figure 8.1. The linear prediction of the prediction signal $x_p(n)$ uses the reconstructed signal $x_r(n)$:

$$x_p(n) = \sum_{j=1}^{M} a_j \cdot x_r(n-j), \tag{8.4}$$

where a_j denote the prediction coefficients for j = 1, 2, ..., M. The prediction coefficients are calculated such that the MSE between the original and compressed signal is minimized. The extension from the above 1-D prediction model to 2-D is straightforward. The very first video coder, developed in the European COST211 project and standardized by ITU-T (then called CCITT) as the H.120 standard in the early 1980s, uses spatial DPCM working on video frames, at 2 Mbit/s compressed bit rate.

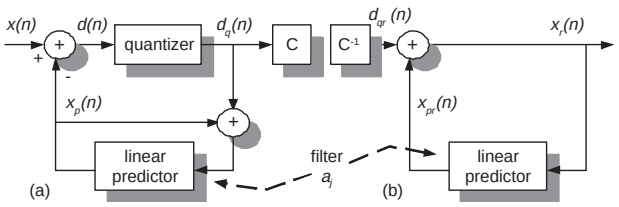


Figure 8.1: *Basic predictive compression structure, called Differential PCM (DPCM).*

In spatial DPCM, the image quality is far from optimal because (i) temporal correlation is ignored; (ii) the compression factor and quality is limited by the pixel-by-pixel operation of the scalar quantizer. In order to improve the quality, two research and development directions were vigorously pursued, namely *block-based transform coding*, which aims at exploiting spatial correlation, and at the same time reach fractional bit rate per pixel, and *motion estimation and compensation* to exploit temporal correlation along the motion trajectories. These developments led to the design of the block-based image coders and the motion-compensated block-based video coders in the late 1980s, which form the foundation of the success of today's image and video compression standards such as JPEG, MPEG, and H.263/H.264.

During the late 1980s, a large number of block-based transform coding proposals for video conferencing were submitted to ITU-T. Except for one, all the proposals were based on the Discrete Cosine Transform (DCT). In parallel to ITU-T's investigation during 1984-1988, the Joint Photographic Experts Group (JPEG) was also interested in compression of still images. They chose the DCT on blocks of 8×8 pixels as the operation for decorrelation. The decision of the JPEG group undoubtedly influenced the ITU-T to also select the 8×8 DCT for spatial decorrelation as a basis for its video compression standard known as H.261.

A DCT decomposes a block of image pixels onto a set of basis functions, typically called basis images in image and video compression. The basis images for the 8×8 DCT are shown in Figure 8.2. The weights of the individual DCT basis images, called DCT coefficients, are quantized, entropy encoded, and sent to the decoder. Because of the importance of block-transforms, we expand on this subject in Section 8.2.

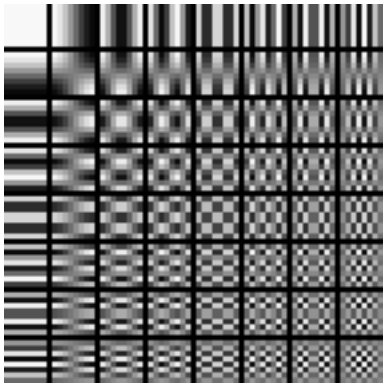


Figure 8.2: *Basis functions of the DCT (8x8 blocks of pixels).*

An alternative to the DCT decomposition is a subband or wavelet transformation. Since these schemes are somewhat more complex, they developed more gradually. Efficient implementations for subband/wavelet decomposition now exist, based on "lifting" schemes, and a variety of ways has been found for making quantization of subband/wavelet coefficients as locally adaptive as DCT-based systems, using for instance zero-tree representations. Subband/wavelet decompositions are currently found in the JPEG2000 image compression standard, and audio compression standards such as MP3 and AAC.

Due to the popularity of motion-compensated DCT systems (1985–1995), motion estimation developed strongly, yielding both theoretical concepts such as the optical flow equation, and a wide variety of practical motion-estimation algorithms. In a video compression context, a temporal DPCM system is used, where a motion-compensated block-based prediction of the current video frame is created based

on the previous video frame. The difference between the motion-compensated prediction and the actual pixel information, called the prediction difference, is spatially compressed and sent to the decoder. Motion estimation is relevant to many problems in video processing, such as noise removal, format conversion, computer vision, and compression. In video compression, motion estimators are relatively simple block-based searching procedures, because they need to operate on real-time video speed. In search of computationally efficient block-based motion estimators, different solutions have been found, ranging from efficient search patterns, to hierarchical and recursive block-matching motion estimators. The first standardized *motion-compensated DCT-based* video coder for video conferencing is known as the H.261 video coder, which operates at bit rates between 384 kbit/s and 1.15 Mbit/s.

In the early 1990s, the ISO Moving Picture Experts Group (MPEG) started investigating compression techniques for storage of video, such as CD-Is and CD-ROMs. The resulting standard, known as MPEG-1, has been very successful. MPEG-1 encoders and decoders/players are widely used on multimedia computers and for video playback in Asia. Since MPEG-1 lacked efficient compression for interlaced signals, its successor MPEG-2 became the standard for broadcasting digital standard TV signals (DVB based on MPEG-2) and storage of TV signals (DVD). The ISO MPEG-2 standard is also known as the ITU-T H.262 standard.

After the success of MPEG-2, development in compression technology has taken four different paths:

- *Higher compression factor* at the same quality. This has resulted in the H.263 and the recent H.264 video compression standard. Alternative compression systems also exist, either as specific products (e.g., RealVideo) or as "hacked DVD" formats (e.g., DivX, Xvid). Although the produced bit streams are incompatible with any standard, the heart of the underlying compression system is still a motion-compensated DCT-based encoder.

- Application in Internet or wireless communication scenarios, in which case the communication channel may *corrupt the compressed bit stream* in various ways. Error-robust, scalable and joint source-channel compression systems were developed as an answer to these channel-induced challenges.

- Numerical and perceptual *optimization strategies* for optimally controlling the many options in motion-compensated video compression systems.

- *Region or object-based* compression. This is the most revolutionary step away from the DCT-based compression philosophy. The basic unit for motion estimation and decorrelation is no longer an 8×8 or 16×16 block of pixels, but an arbitrarily-shaped area of pixels that is homogeneous or correlated in a more meaningful way. The MPEG-4 standardization has contributed significantly to research into region/object-based image and video compression.

8.2 Decorrelation Techniques

As we have seen in the previous section, decorrelation is the first step for efficient compression of an image/video signal. In any compression system, decorrelation precedes the quantization and entropy stages. Decorrelation techniques have evolved in complexity over the past decades, leading to higher video compression at the expense of more signal processing complexity. In this section we elaborate on three important decorrelation techniques, namely transform coding, motion compensated temporal prediction, and subband/wavelet coding. We summarize the underlying principles, and discuss the contributions by Information Theory researchers in the Benelux.

8.2.1 Transform Coding and the DCT

Transform coding techniques form the cornerstone of modern digital compression standards, such as JPEG and MPEG. Signal transforms explicitly aim to spatially decorrelate the image/video signal. Instead of predicting the signal sample-by-sample, *blocks* of samples are taken from the image/video frame and transformed into a "frequency"-domain representation. The resulting transformed signal components are then quantized and entropy encoded. An important motivation for using a transform is that it enables perceptual optimization of compression systems. The quantization errors can sometimes be better hidden when using a signal transform. For example, a Fourier transform enables the introduction of selective quantization noise for the higher frequencies only. This property is exploited by using frequency weighting in the quantization of transformed signal components.

Transform coding operates on blocks of samples, instead of individual samples such as in predictive coding. Because blocks are processed and mostly jointly compressed, the potential efficiency and coding gain of transform coding is higher than that of predictive coding. The result after transforming a block of signal values is called a block of transform coefficients. After applying the transform matrix

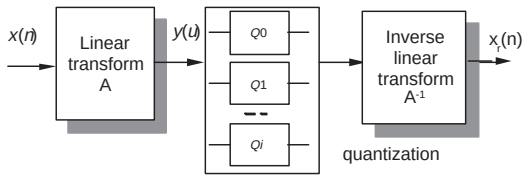


Figure 8.3: *Block diagram of transform coding. The quantization stage is modeled as a bank of scalar quantizers.*

$\mathbf{A}$ and the quantization in Figure 8.3, the reconstruction of the input signal $x(n)$ occurs by applying an inverse transform with matrix $\mathbf{A}^{-1}$ to a group of N quantized coefficients, resulting in the signal $x_r(n)$. In transform coding, the $N \times N$ matrix $\mathbf{A}$ is chosen to be orthogonal, so that $\mathbf{A}^{-1} = \mathbf{A}^T$.

The *Discrete Cosine Transform (DCT)* uses transforms derived from sampled and

modulated cosine functions. The DCT is currently the most popular real-valued transform and is used in many standards, such as MPEG, JPEG and DV (digital camcording). The success of the DCT as decorrelating transform lies in the fact that it closely approximates the optimally decorrelating Karhunen-Loeve transform for natural images and video. A drawback of the DCT is its complexity of implementation, because modulated cosine functions require several real numbers in a reasonable accuracy. The definition of a one-dimensional N-point DCT is

$$\begin{aligned} y(u) &= \sqrt{\frac{2}{N}} C(u) \cdot \sum_{i=0}^{N-1} x(i) \cdot \cos[\frac{(2i+1)u\pi}{2N}] \text{ where} \\ C(0) &= \frac{1}{\sqrt{2}} \quad \text{and} \quad C(u) = 1 \quad \text{for} \quad u = 1, 2, ..., N-1. \end{aligned} \tag{8.5}$$

Due to the orthogonality of the transform, the inverse DCT is defined in nearly the same way, except for several normalizing factors. Despite the rather complex definition of the basis vectors, the DCT uses a limited set of real numbers for making the basis waveforms cosine-based. This is due to the rotational symmetry of the cosine function in the complex plane. This phenomenon can be exploited to design fast DCT implementations, i.e. performing the computation with a reduced number of additions and multiplications.

Two-dimensional transforms are used in practical situations by extracting square blocks of $N \times N$ samples of an image or video frame. Typical values for the block size in image/video compression are N = 4, 8, or 16. Although the square blocks are commonly taken from a single image, for interlaced video signals this is not always the case. The 2-D transform, such as the 2-D DCT, is implemented in a separable way. Effectively, *separability* separates horizontal (row) and vertical (column) operations. A 2-D DCT can then be performed conveniently in two phases, each of which involves N 1-D DCTs, resulting in the basis functions (basis images) shown in Figure 8.2

The first two standards that make use of DCT-based image compression are the JPEG and DV systems. JPEG compresses still pictures (photos), while DV compresses independently consecutive video frames taken from a moving video sequence. The JPEG standard [92] applies an 8×8 DCT transform, adaptive quantization and variable-length coding. The coarseness of quantization is controlled by a user-selectable quality parameter. The quantization itself is based on adaptive uniform quantization using coefficient weighting based on properties of the human visual system. The variable-length coding (VLC) combines the coding of runs of zero-valued DCT coefficients and Huffman coding of nonzero DCT coefficients. The JPEG standard can compress video images between lossless (yielding a compression factor of approximately 1.5–1.7) and up to a factor of 20–25 (0.5 bit/sample).

The DV compact and pocketable camcorder system has been dimensioned for compression of SDTV and HDTV for home use [104]. Similar to JPEG, an 8×8 DCT is used in combination with quantization and VLC coding. Intraframe com-

pression with block shuffling is used because of editability and trick modi (e.g., fast forward). The DV system operates with luminance and chrominance color components, where the color-difference signals (Cr and Cb) are subsampled either horizontally with an extra factor two (4:1:1) for 60 Hz, or vertically with a factor two (4:2:0) for 50-Hz systems. The DV system operates well using compression factors 5–8 (1 bit/sample), yielding a compressed bit rate of 25 Mbit/s.

8.2.2 Motion-compensated Transform Coding and MPEG

Considerable temporal redundancy exists between consecutive video frames that can be exploited with prediction of the motion of objects [98]. The combination results in a *hybrid or motion-compensated transform coder* that is based on transform coding in the spatial domain and predictive coding in the temporal domain:

- *Spatial redundancy* is found in individual pictures within a video sequence. Similar to still picture compression standards, spatial redundancy is exploited by transforming picture blocks to the transform domain using the DCT.
- *Temporal redundancy* is found between successive frames of a video sequence. The redundancy is exploited by compressing frame differences instead of complete frames. A higher compression rate is achieved by predicting spatial frequencies using *motion estimation* (ME) and *compensation* (MC) techniques.

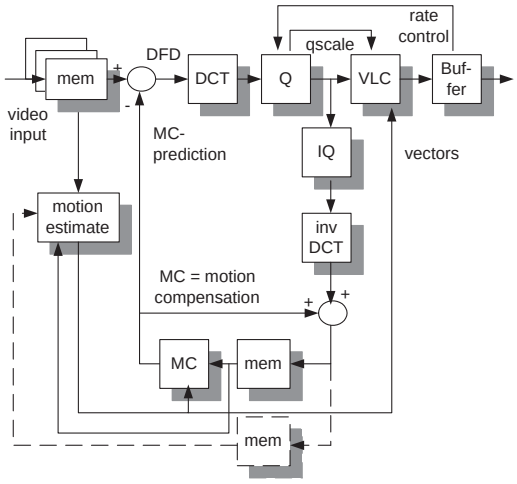


Figure 8.4: *Architecture of hybrid interframe DCT compression system.*

The block diagram of a hybrid MC-DCT encoder is shown in Figure 8.4. The diagram portrays a predictive coding loop in the vertical direction, where the previously compressed frame(s) is (are) stored in frame memories at the bottom of

the diagram. The ME processing computes the *motion vectors* of each block by searching the actual block in the frame memories at the corresponding position within a predetermined search window. If a close "copy" of the actual block is found, that motion vector is adopted.

The motion compensation (MC) uses the final selected vector of the ME to selectively read the indicated block from the reference frame memories at the bottom of Figure 8.4. The reading may involve linear interpolation of past and future data in the case of bidirectional ME. Proposing the reconstructed or read block as a prediction is motion compensation, and the prediction is now called motion-compensated prediction. The MC prediction is subtracted from the actual block, thereby yielding usually a small difference block, i.e. the *displaced frame difference* (DFD). If the ME and MC work well, only a small difference signal is added for reconstruction. This difference block is compressed with the DCT coding steps. If the difference signal is expected to be large, which can be deduced from the variance computations of the ME, the system may ignore the prediction on a block basis (set prediction to zero) and code the original input. During interframe compression, this decision is called "fallback" coding.

The above described motion-compensated video compression systems led to the MPEG-1 video compression standard [98]. At a resolution of 352×240 pixels (SIF), a compressed bit rate of 1.5 Mbit/s was achieved, which makes it possible to store one hour of video and audio on a CD (still known as Video CD). The successor of MPEG-1, called MPEG-2 [103], is based on the same motion-compensated transform coder. However, it is optimized for higher resolutions (SDTV and HDTV) and interlaced video signals, yielding bit rates of 3–7 Mbit/s for SDTV and 19 Mbit/s for HDTV in the USA.

MPEG obtains a fairly large compression factor of 25–30 by using bidirectional ME/MC in at least half or more of the pictures of a video sequence. Since for bidirectional pictures also near future pictures are required, intermediate reference pictures are periodically included. This leads to a particular structure for a sequence of pictures, which can be seen in a Group-Of-Pictures (GOP). An example GOP structure is given in Figure 8.5, which shows various types of pictures.

I-Frames are compressed as completely independent (intraframe) frames, thus only spatial redundancy is exploited for compression. For P- and B- frames (the *inter* frames), temporal redundancy is exploited, where P-frames use one temporal reference, namely the previous reference frame. *B-frames* use both the previous and the upcoming reference frame, where reference frames are I-frames and P-frames. The top of Figure 8.5 shows the transmission order of the pictures. Further, the size of rectangular blocks at the bottom of the figure indicates the amount of bits contained for each picture type. As can be seen, B-pictures are most compressed and are never used as a reference for other pictures. A *Group Of Pictures* (GOP) implicitly defines the processing order of the video frames. Since B-frames refer to future references frames, they cannot be (en/de)coded before this reference frame has been received and processed by the coder (encoder or decoder). There-

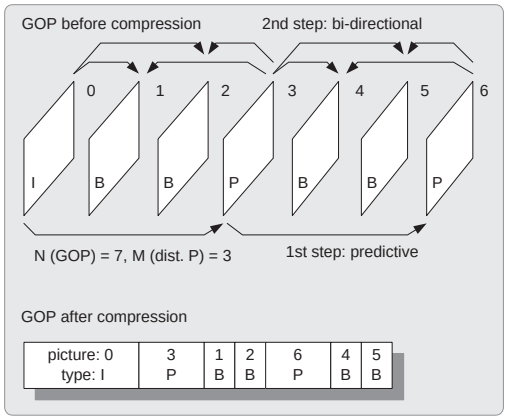


Figure 8.5: *Picture types in a Group-Of-Pictures (GOP) used in MPEG.*

fore, the video frames are processed in a reordered way, e.g., "IPBB" (transmit order) instead of "IBBP" (display order).

In 1984, Plompen and Booman [668] provide an overview of the picture compression techniques as explored at the Neher Laboratory. A key project that is described for development of the first professional compression system was the COST211bis project, which proposed a DPCM coding system with a frame memory, 4-bit quantizer and Variable-Length Coding (VLC). The system operated at three modes: 64–256 kbit/s for still picture or slow scan, $n \times 384$ kbit/s for video conferencing and 34 Mbit/s for video distribution. In 1987, Plompen, Biemond and Heideman [682] describe the COST211bis codec in more detail. In the system, motion compensation is added and a prediction filter is added in the loop, after the frame memory. The authors provide metrics for performance of moving sequences, such as the mean quantizer step size, the mean value of zeros prior to a nonzero coefficient and the mean value of nonzero coefficients.

For a similar compression system (H.261), Barnard, Sankur and Van der Lubbe [704] study the statistics of the transform coefficients. The main conclusion is that for a hybrid coder in inter mode, the DCT coefficients can best be modeled by a Generalized Gaussian distribution. The authors conclude that 16-level Lloyd-Max quantizers with different design parameter setting are robust for real image sequence data.

The DCT was also studied by Van der Schaar and De With [730] in 1997, where they compared several fast DCT algorithms. Several complexity criteria are used, such as the number of stages, registers and the resulting SNR quality. A new multiplication-free DCT is proposed for low-cost or low-rate systems, yielding a moderate quality at a much lower complexity (fewer registers, low delay and no multipliers).

Hekstra [717] studies the duality between filter design and frequency-based transforms such as the DCT and Fourier transforms. He presents an idea to design the filter basis functions in an alternative way. Instead of making all coefficients zero outside the block and leaving the coefficients inside the block unchanged, he proposes to use linear programming to compute those remaining frequency coefficients such that they give mini-max error in the spatial domain.

Van der Vleuten and Oomen [723] compare the coding gain of 512-band transform coding and 64-band subband coding with prediction with filters of 1024 length. It is proven that subband coding with prediction performs close to optimal, and if sufficient prediction coefficients are applied, the subband coder outperforms the transform coder. The cross-over point is at approximately 80-100 coefficients.

8.2.3 Motion Estimation Algorithms

For motion estimation within a hybrid coder, block matching is commonly used. The block size for ME is usually 16×16 pixels. The metric for comparing blocks in order to find the best vector is typically *Sum of Absolute Differences* (SAD). The block given by a minimum SAD value yields the best vector for motion estimation. This vector represents a translation model for the motion. More advanced standards also allow rotation as motion, such as the affine motion model in MPEG-4. If all possible vectors are evaluated, the technique is called full-search block matching. Currently, a multitude of *fast block-matching* algorithms have been published. Popular examples of such algorithms for hybrid DCT coding are Three-Step-Search (TSS) in various forms, Logarithmic Search, One-Time-at-a-Search (OTS) and recursive block matching. Such algorithms typically evaluate only 10-25 vectors (or even less), instead of a few hundred for full-search ME. It is emphasized here that ME and MC can be performed using previous pictures only, or using both past and near future pictures (*bidirectional ME/MC*, see the earlier discussion on MPEG).

The initial research on motion estimation concentrated on block matching algorithms for emerging video standards (e.g., H.261). In 1985, Plompen and Boekee [672] compare three different motion estimators for a hybrid video conferencing system. The estimators are cross-estimator, the One-at-a-Time-Search (OTS) and a Truncated Brute Force search (TBF) technique. All estimators perform equally well for artificial data, but for real video data, the cross-estimator performs reasonably, while OTS is unacceptable and TBF behaves correctly.

Plompen, Groenveld and Boekee [677] exploit the concept of motion estimation in the transform domain and compare this with the regular hybrid codec. This idea potentially saves the inverse transform in the encoder prediction loop. The paper addresses the measurement of displacement in the transform domain via a decomposition into sparse matrices using the ordered Hadamard transform. A transform weighting function is also incorporated. The obtained results do not yield any performance improvement.

Queluz and Macq [707] propose an improved block-matching for motion compensation by taking a region-based approach. The regions are found with a binary mask function that is created by pixel-based frame differences. Median filtering of the motion field at the end provides a much more homogeneous motion field. The algorithm distinguishes itself with a low cost of transmitting the compressed motion field.

An alternative class of motion estimation algorithms is formed by pixel-based motion estimators. This class offers an increased prediction accuracy of the real video signal. Attention is also payed to obtaining homogeneous motion vector fields. Biemond, Looijenga and Boekee [679] study a more advanced form of motion estimation: a *pixel-recursive* Wiener-based displacement estimation algorithm. The concept is that the recursive (displacement) update and the linearization error are assumed to be samples of stochastic processes. The process can then provide a least-squares estimate of the update using N observations. The proposal was successfully evaluated in a video conferencing compression system and compared with other pixel-recursive algorithms, e.g., with processes without initial estimate.

Driessen and Biemond [693] improve a Kalman-based estimator for the motion field between two images. The improvement is on reducing the estimation rate to reduce the sensitivity of the algorithm for local linearization errors. The proposal is tested with a textured image and introducing artificial motion. Ter Horst [694] discusses briefly multi-resolution compression, and conjectures that with a reduced number of signal components, a motion compensated prediction for a signal component can still be obtained. The loss of prediction quality largely depends on the type of filters in the filter bank.

Franich, Lagendijk and Biemond [715] have an alternative to come to homogeneous vector fields. They suggest using genetic algorithms to grow homogeneous fields with actual motion fields as a *chromosome* input signal. First comparisons with full-search matching show similar MSE values. The application is related to stereo video image sequences. The same authors come back on stereoscopic imagery in [721], where they propose a technique for estimating disparity errors. The model for a disparity space image (DSI) is introduced. The problem focuses on finding a path in the DSI using a genetic algorithm. Experimental results show that stable paths in the DSI can be found after limited iterations without any spurious disparity errors.

With respect to implementations, Frimout, Driessen and Deprettere [708] propose a parallel architecture for a pixel-recursive motion estimation algorithm. The system is an array of processors, where each processor consists of initialization, a data-routing part for accessing previous frames and an updating part. The initialization performs a prediction of the motion vector. The benefit of the proposal is the parameterized and structured design of the system.

Kleihorst and Cabrera [733] study the VLSI realization of motion estimation where the reference images are stored in the compressed DCT domain. As a result, the

motion estimation and compensation is performed in the DCT domain. They analyze the first row and column of the DCT coefficient matrix for a limited number of vector candidates. However, a clear ME algorithm is not presented. The authors claim that the hardware efficiency is comparable to existing solutions but offers other advantages.

In 2002, Mietens, De With and Hentschel [749] address another design parameter to motion estimation, called complexity *scalability*. They study MPEG algorithms that are suited for a wide range of applications including mobile devices with limited computing power and memory. At the first stage in the encoder, a simple recursive ME is performed on a frame-by-frame basis to have an early estimate of the motion. Secondly, the obtained vector fields are scaled and combined to find the vectors that refer to usual the MPEG encoder processing order. The optional third stage refines the found vectors. Experiments show that in high-quality operation, the system is comparable to full-search block matching (32×32), although with a much lower computational effort.

8.2.4 Subband Coding

During the 1980s and with the growing importance of HDTV, an alternative decorrelation technique emerged, called *subband coding* [111]. Instead of using non-overlapping pixel blocks for signal transformation, the video (or audio) signal spectrum is decomposed in the encoder into subbands by using filter banks. Each spectral band is critically (re-)sampled such that the resulting subband data is of the same size as the original signal. Subsequently, each band is individually quantized and compressed. The decoder decodes these streams and performs up-sampling and interpolative filtering, using the appropriate filters matching with the encoder filters. A simple example is readily understood as follows. Each spectrum is halved and divided into a low-pass part and a complementary high-pass frequency part, using so called half-band filters. This splitting can be repeated for each subband, leading to an increased number of bands where appropriate.

Figure 8.6 portrays a split of both the horizontal and vertical spectrum, using a two-stage filter bank. This four-band system has been popular for experiments with HDTV signals, because the LL-band offers a signal that resembles a standard TV signal. It should be noted that the filters need to be carefully chosen and matched with each other. For example, it is required that the overall spectra add to unity response over the total signal spectrum, despite the use of non-ideal filters with a finite impulse response and reasonable transition band. A special class of filters satisfying this is the *Quadrature Mirror Filter* (QMF) class [111], which generates alias components in each band, but in such a way that when the bands are added, the alias components are mutually canceled by neighboring bands.

The compression of each band is carried out as follows. Firstly, the low-frequency band contains again a picture, but now smaller in size and with a restricted spectrum. This picture is typically compressed with DPCM or transform coding. Secondly, the sidebands and high-frequency bands contain refinement or residual

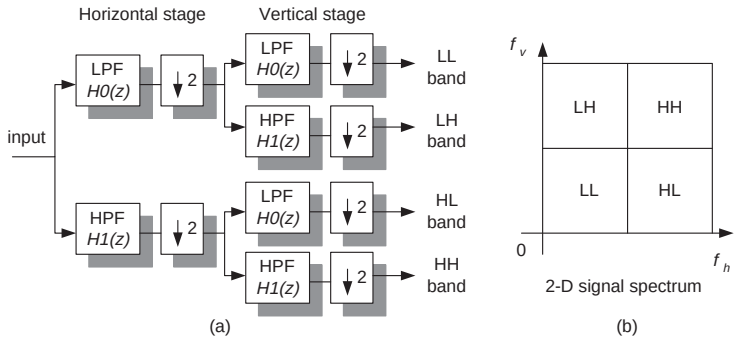


Figure 8.6: *(a) Two-stage filter bank with half-band filters and (b) corresponding 4-band video spectrum.*

high-frequency components. These bands are commonly quantized and compressed only, since they are spectrally white, and hence uncorrelated. The contents of these sideband signals are rather noisy, with this noise concentrated on edges or textured areas. Since these bands contain many zeros, run-length coding is typically used for such signals.

The subband coding principle is sometimes extended with motion compensation in order to compress in three dimensions. Alternatively, temporal decomposition can be performed, but due to the filter length, this becomes complex rather quickly. The attractive aspect of subband coding is the scalable frequency representation of the video signal. The decomposition is scalable from nature, and the quality of the signal can be smoothly changed with the number of subbands that are actively used or transmitted.

In the context of Information Theory research in the Benelux, the first paper on subband coding is from Westerink, Woods and Boekee [678] in 1986. They present a new two-dimensional subband coding system that splits the signal into 16 parallel subbands. The sample outputs are jointly combined into a vector that is then compressed with Vector Quantization (VQ) trained with the LBG algorithm. Three systems are compared, one with DPCM in each subband, one using adaptive DPCM and a system employing VQ. The latter performs significantly better by several dBs in SNR; it can also operate at about 0.5-0.6 bit/pixel for obtaining 30 dB SNR.

In 1987, Westerink, Biemond and Boekee [685] report on the same system with an improved approach, where they integrate the DPCM principle for predicting the 16-element vector. The VQ is subsequently applied to the difference signal after prediction. The performance leads to good quality pictures in the area of 0.4-0.7 bit/pixel.

Westerink, Biemond and Boekee [687] continue this research with a detailed analysis of the quantization errors in a subband coding system. The paper considers

errors of the QMF filter bank, signal errors, random errors and aliasing errors. The QMF error is small, and the aliasing errors are also small for 16-tap filters. The signal error comes from the subband decomposition and relates to sharpness. The random error appears everywhere (around edges more pronounced due to quantization) and, though smaller than the signal error in the MSE sense, is equally important to that error in the perceptual sense. The study proposes a reduction of the random error by applying adaptive quantization.

Van der Waal, Breeuwer and Veldhuis [684] apply subband coding for compression of music signals. The music signal is divided in 26 bands, based on Quadrature Mirror Filter (QMF) cells, using sets with complementary low-pass and high-pass filters. The quantization is however essentially different, because it is based on the masking properties of the human auditory system. The authors explain the possibilities of both simultaneous (frequency) masking and temporal masking. The bit allocation is dependent on the chosen quantization. Each subband signal is based on block companded quantization (BCPCM), relying on stationarity in the block of samples (32 samples). The scaling is expressed with 8 bits. The bit-allocation for each band is chosen such that it depends on neighboring band contents. The bit rate per band varies between 4.64 bits/sample for low frequencies and 1.58 bits/sample for high frequencies.

The concept of subbands can be generalized with *wavelets* as basis waveforms, leading to wavelet coding [111] for video compression. A wavelet is a basis waveform that is scaled and shifted to form a basis for signal decomposition. The wavelet can be chosen to match particularly with the signal statistics, so that potentially a higher compression can be obtained. This appears to be true in practice as well, and therefore wavelet coding has been adopted in the new JPEG2000 standard (successor of the regular JPEG standard [92]) and the MPEG-4 still picture compression standard [113].

In [751], Iregui, Meessen, Chevalier and Macq discuss an efficient way for delivering JPEG2000 data in a client-server architecture. They propose a bandwidth adaptive parsing of JPEG2000 compressed data streams such that users can efficiently browse compressed images. The inherent spatial scalability of wavelet/subband decomposed images greatly eases the implementation of server/client-efficient browsing scenarios.

8.2.5 Segmentation-based Compression

In image processing, data regions are clustered such that segments of similar statistics are obtained. These properties may be exploited for image compression. One of the first standards exploiting this actively is the MPEG-4 standard [113], in which video *objects* can be compressed and manipulated independently. The following papers gradually grow towards this standard.

Renes and De Pagter [669] exploit spline approximation and segmentation for

studying new forms for image data compression. The application area is remote sensing, multi-spectral imaging from aircraft or satellites. Some forms of pixel interval classifications are given for simple segmentation. For compression of the segments, the vertex definitions and *a-priori* information for continuations of edges are presented, to come to efficient compression of the closed contours. The paper gives first results using 240×240 pixels of 4 colors and comes to at least 2 bit/pixel.

Vanroose [729] studies image understanding concepts with the aim to improve image compression. In a historical overview, the author comes to the logical conclusion that understanding and finding objects is relevant. Afterward, the IUE (Image Understanding Environment) of an American ARPA project is described which contains e.g., a toolbox with segmentation algorithms. At the end, an experiment is shown where an object is segmented (IUE) and compressed with only 600 Bytes. For comparison, the JPEG compression algorithm was also applied to the same image, requiring between 1.5 and 10 kBytes.

Desmet, Deknuydt, Van Eycken and Oosterlinck [727] employ motion estimation for segmentation. The estimation process leads to a low-resolution block-based segmentation. This low-resolution step is followed by a high-resolution segmentation on pixel basis. The pixel assignment follows from a cost function incorporating shape, motion and color information. The segmentation is based on region growing. The compression system employs motion-compensated prediction and an Optimum Level Allocation (OLA) algorithm with arithmetic coding. The results are still immature, but announce the upcoming MPEG-4 standard for object-oriented compression.

Wuyts, Van Eycken and Oosterlinck [728] follow the same line of research and work with motion estimation for object-based compression as well. The motion estimation is extended to five dimensions (2 translation, 1 rotation and 2 for 2-D stretching). The final step involves calculating cost functions for all objects. Each pixel gets as cost the maximum of the cost of neighboring pixels and its own displaced frame difference. The segmentation algorithm shows a limited performance for fast moving backgrounds and the cost function is problematic in flat regions. The authors conclude correctly that temporal tracking should be included for improved segmentation results with more stability.

In [738], Desmet, DeKnuydt, Van Gool and Van Eycken re-use the OLA scheme for the compression of texture in 3-D scenes. They introduce view-dependent compression of dynamic textures for e.g., 3-D games or simulations of dynamic systems. The model set-up applies a mapping of the 3-D world onto the image plane using the distance, slant and tilt angles (d, s, t). The system codes iteratively until a quality criterion is satisfied. A Gaussian directional subsampling filter improves the quality further. The authors report on an experimental simulation of a virtual room walk-through where they require only 0.79 Mbit/s for the dynamic texture, whereas MPEG-2 video would require 3.17 Mbit/s with the same quality.

Finally, Farin, De With and Effelsberg [753] study efficient compression of the background for MPEG-4 compression with *sprites*. Instead of one large sprite, they use a counter-intuitive approach where they split the background reconstruction into several independent parts. The optimal partitioning is found by considering the perspective distortion when the camera pans far away in a side direction and introducing scaling factors for the video data. The authors report on achieving a factor three less video data for background compression than the recommended standard MPEG-4 sprite model.

8.3 Quantization Strategies

In this section we first summarize papers that have contributed to the development of theory and practice of scalar and vector quantization strategies. We then describe those papers that consider the optimal usage of quantizers in combination with decorrelating transforms, i.e. the bit allocation problem.

8.3.1 Scalar and Vector Quantization

The development of scalar quantization techniques has a long history, as was already referred to in Section 8.1. Especially the optimality of certain type of quantizers has been a problem thoroughly investigated. In 1990, Györfi, Linder and Van der Meulen [691] address the problem of asymptotic optimality of quantizers. In particular they consider nonuniform quantizers with an infinite number of quantization representation levels. They generalize a well-known theorem by Gish and Pierce on asymptotic optimal quantization by proving that the conditions on the density of the PDF of the signal being quantized are less strict than assumed by Gish and Pierce.

Multiple description (MDC) quantization is the approach where a single source is quantized using two (or more) separate and independent quantizers at rate R_1 and R_2. The MDC quantizers are such that they individually perform close to rate-distortion optimality, but at the same time the combination of the two descriptions also gives a close to rate-distortion optimal result, in this case at rate $R_1 + R_2$. In 2002 and 2003, Cardinal [748, 755] investigates the problem of entropy-constrained assignment of quantizer indices, building on the earlier work of Vaishampayan, among others. The author proposes an optimization procedure to find the multiple description quantizer index assignment, given entropy constraints on the MDC quantizers. The resulting MDC quantizers outperform earlier published results in international literature.

Vector quantization has been a re-appearing theme not only in international literature, but also at the WIC Benelux Symposium. Over the years interest has been focused on how to apply vector quantization as a stand-alone technique, in combination with decorrelating methods such as DCT and subband compression, or even within a standard motion-compensated video compression system. Also

some work appeared dealing with reducing the complexity of vector quantization.

In 1984, Boekee and Van Helden [667] addressed the problem of efficient searching in vector quantization codebooks. One of the problems in searching for the best vector from the codebook is the unstructured nature of the VQ codebook. Essentially this requires the evaluation of each and every codebook vector as possible compressed representation of the (uncoded) vector under consideration (full-search VQ). The complexity of full-search VQ is exponential in size of the codebook. To limit the complexity of the searching process, the authors propose to use a tree-structured codebook (TSVQ) representation. In TSVQ, the codebook is organized in a tree, each node of which contains a codevector. The actual codebook is defined as the set of codevectors contained in the leaf nodes. The search begins at the root node, and progresses along child nodes until the best leaf node is reached. Thanks to this structure, the complexity of TSVQ is considerably reduced, compared to full search VQ. Although the paper itself lacks experimental validation, many other authors have put forward similar and more elaborate ideas to reduce the VQ encoder complexity [66].

In [745], Cardinal also addresses the problem of complexity of tree-structured vector quantizers (TSVQ). The unique perspective offered in this paper is that not only the user specifies a bit-rate constraint R – as is usually done – but also a computational complexity constraint C. The author defines a complexity-distortion curve $D(C)$ as the curve of minimal distortion that can be obtained by a coder with average complexity C at rate R. If the complexity is infinite, the usual rate-distortion curve is obtained. The author investigates properties of the complexity-distortion curve, and proposes a way to solve the optimization problem encountered in the practical usage of the complexity-distortion concept. Experimental results show the feasibility of the concept, yet the author concludes that the complexity of the optimization might be prohibitive in practical cases of interest.

Another approach to reduce search complexity in VQ is addressed by Cardinal in [737]. The proposed approach encompasses mean-shape-gain VQ. Mean-shape-gain VQ encodes separately the mean and the length – or gain – of the vector using two scalar quantizers. The mean-removed normalized vector is called the shape, which is encoded by an index in a shape codebook. The author proposes efficient strategies for finding the proper entry in the shape codebook, using angular and spherical constraints.

Research by Van der Vleuten and Weber [701, 709] around 1992–1993 considers other vector quantization variations known as trellis waveform coding (TWC) and trellis-coded vector quantization (TCVQ). As in all trellis coding approaches, the waveform coding or quantization operation is carried out by a finite state machine, where state transitions specify the codebook symbols to use for representing the source symbols. In the work of Van der Vleuten and Weber the focus is on finding constructive design methods for these trellises. The resulting construction methods are practical and – at the same computational complexity – give a higher performance than the ones proposed up to that moment.

In the period 1985–1994, several papers have appeared that address VQ in combination with other compression techniques [671, 678, 674, 680, 718]. In [671], Van Helden and Boekee describe a video compression technique based on inter-frame conditional replenishment, and intra-frame VQ. Parts of a video sequence that do not change substantially, are copied from the previous frame; since the introduction of MPEG, these two block types are known as non-motion compensated non-coded macroblocks, and intra-coded macroblocks, respectively. The difference with today's MPEG standard is that the authors propose to compress the intra-coded macroblocks with vector quantization.

Woods and Hang propose the unification of predictive compression and vector quantization in [674]. A predictive tree encoder is used, in which the ordinary scalar quantizer is replaced by a vector quantizer. The basic idea of predictive VQ is to use a predictive filter to remove predictable redundancy in the image data – much like DPCM on block basis –, and then encode the resulting prediction error. In order to remain computationally feasible, two implementation variations were proposed, namely sliding block VQ and block-tree VQ. The latter is essentially a TSVQ scheme operating on image blocks.

In [680], Breeuwer proposes to quantize DCT coefficients of 8×8 blocks using VQ. The size of the VQ vectors is identical to the 8×8 size of the DCT blocks. To limit the complexity of the VQ codebook search, cascaded VQ (CVQ) is used. In CVQ, the 64 DCT coefficients are quantized with a cascade of VQs, each of which has a low complexity. Furthermore, the energy of the DCT blocks is used as a means for adaptively selecting the number of stages in the cascade and the particular DCT coefficients to be represented in the VQ vector.

Finally, Shi and Macq [718] propose to use vector quantization in the transform domain. Rather than using the DCT transform, the authors propose to use a non-separable transform that respects edges in images and avoids blocking artifacts. The authors also propose to design the non-separable transform using a genetic algorithm. In the paper, conceptual solutions are worked out, but concrete results are left for future research.

8.3.2 Video Quality and Optimal Bit Allocation

Given the structure of a certain compression system – be it a subband compression system or a motion-compensated DCT-based video encoder – the challenge is to perform quantization of the (usually transformed) image data such that an optimal trade-off between rate and (visible) distortion is obtained. We summarize here two categories of papers, namely (i) papers that focus on the quality assessment of image/video compression systems in a particular application, and (ii) papers that aim at finding ways to (perceptually) optimize DCT-based compression methods.

In [665, 673], Huisman evaluates the performance of several transform-based im-

age compression techniques for spaceborn imagery. These algorithms, some of which were developed by ESTEC/NLR a number of years before JPEG was standardized, first describe the mathematics of transform compression and the effects of quantizing the transform coefficients. On the basis of these mathematical models, procedures for optimal bit allocation are proposed. Theory is verified with experimental results on synthetically generated Gauss-Markov random fields. Over the years, the theory described in these and similar papers has become basic knowledge of the modern image and video compression engineer.

Image compression is never a stand-alone operation, but it is usually part of a much larger image acquisition and processing system. In 1993 and 1996, Slump [711] and De Bruijn, Van Heerde and Slump [722] describe a physical model for the image formation and rendering in a cardiovascular X-ray imaging system. Based on the modulation transfer function of the imaging system and a Poisson model for the acquisition noise, relevant parameters could be quantified, such as the maximum spatial resolution and signal-to-noise ratio of the imagery before compression. Based on these parameters, the appropriate JPEG compression options and preprocessing (subsampling, interpolation) could be selected, and bounds on the achievable compression were proposed. Visual studies were done to evaluate the quality of the resulting compressed images.

In order to perceptually optimize the performance of video compression algorithms, an objective perceptual image/video quality model is required. Several approaches have been published that base the quality model on known spatio-temporal signal processing properties of the human visual system [686, 688, 720, 732], but alternative approaches avoiding the explicit modeling of the human visual system also exist [731]. In [688], Macq and Delogne investigate the use of spatial frequency-weighting in developing a measure of video quality. They first propose weighting functions for luminance and chrominance color components. They then use these functions for defining weights of DCT/Fourier coefficients, much like this is routinely these days done in JPEG and MPEG-compression systems. The main contribution of the paper is the compatible extension of the (then often used) ITU-T recommendation 451-2 for measuring analog television quality to digital video frames.

Stuifbergen and Heideman [683, 686] also propose frequency-weighting models, but they do not limit their models to spatial processing only, but propose to also include temporal processing of the human visual system in the model. Their models have the opportunity to use specific sensitivity properties of the human visual system in different spatio-temporal frequency bands. The focus of the work is defining spatio-temporal frequency bands such that moving smooth edges are properly represented and that motion of a smooth edge can reliably be estimated.

In [720], Westen, Lagendijk and Biemond propose a spatio-temporal quality model that includes linear and nonlinear processing effects in human vision. Properties that are included in the model are (i) the gamma of the display device, (ii) the transfer function of the eye's optics, (iii) the temporal integration in retinal nerve

cells, (iv) nerve cell inhibition, and (v) saturation effects. The quality of a compressed image/video is then defined as the quadratic difference between the output of the model when the original image/video sequence and the compressed original image/video sequence as input. The proposed model is evaluated by correlating numerical model scores and test panel scores using MPEG compressed video.

Westen, Lagendijk and Biemond [732] extend their work by including non-orthogonal spatial-frequency decomposition into the quality model, based on the work of Simoncelli and Adelson. Contrast sensitivity and spatial masking are made frequency dependent by including a sensitivity and masking model specialized to each frequency band. Furthermore, their model includes the notion of "smooth pursuit eye movement (SPEM)", which is the capability of the human visual system to stabilize moving objects on the retina by tracking the movements. Since SPEM have considerable influence on the perceived temporal frequencies, motion estimation needs to be included in the quality model as a means to emulate SPEM.

The work by Beerends and Hekstra [731] defines an objective video quality model without explicitly modeling the human visual system. Departing rather radically from common approaches to image/video quality modeling, they propose to first measure a large number of simple low-level spatio-temporal features from original and compressed video, and then to select the smallest number of (combinations of) features that best predicts the image/video quality as assessed by test panels. This selection process is similar to feature selection, linear regression, and dimension reduction in pattern recognition. The authors compare their model with a ANSI model, and provide experimental evidence that the proposed model is more feasible and superior.

The final category in this section is the one dealing with papers that describe algorithms for achieving optimal (numerical or perceptual) quality of (DCT-)compressed images [700, 724] or MPEG-video [695, 710, 713, 719, 750]. In 1996, Westen, Lagendijk and Biemond [724] propose the Transform Coding Quantization Feedback (TCQF) algorithm for DCT-based compression systems. The TCQF algorithm can be used for spatial noise shaping, as opposed to the usual frequency noise shaping realized by weighting the quantization noise on DCT coefficients. Spatial noise shaping allows for placing quantization noise at those pixels positions (in DCT blocks) where it is visually least disturbing, e.g., textured areas. Although the thus formulated quantization problem is computationally complex, an efficient optimization algorithm is proposed by the authors. Results show that the algorithm greatly reduces "mosquito" quantization noise in JPEG compressed images, while decoder compatibility is maintained.

Keesman [695] proposed to see the bit-assignment problem as a constrained optimization problem. Making use of Lagrange multiplier theory, the author constructs a quantizer assignment procedure for an image compression technique known as "Adaptive Dynamic Range Control (ADRC)". Although the ADRC compression technique itself has not found practical usage and has been superseded by JPEG and MPEG, the method of Lagrange multipliers has found widespread use in state-

of-the-art signal compression, since many of the compression system rate control problems can be formulated as constrained optimization problems.

For instance, in 2002 Farin, De With and Effelsberg [750] propose to formulate the optimal compression of MPEG I-frames as a Lagrange optimization problem. Three DCT quantization parameters are incorporated into the Lagrange optimization model, namely (i) adaptive quantization, (ii) coefficient thresholding, and (iii) DCT coefficient amplitude reduction (CAR). The authors conclude that, although the resulting Lagrange optimization may be too complex for real-time systems, the compression results are excellent and can be regarded as a reference for lower complexity adaptive quantization procedures.

After the finalization of the MPEG video compression standard, in the period 1990–1995 a lot of attention was devoted to the problem of quantization and associated (constant or variable) rate control in MPEG.

De With and Nijssen [700, 713] consider the problem of rate control within the application contexts of digital video recording and editing. In these contexts it is advantageous for trick play, robustness and error concealment to compress all video segments, frames, or pairs of frames in the same amount of bits. In [700], the authors describe a feedforward buffered DCT-based video compression system. In the proposed intra-frame video compression system, data is analyzed prior to compression such that the number of bits produced by the compression system per video segment (i.e. a part of a video frame) can be accurately predicted, and hence a feedforward buffer control can be implemented. The focus of the work is the trade-off between complexity of the analysis procedure, the size of the video segment, and the resulting SNR quality. Experimental results indicate that for intra-encoded video frames, a feedforward buffer control can perform comparably to a (more conventional) feedback buffer control in case video segments include at least 30–60 DCT blocks.

De With and Nijssen consider a related problem of feedback rate-control in [713]. The aim of this work is to obtain an approximately constant quantization coarseness under the constraint of a fixed bit rate for a frame pair. Two control modes are introduced, namely a “fast mode” for rapidly changing signal statistics, for instance after a scene change, and a “stationary mode” that is active when signal statistics are temporally slowly varying. Experimental results show the feasibility of the proposed rate control.

Research of Van der Meer, Biemond and Lagendijk [710, 719] also focused on constant quality MPEG-compression, in their case without a rate constraint. Consequently, the resulting bit rate is variable in time. In [710], a constant-quality MPEG-1 compression system is proposed. Since video frames are not stationary, the quantizer coarseness also needs to vary spatially to achieve constant quality. The authors propose a “locally weighted SNR” (LWSNR) measure to determine video quality on a DCT-by-DCT block basis. The MPEG quantizer coarseness is then controlled in such a way that the LWSNR measure is spatially and temporally

constant.

Constant quality MPEG encoders produce a variable bit rate (VBR). Networks can exploit the variability in bit rates of multiple sources by using statistical multiplexing. In [719], Van der Meer, Biemond and Lagendijk propose a model for describing VBR MPEG video streams. VBR streams are usually smoothed ("shaped") slightly as so to reduce the very short term variability of the produced bit rate and only expose the long term variability to the network. The authors propose a bit rate smoothing procedure that makes use of knowledge of the MPEG encoding parameters, such as the group-of-pictures (GOP) structure. An analytical model is proposed that describes the smoothed VBR traffic well.

8.4 Hierarchical, Scalable, and Alternative Compression Techniques

Alongside the mainstream research on hybrid DCT compression, substantial efforts have been given to image and video compression within certain application or transmission constraints. We subsequently describe and summarize progress by Information Theory researchers in the Benelux.

- *Hierarchical compression* became popular in the early nineties, because HDTV was widely studied in Europe. For practical reasons, it soon was clear that standard-definition and high-definition television (HDTV) could exist side by side in the same communication infrastructure. This resulted in the ideas of compatible and hierarchical compression.
- With the growing complexity of encoders and decoders (e.g., HDTV), the use and cost of memories increases simultaneously. At the end of the nineties, it becomes appropriate to *embed compression* in video memories.
- The increasing diversity in video products causes *complexity scalable* video compression and processing to become attractive.
- Video compression in *networked environments* is relevant because the Internet and ATM networks in telecommunication systems emerged during the 1990s. The network interface and the overall error robustness of packetized compressed bit streams plays an important role in this research.
- *Alternative compression* techniques, aiming at entirely different compression philosophies or particular application contexts.

8.4.1 Hierarchical Compression

The occurrence of this hierarchical compression technology is closely related to the emerging of HDTV signals in broadcasting and the desire to generate a standard TV signal from this. This means that at least a two-layer compression system is required with a low-quality output signal and an enhancement signal that lifts the total quality to sufficiently high level. Several papers in this area are based on

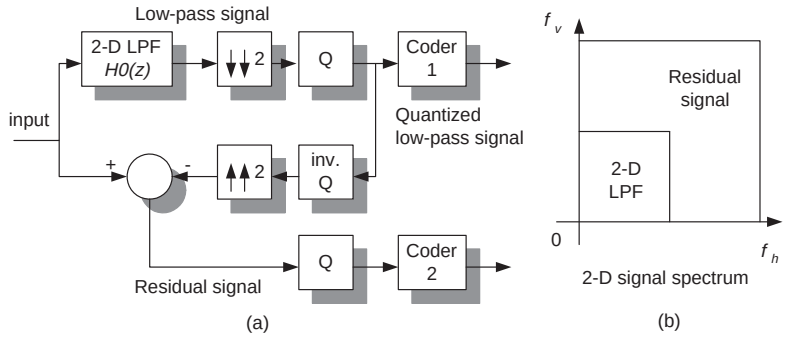


Figure 8.7: *(a) Two-stage pyramidal compression and (b) corresponding two-layer video spectrum.*

Laplacian pyramid compression. A simple two-layer example of this principle is shown in Figure 8.7.

The video signal is 2-D low-pass filtered and down-sampled two-dimensionally. The low-quality base layer signal globally represents TV quality, which is quantized and compressed accordingly. The signal is reconstructed and up-sampled to the higher resolution again. At this level, the low-frequency part is subtracted from the total spectrum, yielding a residual signal that has basically energy in the high-frequency areas of the spectrum. The advantage of this approach is that the errors of the base layer occur in the enhancement layer, so that the total quality is ensured. On the other hand, due to the compatible compression in layers, more sample processing and memory is required (especially when combined with motion estimation) than in the original case, because the base layer is compressed twice.

In 1990, Bosveld, Lagendijk and Biemond [692] study hierarchical compression of images for B-ISDN, where it is likely that extended-quality TV (EQTV) and HDTV are both communicated in the same system. The paper deals with two progressive 28-band subband coding schemes for HDTV, the Refinement and Selection system. Both schemes code HDTV in 135 Mbit/s, while the EQTV signal is compressed with e.g., 45 Mbit/s. The Refinement takes the low-frequency part (for EQTV) as a prediction for the total signal. In the selection system, HDTV is compressed without compromises. The EQTV signal is derived via a selection of suitable subbands. The performance of subband decomposition is compared with DCT transformation.

Vandendorpe and Macq [696] address the compression of moving video with hierarchical subband and entropy coding. Each band is separately compressed from others, even when motion compensation is considered in addition. The authors emphasize compatible transmission with progressive transmission and universality of the entropy coder. A special Universal VLC coder is designed that codes the

MSBs from all corresponding bands in a sequence. The algorithm for the MSBs is truncated runlength coding. The LSBs are not compressed, due to their randomness. At a certain point, the skip step, the coder switches to uncoded data. This paper is an early attempt to the fine-grain scalability compression that was later adopted in MPEG-4.

Bosveld, Lagendijk and Biemond [703] come back on hierarchical compatible compression with new spatio-temporal subband coding schemes. Non-rectangular decompositions are applied, and the schemes can handle both interlaced and progressive video signals. Diamond-shape frequency bands are studied, mainly to get detailed options for the vertical-temporal hierarchy. For the filter banks, QMF filters are used. Experiments show that longer filter lengths provide higher HDTV quality. Despite the flexibility in the vertical-temporal decomposition, the compression performance for the interlaced HDTV signal is much lower. The authors conclude that the full system may be of too high complexity, and a reduced temporal hierarchy would be sufficient.

Belfor, Lagendijk and Biemond [706] also study an alternative to subband coding: sub-Nyquist sampling of HDTV signals. This refers to the MUSE and HD-MAC transmission systems for HDTV, which both have analog output and only rely on advanced filtering and sub-sampling. The paper focuses on the moving parts of the sequence. When having motion, the sub-sampling pattern can support motion velocities in discrete directions. When considering critical velocities, the sub-sampling can be made adaptive. Since the obtained sampling pattern varies locally, this may pose problems when subsequent digital compression would take place. This effect will be reduced when the motion estimator produces a very consistent homogeneous motion field. The results are good if the speed of motion is sufficiently high, otherwise the non-adaptive sampling should be taken.

Leduc [702] also addresses TV and HDTV compression and concentrates on the optimum control of image quality, while also monitoring the buffer occupancy. The best quality is obtained for television if the system reacts slowly to the varying image content. The paper proposes the design of a PID controller for buffer regulation, but now the operation should be tuned to optimum control of both buffer occupancy and the quality level. To this end, the source coding parameters are modeled as stochastic processes (e.g., bit rate) onto which Kalman filtering can be applied. The controller can both learn and control in a dual mode of operation. The learning involves the derivation of correlation coefficients of the state variables. The paper does not provide results of experiments.

8.4.2 Video Compression for Embedded Memories

With the growing complexity of systems and the increased processing in the time domain, the use of memory in video compression systems accounts roughly for half of the system costs. A number of papers deals with compressing the intermediate results, such as reference frames in an MPEG coder, using an alternative technique. The insertion of embedded compression is not trivial, since the embed-

ding should not interfere with the surrounding system.

De With and Van der Schaar [734] are the first in exploring embedded memory compression in MPEG coders. In the MPEG coder, the reference frame memories are compressed using a low-cost block-adaptive prediction system. Using variable quantization with corresponding bit-allocation, the bit cost of compressed data can be easily recovered when the quantization factors are known. Besides this, the data is compressed in fixed-length segments. Both aspects enables easy compressed block data retrieval for motion compensation in the MPEG coder. The authors claim a small reduction of picture quality (compression 2-2.5) and present a remedy for asymmetric quantization, thereby avoiding quality reductions in the coding loop for long GOPs. The system has been implemented in a commercial IC. The same authors come back on this theme for HDTV compression systems in [739]. In the second paper, they improved the system with embedded DCT compression using feedforward buffering for small segments. The new scheme can obtain a compression factor of six. The system can be tuned to several qualities and can efficiently re-use the MPEG quantization parameters.

Kleihorst, Van der Vleuten and Apostolidou [743] propose a scalable compression technique and a hierarchical storage medium for maximum use of the available storage space. If a new image is offered, previously stored images are automatically re-quantized in place, without the need to extract them from the memory. The DCT coefficient data is split in stages from MSB down to LSB. The "swimming-pool" memory uses a hierarchical organization according to the previously mentioned stages. If memory space becomes scarce, the least relevant data is simply overwritten. Experiments show that the PSNR decays from 42 dB for one picture to 25-35 dB for 12 pictures of 512×512 pixels for a 10-stage memory using 32-bit wide data spaces per stage. Van der Vleuten comes back on this research in [752], where he improves the result taking into account visual quality. The new solution improves the minimum quality by 2.3 dB SNR, whereas the average quality never decreases by more than 0.3 dB. The improvement is obtained by using the absolute values of the distortion measures for a data significance decision, so that the image of highest quality is always taken for inserting new data.

8.4.3 Complexity-scalable Compression

This research in complexity-scalable compression and processing has been fueled by the consideration that, with the strong expansion of mobile devices, the application range of video standards has exploded. Mobile devices have limited computing power and memory and limited power consumption. The desire is to design algorithms that can also operate under such circumstances, but if more (computing) power is available, the performance can scale up to standard levels of operation.

Van der Vleuten, Kleihorst and Hentschel [741] propose a new technique for scalable DCT compression without quantization and coding. Instead, the DCT coefficients are compressed bit-plane by bit-plane, starting at the most significant plane.

The individual bit planes are encoded by simple rectangular zones (the adaptive zonal coding technique is a variant that has been studied earlier in recording systems). The experiments show that the performance is similar to JPEG compression, however, with halved hardware complexity, as given by the included estimation table.

Mietens, De With and Hentschel [754] report on a fully scalable MPEG encoder for mobile applications. The processing functions of an MPEG encoder are considered, and the DCT and ME unit are made scalable as they consume the highest computation power and memory. For this purpose, specific algorithms were developed. By controlling the number of computed DCT coefficients, the quantizer and VLC coder become also scalable. The scalable ME algorithm is reported in the hybrid compression sections of this chapter. It was found that the encoder can smoothly reduce to 50% of the operations count or execution time, while the quality varies accordingly between 20 and 48 dB PSNR in average. Another finding is that the DCT has an integrated coefficient selection function that leads to a quality build-up during interframe compression.

Mietens, De With and Hentschel [746] study scalable video processing in a dynamical multi-window TV system. In this case, an array processor platform is used to execute several video windows in parallel. The windows are programmable of size and shape and may vary over time. The paper concentrates on graph programming of TV tasks that together constitute the TV application. The processing platform tasks are programmed at a high level, using a standard RISC core.

Hoeksema, Vermeulen and Slump [747] deal with component and composite compression of residual video signals. The system is scalable since it uses a base layer with MPEG-2 compression and an extension layer based on an M-JPEG compression system. Experiments show that at high quality level of the base layer, it is more attractive to offer a *composite* signal to the residual encoder than a component signal, because in this case the bit rate drops considerably (70 to 45 Mbit/s) for the same quality level. The authors plan to study this unexpected result by using a trans-multiplexing quantizer that exploits the properties in the component-composite conversion.

8.4.4 Networked and Error-robust Video Compression

With the emergence of computer networks and digital broadband telecommunication infrastructures, the transmission of digital (compressed) video over networks has steadily grown as an important research and development issue. The networks usually map data into cells or data packets, taking various measures to improve robustness.

Schinkel and Ter Horst [697] compare a set of H.261 video encoders for an Asynchronous Transfer Mode (ATM) network environment. The comparison concentrates on the selection between Constant Bit-Rate (CBR) or Variable Bit-Rate (VBR) operation. The experiment uses 9 encoders with a video sequence length

of 30 seconds and QCIF resolution. The encoders are e.g., driven with a constant step size $g = 6$, producing VBR output. The SNR varies considerably (5-10 dB) at scene cuts with CBR operation (390 kbit/s), whereas VBR remains nearly constant (2.5 dB variation). The authors propose an adaptation of the packet cell rate to the video sequence, which gives a better overall subjective and objective result.

Hoeksema, Ter Horst, Heideman and Tattje [714] also study H.261 compression in an ATM network and focus on the cell loss. They use a simple Gaussian network model to answer the question whether the effects of cell loss should be controlled by the network or by the video codec. Despite the robustness measures in the video codec (BCH(511,493) FEC code, error concealment), a Gaussian model enables the derivation of analytical expressions for the cell loss ratio, the minimum and maximum case for the number of network users and the average lost cells per user. The simulation of the model at 64, 640 and 1920 kbit/s reveals that a small reduction of the number of users provides a large improvement in cell loss characteristics. The authors suggest further improvements on the model, since the Gauss model tends to overestimate the number of users.

Hekstra and Herrera [725] address the use of data compression in packet-switched networks with channel errors. They study error propagation in the V42bis and MNP5 data compression decoders when used in combination with X.25 packet switched networks. It is explained that channel errors can lead to error propagation at the source decoder that in turn deteriorates the source model. A number of countermeasures are proposed. An extra CRC on the decoder input or CRC on pseudo-random permutations of the decoder output. Also the statistics may be checked, or non-linear checks on the decoder with cryptographic keys are possible.

Bakker and Spaan [735] evaluate the trade-off between error robust network protocols and robust video compression algorithms under CBR operation. The comparison concentrates on the picture quality (SNR). The network protocol is designed such that small fixed-length packets are used, with extra FEC data added to it. The H.263 video codec has error concealment, signals the positions to encoder, which replaces erroneous blocks by intra coded blocks. The paper gives an extensive and detailed description of the experimental environment and settings, however, without any conclusions. The visual experiments provide evidence that the measures are useful.

Compressed data becomes vulnerable to channel errors. It is therefore important to either apply strong enough channel coding techniques to compressed (and packetized) data, or to make the compression system inherently robust against potential channel errors.

Roefs [666] studies an image decompression system for deep-space applications where high robustness is required. The compression is based on transformation (Hadamard, DCT) followed by special entropy coding such as the Rice or Modified Meltzer algorithm. The implementation is based on several parallel programmable 16-bit processors that are connected via a ring bus. The article discusses relevant

aspect such as power consumption, which is key in this type of systems. The system set-up allows the inclusion of new technology in a flexible way.

Simons [675] studies the error sensitivity of compressed data for satellite links (earth observation data and facsimile). For the latter data, the errors generally lead to a loss of a few lines. However, the EOF symbol may be generated by accident leading to the loss of half a 32-kbit frame (75–110 lines). With respect to the images, the use of 2-D DPCM gives error propagation of several lines. If the image is transform coded, with a BER of 10^{-6}, errors may be limited to one block if they fall inside or give propagation in the case of bit deletion of insertion. Generally, the bursty nature of errors is advantageous, since it limits the errors.

Van der Schaaf and Lagendijk [740] investigate the independence of source and channel coding for the progressive transmission of images in mobile communications. Key parameters between the source and channel coding are exchanged at the central interface, which has a Quality of Service (QoS) character. The source builds up encoded variance that more rapidly for images than for video signals. For channel coding, packets with FEC are assumed. The modeling verifies that source and channel coding can be relatively independent, and only a limited set of parameters need to be exchanged: latency, bit-rate and level of protection.

8.4.5 Alternative Compression Techniques

Over time, various alternative compression techniques have been investigated. Rooyackers [670] explores the straight-line approximation of Yan and Sakrison for a three-dimensional video source model. A video signal line is modeled as a concatenation of straight line pieces. The end of an interval is called a breakpoint. The model serves as a prediction for the real signal. The residual signal looks like a stationary Gaussian process. The residual signal still shows correlation, e.g., in the vertical or temporal direction. For this reason, a transform encoder is applied to the difference signal. The encoder sends per scan line the number of segments, a copy/non-copy indication per segment and line position information. Experimental results between 0.5 and 1.3 bit/pixel are reported with low r.m.s. error.

Heideman, Tattje, Van der Linden and Rijks [676] address the use of self-similar hierarchical transforms for video compression to bridge transform coding with the Human Visual System (HVS). The proposed scheme represents a multi-channel sampling model with filter functions of finite impulse response. In the hierarchical extension, the lower filter branches are split into new filter branches with additional subsampling. Using simple filters, the results may lead to the Haar transform of rank M. Self similarity is obtained when at each hierarchy level, the same systems basis functions b_i are used after each sampled low-pass output from the previous level. Impulse responses are then of the same form but with a different scale. This system looks very similar to the wavelet transform.

Simon, Macq and Verleysen [712] employ pyramidal transforms of Burt and Adelson for image compression using neural network interpolators. Instead of linear

interpolator filters assuming stationary unlimited signals, they use a three-layer perceptron for interpolation, in order to cope with non-linearities such as contours and particular textures. A back-propagation algorithm for updating is used. The entropy of the lossless signal for coding drops with 20% compared to linear filters and the picture quality ("Claire", CCIR-601) is clearly better.

For a short time fractals have been a popular research topic, in an attempt to iteratively model texture details in high-quality pictures. Franich, Lagendijk and Biemond [716] study picture compression with fractals. The issue of fractal compression is the finding of iterative functions. An IFS is a set of contractive transformations (usually affine) that maps a region of the image into a smaller region of that same image. The idea can also be inserted into picture sequence compression. Various options are discussed like using IFSs for the displaced frame difference signal. The authors claim similar performance as with DCT coding, where fractals may be slightly advantageous for lower bit rates. It is recognized that DCT coding is faster and components are widely available.

Schelkens, Barbarien and Cornelis [742] explore volumetric data compression based on cube-splitting for medical image data sets. The authors propose the use of 3-D wavelet transforms. When a significant wavelet coefficient is encountered, the cube of transformed data is split into sub-cubes, until the pixel resolution. The cube splitting yields excellent lossy compression results (up to 5 dB improvement in the 0.0625-1.0 bit/pixel range), when compared to multiple 2-D SQP encoding. The lossless compression performance is comparable to linear prediction techniques.

Satellite image data and remote sensing applications have specific statistics, because they have limited colors and typical noise characteristics. In 1987, Okkes and Huisman [681] explore the rate-distortion functions of SAR imagery. For this type of images, speckle noise is a common problem, and this is taken into account in the overall system design. The system is assumed to be an $R(D)$-optimal encoder, preceded or followed by two-dimensional linear complex filters. Assuming no a-priori knowledge about the SAR image statistics, equal distortion to all Fourier coefficients having nonzero allocation should be applied, yielding a suboptimum $R(D)$ bound. The pre-filter is of Wiener type; the derivation of the coefficients from the image statistics is unknown but may be derived from the power spectral density function including also speckle noise. Evaluation results indicate that for typical 4-look SAR imagery with correlation coefficient $\rho = 9$ and $r = 3$, permitting $\leq 1\%$ quantization noise, the bit rate ranges from 0.15 to 0.8 bit/pixel. A practical encoder at ESTEC yields below 0.5 bit/pixel.

Hogendoorn and Kordes [690] present a data compression and encryption system for remote sensing data (satellite Meteosat, 166 kbit/s), called Meteodec and Meteocrypt. For compression, three systems are compared. The ESTEC-1 algorithm encoder consists of a fixed set of Huffman-code tables and selects the table yielding the shortest bit cost. The NLR-Meander algorithm first determines pixel differences, which are assigned classes. Within a class, pixel differences are as-

sumed equiprobable. The classes are then compressed. The third system performs adaptive class assignment, followed by an arithmetic coder. The compression ratios for segments of 250 pixels are between 1.3–71.4 for the ESTEC-1 algorithm, between 1.4–21.1 for the NLR-Meander algorithm and between 1.2–18.0 for the adaptive system. The Meander algorithm provided the best results for the test images and was chosen, while the ESTEC-algorithm was rejected because it gave too much fluctuation in buffering.

8.5 Concluding Remarks

Since the mid 1990s, research and development in image and video compression has been enriched and influenced by several new perspectives and subsequent standards. An overview of the challenges beyond 2000 is given by Biemond in [736]. We mention three important developments, and the associated standards MPEG-4, MPEG-7 and MPEG-21. First, for restricted applications like sport scenes and surveillance imagery, video segmentation is becoming increasingly feasible. The MPEG-4 standard has opened up the exploitation of high-level descriptions of regions and objects of interest in constrained application areas. The MPEG-4 standard already includes compression for facial models, and with improvements in region/object segmentation, attractive perspectives will open up for video compression.

Second, the MPEG-7 “Multimedia Content Description Interface” standard addresses techniques for organizing and searching (compressed) audio-visual materials. Compressing images and video makes easy access to the content more difficult as (partial) decompression may be required before the content can be analyzed.

Finally, with the success of compression technology, Internet and CDs became increasingly affordable ways of distributing hacked multimedia. At the time of writing, illegal music swapping over P2P networks such as KaZaa are taking epidemic forms, and it is not hard to predict that within a few years time the same will be true for video (especially movies). Various bodies and working groups are addressing the development of digital right management systems (DRM), that will on the one hand need to put a stop to these illegal practices, and on the other hand open- p the road to different (Internet-based) distribution models. MPEG-21 aims at defining a framework for multimedia delivery and consumption for use by all the players in the delivery and consumption chain.

Digital video compression has evolved enormously over the past 25 years. A part of the information technology and consumer electronics revolution that we have seen is thanks to digital video compression. Information Theory researchers in the Benelux have contributed substantially to these development, not in the last place because of the role Philips Research and Development Laboratories and the former KPN research laboratory have played in this area. In terms of scientific and practical impact, we like to highlight the research of Westerink, Bosveld

et al. at TU Delft in the area of hierarchical and compatible subband coding [678, 685, 687, 692, 698, 703], the work of Desmet *et al.* at K.U. Leuven in the field of object-based video compression [727, 738], and the domain-constrained compression research of De With *et al.* [730, 739, 746, 750, 754].

References

[1] Fisher, R.A., *Theory of Statistical Estimation*, Proc. Cambridge Phil. Society 22, pp. 700–725, 1925.

[2] Kac, M., *On the Notion of Recurrence in Discrete Stochastic Processes*, Bull. Amer. Math. Soc., vol. 53, pp. 1002 - 1010, Oct. 1947.

[3] Shannon, C.E., *A Mathematical Theory of Communication*, Bell Syst. Tech. J. 27(3,4), pp. 379–423 and 623–656, 1948.

[4] Shannon, C.E., *Communication in the Presence of Noise*, Proc. IRE 37(1), pp. 10–21, 1949.

[5] Shannon, C.E., *Communication Theory of Secrecy Systems*, Bell Syst. Techn. J. 28(4), pp. 656–715, 1949.

[6] Shannon, C.E., *Prediction and Entropy of Printed English*, Bell Syst. Techn. J. 30(1), pp. 50–64, 1951.

[7] Huffman, D.A., *A Method for the Construction of Minimum-Redundancy Codes*, Proc. IRE, vol 40, pp. 1098–1101, Sept. 1952

[8] McMillan, B., *The Basic Theorems of Information Theory*, Ann. Math. Stat. 24, pp. 96–219, 1953.

[9] Elias, P., *Error-Free Coding*, IRE Trans. Inform. Theory, pp. 29–37, 1954.

[10] Shannon, C.E., *The Zero-error Capacity of a Noisy Channel*, IRE Trans. Inform. Th. pp. 8–19, 1956.

[11] Khinchin, A.Ya., *Mathematical Foundations of Information Theory*, Dover Publ., New York, 1957.

[12] Shannon, C.E., *Channels with Side Information at the Transmitter*, IBM J. Res. Devel. 2, pp. 289–293, 1958.

[13] Pinsker, M.S., *Information and Information Stability of Random Variables and Processes*, Izd. Akad. Nauk, 1960.

[14] Shannon, C.E., *Two-way Communication Channels*, Proc. 4th Berkeley Symp. Math. Stat. & Prob. 1, pp. 611–644, 1961.

[15] Gallager, R.G., *Low-Density Parity-Check Codes*, MIT Press, Cambridge, MA, USA, 1963.

[16] Berkoff, M.,*Waveform Compression in NRZI Magnetic Recording*, Proceedings IEEE, vol. 52, pp. 1271-1272, Oct. 1964.

[17] Forney, G.D., *Generalized Minimum Distance Decoding*, IEEE Trans. Inform. Theory, vol. 12, pp. 125–131, April 1966.

[18] Forney, G.D., *Concatenated Codes*, MIT Press, 1966.

[19] Tunstall, B.P., *Synthesis of Noiseless Compression Codes*, Ph.D. dissertation, Georgia Inst. Tech., Atlanta, GA, Sept. 1967.

[20] Jelinek, F., *Buffer Overflow in Variable Length Coding of Fixed Rate Sources*, IEEE Trans. Inform. Theory, vol 14, pp. 490–501, May 1968.

[21] Abramson, N., *The ALOHA System – Another Alternative for Computer Communications*, AFIPS Conf. Proc., Fall Joint Computer Conf. 37, pp. 281–285, 1970.

[22] Chien, T.M., *Upper Bound on the Efficiency of Dc-constrained Codes*, Bell Syst. Tech. J. vol. 49, pp. 2267-2287, Nov. 1970.

[23] Tang, D.T. and L.R. Bahl, *Block Codes for a Class of Constrained Noiseless Channels*, Information and Control, vol. 17, pp. 436-461, 1970.

[24] Berger, T., *Rate Distortion Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1971.

[25] Meulen, E.C. van der, *Three-terminal communication channels*, Advances in Applied Probability 3(1), pp. 120–154, 1971.

[26] Schalkwijk, J.P.M., *A Class of Simple and Optimal Strategies for Block Coding on the Binary Symmetric Channel with Noiseless Feedback*, IEEE Trans. Inform. Theory, vol. 17, pp. 283–287, May 1971.

[27] Cover, T.M., *Broadcast Channels*, IEEE Trans. Inform. Theory, vol. 18, pp. 2–14, Jan. 1972.

[28] Chase, D., *A Class of Algorithms for Decoding Block Codes with Channel Measurement Information*, IEEE Trans. Inform. Theory, vol. 18, pp. 170–182, Jan. 1972.

[29] Schalkwijk, J.P.M., *An Algorithm for Source Coding*, IEEE Trans. Inform. Theory, vol. 18, pp. 395–399, May 1972.

[30] Bell, D.E., LaPadula, L.J., *Secure Computer Systems: Mathematical Foundations*, ESD-TR-73-278, vol. 1-2, ESD/AFSC, Hanscom AFB, Bedford, MA, 1973.

[31] Cover, T.M., *Enumerative Source Coding*, IEEE Trans. Inform. Theory, vol. 19, pp. 73–76, Jan. 1973.

[32] Slepian D. and J.K. Wolf, *A Coding Theorem for Multiple Access Channels with Correlated Sources*, Bell Syst. Tech. J. 52, pp. 1036–1076, 1973.

[33] Varshamov, R.R., *A Class of Codes for Asymmetric Channels and a Problem from the Additive Theory of Numbers*, IEEE Trans. Inform. Theory, pp. 92–95, vol. 19, Jan. 1973.

[34] Bahl, L.R., J. Cooke, F. Jelinek, and J. Raviv, *Optimal Decoding of Linear Codes for Minimimizing Symbol Error Rate*, IEEE Trans. Inform. Theory, vo. 20, pp. 284-287, March 1974.

[35] Kuznetsov, A.V. and B.S. Tsybakov, *Coding for Memories with Defective Cells*, Problemy Peredachi Informatsii 10(2), pp. 52–60, 1974.

[36] Geçkinli, N.C., *Two Corollaries to the Huffman Procedure*, IEEE Trans. Inform. Theory, vol. 21, pp. 342–344, May 1975.

[37] Wyner, A.D., *The Wire-tap Channel*, Bell Syst. Tech. J., vol. 54, no. 8, pp. 1355–1387, 1975.

[38] Knapp, C.H. and C.G. Carter, *The Generalized Correlation Method for Estimation of Time Delay*, IEEE Trans. on Acoustics, Speech and Signal Processing, vol. 24, pp. 320-327, 1976.

[39] Wyner D. and J. Ziv, *The Rate-Distortion Function for Source Coding with Side Information at the Decoder*, IEEE Trans. on Inform. Theory, vol. 22, pp. 1–10, Jan. 1976.

[40] Diffie, W. and M.E. Hellman, *New Directions in Cryptography*, IEEE Trans. Inform. Theory, vol. 22, pp. 644–654, Nov. 1976.

[41] Lawrence, J.C., *A New Universal Coding Scheme for the Binary Memoryless Source*, IEEE Trans. Inform. Theory, vol. 23, pp. 466 - 472, July 1977.

[42] MacWilliams, F.J. and N.J.A. Sloane, *The Theory of Error-Correcting Codes*, North-Holland, 1977.

[43] McEliece, R.J., *The Theory of Information and Coding*, Addison-Wesley, Reading, MA, 1977.

[44] Shirayev, A.N., *Optimal Stopping Rules*, Springer-Verlag, New York, 1977.

[45] Berlekamp, E.R., R.J. McEliece, and H.C.A. van Tilborg, *On the Inherent Intractability of Certain Coding Problems*, IEEE Trans. Inform. Theory, vol. 24, pp. 384–386, May 1978.

[46] Koshelev, V.N., *Multilevel Source Coding and Data Transmission Theorem*, in Proc. VII All-Union Conference on Coding Theory and Data Transmission, part I, pp. 85–92, Vilnius, 1978.

[47] McEliece, R.J., *A Public–key Cryptosystem based on Algebraic Coding Theory*, JPL DSN Progress Report 42–44, pp. 114–116, Jan–Febr. 1978.

[48] Merkle, R.C. and M.E. Hellman, *Hiding Information and Signatures in Trapdoor Knapsacks*, IEEE Trans. Inform. Theory, vol. 24, pp. 525–530, Sept. 1978.

[49] Rivest, R.L., A. Shamir, and L. Adleman, *A Method for Obtaining Digital Signatures and Public Key Cryptosystems*, Comm. ACM, vol. 21, pp. 120–126, Febr. 1978.

[50] Garey, M.R. and D.S. Johnson, *Computers and Intractability: A Guide to the Theory of NP–Completeness*, W.H. Freeman and Co., San Fransisco, 1979.

[51] Shamir, A., *How to share a secret*, Comm. ACM, vol. 22, pp. 612-613, 1979.

[52] El Gamal, A.A. and T.M. Cover, *Multiple User Information Theory*, Proc. IEEE 68(12), pp. 1466–1483, 1980.

[53] Guazzo, M., *A General, Minimum-Redundancy Source-Coding Algorithm*, IEEE Trans. Inform. Theory, vol. 26, Jan. 1980, pp. 15–25.

[54] Hellman, M., *A Cryptanalytic Time-Memory Tradeoff*, IEEE Trans. Inform. Theory, vol. 26, pp. 401–406, 1980.

[55] Csiszár, I. and J. Kőrner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Académiai Kiadó, Budapest, 1981.

[56] Gersho, A. and B. Ramamurthi, *Image Coding Using Vector Quantization*, Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 428-431, 1982.

[57] Krichevsky, R.E. and V.K. Trofimov, *The Performance of Universal Encoding*, IEEE Trans. Inform. Theory, vol. 27, pp. 199-207, March 1981.

[58] Shamir, A., *A Polynomial Time Algorithm for Breaking the Basic Merkle–Hellman Cryptosystem*, Proc. 23rd IEEE Symp. Found. Computer Sci., pp. 145–152, 1982.

[59] Ungerboeck, G., *Channel Coding with Multilevel/Phase Signals*, IEEE Trans. Inform. Theory, vol. 28, pp. 55–67, Jan. 1982.

[60] Beenker, G.F.M. and K.A.S. Immink, *A Generalized Method for Encoding and Decoding Runlength-Limited Binary Sequences*, IEEE Trans. Inform. Theory, vol. 29, no. 5, pp. 751-754, Sept. 1983.

[61] Berger, T. and Z. Zhang, *Minimum Breakdown Degradation of Binary Source Encoding*, IEEE Trans. Inform. Theory, vol. 29, pp. 807–814, Nov. 1983.

[62] Blahut, R.E., *Theory and Practice of Error-control Codes*, Addison-Wesley, 1983.

[63] Costa, M. H. M. .*Writing on Dirty Paper*, IEEE Trans. Inform. Theory, vol. 29, pp. 439–441, May 1983.

[64] Lagarias, J.C. and A.M. Odlyzko, *Solving Low–Density Subset Problems*, Proc. 24th Annual IEEE Symp. on Found. of Comp. Science, pp. 1–10, 1983.

[65] Ferguson, T.J. and J.H. Rabinowitz, *Self-synchronizing Huffman codes*, IEEE Trans. Inform. Theory, vol. 30, pp. 687–693, July 1984.

[66] Gray, R., *Vector Quantization*, IEEE Acoustics, Speech and Signal Processing Magazine, pp. 4–29, April 1984.

[67] Rissanen, J., *Universal Coding, Information, Prediction, and Estimation*, IEEE Trans. Inform. Theory, vol. 30, pp. 629-636, July 1984.

[68] Bouwhuis, G., J. Braat, A. Huijser, J. Pasman, G. van Rosmalen, and K.A.S. Immink, *Principles of Optical Disc Systems*, Adam Hilger Ltd, 1985.

[69] Ahlswede R. and I. Csiszár, *Hypothesis Testing with Communication Constraints*, IEEE Trans. Inform. Theory, vol. 32, pp. 533–542, July 1986.

[70] Montgomery, B.L. and J. Abrahams, *Synchronization of Binary Source Codes*, IEEE Trans. Inform. Theory, vol. 32, pp. 849–854, Nov 1986.

[71] Bertsekas D. and R. Gallager, *Data Networks*, Prentice Hall 1987.

[72] Barron, A.R., *The Convergence in Information of Probability Density Estimators*, IEEE Int. Symp. Inform. Theory, Kobe, Japan, June 19-24, 1988.

[73] Guillou, L.C. and J.-J.Quisquater, *A "Paradoxical" Identity-based Signature Scheme Resulting from Zero-knowledge*, Advances in Cryptology, Proc. of CRYPTO'88 (Ed. S. Goldwasser), LNCS 403, Springer Verlag, 1988.

[74] Lee E.A. and D.G. Messerschmitt, *Digital Communication*, Kluwer Academic Publishers, 1988.

[75] Ahlswede, R. and G. Dueck, *Identification via Channels*, IEEE Trans. Inform. Theory, vol. 35, pp. 15–29, 1989.

[76] Bassalygo, L.A., S.I. Gelfand, and M.S. Pinsker, *Coding for Channels With Localized Errors*, Proc. 4-th Joint Swedish-Soviet Int. Workshop on Information Theory, Gotland, Sweden, pp. 85–89, August 1989.

[77] Quinlan, J.R. and R.L. Rivest, *Inferring Decision Trees Using the Minimum Description Length Principle*, Inform. and Comput., vol. 80, pp. 227-248, 1989.

[78] Ahlswede, R., J.P. Ye, and Z. Zhang, *Creating Order in Sequence Spaces with Simple Machines*, Information and Computation, vol. 89, pp. 47–94, 1990.

[79] Biemond, J., R.L. Lagendijk and R.M. Mersereau, *Iterative Methods for Image Deblurring*, Proc. IEEE, vol. 8, no. 5, pp. 856-883, 1990.

[80] Bingham, J.A.C., *Multicarrier Modulation for Data Transmission: An Idea whose Time has Come*, IEEE Communications Magazine, vol. 28, pp. 7–15, May 1990.

[81] Lagendijk, R.L., J. Biemond and D.E. Boekee, *Identification and Restoration of Noisy Blurred Images Using the Expectation-Maximization Algorithm*, IEEE Trans. Acoustics, Speech and Signal Processing, vol. 38, no. 7, pp. 1180-1191, 1990.

[82] Wiener, M.J., *Cryptanalysis of Short RSA Secret Exponents*, IEEE Trans. Inform. Theory, vol. 36, pp. 553-558, May 1990.

[83] Biglieri, E., D. Divsalar, P. McLane, and M. Simon, *Introduction to Trellis-Coded Modulation with Applications*, Maxwell-Macmillan, 1991.

[84] Cover, T.M. and J.A. Thomas, *Elements of Information Theory*, Wiley series in telecommunication, J.Wiley & Sons, New York, 1991.

[85] Equitz, W.H.R. and T.M. Cover, *Successive Refinement of Information, IEEE Trans. Inform. Theory*, vol. 37, pp. 268–275, 1991.

[86] Immink, K.A.S, *Coding Techniques for Digital Recorders*, Prentice Hall, 1991.

[87] Barron, A.R., L. Györfi, and E.C. van der Meulen, *Distribution Estimation Consistent in Total Variation and in Two Types of Information Divergence*, IEEE Trans. Inform. Theory, vol. 38, pp. 1437 - 1454, Sept. 1992.

[88] Gitlin, R.D., J.F. Hayes, and S.B. Weinstein, *Data Communication Principles*, Plenum Press, 1992.

[89] Alabbadi, M. and S.B. Wicker, *Digital Signature Schemes based on Error-correcting Codes*, IEEE Int. Symp. Inform. Theory, San Antonio, p. 199, 1993.

[90] Berrou, C., A. Glavieux, and P. Thitimajshima, *Near Shannon Limit Error-Correcting Coding and Decoding: Turbo Codes*, Proceedings IEEE ICC, Geneva, Switzerland, pp. 1064–1070, May 1993.

[91] Maurer, U., *Secret Key Agreement by Public Discussion*, IEEE Trans. Inform. Theory, vol. 39, pp. 733-742, May 1993.

[92] Pennebaker, W.B. and J.L. Mitchell, *JPEG Still Image Compression Standard*, Van Nostrand Reinhold, New York, 1993.

[93] Wyner, A.D., J. Ziv, *The Sliding-window Lempel-Ziv Algoritm is Asymptotically Optimal*, Proc. IEEE, vol. 82, pp. 872-877, June 1994.

[94] Best, M.R., M.V. Burnashev, Y. Lévy, A. Rabinovich,, and P.C. Fishburn, *On a Technique to Calculate the Exact Perfomance of a Convolutional Code*, IEEE Trans. Inform. Theory, vol. 41, pp. 441–447, March 1995.

[95] Le Floch B., M. Alard, and C. Berrou, *Coded Orthogonal Frequency Division Multiplex*, Proc. IEEE, vol. 83, pp. 587–592, June 1995.

[96] Willems, F.M.J., Y.M. Shtarkov, and Tj.J. Tjalkens, *The Context-Tree Weighting Method: Basic Properties*, IEEE Trans. Inform. Theory, vol. 41, pp. 653 - 664, May 1995.

[97] Bergmans, J.W.M., *Digital Baseband Transmission and Recording*, Kluwer, 1996.

[98] Haskell, B., A. Puri, and A. Netravali, *Digital Video: An Introduction to MPEG-2*, Chapman and Hall, 1996.

[99] Kocher, P., *Timing Attacks on Implementations of Diffie-Hellman, RSA, DSS and Other Systems*, Advances in Cryptology, Proc. of CRYPTO'96 (Ed. U. Maurer), LNCS 1070, Springer Verlag, 1996.

[100] Shi, Q., *Digital Modulation Techniques*, Digital Electronics Engineering Handbook, chapter 5. McGraw-Hill, 1996.

[101] Wilson, S.G., *Digital Modulation and Coding*, Prenctice Hall, 1996.

[102] Menezes, A.J., P.C. van Oorschot, and S.A. Vanstone, *Handbook of Applied Cryptography*, CRC Press, Boca Raton, 1997.

[103] Pennebaker, W.B., J.L. Mitchell, C. Fogg, and D. LeGall, *MPEG Digital Video Compression Standard*, Chapman and Hall, 1997.

[104] With, P.H.N. de, and Rijckaert, A.M.A., *Design Considerations of the Video Compression System of the New DV Camcorder Standard*, IEEE Trans. Consum. Electron., Vol 43, No. 4, pp. 1160–1179, 1997.

[105] Costello, D.J., J. Hagenauer, H. Imai, and S.B. Wicker, *Applications of Error-Control Coding*, IEEE Trans. Inform. Theory, vol. 44, pp. 2531–2560, Oct. 1998.

[106] Gray, R.M. and D.L. Neuhoff, *Quantization*, IEEE Transactions on Information Theory, vol. 44, pp. 2325–2383, Oct. 1998.

[107] Immink, K.A.S., P.H. Siegel, and J.K. Wolf, *Codes for Digital Recorders*, IEEE Trans. Inform. Theory vol. 44, pp. 2260-2299, Oct. 1998.

[108] Pless, V.S. and W.C. Huffmann (eds.), *Handbook of Coding Theory*, Vols. 1 and 2, Elsevier, 1998.

[109] Immink, K.A.S., *Codes for Mass Data Storage Systems*, Shannon Foundation Publishers, Geldrop, Netherlands, 1999.

[110] Johannesson, R. and K.S. Zigangirov, *Fundamentals of Convolutional Coding*, IEEE Press, 1999.

[111] Sayood, K., *Introduction to Data Compression*, 2nd. Edition, Academic Press, 2000.

[112] Proakis, J.G., *Digital Communications*, McGraw-Hill, fourth edition, 2001.

[113] Pereira, F. and T. Ebrahimi,(eds.), *The MPEG-4 Book*, ISMC Press, 2002.

[114] Immink, K.A.S., J.Y. Kim, S.W. Suh, and S.K. Ahn, *Extremely Efficient Dc-free RLL codes for Optical Recording*, IEEE Trans. Commun., vol. 51, pp. 326-331, March 2003.

WIC Symposium Shannon and Multi-user Information Theory Papers

[115] Boekee, D.E., *Informatie Maten, Fundamentele Begrippen en Enkele Toepassingen*, First SITB (Zoetermeer), pp. 29–32, 1980.

[116] Broekstra, G., *Constraintanalyse: Toepassing van Informatiematen op het Probleem van Structuuridentificatie*, First SITB (Zoetermeer), pp. 39–42, 1980.

[117] Buffart, H. and Collard, R., *Structural Information Theory of Perception*, First SITB (Zoetermeer), pp. 43–46, 1980.

[118] Meulen, E.C. van der, *Een Eenvoudig Bewijs, Gebaseerd op Partities en Typicality, van een Coderingstheorema van Marton voor het Discrete Broadcast Kanaal*, First SITB (Zoetermeer), pp. 105–110, 1980.

[119] Boekee, D.E., *Syntactische Complexiteit en Informatie-Inhoud*, Second SITB (Zoetermeer), pp. 35–40, 1981.

[120] Lubbe, J.C.A. van der, *Een Vergelijkend Onderzoek naar de Informatiematen van Renyi, Daroczy en Arimoto en de Invloed van hun Parameters*, Second SITB (Zoetermeer), pp. 77–85, 1981.

[121] Meulen, E.C. van der, *Overzicht van Recente Resultaten op het Gebied van het Multiple Access Kanaal*, Second SITB (Zoetermeer), pp. 87–98, 1981.

[122] Schalkwijk, J.P.M., *The And-Gate*, Second SITB (Zoetermeer), pp. 103–111, 1981.

[123] Willems, F.M.J., *Codering en Capaciteitsgebied voor het Binary Erasure Multiple Access Kanaal met Feedback*, Second SITB (Zoetermeer), pp. 123–128, 1981.

[124] Willems, F.M.J. and E.C. van der Meulen, *Een Verbetering en Veralgemening van het Transmissiegebied van Ozarow voor het Gaussische Broadcast Kanaal met Feedback*, Second SITB (Zoetermeer), pp. 129–138, 1981.

[125] Collard, R.F.A., *Structural Information Processing: Some Recent Developments*, Third SITB (Zoetermeer), pp. 5–12, 1982.

[126] Meulen, E.C. van der, *Toetsen voor Uniformiteit Gebaseerd op Entropie*, Third SITB (Zoetermeer), pp. 63–75, 1982.

[127] Meulen, E.C. van der, *Overzicht van Recente Resultaten op het Gebied van het Broadcast Kanaal*, Third SITB (Zoetermeer), pp. 77–92, 1982.

[128] Schalkwijk, J.P.M. and Vinck, A.J., *Information Networks — Deterministic Elements*, Third SITB (Zoetermeer), pp. 113–124, 1982.

[129] Willems, F.M.J., *Het Discrete Geheugenloze Multiple Access Kanaal met Gedeeltelijk Coöpererende Encoders*, Third SITB (Zoetermeer), pp. 157–161, 1982.

[130] Willems, F.M.J. and E.C. van der Meulen, *Het Discrete Geheugenloze Multiple Access Kanaal met Afkijkende Encoders*, Third SITB (Zoetermeer), pp. 163–170, 1982.

[131] Coeberg van den Braak, P.A.B.M. and Tilborg, H.C.A. van, *A Set of Uniquely Decodable Codepairs for the 2-Access Binary Adder Channel*, Fourth SITB (Haasrode), pp. 31–38, 1983.

[132] Lubbe, J.C.A. van der, *Applications of Information Theoretical Concepts in Economics*, Fourth SITB (Haasrode), pp. 137–146, 1983.

[133] De Bruyn, K., *Good Codeproducers for the Asymmetric Broadcast Channel*, Fourth SITB (Haasrode), pp. 147–154, 1983.

[134] De Bruyn, K. and E.C. van der Meulen, *Two Codeconstructions for the Asymmetric Multiple Access Channel*, Fourth SITB (Haasrode), pp. 155–162, 1983.

[135] Post, K.A. and Ligtenberg, L.G.T.M., *Coding Strategies for the Binary Multiplying Channel in the Discrete Case*, Fourth SITB (Haasrode), pp. 163–170, 1983.

[136] Schalkwijk, J.P.M., Rooyackers, J.E. and Smeets, B.J.M., *Generalized Shannon Strategies for the Binary Multiplying Channel*, Fourth SITB (Haasrode), pp. 171–178, 1983.

[137] Vinck, A.J., *Constructive Superposition Coding for the Binary Erasure Multiple Access Channel*, Fourth SITB (Haasrode), pp. 179–188, 1983.

[138] Willems, F.M.J., *Two Results for the Multiple Access Channel with Feedback*, Fourth SITB (Haasrode), p. 189–198, 1983.

[139] De Bruyn, K., *Fixed Composition List Codes for Discrete Memoryless One-Way Channels: a Packing Lemma and an Iterative Code Construction*, Fifth SITB (Aalten), pp. 36–44, 1984.

[140] De Bruyn, K. and E.C. van der Meulen, *Feedback Capacity Regions for a Class of Discrete Memoryless Multiple-Access Channels*, Fifth SITB (Aalten), pp. 45–53, 1984.

[141] Gaal, E.W. and Schalkwijk, J.P.M., *Deterministic Binary Two-Way Channels*, Fifth SITB (Aalten), pp. 54–63, 1984.

[142] Hekstra, A.P. and Willems, F.M.J., *Capacity Regions for Multiple-Access Channels with Feedback and Two-Way Channels*, Fifth SITB (Aalten), pp. 73–79, 1984.

[143] Post, K.A., *Construction of a Positive Solution of a Special System of Quadratic Equations*, Fifth SITB (Aalten), pp. 118–122, 1984.

[144] Schalkwijk, J.P.M., *On the Optimality of Coding Strategies for Deterministic Two-Way Channels*, Fifth SITB (Aalten), pp. 131–136, 1984.

[145] Smit, G., *Een Toets voor de Orde van een Markov-Keten welke Gebaseerd is op het Begrip Entropie*, Fifth SITB (Aalten), pp. 162–168, 1984.

[146] Vinck, A.J., Hoeks, W.L.M. and Post, K.A., *Multiple Access with Feedback*, Fifth SITB (Aalten), pp. 187–193, 1984.

[147] De Bruyn, K., Prelov, V.V. and E.C. van der Meulen, *Two Results on the Discrete Memoryless Asymmetric Multiple-Access Channel with Arbitrarily Correlated Sources*, Sixth SITB (Mierlo), pp. 183–192, 1985.

[148] Hekstra, A.P. and Willems, F.M.J., *Dependence Balance Bounds for Multiple Access Channels with Feedback and Equal Output Two-Way Channels*, Sixth SITB (Mierlo), pp. 193–198, 1985.

[149] Schalkwijk, J.P.M., *The Threshold Bound to the Capacity Region of a Two-Way Channel Revisited*, Sixth SITB (Mierlo), pp. 199–206, 1985.

[150] Tolhuizen, L.M.G.M., *Discrete Coding for the BMC, based on Schalkwijk's Strategy*, Sixth SITB (Mierlo), pp. 207–212, 1985.

[151] Schalkwijk, J.P.M., *On Powers of the Defect Channel and Their Equivalence to Noisy Channels with Feedback*, Seventh SITB (Noordwijkerhout), pp. 41–48, 1986.

[152] Willems, F.M.J. and Vinck, A.J., *Repeated Recording for an Optical Disc*, Seventh SITB (Noordwijkerhout), pp. 49–54, 1986.

[153] Kamminga, C., *The Uncertainty Product versus the Sum of Entropies Uncertainty Principle*, Seventh SITB (Noordwijkerhout), pp. 55–60, 1986.

[154] Vanroose, P. and E.C. van der Meulen, *Coding for the Binary Switching Multiple Access Channel*, Seventh SITB (Noordwijkerhout), pp. 183–189, 1986.

[155] Remijn, J.C.C.M., *On Minimum Breakdown Degradation in Binary Multiple Descriptions*, Seventh SITB (Noordwijkerhout), pp. 191–196, 1986.

[156] Barbé, A., *Binary Random Sequences: Derivative Sequences and Multi-level α-Typical Randomness*, Eighth SITB (Deventer), pp. 21–28, 1987.

[157] De Moor, B. and Vandewalle, J., *The Uncertainty Principle of Mathematical Modelling*, Eighth SITB (Deventer), pp. 100–107, 1987.

[158] Overveld, W.M.C.J. van, *Fixed- and Variable Length Strategies are Equivalent*, Eighth SITB (Deventer), pp. 117–123, 1987.

[159] Prelov, V.V. and E.C. van der Meulen, *On the Slepian and Wolf Multiple-Access Channel with Gaussian Noise*, Eighth SITB (Deventer), pp. 132–139, 1987.

[160] Schalkwijk, J.P.M., *The Echo Channel*, Eighth SITB (Deventer), pp. 140–148, 1987.

[161] Vanroose, P., *Techniques for Constructing Codes for the Binary Switching Channel*, Eighth SITB (Deventer), pp. 175–181, 1987.

[162] Verboven, B. and E.C. van der Meulen, *Strong Converses for Multiple-Access Channels*, Eighth SITB (Deventer), pp. 182–188, 1987.

[163] Overveld, W.M.C.J. van and Schmitt, R.J.M., *Generalized Write-Unidirectional Memory Codes*, Ninth SITB (Mierlo), pp. 1–8, 1988.

[164] Shi, G.Q., *On the Characterization of Information Divergence for Two-Terminal Hypothesis Testing with One Sided Data Compression*, Ninth SITB (Mierlo), pp. 171–174, 1988.

[165] Vanroose, P. and E.C. van der Meulen, *Zero-Error Capacity and Quasi-Synchronized Codes for the Binary Switching Channel*, Ninth SITB (Mierlo), pp. 175–181, 1988.

[166] Györfi, L. and E.C. van der Meulen, *The Almost Sure Consistency of a General Class of Entropy Estimators*, Ninth SITB (Mierlo), pp. 183–189, 1988.

[167] Schalkwijk, J.P.M., *Shannon Strategies Revisited*, Tenth SITB (Houthalen), pp. 3–8, 1989.

[168] Verboven, B. and E.C. van der Meulen, *Noiseless Broadcasting for Identification*, Tenth SITB (Houthalen), pp. 9–12, 1989.

[169] Willems, F.M.J., *A Proof of the Coding Theorem for the Additive White Gaussian Noise Channel in Terms of Jointly Typical Sequences*, Tenth SITB (Houthalen), pp. 13–18, 1989.

[170] Jian-Ping Ye, *Progress in Specific Models of Ordering*, Tenth SITB (Houthalen), pp. 19–22, 1989.

[171] Overveld, W.M.C.J. van, *Write-Unidirectional Memory Codes Over Arbitrary Alphabets*, Tenth SITB (Houthalen), pp. 23–30, 1989.

[172] Vanroose, P. and E.C. van der Meulen, *A New Proof of the Zero-Error Capacity Region of the Blackwell Broadcast Channel*, Tenth SITB (Houthalen), pp. 37–44, 1989.

[173] Vanroose, P., *Code Constructions for Deterministic Relay Channels*, Eleventh SITB (Noordwijkerhout), pp. 15–21, 1990.

[174] Schalkwijk, J.P.M., *Another 0.63056*, Eleventh SITB (Noordwijkerhout), pp. 155–161, 1990.

[175] Overveld, W.M.C.J. van and Willems, F.M.J., *An Achievability Proof for Write Unidirectional Memories with Uninformed Encoder and Decoder*, Eleventh SITB (Noordwijkerhout), pp. 162–167, 1990.

[176] Baggen, C.P.M.J. and Wolf, J.K., *An Information Theoretic Approach to Timing Jitter*, Eleventh SITB (Noordwijkerhout), p. 174, 1990.

[177] Baggen, C.P.M.J. and Wolf, J.K., *Timing Jitter: Coding Theorems and Spectral Properties*, Twelfth SITB (Veldhoven), pp. 1–8, 1991.

[178] Hekstra, A.P., *The Discrete Memoryless Timing Jitter Channel and its Capacity in the Case of Weak Synchronisation*, Twelfth SITB (Veldhoven), pp. 9–16, 1991.

[179] Prelov, V.V. and E.C. van der Meulen, *The Capacity Region of the Compound Interference Channel with Additive Almost Gaussian Noise*, Twelfth SITB (Veldhoven), pp. 103–106, 1991.

[180] Schalkwijk, J.P.M., *Upper Bounds for Unit Square Resolution*, Twelfth SITB (Veldhoven), pp. 107–112, 1991.

[181] Salehi, M. and Willems, F.M.J., *Ring Source- and Channel Codes*, Twelfth SITB (Veldhoven), pp. 113–120, 1991.

[182] Vleuten, R.J. van der, *High-Performance Low-Complexity Control of Pure and Slotted Aloha Systems*, Twelfth SITB (Veldhoven), pp. 129–135, 1991.

[183] Schalkwijk, J.P.M., *On Genie Assisted Strategies*, Thirteenth SITB (Enschede), pp. 167–172, 1992.

[184] Bloemen, A.H.A., *Codes for Two-Way Channels Without Feedback*, Thirteenth SITB (Enschede), pp. 173–180, 1992.

[185] Schalkwijk, J.P.M., *Beating Shannon's Inner Bound with Message Percolation*, Fourteenth SITB (Veldhoven), pp. 14–23, 1993.

[186] Bloemen, A.H.A., *Constructing Discrete Strategies for Two-Way Channels*, Fourteenth SITB (Veldhoven), pp. 24–31, 1993.

[187] Meeuwissen, H.B., *New Constructive Coding Strategies for Two-Way Communication*, Fourteenth SITB (Veldhoven), pp. 32–39, 1993.

[188] Kleima, D., *Is There a Foundation for Probability-Theory?*, Fourteenth SITB (Veldhoven), pp. 40–47, 1993.

[189] Prelov, V.V. and E.C. van der Meulen, *The Capacity of a Continuous Alphabet Memoryless Channel with Vector-Valued Weak Input Signals*, Fourteenth SITB (Veldhoven), pp. 48–53, 1993.

[190] Baggen, C.P.M.J. and Wolf, J.K., *On Band-Limited Additive Gaussian Noise Channels in the Presence of Sampling Jitter*, Fourteenth SITB (Veldhoven), pp. 54–61, 1993.

[191] Schalkwijk, J.P.M., Meeuwissen, H.B. and Bloemen, A.H.A., *A Substantial Improvement of the Lower Bound to the Capacity Region of the Binary Multiplying Channel*, Fifteenth SITB (Louvain-la-Neuve), pp. 175–182, 1994.

[192] Györfi, L. and E.C. van der Meulen, *Positive and Negative Findings on the Consistent Estimation of a Probability Density in Information Divergence*, Fifteenth SITB (Louvain-la-Neuve), pp. 183–187, 1994.

[193] Prelov, V.V. and E.C. van der Meulen, *On the Fisher Information of the Sum of Two Independent Random Variables One of which is Small, and an Asymptotic Generalization of de Bruijn's Identity*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 25–32, 1995.

[194] Schalkwijk, J.P.M., Meeuwissen, H.B. and Diederiks, P.J.E., *Two-Way Channels with Delay*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 33–40, 1995.

[195] Vanroose, P., *Code Construction for Non-Cooperative Deterministic Multiuser Channels*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 151–158, 1995.

[196] Tsybakov, B.S. and Weber, J.H., *Conflict-Avoiding Codes*, Seventeenth SITB (Enschede), pp. 49–56, 1996.

[197] Schalkwijk, J.P.M. and Meeuwissen, H.B., *Efficient Coding Strategies From Two-Dimensional Weighting*, Seventeenth SITB (Enschede), pp. 129–136, 1996.

[198] Meeuwissen, H.B. and Schalkwijk, J.P.M., *Some Observations on Two-Way Channels*, Eighteenth SITB (Veldhoven), pp. 57–64, 1997.

[199] Bruin, M.G. de and Kamminga, C., *Normalization Procedures and Shannon's Entropy Measure*, Eighteenth SITB (Veldhoven), pp. 131–141, 1997.

[200] Fu, F.-W. and Vinck, A.J., *On the Capacity of Generalized Write-Once Memory with State Transitions Described by an Arbitrary Directed Acyclic Graph*, Eighteenth SITB (Veldhoven), pp. 150–158, 1997.

[201] Pinsker, M.S., Prelov, V.V. and E.C. van der Meulen, *Information Transmission over Stationary Channels with Additive Non-Gaussian Noise by Means of Weak Input Signals*, Nineteenth SITB (Veldhoven), pp. 143–148, 1998.

[202] Pinsker, M.S., Prelov, V.V. and E.C. van der Meulen, *On Certain Channels with a Random Parameter*, Twentieth SITB (Haasrode), pp. 165–172, 1999.

[203] Koshelev, V.N. and E.C. van der Meulen, *More on the Duality Between Source and Channel Coding*, Twentieth SITB (Haasrode), pp. 181–188, 1999.

[204] Levendovsky, J., Kovács, L., Koller, I. and E.C. van der Meulen, *Optimal Resource Management Algorithm for Adaptive Modelling*, Twentieth SITB (Haasrode), pp. 197–204, 1999.

[205] Tolhuizen, L.M.G.M., *The Binary Multiplying Channel Without Feedback: New Rate Pairs in the Zero-Error Capacity Region*, Twentieth SITB (Haasrode), pp. 215–218, 1999.

[206] Vinck, A.J., *Coding for Random Access Communications*, Twentieth SITB (Haasrode), p. 227, 1999.

[207] Badreddin, E., *Information Theoretic Aspects in the Design of Autonomous Robots*, Twenty-first SITB (Wassenaar), pp. 261–268, 2000.

[208] Pinsker, M.S., Prelov, V.V. and E.C. van der Meulen, *Information Transmission of Slowly Varying Input Signals over Discrete Memoryless Stationary Channels*, Twenty-first SITB (Wassenaar), pp. 277–284, 2000.

[209] Prelov, V.V. and E.C. van der Meulen, *Asymptotic Investigation of the Optimal Filtering Error and Information Rates in Certain Models of Observations and Channels*, Twenty-second SITB (Enschede), pp. 93–100, 2001.

[210] Verdú, S., *New Tools for the Analysis of the Capacity of Very Noisy Channels*, Twenty-second SITB (Enschede), pp. 101–105, 2001.

[211] Prelov, V.V. and E.C. van der Meulen, *Epsilon-Entropy of a Special Class of Ellipsoids in a Hamming Space*, Twenty-third SITB (Louvain-la-Neuve), pp. 37–43, 2002.

[212] Barbé, A. and von Haeseler, F., *Symmetric Codes over Rings*, Twenty-third SITB (Louvain-la-Neuve), pp. 87–95, 2002.

[213] Prelov, V.V. and E.C. van der Meulen, *Asymptotic Expansions of Mutual Information for a General Class of Additive Noise Channels with Small Signal-to-Noise Ratio*, Twenty-fourth SITB (Veldhoven), pp. 165–170, 2003.

WIC Symposium Source Coding Papers

[214] Schalkwijk, J.P.M., *On Petry's Extension of a Source Coding Algorithm*, Second SITB (Zoetermeer), pp. 99-102, 1981.

[215] Desmedt, Y., Vandewalle, J., Govaerts, R., *The Influence of Parallel Coders in the Encoding of a Discrete Source*, Third SITB (Zoetermeer), pp. 13-17, 1982.

[216] Tjalkens, Tj.J., Willems, F.M.J., *Variable to Fixed Length Source Codes for Unifilar Markov Sources*, Fifth SITB (Aalten), pp. 168-177, 1984.

[217] Jansen, P., Oosterlinck, A., *On the Construction of Self-Synchronizing Efficient Encodings*, Sixth SITB (Mierlo), pp. 117-124, 1985.

[218] Vanroose, P., Verbeke, J., *Enkele Beschouwingen bij Optimale Prefix Codes en de Huffman Procedure*, Sixth SITB (Mierlo), pp. 125-132, 1985.

[219] Tjalkens, Tj.J., Willems, F.M.J., *Arithmetic Coding*, Sixth SITB (Mierlo), pp. 141-150, 1985.

[220] Willems, F.M.J., *Repetition Times and Universal Data Compression*, Seventh SITB (Noordwijkerhout), pp. 73-80, 1986.

[221] Tjalkens, Tj.J., *Constructing Arithmetic Source Codes*, Seventh SITB (Noordwijkerhout), pp. 81-88, 1986.

[222] Tjalkens, Tj.J., Willems, F.M.J., *Universal Variable to Fixed Length Source Coding for Binary Memoryless Sources*, Eight SITB (Deventer), pp. 164-170, 1987.

[223] Willems, F.M.J., *Fixed-To-Variable Length Petry Codes*, Eight SITB (Deventer), pp. 214-221, 1987.

[224] Shtarkov, Y.M., Tjalkens, Tj.J., *The Redundancy of the Ziv-Lempel for Memoryless Sources*, Eleventh SITB (Noordwijkerhout), pp. 36-42, 1990.

[225] Tjalkens, Tj.J., Willems, F.M.J., *A Lower Bound on the Asymptotic Redundancy of Universal Variable-To-Fixed Length Codes for Binary Memoryless Sources*, Eleventh SITB (Noordwijkerhout), pp. 43-46, 1990.

[226] With, P.H.N. de, *On the Construction of High-Performance Self-Synchronizing Codes*, Eleventh SITB (Noordwijkerhout), p. 114, 1990.

[227] Barron, A.R., Györfi, L., E.C. van der Meulen, *Universal Source Coding Based on Consistent Distribution Estimation*, Thirteenth SITB (Enschede), pp. 91-97, 1992.

[228] Györfi, L., Pàli, I., E.C. van der Meulen, *Good News and Bad News for Universal Noiseless Source Coding for Infinite Source Alphabet*, Thirteenth SITB (Enschede), p. 99, 1992.

[229] Shtarkov, Y.M., Volkov, S., *Practical Text Compression with Universal Coding*, Thirteenth SITB (Enschede), p. 101, 1992.

[230] Vanroose, P., *On Efficient Tree Representations*, Thirteenth SITB (Enschede), pp. 103-110, 1992.

[231] Willems, F.M.J., Shtarkov, Y.M., Tjalkens, Tj.J., *Context Tree Weighting: General Finite Memory Sources*, Fourteenth SITB (Veldhoven), pp. 120-127, 1993.

[232] Tjalkens, Tj.J., Shtarkov, Y.M., Willems, F.M.J., *Context Tree Weighting: Multi-Alphabet Sources*, Fourteenth SITB (Veldhoven), pp. 128-135, 1993.

[233] Györfi, L., Pàli, I., E.C. van der Meulen, *A General Sufficient Condition for Universal Source Coding for Infinite Alphabets*, Fourteenth SITB (Veldhoven), pp. 136-143, 1993.

[234] Volf, P.J., Willems, F.M.J., *Context Maximizing: Finding MDL Decision Trees*, Fifteenth SITB (Louvain-La-Neuve), pp. 192-199, 1994.

[235] Willems, F.M.J., *The Context Tree Weighting Method: Finite Accuracy Effects*, Fifteenth SITB (Louvain-La-Neuve), pp. 200-207, 1994.

[236] Tjalkens, Tj.J., Willems, F.M.J., Shtarkov, Y.M., *Multi-Alphabet Universal Coding Using a Binary Decomposition Context Tree Weighting Algorithms*, Fifteenth SITB (Louvain-La-Neuve), pp. 259-262, 1994.

[237] Macq, B., Marichal, X., Queluz, M.P., *Entropy Coding of Tree Decompositions*, Fifteenth SITB (Louvain-La-Neuve), pp. 282-289, 1994.

[238] Tjalkens, Tj.J., Willems, F.M.J., *A Comparison of the Lempel-Ziv 1977 and 1978 Universal Data Compression Schemes*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 1-2, 1995.

[239] Volf, P.J.A., Willems, F.M.J., *A Study of the Context Tree Maximizing Methods*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 3-9, 1995.

[240] Gerrits, A.J., Beuker, R.A., Keesman, G.J., *Lossless Compression of Handwritten Signals*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 11-16, 1995.

[241] Vanroose, P., *On Complexity Measures for a Tree*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 41-47, 1995.

[242] Mitrea, M., P.H.N. de With, *A Comparison Between Huffman and Arithmetic Coding for Video Compression*, Seventeenth SITB (Enschede), pp. 25-30, 1996.

[243] Volf, P.A.J., Willems, F.M.J., *Context-Tree Weighting for Extended Tree Sources*, Seventeenth SITB (Enschede), pp. 95-101, 1996.

[244] Keesman, G.J., *Unification of Several Lossless Compression Codes*, Seventeenth SITB (Enschede), pp. 103-109, 1996.

[245] Shtarkov, Y.M., Tjalkens, Tj.J., Willems, F.M.J., *Optimal Universal Coding with Respect to the Relative Redundancy Criterion*, Seventeenth SITB (Enschede), pp. 111-117, 1996.

[246] Schouhamer Immink, K.A., Janssen, A.J.E.M., *Effects of Floating Point Arithmetic in Enumerative Coding*, Eighteenth SITB (Veldhoven), pp. 70-75, 1997.

[247] Volf, P.A.J., Willems, F.M.J., *A Context-Tree Branch-Weighting Algorithm*, Eighteenth SITB (Veldhoven), pp. 115-122, 1997.

[248] Willems, F.M.J., Tjalkens, Tj.J., *Complexity Reduction of the Context-Tree Weighting Method*, Eighteenth SITB (Veldhoven), pp. 123-130, 1997.

[249] Volf, P.A.J., Willems, F.M.J., *The Switching Method: Elaborations*, Nineteenth SITB (Veldhoven), pp. 12-20, 1998.

[250] Vleuten, R.J. van der, Bruekers, A.A.M.L., *Modeling Binary Audio Signals for Lossless Compression*, Nineteenth SITB (Veldhoven), pp. 135-142, 1998.

[251] Balakirsky, V.B., Willems, F.M.J., *Nonasymptotic Lower Bound on the Maximal Cumulative Redundancy of Universal Coding*, Twentieth SITB (Haasrode), pp. 17-24, 1999.

[252] Volf, P.A.J., Willems, F.M.J., Tjalkens, Tj.J., *Complexity Reducing Techniques for the CTW Algorithm*, Twentieth SITB (Haasrode), pp. 25-32, 1999.

[253] Vanroose, P., *Stochastic Language Modelling Using Context Tree Weighting*, Twentieth SITB (Haasrode), pp. 33-38, 1999.

[254] Tjalkens, Tj.J., *The Complexity of Minimum Redundancy Coding*, Twenty-first SITB (Wassenaar), pp. 247-254, 2000.

[255] Nowbakht, A., Tjalkens, Tj.J. Willems, F.M.J., *Coding for Sources Satisfying a Permutation Property*, Twenty-second SITB (Enschede), pp. 77-84, 2001.

[256] Stassen, M.L.A., Tjalkens, Tj.J., *A Parallel Implementation of the CTW Compression Algorithm*, Twenty-second SITB (Enschede), pp. 85-92, 2001.

[257] Nowbakht, A., Willems, F.M.J, *Faster Universal Modeling for Two Source Classes*, Twenty-third SITB (Louvain-La-Neuve), pp. 29-36, 2002.

[258] Hekstra, A.P., *Improvements of the Context Tree Maximizing (CTM) Data Compression Algorithm*, Twenty-third SITB (Louvain-La-Neuve), pp. 123-130, 2002.

[259] Salden, A., Aldershoff, F., Iacob, S., Otte, R., *Web-Enabled Multimedia Categorization*, Twenty-third SITB (Louvain-La-Neuve), pp. 9-16, 2002.

[260] Stasinski R. and G. Ulacha, *Huffman Codes Revisited*, Twenty-fourth SITB (Veldhoven), pp. 63-70, 2003.

WIC Symposium Cryptology Papers

[261] Piret, Ph., *Wire-Tapping of a Binary Symmetric Channel*, First SITB (Zoetermeer), pp. 55–57, 1980.

[262] Desmedt, Y., Vandewalle, J., Govaerts, R., *Critical Analysis of the Security of Knapsack Public Key Algorithms*, Third SITB (Zoetermeer), pp. 19–27, 1982.

[263] Lenstra, H.W., Jr., *Primality and Factorization*, Fourth SITB (Haasrode), pp. 13–17, 1983.

[264] Massey, J.L., *Logarithms in Finite Cyclic Groups Cryptographic Issues*, Fourth SITB (Haasrode), pp. 17–25, 1983.

[265] Desmedt, Y., Vandewalle, J., Govaerts, R., *A General Public Key Cryptographic Knapsack Algorithm Based on Linear Algebra*, Fourth SITB (Haasrode), pp. 55–62, 1983.

[266] Desmedt, Y., Vandewalle, J., Govaerts, R., *The Mathematical Relation Between the Economic, Cryptographic and Information Theoretical Aspects of Authentication*, Fourth SITB (Haasrode), pp. 63–65, 1983.

[267] Jansen, C.J.A., *Classical Key Management*, Fifth SITB (Aalten), pp. 94–101, 1984.

[268] Jansen, C.J.A., *Key Signature Schemes*, Seventh SITB (Noordwijkerhout), pp. 197–205, 1986.

[269] Tilburg, J. van, Boekee, D.E., *The P_e-Security Distance as a Generalized Unicity Distance*, Seventh SITB (Noordwijkerhout), pp. 207–215, 1986.

[270] Jansen, C.J.A., Boekee, D.E., *The Algebraic Normal Form of Arbitrary Functions over Finite Fields*, Eight SITB (Deventer), pp. 69–76, 1987.

[271] Struik, R., Tilburg, J. van, Boly, J-P., *On the Rao-Nam Private-Key Cryptosystem*, Ninth SITB (Mierlo), pp. 137–145, 1988.

[272] Franx, W.G., Jansen, C.J.A., Boekee, D.E., *An Efficient Algorithm for the Generation of De Bruyn Cycles*, Ninth SITB (Mierlo), pp. 147–154, 1988.

[273] Boekee, D.E., Lubbe, J.C.A., van der, *Error Probabilities and Transposition Ciphers*, Ninth SITB (Mierlo), pp. 155–162, 1988.

[274] Willems, F.M.J., *On Gaussian Channels with Side Information at the Transmitter*, Ninth SITB (Mierlo), pp. 129–136, 1988.

[275] Jansen, C.J.A., Boekee, D.E., *Information Theory of Shift Register Sequences*, Tenth SITB (Houthalen), pp. 153–160, 1989.

[276] Jansen, C.J.A., *On the Construction of De Bruijn Sequences*, Eleventh SITB (Noordwijkerhout), pp. 47–51, 1990.

[277] Preneel, B., Van Leekwijck, W., Van Linden, L., Govaerts, R., Vandewalle, J., *An Extension of Higher Order Propagation Criteria for Boolean Functions*, Eleventh SITB (Noordwijkerhout), pp. 52–59, 1990.

[278] Lubbe, J.C.A. van der, Spaanderman, J.J., Boekee, D.E., *On Cryptosystems for Digital Imagery*, Eleventh SITB (Noordwijkerhout), pp. 60–66, 1990.

[279] Verboven, B., *Identification via a Stochastically Varying Channel*, Eleventh SITB (Noordwijkerhout), pp. 168–173, 1990.

[280] Daemen, J., Govaerts, R., Vandewalle, J., *Efficient Pseudorandom Sequence Generation by Cellular Automata*, Twelfth SITB (Veldhoven), pp. 17–24, 1991.

[281] Daemen, J., Van Linden, L., Govaerts, R., Vandewalle, J., *Propagation Properties of Multiplication Modulo 2N-1*, Thirteenth SITB (Enschede), pp. 111–118, 1992.

[282] Preneel, B., Bosselaers, A., Govaerts, R., Vandewalle, J., *A Software Implementation of the McEliece Public-Key Cryptosystem*, Thirteenth SITB (Enschede), pp. 119–126, 1992.

[283] Verschuren, J., Govaerts, R., Vandewalle, J., *Relationship Between the Bell-La Padula Security Policy and Security Services in the OSI-RM*, Thirteenth SITB (Enschede), pp. 127–134, 1992.

[284] Macq, B., Quisquater, J.J., *Lossless Image Encryption*, Fourteenth SITB (Veldhoven), pp. 96–103, 1993.

[285] Delos, O., Quisquater, J.J., *Digital Signature Schemes with Several Cooperating Entities*, Fourteenth SITB (Veldhoven), pp. 104–113, 1993.

[286] Tilburg, J. van, *Cryptanalysis of the Alabbadi-Wicker Digital Signature Scheme*, Fourteenth SITB (Veldhoven), pp. 114–119, 1993.

[287] Harpes, C., Kremer, G.G., Massey, J.L., *Generalized Linear Cryptanalysis and the Applicability of the Piling-Up Lemma*, Fifteenth SITB (Louvain-La-Neuve), pp. 90–99, 1994.

[288] Bosselaers, A., Govaerts, R., Vandewalle, J., *A Fast and Flexible Software Library for Large Number Arithmetic*, Fifteenth SITB (Louvain-La-Neuve), pp. 100–107, 1994.

[289] Daemen, J., Govaerts, R., Vandewalle, J., *An Efficient Nonlinear Shift-Invariant Transformation*, Fifteenth SITB (Louvain-La-Neuve), pp. 108–115, 1994.

[290] Delos, O., Quisquater, J.J., *Schemes for Signature with Bounded Life-Span*, Fifteenth SITB (Louvain-La-Neuve), pp. 116–118, 1994.

[291] Béguin, P., Quisquater, J.J., *Resistant Server-Aided Computations for Public-Key Cryptosystems*, Fifteenth SITB (Louvain-La-Neuve), pp. 127–131, 1994.

[292] Radu, C., Vandenwauver, M., Govaerts, R., Vandewalle, J., *Subject View Access Mechanism in the Personal Database*, Fifteenth SITB (Louvain-La-Neuve), pp. 119–126, 1994.

[293] Boucqueau, J.M., Bruyndonckx, O., Lacroix, S., Mertès, J.Y., Macq, B., Quisquater, J.J., *Access Control and Copyright Protection for Images*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 17–24, 1995.

[294] Dijk, M. van, *Coding Gain Strategies for the Binary Symmetric Broadcast Channel with Confidential Messages*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 53–59, 1995.

[295] Radu, C., Vandenwauver, M., Govaerts, R., Vandewalle, J., *An Efficient Traceable Payment System*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 61–67, 1995.

[296] Tilburg, J. van, *The Fall of the Alabbadi and Wicker Digital Signature Schemes*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 69–72, 1995.

[297] Verschuren, J., *On the Security of OSI-Based Computer Networks*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 73–79, 1995.

[298] Hekstra, A.P., Tilburg, J. van, *An Efficient Scheme Broadcasting Secured Messages*, Seventeenth SITB (Enschede), p. 31, 1996.

[299] Langelaar, G.C., Lubbe, J.C.A. van der, Biemond, J., *Copy Protection for Multimedia Data Based on Labeling Techniques*, Seventeenth SITB (Enschede), pp. 33–39, 1996.

[300] Dijk, M. van, Koppelaar, A., *Quantum Key Agreement*, Eighteenth SITB (Veldhoven), pp. 97–104, 1997.

[301] Langelaar, G.C., Lagendijk, R.L., Biemond, J., *Real-Time Labeling Methods for MPEG Compressed Video*, Eighteenth SITB (Veldhoven), pp. 25–32, 1997.

[302] Vandenwauver, M., Govaerts, R., Vandewalle, J., *An Overview of E-Mail Security Schemes*, Eighteenth SITB (Veldhoven), pp. 105–112, 1997.

[303] Verheul, E.R., Tilborg, H.C.A. van, *Cryptanalysis of Less Short' RSA Secret Exponents*, Eighteenth SITB (Veldhoven), pp. 113–114, 1997.

[304] Kalker, T., *A Security Risk for Publicly Available Watermark Detectors*, Nineteenth SITB (Veldhoven), pp. 119–125, 1998.

[305] Van Rompay, B., Preneel, B., Vandewalle, J., *On the Security of Dedicated Hash Functions*, Nineteenth SITB (Veldhoven), pp. 103–110, 1998.

[306] Borst, J., Preneel, B., Vandewalle, J., *On the Time-Memory Tradeoff Between Exhaustive Key Search and Table Precomputation*, Nineteenth SITB (Veldhoven), pp. 111–118, 1998.

[307] Hachez, G., Koeune, F., Quisquater, J.-J., *Timing Attack: What Can Be Achieved by a Powerful Adversary?*, Twentieth SITB (Haasrode), pp. 63–70, 1999.

[308] Van Rompay, B., Preneel, B., Vandewalle, J., *The Digital Timestamping Problem*, Twentieth SITB (Haasrode), pp. 71–78, 1999.

[309] Massias, H., Serret Avila, X., Quisquater, J.-J., *Design of a Secure Timestamping Service with Minimal Trust Requirement*, Twentieth SITB (Haasrode), pp. 79–86, 1999.

[310] Balakirsky, V.B., *Characterization of the Secrecy of a Common Key Constructed via Data Transmission over the Two-Way "And" Channel*, Twentieth SITB (Haasrode), pp. 87–94, 1999.

[311] Xu, S.-B., Doumen, J., *An Attack Against the Alabbadi-Wicker Scheme*, Twentieth SITB (Haasrode), pp. 95–100, 1999.

[312] Nakahara, J., Jr., Vandewalle, J., Preneel, B., *Diffusion Analysis of Feistel Networks*, Twentieth SITB (Haasrode), pp. 101–108, 1999.

[313] Claessens, J., Preneel, B., Vandewalle, J., *Anonymity Controlled Electronic Payment Systems*, Twentieth SITB (Haasrode), pp. 109–116, 1999.

[314] Kalker, T., Oostveen, J., Linnartz, J.-P., *Maximum Likelihood Detection of Multiplicative Watermarks*, Twenty-first SITB (Wassenaar), pp. 101–108, 2000.

[315] Struik, R., *On One-Pass Combined Encryption and Authentication*, Twenty-first SITB (Wassenaar), pp. 109–113, 2000.

[316] Borst, J., Preneel, B., Vandewalle, J., *Power Analysis: Methods and Countermeasures*, Twenty-first SITB (Wassenaar), pp. 115–120, 2000.

[317] Tilburg, J. van, *Boosting the e-Security of GSM*, Twenty-first SITB (Wassenaar), pp. 121–128, 2000.

[318] Meijer, M.R., Jansen, C.J.A., *Efficient Run Permuted Sequence Generation*, Twenty-first SITB (Wassenaar), pp. 129–137, 2000.

[319] Kremer, S., Markowitch, O., *Optimistic Non-Repudiable Information Exchange*, Twenty-first SITB (Wassenaar), pp. 139–146, 2000.

[320] Willems, F.M.J., *An Information Theoretical Approach to Information Embedding*, Twenty-first SITB (Wassenaar), pp. 255–260, 2000.

[321] Dijk, M. van, Willems, F.M.J., *Embedding Information in Grayscale Images*, Twenty-second SITB (Enschede), pp. 147–154, 2001.

[322] Borne, D. van den, Kalker, T., Willems, F.M.J., *Codes for Writing on Dirty Paper*, Twenty-third SITB (Louvain-La-Neuve), pp. 45–52, 2002.

[323] Diaz, C., Claessens, J., Seys, S., Preneel, B., *Information Theory and Anonymity*, Twenty-third SITB (Louvain-La-Neuve), pp. 179–186, 2002.

[324] Gaddach, A., *A New Group Identification Scheme*, Twenty-third SITB (Louvain-La-Neuve), pp. 53–60, 2002.

[325] Batina, L., Jansen, C.J.A., Muurling, G., Xu, S.-B., *"Almost Montgomery" Base Multiplier in* $\mathrm{GAG}(2^N)$, Twenty-third SITB (Louvain-La-Neuve), pp. 61–68, 2002.

[326] Potgieter, M.J., Dyk, B.J. van, Tjalkens, Tj.J., *A Fast Multiplier for Characteristic-2 Finite Fields*, Twenty-third SITB (Louvain-La-Neuve), pp. 69–74, 2002.

[327] Lefebvre, F., Macq, B., Legat, J.-D., *Agaddis: Authentication and Geometrical Attacks Detection for Digital Image Signature*, Twenty-third SITB (Louvain-La-Neuve), pp. 171–178, 2002.

[328] Nakahara, J., Jr., Barreto, P., Preneel, B., Vandewalle, J., Kim, H., *Square Attacks on Reduced-Round PES and IDEA Block Ciphers*, Twenty-third SITB (Louvain-La-Neuve), pp. 187–195, 2002.

[329] Nikov, V., Nikova, S., Preneel, B., Vandewalle, J., *Applying General Access Structure for Proactive Secret Sharing Schemes*, Twenty-third SITB (Louvain-La-Neuve), pp. 197–206, 2002.

[330] Bechlaghem, M., *Multi-Party Server-Aided Key Distribution Protocols Based on Symmetric Techniques*, Twenty-third SITB (Louvain-La-Neuve), pp. 215–223, 2002.

[331] Batina, L., Jansen, C.J.A., *Secret Exponent Information Leakage for Timing Analyses*, Twenty-third SITB (Louvain-La-Neuve), pp. 225–235, 2002.

[332] Ciet, M., Quisquater, J.-J., Francesco, S., *A Short Note on Irreducible Trinomials in Binary Fields*, Twenty-third SITB (Louvain-La-Neuve), pp. 233–234, 2002.

[333] Canteaut, A., Filiol, E., *On the Influence of the Filtering Function on the Performance of Fast Correlation Attacks on Filter Generators*, Twenty-third SITB (Louvain-La-Neuve), pp. 299–306, 2002.

[334] Carlet, C., Klapper, A., *Upper Bounds on the Numbers of Resilient Functions and of Bent Functions*, Twenty-third SITB (Louvain-La-Neuve), pp. 307–314, 2002.

[335] Ciet, M., Piret, G., Quisquater, J.-J., *Related-Key and Slide Attacks: Analysis, Connections, and Improvements*, Twenty-third SITB (Louvain-La-Neuve), pp. 315–325, 2002.

[336] Moulin, P., *Information Hiding Games*, Twenty-third SITB (Louvain-La-Neuve), pp. 382, 2002.

[337] Batina, L. and Jansen, C.J.A., *Side-Channel Entropy for Modular Exponentiation Algorithms*, Twenty-fourth SITB (Veldhoven), pp. 37–44, 2003.

[338] Laguillaumie F. and Vergnaud D., *Extending the Boneh-Durfee-De Weger's Attack to RSA-like Cryptosystems*, Twenty-fourth SITB (Veldhoven), pp. 45–52, 2003.

[339] Standaert, F.-X., Rouvroy, G., Piret, G., Quisquater, J.-J. and Legat J.-D., *Key-Dependent Approximations in Cryptanalysis – an Application of Multiple* $\mathbb{Z}_4$ *and Non-Linear Approximations*, Twenty-fourth SITB (Veldhoven), pp. 53–62, 2003.

[340] Maas, D., Kalker, T. and Willems, F.M.J. *Capacity of Reversible Information Embedding for Small Distortions*, Twenty-fourth SITB (Veldhoven), pp. 95–102, 2003.

[341] Verbitskiy, E., P. Tuyls, D. Denteneer, and J.P. Linnartz, *Reliable (Robust) Biometric Authentication with Privacy Protection*, Twenty-fourth SITB (Veldhoven), 2003.

[342] Ciet, M., Piret G., and Quisquater, J-J., *A Structure of Block Ciphers Achieving Some Resistance Against Fault Attacks*, Twenty-fourth SITB (Veldhoven), pp. 171–178, 2003.

[343] Kholosha, A., *Tensor Transform of Functions over Finite Fields*, Twenty-fourth SITB (Veldhoven), pp. 179–186, 2003.

[344] Saeednia, S., Kremer S., and Markowitch, O., *Efficient Designated Verifier Signature Schemes*, Twenty-fourth SITB (Veldhoven), pp. 187–194, 2003.

[345] Seys S., and B. Preneel, *Authenticated and Efficient Key Management for Ad-Hoc Networks*, Twenty-fourth SITB (Veldhoven), pp. 195–202, 2003.

WIC Symposium Channel Coding Papers

[346] Post, K.A. , *New Upper Bounds for the First Event Error Probability of Binary Convolutional Codes Using Viterbi Decoding on a Binary Symmetric Channel*, First SITB (Zoetermeer), pp. 59–63, 1980.

[347] Roefs, H.F.A. , *Concatenated Coding; an Investigation for the European Space Agency*, First SITB (Zoetermeer), pp. 65–67, 1980.

[348] Schalkwijk, J.P.M. , *On a Description of the Operation of a Maximum Likelihood Decoder for Convolutional Codes that Allows Exact Evaluation of the Event Error Probability*, First SITB (Zoetermeer), pp. 69–82, 1980.

[349] Tilborg, H.C.A. van, Helleseth, T., *New Results Concerning the Griesmer Bound* , First SITB (Zoetermeer), pp. 93–97, 1980.

[350] Best, M.R. en Roefs, H.F.A. , *Telemetrie-Kanaalcodering met de (256,224) Reed-Solomon Code over GF(257)*, Second SITB (Zoetermeer), pp. 25–33, 1981.

[351] Schalkwijk, J.P.M., Brouwer, J.A.M., *On the Complexity of Sequential Decoders*, Second SITB (Zoetermeer), pp. 113–121, 1981.

[352] Roos, C., *A Result on the Minimum Distance of a Linear Code with Applications to Cyclic Codes*, Third SITB (Zoetermeer), pp. 103–111, 1982.

[353] Best, M.R., *A Convolutional Decoder with Reliability Information*, Fourth SITB (Haasrode), pp. 27–29, 1983.

[354] Vroedt, C. de, *On the Weight Enumerator of Self-Dual Codes*, Fourth SITB (Haasrode), pp. 39–42, 1983.

[355] Piret, Ph., *Binary Codes for Compound Channels*, Fourth SITB (Haasrode), pp. 43–47, 1983.

[356] Pul, C.L.M. van, *Lower Bounds for A(n,4,w)*, Fourth SITB (Haasrode), pp. 49–53, 1983.

[357] Busschbach, P.B., Gerretzen, M.G.L., Tilborg, H.C.A. van, *The Numbers S and ρ of Binary Linear Codes, Meeting the Griesmer Bound with Equality*, Fifth SITB (Aalten), pp. 28–35, 1984.

[358] Schouhamer Immink, K.A., *Performance of DC-Constrained Codes*, Fifth SITB (Aalten), pp. 137–143, 1984.

[359] Simons, H.J., Roefs, H.F.A., *Channel Coding with the (255,255-2T) Reed-Solomon Codes over GF(256)*, Fifth SITB (Aalten), pp. 144–151, 1984.

[360] With, P.H.N. de, *On Performance Criteria for DC-Free Codes*, Fifth SITB (Aalten), pp. 194–201, 1984.

[361] Pul, C.L.M. van, *Computer Memories with Defective Cells*, Sixth SITB (Mierlo), pp. 43–47, 1985.

[362] Baggen, C.P.M.J., *MDS Codes for the Correction of Stuck-at Defects*, Sixth SITB (Mierlo), pp. 49–53, 1985.

[363] Vinck, A.J., *Convolutional Code and Defects*, Sixth SITB (Mierlo), pp. 55–61, 1985.

[364] Gils, W.J. van, *Dot Codes for Product Identification*, Sixth SITB (Mierlo), pp. 63–65, 1985.

[365] Haemers, W., *Een Hammingcode voor de Postcode*, Sixth SITB (Mierlo), pp. 67–73, 1985.

[366] Gils, W.J. van, *Construction and Properties of [3,1] Codes over GF(2^m), m=4,8,16, to be Used in a Fault-Tolerant System Based on Triple Modular Redundancy*, Sixth SITB (Mierlo), pp. 75–79, 1985.

[367] Beenker, G.F.M., Schouhamer Immink, K.A., *On the Number of Codewords of a dc^2-Balanced Code*, Sixth SITB (Mierlo), pp. 133–139, 1985.

[368] Best, M.R., *A Markov Chain Model for a Convolutional Coding Scheme*, Sixth SITB (Mierlo), pp. 151–159, 1985.

[369] Nouwens, W.J.W.M., Verlijsdonk, A.P., *Soft-Decision, R=1/2, Viterbi Decoding*, Sixth SITB (Mierlo), pp. 171–181, 1985.

[370] Blaum, M., Farrell, P.G., Tilborg, H.C.A. van, *A Class of Burst Correcting Codes*, Seventh SITB (Noordwijkerhout), pp. 31–36, 1986.

[371] Gils, W.J. van, *An Error-Control Coding System for Storage of 16-Bit Words in Memory Arrays Composed of Three 9-Bit Wide Units*, Seventh SITB (Noordwijkerhout), pp. 37–40, 1986.

[372] Boly, J-P., Gils, W.J., *On Combined Symbol and Digit Error-Control Codes* , Eight SITB (Deventer), pp. 45–52, 1987.

[373] Moolen, P.C.M. van der, *Decoding with Memory*, Eight SITB (Deventer), pp. 93–99, 1987.

[374] Schouhamer Immink, K.A., *Coding Techniques for Partial-Response Channels*, Eight SITB (Deventer), pp. 149–156, 1987.

[375] Tolhuizen, L.M.G.M., *On the Blokh-Zyablov Construction*, Eight SITB (Deventer), pp. 171–174, 1987.

[376] Vinck, A.J., Post, K.A., *Application of a Combined Test-Error-Correcting Procedure for Memories with Defects*, Eight SITB (Deventer), pp. 189–195, 1987.

[377] Weber, J.H., Vroedt, C. de, Boekee, D.E., *A Construction Method for Codes Correcting Asymmetric Errors*, Eight SITB (Deventer), pp. 203–207, 1987.

[378] Weber, J.H., Vroedt, C. de, Boekee, D.E., *Bounds on the Size of Codes Correcting Unidirectional Errors*, Ninth SITB (Mierlo), pp. 9-15, 1988.

[379] Stevens, P., *Extension of the BCH Decoding Algorithm in Order to Decode Binary Cyclic Codes up to Their Maximum Correction Capacities*, Ninth SITB (Mierlo), pp. 17–23, 1988.

[380] Kapralov, S.N., Tonchev, V.D., *Extremal Doubly-Even Codes of Length 64 Derived from Symmetric Designs*, Ninth SITB (Mierlo), pp. 25–30, 1988.

[381] Schalkwijk, J.P.M., Post, K.A., *Simple and Optimal Coding Strategies for Memories with Known Defects*, Ninth SITB (Mierlo), pp. 49–57, 1988.

[382] Peek, J.A., Vinck, A.J., *Bit Error Rate and Complexity of a New Coding Algorithm for Defect Channels and Erasure Channels*, Ninth SITB (Mierlo), pp. 59–65, 1988.

[383] Vinck, A.J., Vleuten, R. van der, *A Method to Implement Linear Complexity Decoding*, Ninth SITB (Mierlo), pp. 75–80, 1988.

[384] Weber, J.H., Vroedt, C. de, Boekee, D.E., *Conditions on Block Codes for Correction/Detection of Errors of Various Types*, Tenth SITB (Houthalen), pp. 31–36, 1989.

[385] Tolhuizen, L.M.G.M., Baggen, S., *On the Correcting Capabilities of Product Codes*, Tenth SITB (Houthalen), pp. 45–50, 1989.

[386] Ericson, T., *Concatenated Codes – A Survey of Recent Developments*, Tenth SITB (Houthalen), pp. 89–91, 1989.

[387] Tilburg, J. van, *A Probabilistic Decoding Scheme*, Tenth SITB (Houthalen), pp. 147–152, 1989.

[388] Stevens, P., *Two Suggestions to Improve on the Efficiency of the Check Computations in the Banking-System in Belgium*, Tenth SITB (Houthalen), pp. 161–167, 1989.

[389] Weber, J.H., Abdel-Ghaffar, K.A.S., *A Class of Runlength-Limited Error Detecting Codes*, Eleventh SITB (Noordwijkerhout), pp. 22–28, 1990.

[390] Hollmann, H.D.L., Schouhamer Immink, K.A., *Enumeration of Prefix-Synchronized Runlength-Limited Sequences* , Twelfth SITB (Veldhoven), pp. 79–85, 1991.

[391] Hollmann, H.D.L., Tolhuizen, L.M.G.M., *Relaxed Conditions for Successful Generalized Minimum Distance Decoding*, Twelfth SITB (Veldhoven), pp. 87–93, 1991.

[392] Weber, J.H., Abdel-Ghaffar, K.A.S., *Methods for Cascading Runlength-Limited Sequences*, Twelfth SITB (Veldhoven), pp. 95–101, 1991.

[393] Hekstra, A.P., *On the Maximum Difference Between Path Metrics in a Viterbi Decoder*, Thirteenth SITB (Enschede), pp. 47-55, 1992.

[394] Tjalkens, Tj.J., *Glueless Runlength-Limited Sequences*, Thirteenth SITB (Enschede), pp. 57–64, 1992.

[395] Weber, J.H., Abdel-Ghaffar, K.A.S., *Merging Bits for Cascading Runlength-Limited Sequences*, Thirteenth SITB (Enschede), pp. 65–72, 1992.

[396] Veugen, T.H., *Repetition Strategies for the Binary Symmetric Channel with Feedback*, Fourteenth SITB (Veldhoven), pp. 8–13, 1993.

[397] Hekstra, A.P., *Reduction of the Numerical Range of the Path Metrics in a Viterbi Decoder*, Fourteenth SITB (Veldhoven), pp. 70–78, 1993.

[398] Hollmann, H.D.L., *Construction of Bounded-Delay Encodable Modulation Codes by State-Combination and State-Splitting*, Fourteenth SITB (Veldhoven), pp. 80–87, 1993.

[399] Weber, J.H., Kaag, G.H., *A Construction Method for Systematic Codes Correcting/Detecting Asymmetric Errors*, Fourteenth SITB (Veldhoven), pp. 88–94, 1993.

[400] Delsarte, P., *Application and Generalization of the MacWilliams Transform in Coding Theory*, Fifteenth SITB (Louvain-La-Neuve), pp. 9–44, 1994.

[401] Ericson, T., Zinoviev, V., *Spherical Codes from Unsymmetric Alphabets*, Fifteenth SITB (Louvain-La-Neuve), pp. 45–52, 1994.

[402] Gillot, V., *Minimum Weight for Codes Stemmed from Exponential Sums Bounds*, Fifteenth SITB (Louvain-La-Neuve), pp. 53–60, 1994.

[403] Peirani, B., *(U,U+V) Codes of Asymptotic Normal Weight Distribution*, Fifteenth SITB (Louvain-La-Neuve), pp. 61–68, 1994.

[404] Vanroose, P., *In Search of Maximum Distance Separable Codes over the Ring of Integers Modulo M*, Fifteenth SITB (Louvain-La-Neuve), pp. 69–76, 1994.

[405] Weber, J.H., *Asymptotic Results on Symmetric, Unidirectional and Asymmetric Error Control Codes*, Fifteenth SITB (Louvain-La-Neuve), pp. 77–81, 1994.

[406] Veugen, T., *Error Probabilities of Repetition Feedback Strategies with Fixed Delay for Discrete Memoryless Channels*, Fifteenth SITB (Louvain-La-Neuve), pp. 188–191, 1994.

[407] Veugen, T., *Tail Conditions for Multiple Repetition Feedback Block Coding*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 107–113, 1995.

[408] Offermans, G.W.A., Breeuwer, E.J., Weber, J.H., Willigen, D. van, *Error-Correction Strategies for the Eurofix Navigation System*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 115–122, 1995.

[409] Baggen, C.P.M.J., Tolhuizen, L.M.G.M., *On the Diamond Code Construction*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 123–126, 1995.

[410] Tolhuizen, L.M.G.M., Baggen, C.P.M.J., *Block Variations of Diamond Codes* , Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 127–131, 1995.

[411] Hekstra, A.P., *Synchronisation for Codes on Circles*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 175–176, 1995.

[412] Abdel-Ghaffar, K.A.S., Weber, J.H., *Constrained Block Codes for Partial-Response Maximum-Likelihood Channels*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 177–183, 1995.

[413] Willems, F.M.J., Pašić, A., *Minimizing the Packet Error Probability*, Seventeenth SITB (Enschede), pp. 41–48, 1996.

[414] Schalkwijk, J.P.M., Bargh, M.S., *Coding for Channels with Low Rate Noiseless Feedback*, Seventeenth SITB (Enschede), pp. 121–127, 1996.

[415] Heijnen, P., *The Decoding of Binary Quasi-Cyclic Codes*, Eighteenth SITB (Veldhoven), pp. 1–4, 1997.

[416] Keuning, J., *Performance and Complexity of Decoding Algorithms*, Eighteenth SITB (Veldhoven), pp. 5–8, 1997.

[417] Bratatjandra, G.H., Weber, J.H., *Variable-Rate Codes for Multiple Localized Burst Error Correction*, Eighteenth SITB (Veldhoven), pp. 49–56, 1997.

[418] Bart, B. de, *Coping with Ambiguities in the Channel Code for DVB-S*, Eighteenth SITB (Veldhoven), pp. 65–69, 1997.

[419] Tolhuizen, , *A Bound on the State-Complexity of a Binary Linear Block Code*, Eighteenth SITB (Veldhoven), pp. 76–80, 1997.

[420] Weber, J.H., Abdel-Ghaffar, K.A.S., *Inner Decoder Optimization in a Simple Concatenated Coding Scheme with Single-Trial Decoding*, Nineteenth SITB (Veldhoven), pp. 67–74, 1998.

[421] Tolhuizen, L.M.G.M., Hekstra-Nowacka, E., *Some Results on Serially Concatenated Codes*, Nineteenth SITB (Veldhoven), pp. 75–82, 1998.

[422] Dijk, M. van, Keuning, J., *A Quaternary BCH-Code Based Binary Quasi-Cyclic Code Construction*, Nineteenth SITB (Veldhoven), pp. 83–90, 1998.

[423] Bargh, M.S., Schalkwijk, J.P.M., *A Block Retransmission Strategy for Multiple Repetition Feedback Coding Schemes*, Nineteenth SITB (Veldhoven), pp. 91–98, 1998.

[424] Canogar, R., *An Example of Reconstructing the Cells of a Partition Design from its Adjacency Matrix*, Nineteenth SITB (Veldhoven), pp. 99–102, 1998.

[425] Vangheluwe, S., *Experimental Investigation of Bounds on the Rate of Superimposed Codes in* $\mathbb{R}^n$, Nineteenth SITB (Veldhoven), pp. 165–172, 1998.

[426] Stam, M., Vinck, A.J., *On Optical Orthogonal Codes*, Nineteenth SITB (Veldhoven), pp. 185–192, 1998.

[427] Koppelaar, A., *Soft-in Soft-Out Multiplexers as a Building Block in Soft-Output Viterbi Decoders*, Nineteenth SITB (Veldhoven), pp. 193–200, 1998.

[428] Bargh, M.S., Schalkwijk, J.P.M., *Recursive Decoding of Multiple Repetition Feedback Coding Schemes for Binary-Input Soft-Output Discrete Memoryless Channels*, Nineteenth SITB (Veldhoven), pp. 201–208, 1998.

[429] Bargh, M.S., Schalkwijk, J.P.M., *On Error Correction in Information Feedback Schemes*, Twentieth SITB (Haasrode), pp. 173–180, 1999.

[430] Weber, J.H., Abdel-Ghaffar, K.A.S., *Error Correction Capabilities of Concatenated Coding Schemes with Single-Trial Bounded Distance Decoding and Optimized Erasing*, Twentieth SITB (Haasrode), pp. 219–226, 1999.

[431] Weber, J.H., Abdel-Ghaffar, K.A.S., *Single-Trial Generalized Minimum Distance Decoding* , Twenty-first SITB (Wassenaar), pp. 1–8, 2000.

[432] Dielissen, J., Huisken, J., *Implementation Issues of 3rd Generation Mobile Communication Turbo Decoding*, Twenty-first SITB (Wassenaar), pp. 9–16, 2000.

[433] Janssen, A.J.E.M., Koppelaar, A.G.C., *Box-Functions and Mismatched Log-Likelihood Ratios*, Twenty-first SITB (Wassenaar), pp. 17–24, 2000.

[434] Muurling, G., Kleihorst, R.P., Benschop, N.F., Vleuten, R. van der, Simonis, J., *Error Correction for Combinational Logic Circuits*, Twenty-first SITB (Wassenaar), pp. 25–31, 2000.

[435] Balakirsky, V.B., *An Upper Bound on the Expected Number of Computations for Maximum Likelihood Decoding of Low-Density Codes*, Twenty-first SITB (Wassenaar), pp. 285–292, 2000.

[436] Martirosyan, S., Vinck, A.J.H., *On Optical Orthogonal Code Constructions with Correlation 1*, Twenty-second SITB (Enschede), pp. 53–57, 2001.

[437] Weber, J.H., Abdel-Ghaffar, K.A.S., *Error-Correction Radius of Reduced GMD Decoders*, Twenty-second SITB (Enschede), pp. 107–114, 2001.

[438] Dijk, M. van, Baggen, S., Tolhuizen, L.M.G.M., *Coding for Informed Decoders*, Twenty-second SITB (Enschede), pp. 123–128, 2001.

[439] Desset, C., *Error Control Coding for Wireless Personal Area Networks*, Twenty-third SITB (Louvain-La-Neuve), 2002.

[440] Tolhuizen, L.M.G.M., Hekstra, A., Cai, N., Baggen, S., *Two Aspects of Coding for Informed Decoders* , Twenty-third SITB (Louvain-La-Neuve), pp. 25–28, 2002.

[441] Steendam, H., Moeneclaey, M. , *ML-Performance of Low-Density Parity Check Codes*, Twenty-third SITB (Louvain-La-Neuve), pp. 75–77, 2002.

[442] Kossen, F., Weber, J. , *Performance Analysis of Limited-Trial Chase Decoders*, Twenty-third SITB (Louvain-La-Neuve), pp. 79–86, 2002.

[443] Piret, P., Le Bars, P., Le Dantec, C. , *Efficient Algebraic Interleavers for Turbocodes* , Twenty-third SITB (Louvain-La-Neuve), pp. 293–297, 2002.

[444] Delsarte, Ph., *The Hamming Space Viewed as an Association Scheme*, Twenty-third SITB (Louvain-La-Neuve), pp. 329–380, 2002.

[445] Sloane, N., *Recent Progress on Self-Dual Codes and Orthogonal Arrays*, Twenty-third SITB (Louvain-La-Neuve), p. 381, 2002.

[446] Sudan, M. , *List Decoding Algorithms: a Survey*, Twenty-third SITB (Louvain-La-Neuve), p. 383, 2002.

[447] Hekstra, Andries P., *Set Decoding of Convolutional Codes with Application to GSM/GPRS*, Twenty-fourth SITB (Veldhoven), p. 9–16, 2003.

[448] Baggen, S., S. Egner, and B. Vandewiele, *On the Use of the Cut-Off Rate for Determining Optimal Input Quantization of a Viterbi Decoder on Fading Channels*, Twenty-fourth SITB (Veldhoven), pp. 17–26, 2003.

[449] Weber, J.H, *Static and Dynamic Chase-Like Bounded Distance Decoding*, Twenty-fourth SITB (Veldhoven), pp. 27–34, 2003.

[450] Baggen, S., and Balakirsky,V.B., *An Efficient Algorithm for Computing the Entropy of Output Sequences for Bitshift Channels*, Twenty-fourth SITB (Veldhoven), pp. 157–164, 2003.

WIC Symposium Communication and Modulation Papers

[451] Bergmans, J.W.M., *Correlative Level Decision Feedback Equalization*, Seventh SITB (Noordwijkerhout), pp. 161-170, 1986.

[452] Bergmans, J.W.M., Jansen, A.J.E.M., *Robust Decision-Feedback Equalization*, Eight SITB (Deventer), pp. 29-36, 1987.

[453] Wolf, J.K., *Coding for Digital Recording Systems*, Ninth SITB (Mierlo), pp. 31, 1988.

[454] Schouhamer Immink, K.A., *Graceful Degradation of Digital Sound Reproduced from Magnetic Recording Channels*, Ninth SITB (Mierlo), pp. 33-39, 1988.

[455] Dekker, H.J., Smit, G., *Multi-Dimensional Trellis-Coded Modulation*, Ninth SITB (Mierlo), pp. 41-47, 1988.

[456] Vleuten, R.J. van der, Schouhamer Immink, K.A., *A Maximum-Likelihood Detector for a Class IV Partial Response Magnetic Recording System*, Tenth SITB (Houthalen), pp. 117-123, 1989.

[457] Giannakouros, N.P., Laloux, A., *Waiting-Time Approximations for Service Systems with a Polling Table*, Tenth SITB (Houthalen), pp. 139-145, 1989.

[458] Bot, P.G.M. de, Vinck, A.J., *Bandwidth Efficient Coding/Modulation with Low-Complexity Detection/Decoding*, Eleventh SITB (Noordwijkerhout), pp. 1-7, 1990.

[459] Bergmans, J.W.M., *On the SNR Merits of Run-Length-Limited Codes in Feedback-Equalized Digital Recoding Systems*, Eleventh SITB (Noordwijkerhout), pp. 8-14, 1990.

[460] Giannakouros, N.P. and Laloux, A., *Optimization of Service Systems with Deterministic Polling via the Pseudoconservation Law*, Eleventh SITB (Noordwijkerhout), pp. 133–139, 1990.

[461] Bergmans, J.W.M., *Effect of Binary Modulation Codes with Rate R=1/n on Equivalent Discrete-Time Models for Channels with Intersymbol Interference*, Twelfth SITB (Veldhoven), pp. 71-78, 1991.

[462] Bot, P.G.M. de, *A Simple Phase Recovery Algorithm for M-PSK with Asymptotically Maximum Likelihood Detection*, Twelfth SITB (Veldhoven), pp. 121-128, 1991.

[463] Camkerten, H., Arnbak, J.C., Sankur, B., *Optimum Single-User Coherent and Partially Coherent BPSK Receiver Design and Exact Performance Analyses for CDMA Uncorrelated Rayleigh Fading Channels*, Thirteenth SITB (Enschede), pp. 73-80, 1992.

[464] Linden, O.L. van der, Bot, P.G.M. de, Baggen, C.P.M.J., *Performance Analysis of 2-DPSK with Non-Coherent Detection on a Ricean Fading Channel*, Fourteenth SITB (Veldhoven), pp. 198-205, 1993.

[465] Prasad, R., *An Overview of Code Division Multiple Access Techniques for Universal Personal Communications Networks*, Fourteenth SITB (Veldhoven), pp. 206-213, 1993.

[466] Prasad, R., Jansen, M.G., Deursen, J.P. van, *Frequency Hopping Slotted Aloha in a Shadowed Radio Environment*, Fourteenth SITB (Veldhoven), pp. 214-221, 1993.

[467] Ribeiro, M.A., *Optimal Bit-Level Synchronization Strategy for Magnetic Recorders*, Fourteenth SITB (Veldhoven), pp. 222-227, 1993.

[468] Linden, O.L. van, *A Multipath Channel Model for Analytical Evaluation of Coded OFDM-Based Transmission Scheme*, Fourteenth SITB (Veldhoven), pp. 228-235, 1993.

[469] Koppelaar, A.G.C., *Matrix Equalization for OFDM Systems*, Fourteenth SITB (Veldhoven), pp. 236-243, 1993.

[470] Bot, P.G.M. de, *Antenna Diversity with Narrow Band Combining*, Fourteenth SITB (Veldhoven), pp. 244-251, 1993.

[471] Rodrigues, A.J., Vandendorpe, L., Albuquerque, A.A., *Direct-Sequences CDMA Multi-H Cpm in Indoor Mobile Radio Systems with Post-Detection Diversity*, Fifteenth SITB (Louvain-La-Neuve), pp. 138-145, 1994.

[472] Siala, M., Kawas Kaleh, G., *Cut-Off of the One-Track Partial Response Magnetic Recording Channel*, Fifteenth SITB (Louvain-La-Neuve), pp. 146-151, 1994.

[473] Van De Wiel, O., Vandendorpe, L., *A Comparison of Bidimensional RLS and LMS Linear Equalization for OFDM/DS Transmission in an Indoor Environment*, Fifteenth SITB (Louvain-La-Neuve), pp. 152-159, 1994.

[474] Ruszinko, M., Vanroose, P., *A Collision Resolution Protocol of Throughput One, Using Multiplicity Feedback*, Fifteenth SITB (Louvain-La-Neuve), pp. 168-174, 1994.

[475] Vvedenskaya, N.D., *Distribution of Message Delay in a Network with Many Multiple Routes*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 49-52, 1995.

[476] Arnbak, J.C., *Between Information Theory and Communication Practice: Observations from a Stranger*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 133-134, 1995.

[477] Jacquemin, P., Rodriques, A.J., Vandendorpe, L., *Multi-H DS-CDMA in Multipath Rayleigh Fading Channels with Multiuser Interference*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 135-142, 1995.

[478] Krapels, M.J., Jansen, G.J.M., *Comparison of BER Performance of Different Detectors for a Narrowband BPSK Dual-Signal Receiver with Co-Channel Interference Cancellation*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 143-150, 1995.

[479] Vvedenskaya, N.D., Linnartz, J.P.M.G., *Performance of Stack Algorithms in Case of Mutually Interfering Transmissions in Two Cells*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 159-166, 1995.

[480] Bart, B. de, Willems, F.M.J., *Combining Enumerative Shaping Techniques and Block Coded Modulation for ISI Channels*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 167-174, 1995.

[481] Vvedenskaya, N.D., *An Example of Optimal Message Routing in a Complete-Graph Network Model*, Seventeenth SITB (Enschede), pp. 65-72, 1996.

[482] Bargh, M.S., Schalkwijk, J.P.M., *Feedback Coded Modulation*, Eighteenth SITB (Veldhoven), pp. 41-48, 1997.

[483] Boxma, O., *Invited Lecture 2: Stochastic Networks*, Nineteenth SITB (Veldhoven), pp. 173-176, 1998.

[484] Levendovsky, J., Elek, Zs., Meulen, E.C. van der, *CAC Based on Queuing Models in ATM Networks*, Nineteenth SITB (Veldhoven), pp. 177-184, 1998.

[485] Gerrits, A., Koppelaar, A., Taori, R., Sluijter, R., Baggen, C., Hekstra-Nowacka, E., *Proposal for an Adaptive Multi-Rate Coder for GSM*, Twentieth SITB (Haasrode), pp. 133-140, 1999.

[486] Peek, J.B.H., *Multirate Block Codes*, Twentieth SITB (Haasrode), pp. 205-214, 1999.

[487] Bakker, J.-D., Schoute, F.C., *LART: Design and Implementation of an Experimental Wireless Platform*, Twenty-first SITB (Wassenaar), pp. 63-70, 2000.

[488] Vinck, A.J.H., *Codes for Frequency Hopping Communication*, Twenty-first SITB (Wassenaar), pp. 147-154, 2000.

[489] Heideman, G., *A Generalization of a Coherence Multiplex System*, Twenty-first SITB (Wassenaar), pp. 155-156, 2000.

[490] Jansen, G.J.M., Slimana, S.B., *BER Results for a Narrowband Multiuser Receiver Based on Successive Subtraction for M-PSK Modulated Signals*, Twenty-first SITB (Wassenaar), pp. 157-164, 2000.

[491] Haartsen, J.C., *Embedded Connectivity with Bluetooth*, Twenty-second SITB (Enschede), pp. 15, 2001.

[492] Levendovszky, J., Fancsali, A., Vegso, Cs., Meulen, E.C. van der, *CNN Based Algorithm for QoS Routing with Incomplete Information*, Twenty-second SITB (Enschede), pp. 45-52, 2001.

[493] Meijerink, A., Heideman, G.H.L.M., Etten, W.C. van, *Generalization and Performance Improvement of a Coherence Multiplexing System*, Twenty-second SITB (Enschede), pp. 59-68, 2001.

[494] Tang, F., Deneire, L., Engels, M., *On the Optimal Switching Scheme of Link Adaption*, Twenty-second SITB (Enschede), pp. 69-76, 2001.

[495] Gorokhov, A., Dijk, M. van , *Optimised Labelings for Bit-Interleaved Coded Modulation Schemes with Iterative Demodulation*, Twenty-second SITB (Enschede), pp. 157-164, 2001.

[496] Vitale, G., Stassen, M.L.A., Colak, S.B., Pronk, V., *Multipath Diffuse Routing over Heterogeneous Mesh Networks of Web Devices and Sensors*, Twenty-third SITB (Louvain-La-Neuve), pp. 1-8, 2002.

[497] Vanhaverbeke, F., Moeneclaey, M., *Sum Capacity of the OCDMA/OCDMA Signature Sequence Set with Unequal Power Constraints*, Twenty-third SITB (Louvain-La-Neuve), pp. 97-105, 2002.

[498] Bargh, M., Eijk, R. van, Salden, A., *Brokerage of Next Generation Mobile Services*, Twenty-third SITB (Louvain-La-Neuve), pp. 247-254, 2002.

[499] Tauboeck, G., *Rotationally Variant Complex Channels*, Twenty-third SITB (Louvain-La-Neuve), pp. 261-268, 2002.

[500] Meijerink, A., Heideman, G., Etten, W. van , *BER Analysis of a DPSK Phase Diversity Receiver for Coherence Multiplexing*, Twenty-third SITB (Louvain-La-Neuve), pp. 269-276, 2002.

[501] Levendovszky, J., Kovacs, L., Meulen, E.C. van der , *A New Blind Signal Processing Algorithm for Channel Equalization*, Twenty-third SITB (Louvain-La-Neuve), pp. 277-284, 2002.

[502] Levendovszky, J., David, T., Meulen, E.C. van der , *Optimal Stochastic Timers for Feedback Mechanisms in Multicast Communications*, Twenty-third SITB (Louvain-La-Neuve), pp. 285-292, 2002.

[503] Calderbank, R., *Combinatorics, Quantum Computing and Cellular Phones*, Twenty-third SITB (Louvain-La-Neuve), pp. 384, 2002.

[504] Houtum, W. van, *On Understanding the Performance of the IEEE 802.11A WLAN Physical Layer for the Gaussian Channel*, Twenty-fourth SITB (Veldhoven), pp. 1-8, 2003.

[505] Riani, J., J.W.M. Bergmans, S.J.L. van Beneden, W.M.J. Coene, and A.H.J. Immink, *Equalization and Target Response Optimisation for High-Density Two-Dimensional Optical Storage*, Twenty-fourth SITB (Veldhoven), pp. 141-148, 2003.

[506] De Lathauwer, L., J. Vandewalle, and B. De Moor, *An Algebraic Technique for Blind MIMO Deconvolution of Constant Modulus*, Twenty-fourth SITB (Veldhoven), pp. 203-210, 2003.

[507] De Lathauwer, L. A. De Baynast, J. Vandewalle, and B. De Moor, *New Algebraic Techniques for the Separation of DS-CDMA Signals*, Twenty-fourth SITB (Veldhoven), pp. 211-218, 2003.

[508] Cendrillon, R., O. Rousseaux, M. Moonen, E. van den Bogaert, and J. Verlinden, *Power Allocation and Optimal TX/RX Structures for MIMO Systems*, Twenty-fourth SITB (Veldhoven), pp. 219-226, 2003.

[509] Janssen, G.J.M, *A Power-Efficient Compound Modulation Scheme for Addressing Multiple Users in the Downlink*, Twenty-fourth SITB (Veldhoven), pp. 227-234, 2003.

[510] Levendovsky, J., L. Kovacs, A. Olah, D. Varga, and E.C. van der Meulen, *Novel Sampling Method for Increased Spectral Efficiency in Wireless Communication Systems*, Twenty-fourth SITB (Veldhoven), pp. 235-242, 2003.

WIC Symposium Estimation and Detection Papers

[511] Backer, E., *Over Minimale Vervorming in een Gelijksoortigheidsrelaties bij Classificeren Zonder Leraar*, First SITB (Zoetermeer), pp. 7-22, 1980.

[512] Boel, Rene K., *Optimale Schatting van een Diffusieproces dat de Intensiteit van een Waargenomen Poissonproces Bepaalt*, First SITB (Zoetermeer), pp. 33-37, 1980.

[513] Duin, R.P.W., *Needs and Possibilities of Using a Priori Knowledge in Pattern Recognition*, First SITB (Zoetermeer), pp. 47-51, 1980.

[514] Kwakernaak, H., *Estimation of Pulse Heights and Arrival Times* , First SITB (Zoetermeer), pp. 53, 1980.

[515] Schuppen, J.H. van, *Enkele Schattings- en Detectieproblemen*, First SITB (Zoetermeer), pp. 83-84, 1980.

[516] Veelenturf, L.P.J., *Adaptive Identification of Sequential Machines*, First SITB (Zoetermeer), pp. 99-103, 1980.

[517] Duin, R.P.W., *Small Sample Size Considerations in Discriminant Analysis*, Second SITB (Zoetermeer), pp. 49-52, 1981.

[518] Kemp, B., *Schatting en Detektie van Sprongsgewijze Veranderingen in het Electro-Encefalogram: Een Martingaal Aanpak*, Second SITB (Zoetermeer), pp. 71-76, 1981.

[519] Gröneveld, E.W., Kleima, D., *Enkele Opmerkingen over M-Voudige Detectie*, Third SITB (Zoetermeer), pp. 29-37, 1982.

[520] Schripsema, J., Veelenturf, L.P.J., *Petri-Netwerken Als Representatie voor Lerend Gedrag*, Third SITB (Zoetermeer), pp. 125-132, 1982.

[521] Kemp, B., Jaspers, P., *Optimal Detection of a Finite-State Markov Brain Process, Based on Vector EEG Observations*, Fifth SITB (Aalten), pp. 102-108, 1984.

[522] Kleima, D., *Invarianten, waaronder 'Shift-Invariant Functions'*, Fifth SITB (Aalten), pp. 109, 1984.

[523] Liefhebber, F., *Minimum Information and Parametric Modelling*, Sixth SITB (Mierlo), pp. 13-25, 1985.

[524] Moddemeyer, R., *Estimation of Entropy and Mutual Information of Continuous Distribution*, Sixth SITB (Mierlo), pp. 27-34, 1985.

[525] Bergmans, J., *Equalization, Detection and Channel Coding for Digital Transmission and Recoding Systems*, Sixth SITB (Mierlo), pp. 161-169, 1985.

[526] Backer, E., Eijlers, E.J., *CLUSAN1: A Knowledge Base for Cluster Analysis*, Seventh SITB (Noordwijkerhout), pp. 113-120, 1986.

[527] Moddemeijer, R., *An ARMA Model Identification Algorithm* , Seventh SITB (Noordwijkerhout), pp. 151-159, 1986.

[528] Kemp, B., *Optimal Detection of the Rapid-Eye-Movement Brain State* , Seventh SITB (Noordwijkerhout), pp. 175-182, 1986.

[529] Moddemeijer, R., *From Maximum Likelihood to an Entropy Estimate* , Eight SITB (Deventer), pp. 86-92, 1987.

[530] Backer, E., Lubbe, J.C.A. van der, Krijgsman, W., *On Modelling of Uncertainty and Inexactness in Expert Systems*, Ninth SITB (Mierlo), pp. 101-111, 1988.

[531] Moddemeijer, R., *An Information Theoretical Delay Estimator*, Ninth SITB (Mierlo), pp. 121-128, 1988.

[532] De Wilde, Ph., *A Marquardt Learning Algorithm for Neural Networks*, Tenth SITB (Houthalen), pp. 51-57, 1989.

[533] Coolen, A.C.C., Kuyk, F.W., *A Learning Mechanism for Invariant Pattern Recognition in Neural Networks*, Tenth SITB (Houthalen), pp. 59-65, 1989.

[534] Piret, Ph., *Some Properties of a Modified Hebbian Rule*, Tenth SITB (Houthalen), pp. 67-72, 1989.

[535] Verleysen, M., Martin, D., Jespers, P., *A Capacitive Neural Network for Associative Memory*, Tenth SITB (Houthalen), pp. 73-79, 1989.

[536] Vandenberghe, L., Vandewalle, J., *Dynamic Properties of Neural Networks*, Tenth SITB (Houthalen), pp. 81-88, 1989.

[537] Moddemeijer, R., Gröneveld, E.W., *Testing Composite Hypotheses*, Tenth SITB (Houthalen), pp. 133-138, 1989.

[538] Kleihorst, R.P., Hoeks, W.L.M., *Fuzzy OCR*, Eleventh SITB (Noordwijkerhout), pp. 81-88, 1990.

[539] Backer, E., *Approximate Reasoning in Exploratory Data Analysis* , Eleventh SITB (Noordwijkerhout), pp. 115, 1990.

[540] Vanroose, P., *Optimal Decision Trees and Test Algorithms*, Twelfth SITB (Veldhoven), pp. 25-31, 1991.

[541] Heideman G.H.L.M., *Realization of a Maximum Likelihood Classifier by a Learning Process*, Thirteenth SITB (Enschede), pp. 89, 1992.

[542] Lankhorst, M.M., Moddemeijer, R., *Automatic Word Categorization: an Information-Theoretic Approach*, Fourteenth SITB (Veldhoven), pp. 62-69, 1993.

[543] Bruin, F.F.G. de, *How a Feedforward Neural Network Classifies*, Fifteenth SITB (Louvain-La-Neuve), pp. 219-227, 1994.

[544] Levendovszky, J., Mommaerts, W., E.C. van der Meulen, *General Tolerance Analysis for Neural Networks* , Fifteenth SITB (Louvain-La-Neuve), pp. 228-234, 1994.

[545] Vanroose, P., Van Gool, L., Oosterlinck, A., *A Bottom-Up Approach to Pattern Classification*, Fifteenth SITB (Louvain-La-Neuve), pp. 235-242, 1994.

[546] Levendovszky, J., E.C. van der Meulen, Pozsgai, P., *Tail Estimation by Statistical Bounds and Neural Networks*, Seventeenth SITB (Enschede), pp. 137-145, 1996.

[547] Hupkens, E.P., *On the Quickest Detection of Changes in Random Fields*, Seventeenth SITB (Enschede), pp. 147-152, 1996.

[548] Berlinet, A., Györfi, L., E.C. van der Meulen, *The Asymptotic Normally of Centered Information-Divergence in Density Estimation*, Seventeenth SITB (Enschede), pp. 153-157, 1996.

[549] Cremer, F., Veelenturf, L.P.J., *Statistical Signal Detection and Kohonen's Neural Network*, Eighteenth SITB (Veldhoven), pp. 9-16, 1997.

[550] Hupkens, E.P., *Quickest Detection in Random Fields: a Bayesian Approach*, Eighteenth SITB (Veldhoven), pp. 17-24, 1997.

[551] Slump, C.H., *Applications of Information Theory in Optics*, Eighteenth SITB (Veldhoven), pp. 142-149, 1997.

[552] Moddemeijer, R., *Testing Composite Hypotheses Applied to AR Order Estimation; the Akaike-Criterion Revised*, Nineteenth SITB (Veldhoven), pp. 149-156, 1998.

[553] Levendovsky, J., Meszaros, A., E.C. van der Meulen, *Neuron Based Penalty Function Classifiers*, Nineteenth SITB (Veldhoven), pp. 157-164, 1998.

[554] Moddemeijer, R., *An Efficient Algorithm for Selecting Optimal Configurations of AR-Coefficients*, Twentieth SITB (Haasrode), pp. 189-196, 1999.

[555] Someren, E.P. van, Wessels, L.F.A., Reinders, M.J.T., *Information Extraction for Modeling Gene Expressions*, Twenty-first SITB (Wassenaar), pp. 215-222, 2000.

[556] Moddemeijer, R., *The Distribution of Entropy Estimators Based on Maximum Mean Log-Likelihood*, Twenty-first SITB (Wassenaar), pp. 231-238, 2000.

[557] Levendovszky, J., Kovács, L., Jeney, G., E.C. van der Meulen, *A New Blind Signal Processing Algorithm for Multi-User Detection*, Twenty-second SITB (Enschede), pp. 17-24, 2001.

[558] Vellekoop, M.H., *Suboptimal Approximations in Simultaneous Detection and Estimation Problems*, Twenty-second SITB (Enschede), pp. 25-32, 2001.

[559] Reinders, M.J.T., *Analyzing DNA Microarrays to Unravel Gene Function*, Twenty-fourth SITB (Veldhoven), pp. 35-36, 2003.

[560] Veldhuis, R., Bazen, A., and Boersma, M., *Biometric Verification: a Result and an Exotic Example* , Twenty-fourth SITB (Veldhoven), pp. 109-116, 2003.

[561] Goseling, J., Baggen, S., Akkermans, T., *Verification Using Partially Known Biometrics* , Twenty-fourth SITB (Veldhoven), pp. 117-124, 2003.

WIC Symposium Signal Processing and Restoration Papers

[562] Biemond, J., *Recursive Image Models and Model Quality*, First SITB (Zoetermeer), pp. 23–28, 1980.

[563] Spek, G.A. van der , *The Management of Radar Energy and Time in a Phased-Array Radar System*, First SITB (Zoetermeer), pp. 85–91, 1980.

[564] Biemond, J., *Beeldreconstructie als Lineair Filterprobleem (in Dutch)*, Second SITB (Zoetermeer), pp. 5–23, 1981.

[565] Blom, H.A.P., *Implementable Differential Equations for Non-Linear Filtering*, Second SITB (Zoetermeer), pp. 41–48, 1981.

[566] Gerbrands, J.J., *Beeldsegmentatie m.b.v. Probabilistische Relaxatie-Procedures*, Second SITB (Zoetermeer), pp. 53–61, 1981.

[567] Heideman, G.H.L.M., *Een Beeldbeschrijvingsmodel, Gebaseerd op de Structuur van de Primaire Visuele Cortex: Een Waarnemer Gerichte Codeermethode (in Dutch)*, Second SITB (Zoetermeer), pp. 63–69, 1981.

[568] Heideman, G.H.L.M., Veldhuis, R.N.J., *Een Signaaltheoretisch Model voor de Primaire Visuele Cortex; een Beeldbeschrijvingsmodel (in Dutch)*, Third SITB (Zoetermeer), pp. 39–45, 1982.

[569] Kruisbrink, J.C., *Een Parser voor Matrix-Array Grammatikas, Toegepast op Segmentering van Celklompjes (in Dutch)*, Third SITB (Zoetermeer), pp. 47–62, 1982.

[570] Rompelman, O., *Hartritme-Variabiliteit: Meting, Analyse en Interpretatie*, Third SITB (Zoetermeer), pp. 93–102, 1982.

[571] Slump, C.H., Ferwerda, H.A., Hoeders, B.J., *Informatie-Theoretische Aspecten Lage-Dosis Elektronenmicroscopie (in Dutch)*, Third SITB (Zoetermeer), pp. 133–140, 1982.

[572] Veldhuis, R.N.J., Heideman, G.H.L.M., *Een Bemonsteringsmodel voor Ruimtelijk Begrensde Twee-Dimensionale Signalenv(in Dutch)*, Third SITB (Zoetermeer), pp. 1141–155, 1982.

[573] Mars, N.J.I., *An Estimator for Delay Times in a Non-Linear Biological System*, Fourth SITB (Haasrode), pp. 67-73, 1983.

[574] Rompelman, O., *The Assessment of the Bandwidth of Trigger Related Waveforms*, Fourth SITB (Haasrode), pp. 75-81, 1983.

[575] Slump, C.H., Hoenders, B.J., Ferwerda, H.A., *The Determination of the Global Extremum of a Function of Several Variables*, Fourth SITB (Haasrode), pp. 83-91, 1983.

[576] Haas, H.P.A., *Digital Convexity and Straightness on the Hexagonal Grid*, Fourth SITB (Haasrode), pp. 103-114, 1983.

[577] Heideman, G.H.L.M., *An Implicit Sampling Model for Images*, Fourth SITB (Haasrode), pp. 115–120, 1983.

[578] Boekee, D.E., Helden, J. van, *Some Properties of Spectral Distortion Measures*, Fourth SITB (Haasrode), pp. 129–136, 1983.

[579] Wiersma, H., *Bounds on the Sampling Rate for Short-Time Narrowband Signals*, Fourth SITB (Haasrode), pp. 93–102, 1983.

[580] Koenderink, J.J., *Simultaneous Order in the Visual System*, Fifth SITB (Aalten), pp. 5–10, 1984.

[581] Biemond, J., Katsaggelos, A.K., *Iterative Restoration of Noisy Blurred Images*, Fifth SITB (Aalten), pp. 11–20, 1984.

[582] Gerbrands, J.J., Backer, E., *Split-And-Merge Segmentation of SLAR-Imagery: Consistency Problems*, Fifth SITB (Aalten), pp. 64–72, 1984.

[583] Slump, C.H., Ferwerda, H.A., Hoenders, B.J., *Some (Information Theoretical) Aspects of Low-Dose Electron Microscopy*, Fifth SITB (Aalten), pp. 152–161, 1984.

[584] Veldhuis, R.N.J., Jansen, A.J.E.M., Vries, L.B., *Adaptive Restoration of Unknown Samples in Time-Discrete Signals*, Fifth SITB (Aalten), pp. 178–186, 1984.

[585] Lohmann, A.W., *Digital Optical Computing*, Sixth SITB (Mierlo), pp. 9–12, 1985.

[586] Gerbrands, J.J., Backer, E., Hoeven, W.A.G. van der, *Edge Detection by Dynamic Programming*, Sixth SITB (Mierlo), pp. 35–42, 1985.

[587] Otterloo, P.J. van, Rohra, K., Veldhuis, R.N.J., *Motion Blur Due to Field Rate Conversion of Television Signals*, Sixth SITB (Mierlo), pp. 81–89, 1985.

[588] Woods, J.W., *Doubly Stochastic Gaussian Random Field Models for Image Estimation*, Seventh SITB (Noordwijkerhout), pp. 21–29, 1986.

[589] Spek, G.A. van der, *Inverse Synthetic Aperture Radar (ISAR)*, Seventh SITB (Noordwijkerhout), p. 61, 1986.

[590] Mieghem, E.F.P. van, Gerbrands, J.J., Backer, E., *Three-Dimensional Object Recognition by Using Stereo Vision*, Seventh SITB (Noordwijkerhout), pp. 89–93, 1986.

[591] Gerbrands, J.J., Backer, E., Cheng, X.S., *Multiresolutional Cluster/Relaxation in Segmentation*, Seventh SITB (Noordwijkerhout), pp. 95–102, 1986.

[592] Lagendijk, R.L., Biemond, J., *Regularized Iterative Image Restoration*, Seventh SITB (Noordwijkerhout), pp. 103–111, 1986.

[593] Rompelman, O., *Event Series Processing: a Signal Analysis Approach*, Seventh SITB (Noordwijkerhout), pp. 171–174, 1986.

[594] Backer, E., Gerbrands, J.J., *A Flexible and Intelligent System for Fast Measurements in Binary Images for In-Line Robotic Control*, Eight SITB (Deventer), pp. 6–20, 1987.

[595] Braadbaart, J., Kamminga, C., *On Several Definitions of Time Resolution Applied to Bio-Sonar*, Eight SITB (Deventer), pp. 53–60, 1987.

[596] Heideman, G.H.L.M., Hoeksema, F.W., Tattje, H.E.P., *Multi-Channel Sampling (Abstract)*, Eight SITB (Deventer), p. 68, 1987.

[597] Kamminga, C., *Structural Information Theory of Bio-Sonar, the Odontocete Echolocation Signal (Abstract)*, Eight SITB (Deventer), p. 77, 1987.

[598] Lagendijk, R.L., Biemond, J., Boekee, D.E., *Iterative Nonlinear Image Restoration*, Eight SITB (Deventer), pp. 78–85, 1987.

[599] Verbakel, J.M.M., *SILAGE, a Description and Simulation Language for Digital Signal Processing*, Ninth SITB (Mierlo), pp. 67–73, 1988.

[600] Gerbrands, J.J., Backer, E., Hoogeboom, P., Kleijweg, J., *Segmentation of SLAR Imagery Guided by a Priori Knowledge*, Ninth SITB (Mierlo), pp. 81–87, 1988.

[601] Lagendijk, R.L., Biemond, J., *Maximum Likelihood Identification and Restoration of Blurred*, Ninth SITB (Mierlo), pp. 97–103, 1988.

[602] Chen, J., Vandewalle, J.P.L., *A Comparison Between Adaptive IIR and Adaptive FIR Filter*, Ninth SITB (Mierlo), pp. 163–169, 1988.

[603] Callaerts, D., Vandewalle, J., *The Use of SVD-Based Techniques for Signal Separation*, Tenth SITB (Houthalen), pp. 109–115, 1989.

[604] Slump, C.H., *On the Prediction of the Optimal Exposure Timing from ECG Data in Digital Subtraction Angiography (DSA)*, Tenth SITB (Houthalen), pp. 125–131, 1989.

[605] Lagendijk, R.L., Biemond, J., *Advances in the Identification of Noisy Blurred Images*, Eleventh SITB (Noordwijkerhout), pp. 97–103, 1990.

[606] Vlugt, M.J. van der, *PC-Protocol: a System for Collecting and Correcting Ethological Data*, Eleventh SITB (Noordwijkerhout), pp. 116–117, 1990.

[607] Moddemeijer, R., *Sampling and Linear Algebra*, Eleventh SITB (Noordwijkerhout), pp. 118–125, 1990.

[608] Haan, H.G. de, Slump, C.H., *On the Reduction of Alias Distortion in Digital Signal Processing*, Eleventh SITB (Noordwijkerhout), pp. 126–132, 1990.

[609] Kamminga, C., *Some Results on Time Resolution in Delphinid Sonar*, Eleventh SITB (Noordwijkerhout), p. 140, 1990.

[610] Beck, W., *Frequency Estimation by Iterated Total Least Squares*, Eleventh SITB (Noordwijkerhout), pp. 141–147, 1990.

[611] Wurf, P. van der, *Statistical Analysis of Synchronous Random Pulse Trains by Means of Hybrid Correlation Functions*, Eleventh SITB (Noordwijkerhout), pp. 148–154, 1990.

[612] Kleihorst, R.P., Lagendijk, R.L., Biemond, J., *Non-Linear Filtering of Image Sequences Using Order Statistics*, Twelfth SITB (Veldhoven), pp. 49–55, 1991.

[613] Slump, C.H., *On the Reduction of Moiré Pattern Distortion in Digital Diagnostic X-Ray Imaging*, Twelfth SITB (Veldhoven), pp. 57–62, 1991.

[614] Laan, M.D. van der, *Towards Alternative Strategies for Signal-Sampling*, Thirteenth SITB (Enschede), pp. 81–88, 1992.

[615] Lubbers, A.P.G., Slump, C.H., Storm, C.J., *Digital Densitometric Determination of Relative Coronary Flow Distributions*, Thirteenth SITB (Enschede), pp. 181–188, 1992.

[616] Hoeksema, F.W., *Two Solutions to the Problem of Matrixing for Non-Ideal Camera Transmission Filters*, Thirteenth SITB (Enschede), pp. 189–196, 1992.

[617] Kleihorst, R.P., Haan, G. de, Lagendijk, R.L., Biemond, J., *Noise Filtering of Image Sequences with Double Compensation for Motion*, Thirteenth SITB (Enschede), pp. 197–204, 1992.

[618] Cohen Stuart, A.B., *Correlating Two Sonar Signals with Different Dominant Frequencies*, Fifteenth SITB (Louvain-La-Neuve), pp. 132–137, 1994.

[619] Cohen Stuart, A.B., Kamminga, C., *Modelling the Polycyclic Sonar Waveform of the Phoecena Phoecena Using Gabor's Elementary Signal*, Fifteenth SITB (Louvain-La-Neuve), pp. 160–167, 1994.

[620] Piret, P., *Caricatures by Means of Informational Divergence*, Fifteenth SITB (Louvain-La-Neuve), p. 218, 1994.

[621] Simon, B., *Smooth Non-Symmetrical Interpolation Functions for Quadtree Representation of Images*, Fifteenth SITB (Louvain-La-Neuve), pp. 252–258, 1994.

[622] Kamminga, C., Bruin, M.G. de, *A Time-Frequency Entropy Measure of Uncertainty Applied to Echolocation Signals*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 89–98, 1995.

[623] Vanroose, P., Van Gool, L., Oosterlinck, A., *Localization and Identification of Plane Objects in a Complex Scene*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 99–105, 1995.

[624] Hanjalic, A., Lagendijk, R.L., Biemond, J., *Achievements and Challenges in Visual Search of Video*, Seventeenth SITB (Enschede), pp. 159–165, 1996.

[625] Bruijn, F.J. de, Schrijver, M., Slump, C.H., *Compression of Cardiac X-Ray Images Based on Acquisition Noise*, Nineteenth SITB (Veldhoven), pp. 45–52, 1998.

[626] Vanroose, P., *Information Flow and Spatial Locality of Image Processing Operators*, Nineteenth SITB (Veldhoven), pp. 53–57, 1998.

[627] Slump, C.H., *On Information Theoretical Aspects of Speech Transmission*, Nineteenth SITB (Veldhoven), pp. 127–134, 1998.

[628] Hermus, K., Wambacq, P., Van Compernolle, D., *Improved Noise Robustness for Speech Recognition by Adaptive SVD-Based Filtering*, Twentieth SITB (Haasrode), pp. 117–124, 1999.

[629] Slump, C.H., Bont, T. de, Mertens, A.M., Verwey, K., *On the Objective Speech Quality of the TETRA System*, Twentieth SITB (Haasrode), pp. 125–132, 1999.

[630] Demuynck, K., Wambacq, P., *Linear Feature Transformations Based on MCE and MMI*, Twentieth SITB (Haasrode), pp. 141–148, 1999.

[631] Lerouge, E., Van Huffel, S., *Generalization Capacity of Neural Networks for the Classification of Ovarium Tumours*, Twentieth SITB (Haasrode), pp. 149–156, 1999.

[632] Mindru, F., Moons, T., Van Gool, L., *Generalized Moment Invariants for Viewpoint and Illumination Independent Color Pattern Recognition*, Twentieth SITB (Haasrode), pp. 157–164, 1999.

[633] Lagendijk, R.L., *The TU Delft Research Program 'Ubiquitous Communications'*, Twenty-first SITB (Wassenaar), pp. 33–43, 2000.

[634] Pasman, W., Jansen, F.W., *Latency Layered Rendering for Mobile Augmented Reality*, Twenty-first SITB (Wassenaar), pp. 45–54, 2000.

[635] Persa, S., Jonker, P., *Human-Computer Interaction Using Real-Time 3D Hand Tracking*, Twenty-first SITB (Wassenaar), pp. 71–75, 2000.

[636] Vos, K., Heusdens, R., *Rate-Distortion Optimal Exponential Modeling of Audio and Speech Signals*, Twenty-first SITB (Wassenaar), pp. 77–84, 2000.

[637] Farin, D., P.H.N. de With, *Towards Real-Time MPEG-4 Segmentation: a Fast Implementation of Region-Merging*, Twenty-first SITB (Wassenaar), pp. 173–180, 2000.

[638] Haan, G. de, *Video Processing for Multimedia Systems*, Twenty-first SITB (Wassenaar), pp. 189–198, 2000.

[639] Vanroose, P., *Information Measures for 3-D Scene Modeling*, Twenty-first SITB (Wassenaar), pp. 199–203, 2000.

[640] Lei, B.J., Hendriks, E.A., *Eigen Finder: an Extended Approach to Unify Low-Level Feature Extraction*, Twenty-first SITB (Wassenaar), pp. 205–213, 2000.

[641] Rares, A., Reinders, M.J.T., *Adaptive Mixtures for Object Tracking*, Twenty-first SITB (Wassenaar), pp. 223–230, 2000.

[642] Bruin M.G. de, Kamminga, C., *Minimizing the Uncertainty Product with Composite Signals*, Twenty-first SITB (Wassenaar), pp. 269–276, 2000.

[643] Burazerović, D., Gerrits, A., Taori, R., Ritzerfeld, J., *Time-Scale Modification for Speech Coding*, Twenty-second SITB (Enschede), pp. 1–8, 2001.

[644] Vanroose, P., *Part-Of-Speech Tagging from an Information-Theoretic Point of View*, Twenty-second SITB (Enschede), pp. 33–38, 2001.

[645] Ravyse, I., Sahli, H., Cornelis, J., *Head Detection, Tracking and Pose Estimation*, Twenty-second SITB (Enschede), pp. 39–44, 2001.

[646] Gonzalez, O., Katartzis, A., Sahli, H., Cornelis, J., *Pre-Processing of Polarimetric Ir Images for Land Mine Detection*, Twenty-second SITB (Enschede), 2001.

[647] Benschop, N.F., *Symmetric Logic Synthesis with Phase Assignment*, Twenty-second SITB (Enschede), pp. 115–122, 2001.

[648] Mindru, F., Moons, T., Van Gool, L., *Changes in Color Images*, Twenty-second SITB (Enschede), pp. 131–138, 2001.

[649] Slump, C.H., Schiphorst, R., Hoeksema, F.W., Nauta, B., Arkesteijn, V., Klumperink, E., *On AD Conversion for Telecommunications (Abstract)*, Twenty-second SITB (Enschede), p. 155, 2001.

[650] Jensen, J., Heusdens, R., Veenman, C., *Optimal Time- Differential Encoding of Sinusoidal Model Parameters*, Twenty-second SITB (Enschede), pp. 165–172, 2001.

[651] Hermus, K., Verhelst, W., Warnbacq, P., *A Scheme for Perceptual Speech and Audio Coding with Damped Sinusoids Based on Total Least Squares Algorithms*, Twenty-second SITB (Enschede), pp. 173–180, 2001.

[652] Hanjalic, A., XU, L.-Q., *An Approach to Affective Video Content Extraction*, Twenty-second SITB (Enschede), pp. 181–188, 2001.

[653] Brox, T., D. Farin, P.H.N. de With, *Multi-Stage Region Merging for Image Segmentation*, Twenty-second SITB (Enschede), pp. 189–196, 2001.

[654] Rares, A., Reinders, M.J.T., Biemond, J., *A Motion-Based Analysis of Fast-Changing Image Content*, Twenty-second SITB (Enschede), pp. 197–204, 2001.

[655] Srinivasan, R , *Fast Simulation and Applications in Communications and Signal Processing*, Twenty-second SITB (Enschede), pp. 129–130, 2001.

[656] Albu, F., Fagan, A., *Fast Affine Projection Algorithm Using the Successive Over-Relaxation Method*, Twenty-third SITB (Louvain-La-Neuve), pp. 147–154, 2002.

[657] De Bie, T. De Moor, B., *On Two New Classes of Alternatives to Canonical Correlation Analysis*, Twenty-third SITB (Louvain-La-Neuve), pp. 163–170, 2002.

[658] De Lathauwer, L., Fevotte, C., De Moor, B., Vandewalle, J., *Jacobi Algorithm for Joint Block Diagonalization in Blind Identification*, Twenty-third SITB (Louvain-La-Neuve), pp. 155-162, 2002.

[659] Zuo, F., With, P. de , *Automatic Human Face Detection for Home Surveillance Application*, Twenty-third SITB (Louvain-La-Neuve), pp. 207–214, 2002.

[660] De Lathauwer, L., De Moor, B., Vandewalle, J., *An Algorithm for Joint Diagonalization by a Congruence Transformation*, Twenty-third SITB (Louvain-La-Neuve), pp. 235–240, 2002.

[661] De Lathauwer, L., De Moor, B., Vandewalle, J., *An Algebraic Algorithm for Blind Identification with More Inputs Than Outputs*, Twenty-third SITB (Louvain-La-Neuve), pp. 241-246, 2002.

[662] Vanroose, P., Kalberer, G., Wambacq, P., Van Gool, L., *From Speech to 3D Face Animation*, Twenty-third SITB (Louvain-La-Neuve), pp. 255–260, 2002.

[663] Vanroose, P., *Blind Source Separation of Speech and Background Music for Improved Speech Recognition*, Twenty-fourth SITB (Veldhoven), pp. 103–108, 2003.

[664] Zuo F., and P.H.N. de With, *Experimenting with Face Detection and Recognition for Home Surveillance: a Status Report*, Twenty-fourth SITB (Veldhoven), pp. 133–140, 2003.

WIC Symposium Image and Video Compression Papers

[665] Huisman, W.C., *Three Image Compression Algorithms for CADISS*, Fourth SITB (Haasrode), pp. 80–93, 1983.

[666] Roefs, H.F.A., *CADISS: An Image (De)Compression System for Deep Space Application*, Fourth SITB (Haasrode), pp. 121–127, 1983.

[667] Boekee, D.E., Helden, J. van, *Vector Quantization of Images Using a Generalized Tree-Search Technique*, Fifth SITB (Aalten), pp. 21–27, 1984.

[668] Plompen, R.H.J.M., Booman, F., *Broncodering van Video Signalen op het Dr. Neher Laboratorium (in Dutch)*, Fifth SITB (Aalten), pp. 110-117, 1984.

[669] Renes, J.J., Pagter, P.J. de, *Image Data Compression with Spline Approximation and Segmentation*, Fifth SITB (Aalten), pp. 123–130, 1984.

[670] Rooyackers, J., *An Interframe Video Codec with Straight-Line Approximation*, Sixth SITB (Mierlo), pp. 91–97, 1985.

[671] Helden, J. van, Boekee, D.E., *A 384 Kbits/s Videoconferencing Coding Scheme Based Upon Vector Quantization*, Sixth SITB (Mierlo), pp. 99–107, 1985.

[672] Plompen, R.H.J.M., Boekee, D.E., *Motion Estimation in a Hybrid Coding Configuration*, Sixth SITB (Mierlo), pp. 109–115, 1985.

[673] Huisman, W.C., *Rate Distortion Characteristics of Two Adaptive Data Compression Algorithms*, Sixth SITB (Mierlo), pp. 213-222, 1985.

[674] Woods, J.W., H.M. Hang, *Predictive Vector Quantization of Images*, Seventh SITB (Noordwijkerhout), pp. 11–19, 1986.

[675] Simons, H.J., *Error Sensitivity of Compressed Image Data Satellite Communication Links*, Seventh SITB (Noordwijkerhout), pp. 63–72, 1986.

[676] Heideman, G.H.L.M., Tattje, H.E.P., Linden, E.A.R. van der, Rijks, D., *Self Similar Hierarchical Transforms: a Bridge Between Block-Transform Coding and Coding with a Model of the Human Visual System*, Seventh SITB (Noordwijkerhout), pp. 121–130, 1986.

[677] Plompen, R.H.J.M., Groenveld, J.G.P., Boekee, D.E., *Properties of Motion Estimation in the Transform Domain*, Seventh SITB (Noordwijkerhout), pp. 133–141, 1986.

[678] Westerink, P.H., Woods, J.W., Boekee, D.E., *Sub-Band Coding of Images Using Vector Quantization*, Seventh SITB (Noordwijkerhout), pp. 143–150, 1986.

[679] Biemond, J., Looijenga, L., Boekee, D.E., *A New Pel-Recursive Displacement Estimation Algorithm for Video-Conferencing Purposes*, Eight SITB (Deventer), pp. 37–44, 1987.

[680] Breeuwer, M., *Adaptive Transform Coding Using Cascaded Vector Quantisation*, Eight SITB (Deventer), pp. 61–67, 1987.

[681] Okkes, R.W., Huisman, W.C., *Rate Distortion Functions of SAR Imagery*, Eight SITB (Deventer), pp. 108–116, 1987.

[682] Plompen, R.H.J.M., Biemond, J., Heideman, G.H.L.M., *The Evaluation of a Hybrid DPCM/Transform Codec for Low Bitrates*, Eight SITB (Deventer), pp. 124–131, 1987.

[683] Stuifbergen, J.A.M., Heideman, G.H.L.M., *A Model for Moving Images Based on the Human Visual System*, Eight SITB (Deventer), pp. 157–163, 1987.

[684] Waal, R.G. van der, Breeuwer, M., Veldhuis, R.N.J., *Subband Coding of Music Signals Without Loss of Quality*, Eight SITB (Deventer), pp. 196–202, 1987.

[685] Westerink, P.H., Biemond, J., Boekee, D.E., *Sub-Band Coding of Images Using a Vector Equivalent of DPCM*, Eight SITB (Deventer), pp. 208–213, 1987.

[686] Stuifbergen, J.A.M., Heideman, G.H.L.M., *A Comparison of Two 3-D Models for Image Coding Based on Processing*, Ninth SITB (Mierlo), pp. 89–96, 1988.

[687] Westerink, P.H., Biemond, J., Boekee, D.E., *Image Subband Coding: a Quantization Error Analysis*, Ninth SITB (Mierlo), pp. 113–119, 1988.

[688] Macq, B., Delogne, P., *In Search of a Human Visual Quality Criterion for Image Data Compression*, Tenth SITB (Houthalen), pp. 93–100, 1989.

[689] Stuifbergen, J.A.M., *Estimation of the Velocity of Contours in a Moving Image by Minimization of the Change of the Velocity Field in a Hierarchical Spatio-Temporal Image Model*, Tenth SITB (Houthalen), pp. 101–107, 1989.

[690] Hogendoorn, R.A., Kordes, F.L.G., *METEODEC/METEOCRYPT: A Demonstration of Data Compression and Encryption for Operational Remote-Sensing*, Tenth SITB (Houthalen), pp. 169–175, 1989.

[691] Györfi, L., Linder, T., E.C. van der Meulen, *On the Asymptotic Optimality of Quantizers*, Eleventh SITB (Noordwijkerhout), pp. 29–35, 1990.

[692] Bosveld, F., Lagendijk, R.L., Biemond, J., *Hierarchical Coding Schemes for HDTV Using SBC and DCT*, Eleventh SITB (Noordwijkerhout), pp. 67–73, 1990.

[693] Driessen, J.N., Biemond, J., *Reduced Resolution Motion Field Estimation by 2-D Kalman Filtering*, Eleventh SITB (Noordwijkerhout), pp. 74-80, 1990.

[694] Horst, R. ter, *Motion Compensation for Multi Resolution Video Coding (Abstract)*, Eleventh SITB (Noordwijkerhout), pp. 89–89, 1990.

[695] Keesman, G., *Bit Assignment Method and its Application to Adaptive Dynamic Range Coding*, Eleventh SITB (Noordwijkerhout), pp. 90–96, 1990.

[696] Vandendorpe, L., Macq, B., *Hierarchical Subband and Entropy Coding*, Eleventh SITB (Noordwijkerhout), pp. 104–110, 1990.

[697] Schinkel, D., Horst, R. ter, *Coding of Multiple Video Sequences in an ATM Environment*, Eleventh SITB (Noordwijkerhout), pp. 111–113, 1990.

[698] Bosveld, F., Lagendijk, R.L., Biemond, J., *Hierarchical HDTV Coding for ATM Networks*, Twelfth SITB (Veldhoven), pp. 33–39, 1991.

[699] Klerk, P.P.C. de, Horst, R. ter, *Variable Length Coding in a Hybrid DCT Codec*, Twelfth SITB (Veldhoven), pp. 41–47, 1991.

[700] With, P.H.N. de , Nijssen, S.J.J., *An Intraframe Feedforward Coding System*, Twelfth SITB (Veldhoven), pp. 63–69, 1991.

[701] Vleuten, R.J. van der, Weber, J.H., *A New Constructive Design Method for Trellis Waveform Coders*, Thirteenth SITB (Enschede), pp. 15–22, 1992.

[702] Leduc, J.P., *Optimum Control of the Image Quality for Digital TV and HDTV Codecs*, Thirteenth SITB (Enschede), pp. 23–30 1992.

[703] Bosveld, F., Lagendijk, R.L., Biemond, J., *Compatible Video Transmission Using Spatio-Temporal Subband Coding Schemes*, Thirteenth SITB (Enschede), pp. 31–38, 1992.

[704] Barnard, H.J., Sankur, B., Lubbe, J.C.A. van der, *Statistics of DCT Coefficients in a Hybrid Video Codec*, Thirteenth SITB (Enschede), pp. 39–46, 1992.

[705] Stuifbergen, J.A.M., *A Scheme for Displacement Estimation in Image Coding*, Thirteenth SITB (Enschede), pp. 135-141, 1992.

[706] Belfor, R.A.F., Lagendijk, R.L., Biemond, J., *Sub-Nyquist Sampling of HDTV Using Motion Information*, Thirteenth SITB (Enschede), pp. 143–150, 1992.

[707] Queluz, M.P., Macq, B., *An Improved Block-Matching, Region-Oriented Motion Compensation Technique*, Thirteenth SITB (Enschede), pp. 151–158, 1992.

[708] Frimout, E.D., Driessen, J.N., Deprettere, E.F., *Parallel Architecture for a Pel-Recursive Motion Estimation Algorithm*, Thirteenth SITB (Enschede), pp. 159–166, 1992.

[709] Vleuten, R.J. van der, Weber, J.H., *A New Construction of Trellis-Coded Vector Quantizers*, Fourteenth SITB (Veldhoven), pp. 144–151, 1993.

[710] Meer, P.J. van der, Biemond, J., Lagendijk, R.L., *A Constant Quality MPEG Codec*, Fourteenth SITB (Veldhoven), pp. 152–159, 1993.

[711] Slump, C.H., *On Image Compression Related to Image Formation, Capture and Quality*, Fourteenth SITB (Veldhoven), pp. 160–167, 1993.

[712] Simon, B., Macq, B., Verleysen, M., *Pyramids for Image Compression with Neural Networks Interpolators*, Fourteenth SITB (Veldhoven), pp. 168–174, 1993.

[713] With, P.H.N. de, Nijssen, S.J.J., *A Buffer Regulation Concept for MC-DCT Systems Tuning to Constant Quantization*, Fourteenth SITB (Veldhoven), pp. 176–182, 1993.

[714] Hoeksema, F., Horst, R. ter, Heideman, G., Tatje, H., *Evaluation of a H.261 Video Codec in an ATM Network Using a Gaussian Model*, Fourteenth SITB (Veldhoven), pp. 184-191, 1993.

[715] Franich, R., Lagendijk, R.L., Biemond, J., *A Genetic Algorithm for Smooth Vector Field Estimation*, Fourteenth SITB (Veldhoven), pp. 192–197, 1993.

[716] Franich, R.E.H., Lagendijk, R.L., Biemond, J., *Fractal Picture Sequence Coding: Finding the Effective Search*, Fifteenth SITB (Louvain-La-Neuve), pp. 209–215, 1994.

[717] Hekstra, A.P., *On the Duality of Filter Design and Frequency Transform Based Video Coding (abstract)*, Fifteenth SITB (Louvain-La-Neuve), pp. 216–217, 1994.

[718] Shi, H.Q., Macq, B., *Vector Quantization with Orientation Discrimination*, Fifteenth SITB (Louvain-La-Neuve), pp. 243–251, 1994.

[719] Meer, P.J. van der, Biemond, J., Lagendijk, R.L., *Modeling of Variable Bit Rate Video Streams*, Fifteenth SITB (Louvain-La-Neuve), pp. 266–273, 1994.

[720] Westen, S.J.P., Lagendijk, R.L., Biemond, J., *Visibility Thresholds of Quantization Noise in Compressed Digital Image Sequences*, Fifteenth SITB (Louvain-La-Neuve), pp. 274–281, 1994.

[721] Franich, R.E.H., Lagendijk, R.L., Biemond, J., *A Path Through the Disparity Space Image*, Sixteenth SITB (Nieuwerkerk a/d IJssel), pp. 81-88, 1995.

[722] Bruijn, F.J. de, Heerde, C.J.E. van, Slump, C.H., *Medical Image Compression Boundaries Based on the Image Acquisition Process*, Seventeenth SITB (Enschede), pp. 1–7, 1996.

[723] Vleuten, R.J. van der, Oomen, A.W.J., *A Comparison of Subband Coding Gain and Transform Coding Gain*, Seventeenth SITB (Enschede), pp. 9–15, 1996.

[724] Westen, S.J.P., Lagendijk, R.L., Biemond, J., *The TCQF Algorithm: An Encoder Based Noise Shaping Technique for Image Coding*, Seventeenth SITB (Enschede), pp. 17–23, 1996.

[725] Hekstra, A.P., Herrera, J.M., *On Data Compression in Packet Switched Networks with Channel Error*, Seventeenth SITB (Enschede), pp. 57–64, 1996.

[726] Heideman, G.H.L.M., *Minimum Entropy-Representations and Decorrelation*, Seventeenth SITB (Enschede), 1996.

[727] Desmet, S., DeKnuydt, B., Van Eycken, L. Oosterlinck, A., *A Segmentation-Based Video Codec*, Seventeenth SITB (Enschede), pp. 73–79, 1996.

[728] Wuyts, T., Van Eycken, L., Oosterlinck, A., *Combined Motion Estimation and Segmentation for Object-Based Very Low Bitrate Coding*, Seventeenth SITB (Enschede), pp. 81–86, 1996.

[729] Vanroose, P., *Image Understanding Concepts for Improved Image Compression*, Seventeenth SITB (Enschede), pp. 87–93, 1996.

[730] Schaar-Mitrea, M. v.d., P.H.N. de With, *On the Application of Fast DCT Transforms Combined SW/HW Implementation*, Eighteenth SITB (Veldhoven), pp. 33–40, 1997.

[731] Beerends, J.G., Hekstra, A.P., *Objective Measurement of Video Quality*, Eighteenth SITB (Veldhoven), pp. 81–88, 1997.

[732] Westen, S.J.P., Lagendijk, R.L., Biemond, J., *An Eye Movement Compensated Spatio-Temporal Model for Predicting Distortion Visibility in Digital Image Sequences*, Eighteenth SITB (Veldhoven), pp. 89–96, 1997.

[733] Kleihorst, R.P., Cabrera, F., *VLSI Implememention of DCT-Domain Motion Estimation and Compensation*, Nineteenth SITB (Veldhoven), pp. 21–28, 1998.

[734] With, P.H.N. de, Schaar-Mitrea, M. v.d., *Low-Cost Embedded Compression for Memory Reduction in MPEG Decoding*, Nineteenth SITB (Veldhoven), pp. 29–36, 1998.

[735] Bakker, J.-D., Spaan, F.H.P., *Establishing a Trade-Off Between Error Robust Network Protocols and Error Robust Video Compression Algorithms*, Nineteenth SITB (Veldhoven), pp. 37–43, 1998.

[736] Biemond, J., *Video Compression Beyond 2000*, Nineteenth SITB (Veldhoven), pp. 58–66, 1998.

[737] Cardinal, J., *A Fast Full Search Equivalent for Mean-Shape-Gain Vector Quantizers*, Twentieth SITB (Haasrode), pp. 39–46, 1999.

[738] Desmet, S., DeKnuydt, B., Van Gool, L., Van Eycken, L., *Efficient Coding of Non-Static Texture in 3D Scenes*, Twentieth SITB (Haasrode), pp. 47–54, 1999.

[739] Schaar-Mitrea, M. v.d., P.H.N. de With, *High-Quality Embedded Compression for Digital TV*, Twentieth SITB (Haasrode), pp. 55–62, 1999.

[740] Schaaf, A. van der, Lagendijk, R.L., *Independence of Source and Channel Coding for Progressive Image and Video Data in Mobile Communications*, Twenty-first SITB (Wassenaar), pp. 55–62, 2000.

[741] Vleuten, R.J. van der, Kleihorst, R.P., Hentschel, C., *Low-Complexity Scalable Image Compression Using the DCT*, Twenty-first SITB (Wassenaar), pp. 85–92, 2000.

[742] Schelkens, P., Barbarien, J., Cornelis, J., *Volumetric Data Compression Based on Cube-Splitting*, Twenty-first SITB (Wassenaar), pp. 93–100, 2000.

[743] Kleihorst, R.P., Vleuten, R.J. van der, Apostolidou, M., *Swimming Pool Memories for Image Storage*, Twenty-first SITB (Wassenaar), pp. 165–172, 2000.

[744] Hunger, A., Werner, S., Akbarov, I., *Improvement and Implementation of Real-Time Video Compression Method for CSCL Software*, Twenty-first SITB (Wassenaar), pp. 181–188, 2000.

[745] Cardinal, J., *Complexity-Constrained Tree-Structured Vector Quantizers*, Twenty-first SITB (Wassenaar), pp. 239–246, 2000.

[746] Mietens, S., P.H.N. de With, Hentschel, C., *Implementation of a Dynamic Multi-Window TV System*, Twenty-second SITB (Enschede), pp. 139–146, 2001.

[747] Hoeksema, F., Vermeulen, H., Slump, K., *Component and Composite Coding of Residual Video Signals: Trans-Multiplexing Quantization?*, Twenty-second SITB (Enschede), pp. 205–212, 2001.

[748] Cardinal, J., *Entropy-Constrained Index Assignments for Multiple Description*, Twenty-third SITB (Louvain-La-Neuve), pp. 17–24, 2002.

[749] Mietens, S., With, P. de, Hentschel, C., *Frame Reordered Multi-Temporal Motion Estimation for Scalable MPEG*, Twenty-third SITB (Louvain-La-Neuve), pp. 115-121, 2002.

[750] Farin, D., Käsemann, M., P.H.N. de With, Effelsberg, W., *Rate-Distortion Optimal Adaptive Quantization and Coefficient Thresholding for MPEG Coding*, Twenty-third SITB (Louvain-La-Neuve), pp. 131-138, 2002.

[751] Iregui, M., Meessen, J., Chevalier, P., Macq, B., *Flexible Access to JPEG2000 Codestreams*, Twenty-third SITB (Louvain-La-Neuve), pp. 139-146, 2002.

[752] Vleuten, R.J. van der, *Improved Elastic Storage of Digital Still Images*, Twenty-fourth SITB (Veldhoven), pp. 71–78, 2003.

[753] Farin, D., P.H.N. de With, and W. Effelsberg , *Optimal Partitioning of Video Sequences for MPEG-4 Sprite Encoding*, Twenty-fourth SITB (Veldhoven), pp. 79–86, 2003.

[754] Mietens, S. P.H.N. de With, and C. Hentschel, *A SW-Based Complexity Scalable MPEG Encoder for Mobile Consumer Equipment*, Twenty-fourth SITB (Veldhoven), pp. 87–94, 2003.

[755] Cardinal, J., *Index Assignment Schemes for M-Description Coding*, Twenty-fourth SITB (Veldhoven), 2003.