

System characterization and reception techniques for two-dimensional optical storage

Citation for published version (APA):

Van Beneden, S. J. L. (2008). *System characterization and reception techniques for two-dimensional optical storage*. [Phd Thesis 1 (Research TU/e / Graduation TU/e), Electrical Engineering]. Technische Universiteit Eindhoven. <https://doi.org/10.6100/IR638352>

DOI:

[10.6100/IR638352](https://doi.org/10.6100/IR638352)

Document status and date:

Published: 01/01/2008

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

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

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

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

System Characterization and Reception Techniques for Two-Dimensional Optical Storage

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Technische
Universiteit Eindhoven, op gezag van de Rector Magnificus,
prof.dr.ir. C.J. van Duijn, voor een commissie aangewezen door
het College voor Promoties in het openbaar te verdedigen op
woensdag 19 november 2008 om 16.00 uur

door

Steven Jean-Marie Lucie Van Beneden

geboren te Deurne, België

Dit proefschrift is goedgekeurd door de promotor:

prof.dr.ir. J.W.M. Bergmans

Copromotor:

dr. W.M.J. Coene

CIP-DATA LIBRARY TECHNISCHE UNIVERSITEIT EINDHOVEN

Van Beneden, Steven

System Characterization and Reception Techniques for Two-Dimensional Optical Storage / by Steven Van Beneden. - Eindhoven : Technische Universiteit Eindhoven, 2008.

Proefschrift. - ISBN 978-90-386-1437-3

NUR 959

Subject headings: optical storage / signal processing / Viterbi detection / multidimensional systems / adaptive equalisers / modulation coding.

This work has been supported in part via a European IST-project called ‘TwoDOS’
(Project Nr. IST-2001-34168)

Committee:

Prof. dr. ir. J.W.M. Bergmans
Eindhoven University of Technology, The Netherlands

dr. W.M.J. Coene
Philips Research, Eindhoven, The Netherlands

Prof. dr. Dirk T.M. Slock
Institut Eurecom, Sophia Antopolis cedex, France

dr. Haibin Zhang
Shanghai Jiaotong University, China, and TNO Telecom, Delft, The Netherlands

Prof. Dr. Ir. P.G.M. Baltus
Eindhoven University of Technology, The Netherlands

System Characterization and Reception Techniques for Two-Dimensional Optical Storage

The digital revolution has spurred a tremendous growth in the distribution and storage of digital information worldwide. To support this growth, capacities and data rates of storage technologies have had to grow rapidly, and must continue to grow rapidly. Storage systems convert digital information into physical effects on a storage medium such as a magnetic or optical disk, and reconvert these effects into an electrical signal when reading out the stored information. A data receiver then operates on this analog read-out signal so as to recover the information. For this receiver to work properly, it must exploit detailed prior knowledge about the behavior of the storage channel, including the electrical-to-physical and physical-to-electrical conversion. During the development of a new storage system, this knowledge is obtained through construction of a channel model that describes the behavior and artifacts of the channel, and through channel characterization techniques that permit experimental validation and iterative refinement of the channel model.

In existing optical storage systems such as CD, DVD and Blu-Ray disc, information is stored on the disc in a spiral with a single data track, and with a sufficiently large spacing between adjacent rotations of the spiral to avoid intertrack interference. This is a one-dimensional (1-D) storage format in that data symbols are packed tightly (and interfere) only in the along-track direction. In order to increase storage density and data rates, data can instead be stored in a so-called broad spiral that encompasses multiple data tracks, with no intertrack spacing. This format is two-dimensional (2-D) in that data symbols are now packed tightly both in the along-track and across-track directions. Because of this tight packing, storage densities can increase significantly. Furthermore, by using a set of parallel laser beams, all tracks in the broad spiral can be read out simultaneously, thereby dramatically increasing data rates. The key disadvantage of 2-D *vis a vis* 1-D optical storage stems from the much higher storage density, which induces strong 2-D intersymbol interference (ISI), and simultaneously increases the sensitivity of the receiver to interferences and artifacts. For this reason, accurate channel characterization becomes essential, and the receiver must be accurately tailored to the key channel artifacts. This thesis addresses these two challenges. As a basis of reference it uses the so-called TwoDOS system, the first fully operational 2-D optical storage system developed to date. The developed techniques are, however, generically applicable to 2-D optical and magnetic storage systems.

The thesis sets out with a comprehensive study of the key characteristics of the TwoDOS channel, including linear and nonlinear ISI, various types of noise, and temporal variations. The salient characteristics are described in terms of simple models, which are validated experimentally. Special attention is devoted to the characteriza-

tion of noise sources. As storage capacities increase, media noise becomes increasingly prevalent in both optical and magnetic storage systems. For this reason a noise characterization scheme that efficiently decomposes combinations of media and additive noise and subsequently estimates the key noise parameters, is highly desirable. In this thesis, such an adaptive noise decomposition scheme is proposed and analyzed. Simulation results show that high estimation accuracies are obtained at low computational complexity.

The thesis proceeds to develop reception techniques that exploit some of the key characteristics of the TwoDOS system. It focuses on two critical receiver building blocks, namely the bit detector, which reconstructs the recorded bits, and the adaptation loops, which continuously keep track of the system parameters.

In 2-D systems, the bit detector tends to be highly complicated because it must be two-dimensional in nature. To simplify the detector, an adaptive equalizer commonly precedes the bit detector in order to limit the span of the 2-D ISI. Since detector complexity tends to grow exponentially with this ISI span, the use of an adaptive equalizer permits dramatic simplifications of the detector. At high storage densities, however, our characterization results suggest that significant ISI is left outside the span that the detector can handle. This *residual* ISI causes a significant performance deterioration. To overcome this deterioration, an innovative 2-D ISI cancellation scheme is developed. At the heart of this scheme is a 2-D filter that ideally produces a replica of the residual ISI. Subtraction of the filter output from the detector input produces a new input that ideally contains no residual ISI. The 2-D filter is excited by tentative bit decisions. These are readily available in many 2-D systems as the detector typically uses several iterations, and the decisions produced in the first iterations can be earmarked as tentative. In the thesis it is shown analytically, through simulations and experimentally that the application of a 2-D ISI cancellation scheme can yield substantial performance improvements at very modest hardware cost.

Even with an adaptive equalizer, 2-D bit detectors tend to be highly complex. A typical strategy to further lower complexity is to split the detection problem up into a succession of smaller tasks, each typically covering a limited number of tracks. This subdivision invariably leads to a larger detection latency, as these smaller tasks are carried out consecutively, with the result of one task serving as input for the next. Unfortunately the tracking capabilities of the adaptation loops within the receiver depend heavily on this latency, and tend to become inadequate to track rapid variations of e.g. DC, amplitude and timing parameters. In this thesis a scheme is proposed that overcomes this problem by exploiting the fact that the bulk of these variations is common across all the tracks. Accordingly, control information for the common part of the variations can be extracted from the tracks for which detection latency is smallest. Simulations and experimental results confirm the effectiveness of the developed scheme.

Contents

1	Introduction	1
1.1	Digital Data Storage: History and Trends	1
1.2	Optical Storage	4
1.2.1	Single-Spiral Optical Discs	5
1.2.2	Two-Dimensional Optical Storage	7
1.2.3	New Technologies	10
1.3	Magnetic Storage	12
1.3.1	Longitudinal and Perpendicular Storage	13
1.3.2	Two-Dimensional Magnetic Storage	14
1.4	Basic Signal Processing for 1D Storage System	15
1.4.1	Modulation Codes	17
1.4.2	Detection Principles	18
1.4.3	Viterbi Detection	21
1.4.4	Adaptation	23
1.5	Motivation and Content of this Thesis	29
1.6	List of publications and Patents	30
1.6.1	Papers	30
1.6.2	Patents	33
2	Two-Dimensional Optical Data Storage	35
2.1	Two-Dimensional Disc Format	35
2.1.1	2-D Lattice Characteristics	36
2.1.2	Manufacturing of TwoDOS discs	38
2.1.3	Test Format	40
2.2	Read Out of a TwoDOS disc	41
2.3	Optical Storage Channel Models	43
2.3.1	Intersymbol Interference Model	43
2.3.2	Noise Model	47
2.4	Signal Processing Principles	49
2.4.1	Modulation Code	49
2.4.2	Receiver principles	51
2.5	Data Receiver	52
2.5.1	Basic operation	52

2.5.2	Equalization	55
2.5.3	Timing Recovery	57
2.5.4	DC and gain control	58
2.5.5	Interaction between adaptation loops	59
2.6	Bit Detection Techniques	59
2.7	Conclusions	62
3	Characterization of Experimental TwoDOS PRML System	65
3.1	Introduction	65
3.2	Intersymbol interference Characterization	67
3.2.1	Linear ISI Model	69
3.2.2	Bilinear ISI Model	73
3.2.3	Look-Up Table Model	74
3.2.4	Residual ISI Model	75
3.2.5	Experimental Results	76
3.3	Noise Characterization	79
3.3.1	Correlated Gaussian Noise Model	80
3.3.2	Data-Dependent Auto-Regressive Noise Model	80
3.3.3	Experimental Results	84
3.3.4	Media Noise	86
3.4	Time Variations	89
3.4.1	Adaptive Data-Aided Parameter Estimation for the Channel Characterization	89
3.4.2	Time-varying Channel Artifacts	93
3.5	Conclusions	97
4	Adaptive Decomposition of Noise Sources in Digital Storage Systems with Media Noise.	101
4.1	Introduction	101
4.2	Media Noise in Optical Storage	105
4.2.1	Data-Dependent Media Noise Characterization	106
4.2.2	Adaptive Estimation Scheme	107
4.2.3	Simulation results	110
4.3	Magnetic Storage	116
4.3.1	Media Noise Model	117
4.3.2	Adaptive Estimation Scheme	118
4.3.3	Simulation Results	119
4.4	Test Pattern Design	121
4.5	Conclusions	122

5	Cancellation of Linear Intersymbol Interference for Two-Dimensional Storage Systems	127
5.1	Introduction	127
5.2	Overview of ISI Cancellation	129
5.3	Linear ISI Cancellation in 2-D Systems	131
5.3.1	Probability of Error of a Viterbi Detector in the presence of RISI	132
5.3.2	Probability of Error of the ISI cancellation scheme	134
5.3.3	Error Propagation in the Receiver using Tentative Decisions for ISI Cancellation	136
5.3.4	Examples	137
5.4	Experimental Results for TwoDOS	141
5.4.1	SWVD with Two Detection Iterations	144
5.4.2	SWVD with Three Detection Iterations	145
5.4.3	Cross-Talk Cancellation	146
5.5	Conclusions	150
6	Minimum-Latency Tracking of Rapid Variations in Two-Dimensional Storage Systems.	151
6.1	Introduction	151
6.2	Receiver Model	154
6.3	Effect of latency on loop behavior	155
6.3.1	Loop Behavior	155
6.3.2	Gradient Noise	157
6.4	Minimum-Latency Adaptation	158
6.5	First-Order Minimum-Latency Adaptation Loops	159
6.5.1	Basic Behavior	161
6.5.2	Gradient Noise	162
6.5.3	Behavior of the Inner Loop with Latency	163
6.5.4	Simulation Results	164
6.6	Minimum-Latency Timing Recovery	167
6.6.1	Basic Behavior	169
6.6.2	Gradient noise	171
6.6.3	Behavior of Inner Loop with Latency	172
6.7	Experimental Results for the TwoDOS system	173
6.8	Conclusion	176
7	Conclusion and Recommendations for Future Work	177
7.1	Conclusions	177
7.2	Recommendations for Future Work	180

Bibliography	182
Acknowledgment	203
Curriculum Vitae	205

List of Abbreviations

ADC:	Analog to Digital Converter
AGC:	Automatic Gain Control
AR:	Auto-Regressive
ASIC:	Application Specific Integrated Circuit
AWGN:	Additive White Gaussian Noise
BD:	Blu-Ray Disc
BER:	Bit-Error Rate
CD:	Compact Disc
DA:	Data-Aided
DD:	Decision-Directed
DFE:	Decision-Feedback Equalizer
DL:	Dual Layer
DVD:	Digital Versatile Disc
EBR:	Electron-Beam Recording
ECC:	Error Correction Coding
EFM:	Eight-to-Fourteen Modulation
EPRML:	Extended Partial Response Maximum Likelihood
FDTS:	Fixed-Depth Tree-Search
FFT:	Fast Fourier Transform
FIR:	Finite Impulse Response
FPGA:	Field Programmable Gate Array
GB:	Giga Byte
HF:	High Frequency
HDD:	Hard Disk Drive
IC:	Integrated Circuit
IFFT:	Inverse FFT
ISI:	Intersymbol Interference
LBR:	Laser Beam Recording
LE:	Linear Equalizer
LF:	Loop Filter
LIM:	Liquid Immersion Mastering
LMS:	Least Mean Square
LS:	Least Square
LUT:	Look-Up Table
MAP:	Maximum A-Posteriori
MIMO:	Multiple-Input Multiple-Output
MB:	Mega Byte
ML:	Maximum Likelihood

MLSD:	Maximum Likelihood Sequence Detection
MMSE:	Minimum Mean Square Error
MNP:	Media Noise Percentage
MR:	Magnetic Resonance
MSE:	Mean Square Error
MTF:	Modulation Transfer Function
MTR:	Maximum Transition Length
MVA:	Multi-track Viterbi Algorithm
NA:	Numerical Aperture
NCO:	Numerically Controlled Oscillator
NEA:	Normalized Estimation Accuracy
NF:	Near-Field
NLC:	Non-linearity Compensation
NPML:	Noise-Predictive Maximum Likelihood
OSR:	Oversampling Ratio
PDIC:	Photo Diode Integrated Circuit
PID:	Proportional, Integrating and Differentiating
PLL:	Phase-Locked Loop
PRML:	Partial Response Maximum Likelihood
PSD:	Power Spectral Density
RISI:	Residual Intersymbol Interference
RLL:	Run-Length Limited
SEM:	Scanning Electron Microscope
SIL:	Soli Immersion Lens
SNR:	Signal-to-Noise Ratio
SOVA:	Soft-Output Viterbi Algorithm
SP:	Signal Processing
SRC:	Sample Rate Convertor
SWVD:	Stripe-Wise Viterbi Detector
TB:	Tera Byte
TED:	Timing Error Detector
TwoDOS:	Two-Dimensional Optical Storage
UV:	Ultra-Violet
VCO:	Voltage Controlled Oscillator
VA:	Viterbi Algorithm
VD:	Viterbi Detector
VGA:	Variable Gain Amplifier
XTC:	Cross-Talk Cancellation
ZF:	Zero-Forcing

List of Symbols

Notational Conventions

a	scalar value.
\mathbf{a}	vector.
\mathbf{A}	matrix.
a_k	value at time instant k .
a_k^l	value at time instant k for track l .
\mathbf{a}_k	vector at time instant k .
\tilde{a}	estimate of a produced by an adaptation loop.
\hat{a}	binary estimate as produced by a bit-detector.
\mathbf{A}^T	transpose of matrix \mathbf{A} .

Often Used Symbols

a_H	lattice constant, i.e. grid spacing.
a_k	RLL constrained bit sequence.
\hat{a}_k	detected bit sequence.
b_H	diameter of pits in the TwoDOS discs.
c_k	gain parameter sequence as produced by the AGC.
d^2	Euclidean detection distance.
$d(\epsilon)$	Euclidian weight of a particular error event (bit-error sequence) ϵ .
D	number of symbols delay in an adaptation loop.
d_k	desired detector input sequence according to g_k .
e_k	error sequence.
f_k	RISI impulse response.
f_c	cut-off frequency of the optical channel.
f_s	sampling frequency.
G	target impulse response length.
$G_\phi(z)$	transfer function of an adaptation loop in the parameter domain.
g_k	target impulse response.
h_k	discrete-time channel impulse response.
$i^a(k, l)$	RISI sequence at the detector input due to the RISI filter \mathbf{f}_k .
I	half length of the data-dependence window used in ISI characterization.
k	discrete-time index in units T , i.e. synchronous to the baud rate.
K	total number of samples in an input sequence.
K_t	total gain of a first-order adaptation loop.
K_{ti}	gain of integrating branch of a second-order adaptation loop.

L	number of parallel tracks in a 2-D storage system.
$L(z)$	transfer function of the loop filter in an adaptation loop.
M	memory length of the channel.
N	memory length of the noise in the noise characterization.
n	discrete-time index in units T_s , i.e. asynchronous to the baud rate.
n_k	equivalent noise sequence at the detector input.
p_k	discrete-time impulse response at the detector input.
q_k	impulse response of the sampled derivative of a target response g_k .
R	rate of a modulation code.
$R_e(i, j)$	2-D autocorrelation function of the error signal \mathbf{e}_k .
$r(t)$	continuous-time read-out (replay) signal at the channel output.
r_k	discrete-time read-out (replay) signal sampled at the rate $1/T_s$.
S	set of admissible data patterns as defined by the modulation code.
$S_e(\Omega_x, \Omega_y)$	Power Spectral Density of the error signal \mathbf{e}_k
s_k	noiseless channel output signal.
S	total number of possible states.
T	channel bit period.
T_s	sampling period.
u_k	media noise sequence: pit-size noise for optical storage and position-jitter for magnetic storage.
V	delay of a VD.
v_k	additive noise sequence at the channel output.
W	half of the equalizer length.
w_k	equalizer impulse response.
$w_H(\epsilon)$	the number of symbol errors in the error event ϵ .
y_k	detector input sequence.
α	leakage factor in a ZF adaptation loop.
$\beta(s_{k-1}, s_k)$	branch metric for going from state s_{k-1} to state s_k .
Δ_k	sequence of mismatch values between estimated parameter values ϕ_k and actual parameter values θ_k .
ϵ	bit-error sequence.
ζ	damping factor of a second-order adaptation loop.
λ	laser wavelength.
κ	number of postcursive ISI components.
$\lambda\{s_{k-1}, s_k\}$	path metric for going from state s_{k-1} to state s_k .
$\hat{\lambda}^{s_k}$	smallest path metric leading to state s_k .
μ	general adaptation constant.
ν	input-referred noise in an adaptation loop.
ϕ_{Airy}	airy distance: full-width at half distance of laser spot.
ϕ_k	sequence of estimated parameter values in an adaptation loop.

$\mathbf{\rho}$	position vector on the optical disc.
σ^2	variance.
τ	time constant of an adaptation loop.
θ_k	sequence of actual parameter values.
\mathbf{v}_k	impulse response of the filter used in a XTC scheme.
$\omega_n T$	natural frequency of a second-order adaptation loop.
Ω_c	normalized cut-off frequency of an adaptation loop.
χ_k	timing error sequence generated by a TED.

Chapter 1

Introduction

“Faster and larger” is the comment you often hear, when people are talking about the evolution of data rate and capacity of storage systems. This statement indeed indicates one of the commercial requirements for new storage devices. The need for storing tremendous amounts of digital data has prompted the development of various storage systems. Recently two-dimensional (2-D) storage systems have been proposed as a candidate for next generation storage systems [1]. These 2-D systems are based on reading and processing several data streams in parallel, in optical systems by using parallel laser beams and in magnetic systems by using an array of read heads. The exploitation of parallelism results in an increased data rate and enables an increased capacity, which are effectively achieved by applying innovative 2-D channel coding and advanced 2-D signal processing techniques.

This work aims at the development of advanced signal processing algorithms that overcome some of the main bottlenecks in 2-D systems. Bottlenecks should be understood as issues that seriously hamper the performance of the 2-D system. An accurate characterization of the 2-D system is essential for the identification of these bottlenecks. Subsequently advanced signal processing algorithms can be designed to resolve these bottlenecks.

In this chapter the background, the motivation and the organization of the thesis are presented. In Section 1.1 an overview of the history of storage systems and an explanation for its great market success are given. Section 1.2 discusses the principles of optical storage systems and also the extension of the conventional one-dimensional (1-D) system to its 2-D equivalent. In Section 1.3 a similar discussion is given for magnetic storage systems. An overview of the signal processing algorithms involved in 1-D storage systems is presented in Section 1.4. Finally, in Section 1.5 an outline of the thesis is given.

1.1 Digital Data Storage: History and Trends

During the emergence of the digital era, the fast growth of information technology demanded the transmission and the storage of digital data in huge volumes and at high

speed. As a result the development of improved communication and storage systems and this growth came together. Whereas communication systems transport information from one location to another, storage systems do it from one time to another. The common goal is to eventually retrieve the stored or transmitted information as reliably as possible. The information is represented in the form of digital binary data. Although communication systems and storage systems have a lot of similarities, in this work the focus is on storage systems.

In Fig. 1.1 a schematic overview of a general digital storage system is shown. In general the functionality of the systems can be described as storing (accomplished by the write channel) information on a specific media at one point in time and retrieving (accomplished by the read channel) it from the media at another point in time. Hence the storage system can be considered to consist of three distinct parts: the write channel, the physical channel and the read channel. The write channel generates an analog write signal based on the binary input data. The physical channel consists of the combination of media and physical components to read/write information on the media. Its input is the analog write signal and its output is an analog read signal. In magnetic storage the read/write head is the main physical components, while in optical storage the laser, the optics (lenses) and the Photo Detector Integrated Circuit (PDIC) are the physical components of main interest. The read channel recovers the original data by processing the read signal in accordance with certain algorithms. Typical signal processing techniques that are utilized in the read channel are equalization, bit-detection and timing recovery.

Several types of storage systems can be identified depending on the type of media used to store the information. The three main types are: magnetic storage, optical storage and solid-state storage. Also a combination of magnetic and optical storage has been proposed: the magneto-optical system [2]. In this work only systems that are based on rotating discs will be discussed, both of the magnetic and optical type. The evolution of densities and data rates of magnetic and optical storage systems is shown in Fig. 1.2. In the left plot the densities of the different standardized optical storage systems are shown: Compact Disc (CD), Digital Versatile Disc (DVD), DVD Dual-Layer (DVD-DL), and finally Blu-Ray disc. Furthermore the densities for the experimental Two-Dimensional Optical Storage (TwoDOS) system and for the magnetic Hard Disk Drive (HDD) are shown. In the right plot (evolution of data rates) the

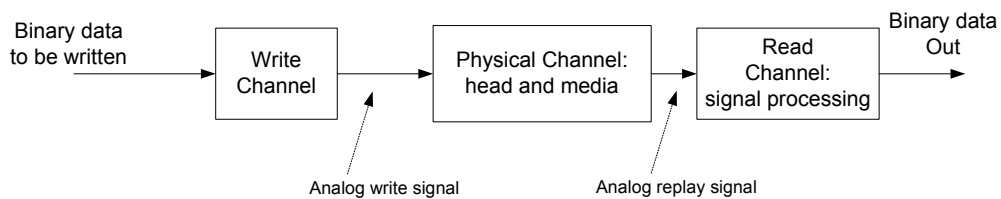


Figure 1.1: Schematic overview of a general storage system.

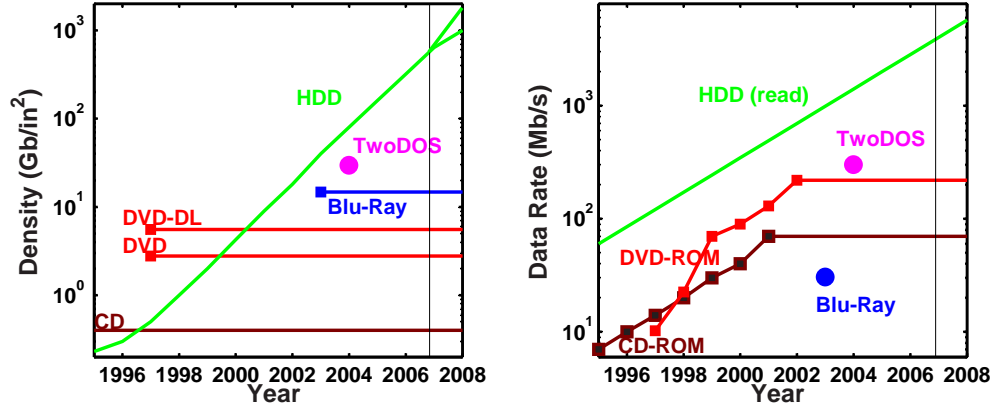


Figure 1.2: Evolution of areal density and data rate in storage systems.

Read-Only-Memory (ROM) versions of the CD and DVD are compared to the read versions of Blu-Ray, TwoDOS and HDD.

Although the depicted evolution has been mainly due to technological improvements made in the physical channel (improved design of media and/or physical components such as heads, lasers, etc.), sophisticated coding and signal processing algorithms have also played an important role [3]. For a given head/media combination the use of more advanced signal processing techniques allows the bits to be more closely packed together on the media, resulting in an increased areal density. A classical example of this is the increase in capacity due to the replacement of peak detection techniques by partial response techniques in hard disk drives in the early 1990s [4–6].

The most striking performance lag of conventional optical storage technology compared to hard disk drives is its lower data rate, and, to a lesser extent, its lower storage density (see Fig. 1.2) [7]. On the other hand, a major advantage of optical storage over magnetic storage is the removability of the optical media (discs) [8]. At the time of the introduction of the CD, its capacity greatly exceeded that of HDDs. Unlike in optical storage, the media used in magnetic storage cannot be removed from the player and is not standardized. As a result its capacity and data rate can grow continuously due to incremental innovations [9]. Another reason for the comparatively slow evolution of optical storage densities, besides the inertia that arises with standardization processes, is the slow pace at which the wavelength of laser diodes has improved. Fig. 1.2 shows the current status of HDD, with a data rate of 1 Gb/s, and a density of 100 Gb/in², whereas Blu-Ray achieves a data rate of 35 Mb/s and a density of 14.7 Gb/in². Fundamental physical limitations exist in magnetic storage, which ultimately restrict achievable densities in the order of 1 Tb/in² and data rates in the order of 10 Gb/s [10, 11]. In the following two sections, a more detailed discussion will be given about the evolution of optical and magnetic storage technologies.

1.2 Optical Storage

Although optical storage dates back to the early 1970s [12], the first commercial success was achieved with the introduction of the CD in 1983 [13]. At that time, the CD provided an alternative for magnetic storage systems with the following advantages: high capacity (680 MB on a disc with a diameter of 12 cm), removability of the disc without risk of damaging the data and finally its reliability (there is no risk of erasure of bits and the addition of a transparent protective layer avoids head crashes like they occur in magnetic disc systems). CD uses prerecorded, replicated discs (so-called CD-ROM, read-only memory) to store digital audio at an information density of about $1\mu\text{m}^2/\text{bit}$. This information density is directly related to the size of the optical spot which is diffraction limited. This size depends only on the wavelength of the laser and the numerical aperture (NA) of the objective lens where NA is defined as the sine of the opening angle of the light cone that is focused on the storage medium. For CD, an infrared laser is used with wavelength $\lambda = 780\text{ nm}$ and furthermore $\text{NA} = 0.45$. The thickness of the transparent disc (that serves as the protecting cover-layer for the data) is 1.2 mm. Despite its gigantic success, the CD suffered from one major disadvantage with respect to magnetic storage: it does not permit information to be written and/or erased. This disadvantage was circumvented later on with the introduction of the CD-RW (rewritable), which is based on phase-change techniques [14].

In Fig. 1.3 the evolution of optical storage systems is shown together with the corresponding capacities and physical parameters. Furthermore also the disc formats are shown on which the distance between two neighboring tracks is indicated. In 1996 the successor of the CD standard, known as digital versatile disc (DVD), was introduced. DVD has a storage capacity of 4.7 GB. This enlarged capacity was achieved by exploiting improved physical components: a red laser with wavelength $\lambda = 650\text{ nm}$, an objective lens with $\text{NA} = 0.60$ and a substrate thickness of 0.6 mm. The main field of application for DVD is digital movie, whereas CD focused on digital audio. In conjunction with the breakthrough of high-definition television, the need for even higher storage capacities emerged in the development of new optical disc systems.

Currently, two standards are competing to be the third generation optical storage system: the Blu-Ray disc (BD) [8] and the high-definition digital versatile disc (HD-DVD) [15]. In Fig. 1.3 only BD is depicted but properties of HD-DVD are similar. Both standards use blue laser light with a wavelength of 405 nm. The BD format is based on a NA of 0.85 and a cover layer of 0.1 mm thickness. It achieves a capacity of 23.3, 25 or 27 GB on a single storage layer. The HD-DVD format is based on a NA of 0.65 and a cover layer of 0.6 mm thickness. It achieves a capacity of 15 GB for ROM and 20 GB for RW. Although the capacity of HD-DVD is lower than that of BD, HD-DVD is less sensitive to dust and scratches compared with BD, due

to the use of a thicker cover layer. Furthermore, the 0.6 mm cover-layer fabrication process of HD-DVD is similar to the conventional DVD technology, which results in a lower overall fabrication cost. At this point in time, it is not clear which of the two standards will be the winner of the competition.

Beyond these standardized products, much new research is in progress for the development of optical systems with capacities and data rates beyond those of BD/HD-DVD [16]. In Fig. 1.3 one of these systems is depicted, namely the TwoDOS system. It utilizes the same physics as the BD system but, based on innovative signal processing techniques, it achieves a capacity of 50 GB at a data rate which is 10 times the data rate of BD. The basic operation of this experimental system will be explained in more detail in Section 1.2.2, whereas the basic operation of the standardized products will be discussed in Section 1.2.1.

Summarizing, optical storage is the preferred technology when high-density storage on removable storage media is required, for a number of different reasons: low cost, exchangeability between all drives from different brands (obtained through standardization), and, last but not least, robustness. This technology is ideal for content distribution because of its low-cost replication, and plays a key role in the archival of data. In the near future, optical storage devices will continue to form an integral part of the daily life of both consumers and specialist users.

1.2.1 Single-Spiral Optical Discs

In general, an optical storage system operates based on different intensities of reflected light for the ones and the zeros that are to be recorded. On a read-only disc, microscopically small lands and pits are arranged in a spiral path. The lands and pits represent the digital binary data. The pits on the disc scatter the light and result in a

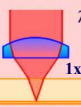
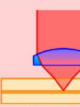


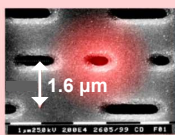
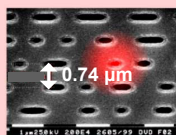
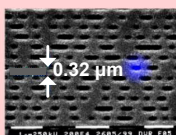
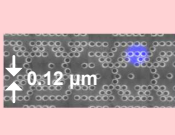
<i>CD (650 MB)</i>	<i>DVD (4.7 GB)</i>	<i>BD (25 GB)</i>	<i>TwoDOS (50 GB)</i>
 $\lambda = 780 \text{ nm}$ $NA = 0.45$ 0.4 Gb/in^2 $1x: 1.2 \text{ Mb/s}$ 1.2 mm substrate	 $\lambda = 650 \text{ nm}$ $NA = 0.6$ 2.8 Gb/in^2 $1x: 11 \text{ Mb/s}$ 0.6 mm substrate	 $\lambda = 405 \text{ nm}$ $NA = 0.85$ 14.7 Gb/in^2 $1x: 35 \text{ Mb/s}$ 0.1 mm cover layer	 $\lambda = 405 \text{ nm}$ $NA = 0.85$ 29.4 Gb/in^2 $1x: 300 \text{ Mb/s}$ 0.1 mm cover layer
 1.6 μm	 0.74 μm	 0.32 μm	 0.12 μm

Figure 1.3: Generations of Optical Storage. The parameters shown are: disc capacity, physical parameters (wavelength, λ and Numerical Aperture, NA), areal density and substrate thickness.

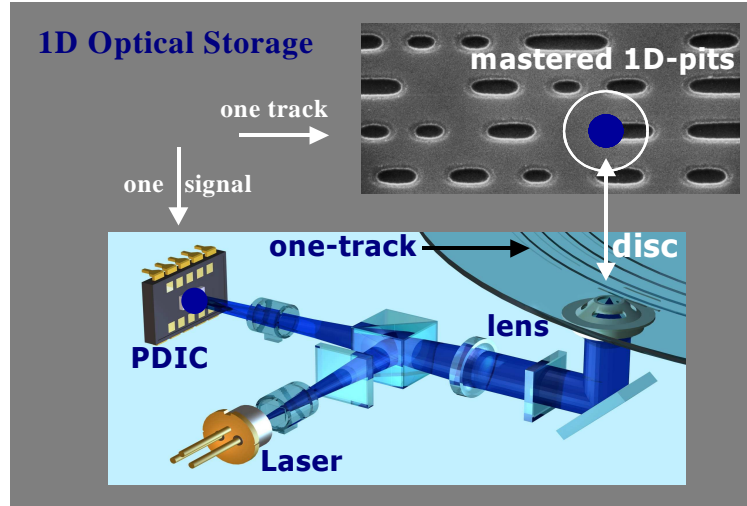


Figure 1.4: Schematic overview of an optical storage system.

low reflectivity. The lands however have a high reflectivity. As a result pits and lands cause different reflected light intensities, making them distinguishable at the receiver end of the system.

The production of the read-only discs is called mastering and is achieved by mechanically impressing a negative image of a master stamper. For rewritable media the information is not stored by lands and pits but by areas where the state of the alloy is different. In its preferred crystalline state, the alloy reflects light in a unique direction. In the amorphous state the light is reflected equally in all directions. As a result the two states cause different reflection intensities, again making them distinguishable at the receiver end.

Fig. 1.4 depicts some of the basic elements of a conventional 1-D optical disc player. The light beam, generated by a semiconductor laser diode, is focused on the disc by a beam splitter and an objective lens. A servo procedure (not shown) ensures that the optical spot is centered on a single track of mastered pits. The reflected light is focussed on the Photo Detector Integrated Circuit (PDIC). This PDIC generates an electrical signal according to the intensity of the light. As a result pits (or amorphous areas for RW discs) on the disc result in electrical signals with a low amplitude while lands (or crystalline areas) result in signals with a high amplitude. Based on this information the read channel is able to retrieve the information that was written on the disc.

As mentioned before, the light is focussed by the objective lens down to the limits of diffraction, resulting in the well-known airy light intensity profile [17]. The full-width at half-maximum of the light profile is known as the airy distance and is of major importance to optical storage since the achievable density of the disc will be

governed by this parameter. In general, the spot size is determined by two parameters: λ and NA. The airy distance ϕ_{Airy} (which is half of the spot size) is defined as

$$\phi_{Airy} = \frac{\lambda}{2NA} \quad (1.1)$$

and is a fundamental quantity because the permissible distance between two adjacent bits is ruled by ϕ_{Airy} . If the distance is smaller than the airy distance, Intersymbol Interference (ISI) arises, i.e. light is reflected not only by the current bit but also by bits adjacent to the current one. In the tangential direction (along the track) ISI is allowed because the receiver is able to deal with it up to a limited amount. As a result, the tangential distance can be reduced below the airy distance but only to a limited amount. In the radial direction (orthogonal to the track), cross-talk cannot be handled in conventional single-spiral disc systems. Hence the intertrack distance must be larger than the airy distance. As a result the total amount of bits that can be stored in a specific area is limited by the airy distance. More precisely the area of a user bit cell scales proportionally to $(\lambda/NA)^2$.

Besides the limit in achievable capacity for a given combination of laser and optics, there exists also a limit on the achievable data rate. The maximum achievable data rate is limited by the maximum rotation velocity of the disc, which is limited by the maximum centrifugal forces the polycarbonate disc can endure without breaking. Experiments have shown that the ultimate linear velocity at the outer radius of a standard 12 cm disc is approximately 56 m/s. Whereas the capacity scales proportional to $(\lambda/NA)^2$, the maximum data-rate depends on the tangential bit size only and scales linearly with λ/NA . Hence, in optical storage, the maximum data rate does not keep pace with the growth in storage capacity. While the time to record a full DVD at the maximum rate amounts 5 minutes, it takes about 12 minutes to write a full BD [18]. One possible solution for this problem is the parallel writing/reading of tracks on a disc. This parallel access will be referred to as 2-D storage and is the topic of the next section.

1.2.2 Two-Dimensional Optical Storage

The first and so far only 2-D system on the market has been introduced by Zen-Kenwood around 1997 under the TrueX trademark [19–21]. A schematic representation of the system is shown in Fig. 1.5. Because of the strong market position of the standardized CD and DVD format, the TrueX system was bound to read these formats (single-spiral discs). Although the TrueX system uses 7 laser spots, the gain in data rate did not amount to the same factor for these discs. Basically there are two reasons for this. The first reason is the fact that the data is read discontinuously and when one of the beams reaches a zone that was read previously by another beam, it

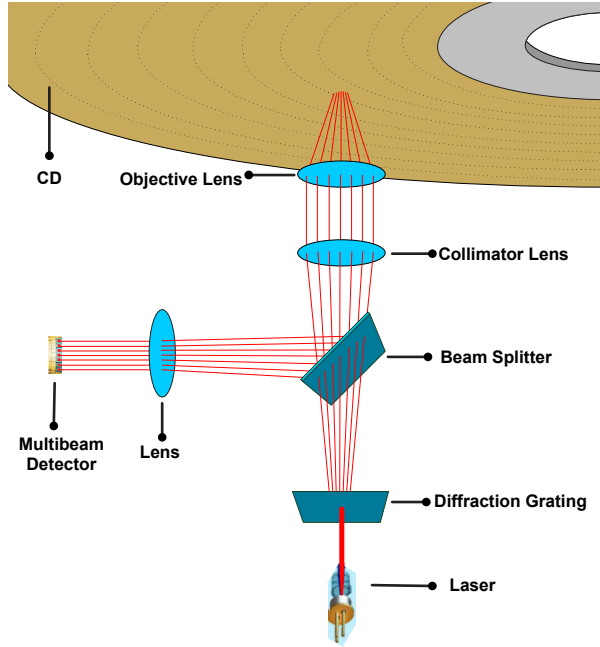


Figure 1.5: Schematic representation of the Zen-Kenwood technology.

has to jump to another zone of the disc. That jump is quite time consuming. The second reason is that data are not read in the correct order, and hence must be re-ordered to re-form logic data blocks and to ensure correct reading of the disc.

To really benefit from the multiple spots in the system, the format of the disc should be adapted accordingly. This results in a multi-track disc in which the jump necessary in the TrueX system becomes unnecessary. The combination of multi-track recording and multi-spot reading allows the data rate to be increased by a factor equal to the number of tracks on the disc, i.e. the number of parallel laser beams. An experimental system, called Two-Dimensional Optical Storage (TwoDOS), utilizes this combination to achieve an increase in data rate with a factor of 10 with respect to the Blu-Ray system using the same physics. In Fig. 1.6 the disc formats of Blu-Ray and TwoDOS are shown. In contrast with conventional optical storage, where the bits are stored in a single spiral (a 1-D sequence of bits), in TwoDOS the bits are organized in a so-called broad spiral [18]. Within a single rotation of this broad spiral, a number of L bit-tracks are placed besides each other to form a hexagonal structure. Adjacent rotations of the broad spiral are separated by a guard band consisting of a bit-track without any pits. The data is read out with an array of L laser spots arranged such that each spot is centered on one of the bit-tracks within the broad spiral.

An additional advantage of multi-track storage is that it can lead to an increase of areal density [22]. As mentioned earlier, traditional 1-D systems treat cross-talk as

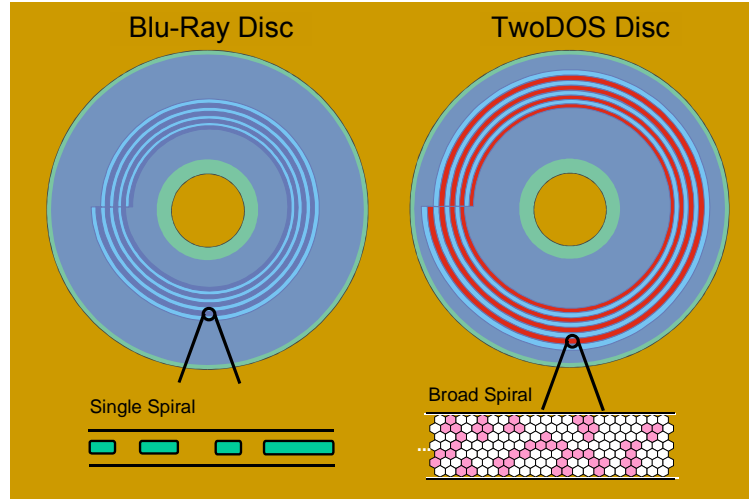


Figure 1.6: Disc formats of Blu-Ray disc and TwoDOS systems.

undesired interference, and efforts in advanced signal processing have mainly been directed towards an increase of the tangential linear density, while increases of radial density (number of tracks that can be packed in the radial direction) have generally been neglected. As a result the tracks are separated by a distance that makes radial ISI negligible. Recently, however, crosstalk cancellers have been used to permit an increase of radial density [23]. Also the use of 2-D modulation codes can allow a reduction of the impact of radial ISI as demonstrated in [24]. Nevertheless, as radial ISI increases, all those methods deteriorate quite quickly. As a result, to increase the radial density substantially, radial ISI should not be treated as undesired crosstalk but as information.

In 2-D systems, the different tracks are read out simultaneously, resulting in an array of signals. This array is used as input of the read channel which performs parallel processing on the signals of the array. The parallel processing makes it possible to treat the radial ISI present in the different signals, as information in the detection process. Because the data is stored on a fixed hexagonal lattice, a stationary 2-D bit configuration is present under every spot at each detection instant. By accounting for all ISI (tangential and radial ISI) in the read channel, the distance between the tracks can be narrowed within the broad spiral (equivalently increasing the radial density), hence an increase of the quantity of information stored on the disc.

The improvements in data rate and in capacity of 2-D optical storage are made possible by advanced signal processing techniques, which are necessary to reliably detect the bits of the new disc format. An additional advantage of 2-D systems is the fact that it can be implemented on top of the capacity/data rate improvements that are obtained by further improving the laser wavelength λ and the NA of the objective

lens. Hence the 2-D approach is orthogonal to other approaches for improving the capacity and the data rate of optical storage systems. For this reason the 2-D approach was discussed separately from other new optical storage technologies, which are discussed in the next section. Two-dimensional optical storage is the main topic of this work and a more comprehensive introduction to it is given in chapter 2.

1.2.3 New Technologies

By extrapolating the parameters of the conventional standardized optical storage systems (see Fig. 1.3), the capacity and the data rate of the next standard are expected to be around a Quarter Terabyte (QTB) per layer and 150 Mb/s (1x), respectively. The exact capacity and data rate has always been defined by a particular target application for each new format: CD for audio (74 minutes in digital format), DVD for a full-length movie (about two hours in MPEG2 coding) and finally BD (or HD-DVD) for movies in high-definition format. Besides these specific target applications, several other application fields continuously demanded higher storage capacities and higher data rates: data archiving and software distribution, in particular PC-games with a high video resolution.

At this moment the target application of the 4th generation system is less clear. There is a general trend to store digital content (audio and video) on hard-disks, witness the commercial success of the Apple iPod. The device containing the hard-disk is becoming the multimedia center of the home, containing digital audio, digital video and personal data. This multimedia center generally also contains an optical drive to archive data. This optical drive should have a reasonable storage capacity but more importantly the data rate will be a key factor, as already explained before. Other possible applications for the new generation of optical drives are 3-dimensional video, interactive video and gaming with full-resolution video content.

In search of the next generation of optical storage devices, many research efforts have been established, each of them with its advantages and disadvantages. In this section a brief overview will be given of the different approaches.

Near-field Storage

One possible solution for increasing the capacity, is increasing the NA of the lens beyond 1. This is possible by reading data through a “solid immersion lens” (SIL) [25]. This type of optics is already used in microscopes and in lithography equipment for semiconductor production. The SIL uses the different refractive indices of glass and air to achieve a high NA. The SIL optical head is composed of a hemisphere which is made of high refractive index glass and high NA focusing objective lens [18].

The attribute “near-field” refers to the extremely short distance between the read/write head and the disc surface. Since the intensity of the reflected light is very

sensitive to the distance between the head and the disc, the SIL should be allowed to fly over the disc at only a few ten nanometers from the disc [26]. A system using an actuator accomplishes this by carrying the head and containing the SIL. The roughly 25 nm gap is directly comparable to the distance between the head and the disk surface in hard-disk assemblies. In literature experimental systems based on near-field technology, have been described that have a capacity of 150 GB per layer [27, 28].

Multi-layer Storage

Commercial optical discs are now available in dual-layer formats, where the two layers are separated by a distance that is relatively large compared to the focal depth of the laser beam. To increase capacity, it would be desirable to increase the number of layers. It is well known that the amount of spherical aberration increases considerably with the number of layers [28]. This aberration originates from crosstalk from adjacent layers and interference of out-of-focus tracks, and causes great difficulty in the read-out electronics [29]. However, for a limited amount of layers the aberrations can be controlled and reliable read-out can be realized. One of the main reasons that discs with more than two layers have not been commercially available is the increased production cost for these types of discs. Experimental systems have been demonstrated that use 4 or 8 layers [28]. Results show that 100 or more layers may be possible with conventional thin-film technology if sufficient read-out signal-to-noise ratio is obtained [29, 30].

Multi-level Storage

On conventional optical discs data is stored only via a binary alphabet. A natural and immediate idea to increase the capacity of disc is to use a larger alphabet. This idea has, of course, been extensively studied in the past, but has not been very successful. A rewritable multi-level system can be realized by recording marks with different sizes [31]. For read-only systems the reduction in mark size is difficult to achieve because the pits have to be mastered and replicated. To overcome this problem pit-edge modulation has been proposed: a multi-level signal is generated by shifting the rising and falling edge of the binary modulated signal in discrete steps during mastering of the disc [32–34, 34]. Another option is to modulate the pits in the radial direction as discussed in [35]. Also pit-depth modulation has been proposed [36]. Besides the problem of mastering, the multi-level approach is not compatible with the existing formats. For all those reasons and the need of high Signal-To-Noise Ratio (SNR), multi-level storage has not been a great success.

Holographic Storage

In optical holography, data is stored throughout the volume of the recording medium, as opposed to on the surface for disc storage systems. Data are impressed onto an optical coherent beam using a spatial light modulator or page composer. The signal-bearing beam interferes with a reference beam inside the recording medium to produce an interference grating, representing a data page. Multiple gratings are superimposed by varying the optical properties of the reference beam, a process referred to as multiplexing. Upon data retrieval or read-out, a single reference beam is incident on the medium under the same conditions as used for storage, producing a diffracted beam representing the stored data page. The diffracted beam is detected by a detector array, which allows extraction of the stored data bits from the measured intensity pattern.

Since data can be accessed through large pages, holographic memories can offer extremely high data rate, as fast as 10 Gb/s. An important limitation to holographic memory developments is that the power of the refracted signal is reduced with the number of holograms that are superimposed. As a result, at high density numerous pages are superimposed which leads to low reflection and the information cannot be accessed at high rate with high reliability. Therefore holographic memories are facing the problem of having to realize a trade-off between access speed (which optical memory generally lacks) and capacity. To be competitive, holographic memory needs to achieve 500 Mb/s and 250 GB on a 12 cm disc. This technology is not available yet but with improvements of medium and read-out techniques this could be achieved in the next few years.

1.3 Magnetic Storage

Digital magnetic storage systems originated after the second world war, closely linked to the development of the first digital computers [9]. IBM's 350 was the first disk drive system and was invented by Johnson in 1956 [37]. The drive consisted of 50 disks of 24 inch and could contain about 4.4 MB of data. From that point on, storage capacities, data rates and price per bit have undergone a rapid and continuous growth. In 2008 disk drives with a capacity of 1 TB are commercially available containing 4 disks of 3.5 inch. In the 21st century applications for hard disks have expanded beyond computers to include digital video recorders, digital audio players, personal digital assistants, and digital cameras. In 2005, the first mobile phones to include hard disks were introduced.

Continuous improvements in both recording/reading heads and magnetic media (the disk itself) have been the key enablers of this evolution. The heads have been made considerably smaller and more sensitive, allowing writing and reading of smaller bits. Also the distance between the disk and the flying heads has been re-

duced from about 6.35 mm for the IBM 350 to 10 nm and below nowadays and as this distance is one of the determining factors of the achievable resolution (i.e. the smallest disk region that reliably can be written or read), a higher storage density can be achieved. The main media improvements were the reduction of the substrate coating thickness, the improved quality of the coating (flatness, robustness) and the improved thermal stability of the magnetic material (to avoid the problem of thermal bit erasure).

Besides these improvements in the physics of the system, also improvements in the signal processing algorithms have led directly to increased storage capacities. The continuously increasing capabilities of digital electronics allowed the signal processing algorithms to be more complex. In Section 1.4, the developments in the signal processing algorithms utilized for data storage will be discussed in more detail. In the next section the operation of magnetic storage systems will be discussed.

1.3.1 Longitudinal and Perpendicular Storage

Figure 1.7 depicts schematic views of both longitudinal and perpendicular disk drives. In both types of magnetic drives, the information is stored in the recording layer of the magnetic disk in the form of small regions with a magnetization in either one of the two opposite directions. These regions are denoted as magnetic elements and the direction of magnetization of these elements represents the bits. In longitudinal magnetic storage the medium is magnetized in the direction of the disk motion, whereas in perpendicular storage the medium is magnetized vertically, i.e. perpendicular to the direction of disk motion [38]. The recording (i.e. writing) of the information is accomplished by applying a signal current to the windings of the recording head. This current magnetizes the head and causes a flux pattern that follows the head poles and fringes from the head due to the presence of an air gap. The fringing head flux magnetizes the media. A very small distance of the head to the media is a prerequisite for high information densities as it determines the achievable resolution. To achieve a very small distance, the magnetic head is a flying head that uses air bearing to levitate at a constant height over the disc (called flying height).

During the read-out process the magnetization of a bit region causes a flux in the head, resulting in a voltage across the windings of the head. The detection of the bits is realized by monitoring this voltage. This type of head is called an inductive head and is also depicted in the figure. In replacement of these inductive heads, magneto-resistive (MR) heads have been introduced for reading. MR heads use a sensor of magneto-resistive material that is placed between two shields. The excellent sensitivity of these MR heads has been a key factor in the density improvements after 1992.

Up to a few years ago, all commercial HDDs used longitudinal storage while perpendicular storage received a lot of scientific attention but was not commercial-

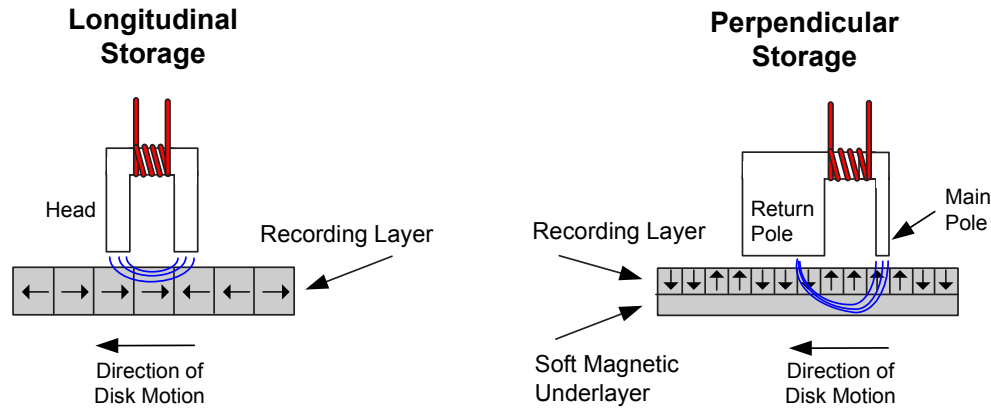


Figure 1.7: Comparison of longitudinal and perpendicular magnetic storage.

ized [39]. However in 2006, the first commercial HDD based on a perpendicular arrangement was introduced allowing higher densities (in 2007 the first commercial HDD with a capacity of 1 TB was introduced based on perpendicular storage). Perpendicular storage achieves a higher density because the alignment of the bits in this manner takes less platter area than in the longitudinal system [40]. Hence bits can be placed closer together on the platter, increasing the number of magnetic elements that can be stored in a given area. Another reason for the increased capacity of perpendicular systems is the higher coercivity of the magnetic material. This is possible due to the fact that in a perpendicular arrangement the magnetic flux is guided through a magnetically soft underlayer underneath the hard magnetic media films. This magnetically soft underlayer can be effectively considered as a part of the write head, making the write head more efficient, thus making it possible to produce a stronger write field gradient with essentially the same head materials as for longitudinal heads, and therefore allowing for the use of the higher coercivity magnetic storage media.

1.3.2 Two-Dimensional Magnetic Storage

The motivation of multi-track recording/multi-head reading is twofold: increasing the overall density and increasing the data rate. As already stated before, a conventional 1-D system does not treat the radial ISI as information but considers it as interference. However in a 2-D system the radial ISI can be treated as a source of information. Hence the radial density can be increased considerably by placing the different tracks next to each other without any guard space in between them [41]. Besides the increase in density, also an increase in data rate is achieved by using an array of heads to read out the information on the disk.

A general multi-track/multi-head configuration is depicted in Fig. 1.8, together

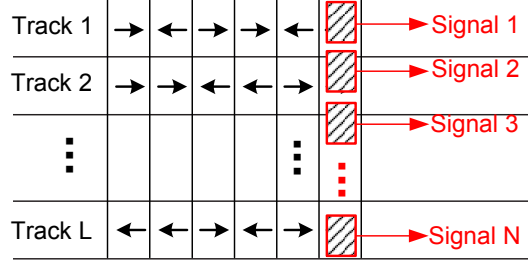


Figure 1.8: General multi-track multi-head configuration in magnetic storage system.

with the disk format [42]. The basic idea of such a configuration is to adapt the format of the disk to the multi-head read process [43]. Basically the information on the disk is stored in a multi-track format that is then accessed by parallel reading. For magnetic storage the parallel tracks are contained within concentric meta-tracks. The parallel multi-track reading over L tracks is performed by an array of N magnetic read heads.

The detection process should be based on multi-channel signal processing theory. Hence a multi-input, multi-output (MIMO) problem statement can be used as basis for simultaneous detection of the read-out signals from the interfering magnetic tracks [22]. In Chapter 2 receiver structures will be discussed that can be applied for bit-detection in these 2-D magnetic storage systems.

1.4 Basic Signal Processing for 1D Storage System

In Fig. 1.9 a more detailed overview of a storage channel is depicted (see Fig. 1.1 for the general overview). As already stated before, the user data is stored via the write channel on a physical media. The media together with the physical components to read and write, form the physical channel. Finally the read channel recovers the stored user data.

Error-Correction Coding (ECC) is first applied to the user data to prevent burst errors. To this end the ECC encoding step adds some redundant information [44]. In many commercial HDDs, Reed-Solomon codes with certain degrees of interleaving are used [45]. The encoded bits are then subject to another type of coding, namely, modulation coding [44]. The purpose of modulation coding is to match the data to the characteristics of the physical channel and to help in the operation of various adaptation loops [46, 47]. Many types of modulation codes are used depending on the specific needs (see Section 1.4.1). The modulation-encoded bits are the actual bits that are stored on the media. The pulse modulation block converts these bits into an appropriate write-current waveform which can be used by the physical storage channel. For example in magnetic storage, each current pulse is properly shaped

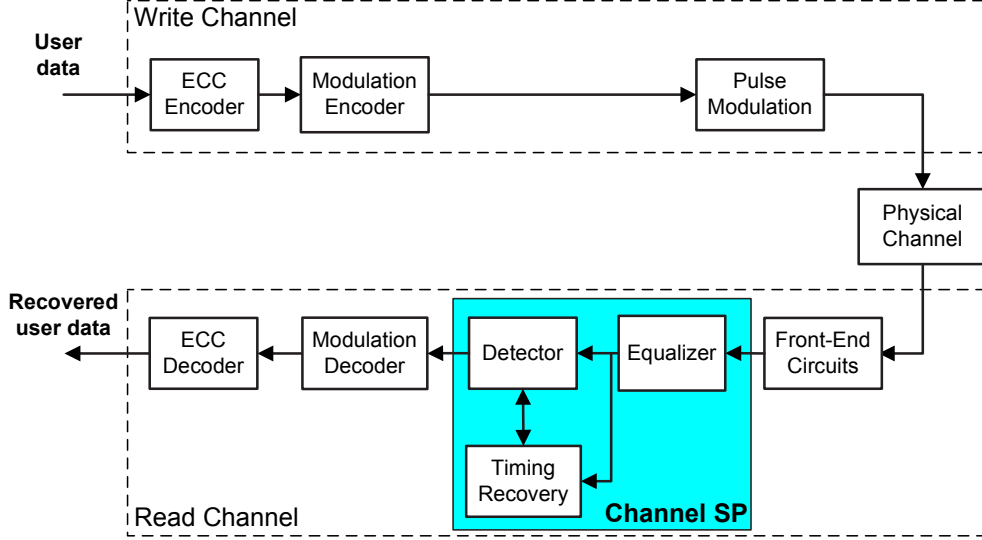


Figure 1.9: Block diagram of digital optical storage system.

and positioned (by means of pulse shaping and write precompensation) to counteract nonlinear distortions in the recording process. These operations are performed in the pulse modulation block.

In the overall block diagram, every block in the write channel has a counterpart in the read channel. The analog signal that is generated by the physical channel during read-out, is processed by front-end circuits (e.g. amplifier, bandlimiting filter, analog to digital convertor, ADC, etc.) which condition the replay signal prior to the channel Signal Processing (SP) block [48]. The channel SP block aims at recovering the data written on the disc as reliably as possible. To this end, an equalizer shapes the signal according to certain pre-chosen criteria so that a data detector is able to recover the binary data with as few errors as possible, while a timing recovery block ensures the detector operates at a digital signal that is sampled synchronously with respect to the recorded bits [3, 6]. The detected data sequence is then applied to a modulation decoder and finally to an error-correction decoder. The resulting recovered data sequence is the best estimate of the user data at the input of the storage system.

In this section we describe the state-of-the-art in signal processing algorithms for storage systems. Although the SP algorithms in optical storage (see [49, 50]) and in magnetic storage (see [6, 9]) are tailored to different applications, they are very similar. As a result, the SP algorithms described in this section, are applicable to both systems. If a specific algorithm is tailored to one of both systems, this will be explicitly mentioned. The focus is on the advanced digital signal processing algorithms that are of particular interest for the remainder of this work. The main important terms are discussed in more detail: modulation codes, detection principles and other important

Table 1.1: Part of the EFM conversion table

User data sequence	EFM sequence
00000000	01111000111111
01010101	00000011111000
01111100	01111111111110
11010010	11100011100001
11111111	00111111100011

signal processing techniques (e.g. equalization and timing recovery).

1.4.1 Modulation Codes

As already stated above, the task of the modulation encoder is to convert its input data into a constrained sequence which is suitable for the physical storage channel. Run-Length-Limited (RLL) codes are widely used for this purpose in digital magnetic and optical storage systems. They are also known as (d, k) codes, where $d + 1$ and $k + 1$ are respectively the minimum and the maximum lengths of strings of identical symbols in the encoder output stream. The d -constraint controls the highest transition frequency and thus has a bearing on intersymbol interference when a bandwidth-limited channel is considered. The k -constraint limits the maximum transition spacing and ensures that the adaptation loops are updated frequently enough. For example, timing is commonly recovered with a phase-locked loop which adjusts the phase according to observed transitions in the waveform, and the k -constraint ensures an adequate number of transitions for synchronization of the read clock.

The benefits of RLL codes come at a cost in the form of redundancy that is added in the data stream. On the average, p source symbols are translated into q channel symbols. The rate R of the modulation code is given by $R = p/q$. Clearly $0 < R < 1$. In general RLL codes will decrease the overall throughput of the system, resulting in either a lower data rate, or a lower Signal-To-Noise Ratio (SNR) in case the baud rate $1/T$ was enlarged to achieve the same overall user data rate [3]. In general, the baud rate $1/T$ is defined as the number of modulated bits read from the media per unit of time, i.e. it takes T seconds to read one modulated bit from the physical media.

In practical storage systems, the d -constraint is restricted to 0,1 or 2; and the k -constraint ranges between 2 and 10. For example, in the CD system the Eight-to-Fourteen Modulation (EFM) code is used, which is a (2,10) code (with $R = 8/17$). In Table 1.1 some examples of the conversion of the user data to the modulation encoded bits are given for the EFM code. In this code 8 user bits are converted into 14 modulation encoded bits [51]. In magnetic storage systems, the rates of $d = 0$ codes have been steadily increasing over the years, from initially, rate 8/9 to 19/20 and 64/65 [52].

Besides these improvements in rates, further enhancements have been introduced

in recent years: the combination of RLL codes with parity bits in parity-based post processing schemes [52] and a Maximum Transition Run (MTR) constraint which eliminates the critical bit patterns that cause most errors in sequence detectors [53]. For optical storage, modulation codes often also need to have the DC-free property, i.e. they should have almost no content at very low frequencies [54]. This DC-free constraint significantly reduces interference between data and servo signals. Furthermore it facilitates filtering of low-frequency disc noise, such as finger marks on the disc surface.

1.4.2 Detection Principles

The objective of the channel SP block (see Fig. 1.9) is to recover the data written on the disk as reliably as possible. In the remainder of this work this block will be denoted as the data receiver. The analog replay signal coming from the physical channel is preprocessed by front-end circuits. These circuits comprise an anti-aliasing filter and an ADC to convert the analog signal into a digital signal. This digital signal is subsequently used as the input of the data receiver. Generally, the receiver can be considered to consist of two parts: a detector and a preprocessing part. The preprocessing part aims at transforming the receiver input signal into a signal with properties that are desired by the bit-detector. Typically this part consists of an equalizer to shape the ISI structure, a timing recovery circuit to make the detector input signal synchronous with respect to the baud rate $1/T$, and some additional signal processing blocks with a specific purpose (e.g. offset and gain control). The type of detector that is used, determines the desired operation of the preprocessing blocks. Therefore the different types of detectors will be discussed first and in the next subsection the other signal processing blocks will be discussed in more detail.

Strictly speaking, detectors come in two categories: symbol-by-symbol detectors and sequence detectors. Symbol-by-symbol detectors essentially make a memoryless mapping of the detector input into detected bits. Peak detectors are a typical example of this type of detectors and were the universal choice for data detection in magnetic storage until the 90s [55]. In optical storage systems, a slicer is a typical example of a symbol-by-symbol detector that is extensively used in CD systems. To account for the ISI in the system, symbol-by-symbol detectors should be properly combined with RLL coding. In optical storage systems the use of a nonlinear equalizer called limit equalizer [56] and a post-processing scheme to correct dominant errors in the threshold detector output [57] have been proposed to improve the performance. These additional schemes make the receiver more robust against ISI and other artifacts, such as media noise. However, as tangential storage densities increases, the overlap between neighboring pulses becomes severe and the peak detector performance deteriorates significantly, even with the use of these additional mechanisms [58].

Sequence detectors make a symbol decision based on observation of signals over

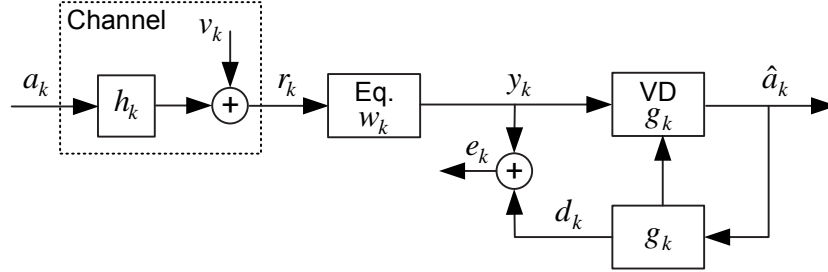


Figure 1.10: Schematic overview of a PRML system.

many symbol intervals and, as a result, they can perform considerably better than symbol-by-symbol detectors in the presence of large amounts of ISI. Maximum Likelihood Sequence Detectors (MLSD) are the most prominent example of sequence detectors and are typically implemented as a Viterbi detector (see Section 1.4.3 for more explanation about Viterbi detection) [59–62]. The optimal detector can generally not be realized because of its excessive complexity. Hence, in real applications, sequence detectors are invariably used in combination with an equalizer, which reduces the effect of ISI to a certain extent [63]. Receivers of this type are known as Partial-Response Maximum-Likelihood (PRML) receivers [4] and are widely used in magnetic storage systems since 1990 and also in the Blu-Ray Disc system [16]. Fig. 1.10 depicts a PRML system. The data a_k is corrupted by the channel which is characterized by an unknown impulse response h_k and an additive noise sequence v_k . The received signal r_k is input of the equalizer with impulse response w_k . The equalized signal y_k is input of the VD which is designed for a predefined target response. In general, a target response defines the expected ISI structure at the detector input and may be characterized by the impulse response g_k . The memory length G of the target response influences the complexity of the VD following an exponential rule (2^G), see Section 1.4.3 for further explanation about the VD. The equalizer serves, roughly speaking, to transform h_k into g_k . The equalizer impulse response w_k is commonly adapted based on the error signal e_k which is the difference between y_k and d_k , where d_k is the desired detector input signal. The adaptation of w_k is discussed in Section 1.4.4. The VD produces bit decisions \hat{a}_k . The operation of the VD will be explained in Section 1.4.3.

The choice of the target response g_k is crucial for guaranteeing optimal system performance [64]. Hence, in literature, numerous methods have been proposed for choosing a target response based on several criteria [65–69]. In general Viterbi detectors are optimal in case there is no residual ISI (RISI) at the detector input and the noise is spectrally flat (i.e. white). As we will see in Chapter 3, RISI and non-white noise will be key problems in 2-D systems which each require designated solutions to guarantee acceptable system performance. The most favorable choice for g_k would be one which yields a limited amount of ISI components to limit the complexity of

the VD, and which has an amplitude spectrum that is similar to the one of the channel to minimize noise enhancement.

Mismatches between channel and target causes the noise to be colored at the VD input and the VD becomes a suboptimal detector. Several modifications have been proposed to improve the performance based on the noise characteristics: for colored noise [70–72], for data-dependent noise [73] and for data-dependent colored noise [74, 75]. Basically all these modifications can be divided into two groups: techniques where the target is adapted such that the noise is as white as possible [76] and techniques where noise prediction within the VD is used to effectively whiten the noise [77–79].

Besides these modifications for noise characteristics, also modifications for nonlinear channels have been proposed. By employing a linear target response, it is not possible to cover nonlinear ISI components, which can be significant especially at high storage densities. As a result significant Residual ISI (RISI) remains at the detector input, which deteriorates the performance of the detector significantly. For this problem, researchers have proposed two types of solutions: a nonlinear equalizer to minimize nonlinear ISI at the detector input, and a modified detector that accounts for the nonlinear ISI at its input. In the latter solution the VD is not matched anymore to the linear target response g_k on which the equalizer is based but it is matched to a nonlinear target response, often described by a look-up table, denoted as the VD ideal values table. In this table, for every possible combination of bits within a pre-defined window (normally with the same length as the linear target response g_k), an entry is stored that represents the VD ideal input value. This table is used in the VD for the computation of the branch metrics, see Section 1.4.3 for the definition and the computation of branch metrics. The entries in the table account for all nonlinear ISI which is still present after equalization [76, 80]. These ideal values are adapted based on the Least Mean Square (LMS) algorithm and are used in the VD for the computation of the branch metrics. Significant performance gains are possible when such measures are employed [74, 81–83].

Another detection technique of interest is the Decision-Feedback Equalizer (DFE) [84, 85]. In Fig. 1.11 a schematic overview of a DFE is depicted. The DFE consists of a feedforward filter, a feedback filter and a slicer. The feedforward filter equalizes the signal into a target response which is constrained to be causal so that precursive ISI, i.e. interference due to symbols which are not yet detected, is absent. The feedback filter cancels all postcursive or trailing ISI, i.e. interference due to symbols that have already been detected, based on past decisions such that at the slicer input only ISI due to the current symbol a_k is present. To do perfect cancellation, the impulse response of the feedback filter should contain all postcursive ISI components of the target response. The slicer makes bit-decision \hat{a}_k on a symbol-by-symbol basis. The cancellation of postcursive ISI comes for free in terms of noise enhancement [3]. As

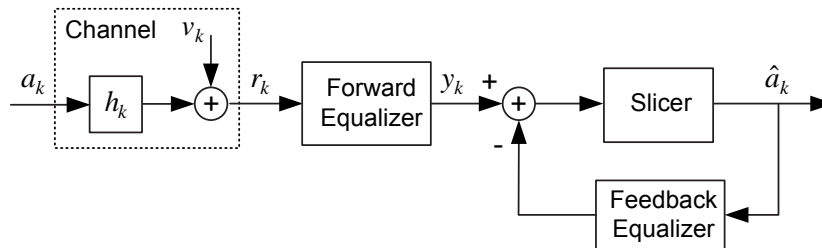


Figure 1.11: Schematic overview of a DFE system.

a result, the DFE is optimal in the sense of minimal noise enhancement. In fact it is the optimum receiver with no detection delay. A disadvantage of the DFE is the problem of error propagation, i.e. erroneous decisions propagate via the feedback filter and can corrupt subsequent decisions. Fortunately, error propagation is typically only a minor problem. Also for DFE systems, improvements have been proposed for data-dependent noise [86] and for nonlinear channels [87].

To improve the performance further, several modifications to the basic DFE have been proposed: for example parallel DFE [88], dual DFE [89] and multi-level DFE [90]. Another key proposal has been to use a combination of decision-feedback and a fixed-depth tree-search (FDTS) based detector, this combination is called fixed-depth tree-search with decision-feedback (FDTS/DF) [5]. The FDTS detector is a depth-limited exhaustive tree search algorithm, A good comparison of the performance of the different types of detection techniques (PRML, DFE and FDTS) is given in [91] and in [92].

Besides these traditional detection techniques, new techniques have been proposed to improve system performance even further. One of these new techniques is iterative/turbo detection [93–96], which has been applied to optical systems [97, 98] and to magnetic systems [99–101]. Another type of technique is the postprocessing technique which try to improve the reliability of the bit decisions based on knowledge of the nonlinear channel [102] or of the noise characteristics [50, 103, 104]. As these new techniques are not of primary interest for this work, we will not further elaborate on them.

1.4.3 Viterbi Detection

In this work the PRML receiver will be mainly used. In this section, the VD which is used in the PRML receiver will be discussed in more detail. The very well-known algorithm introduced by Viterbi in 1967 [105] is the most extensively used detection algorithm and had a massive impact on digital communication. A Viterbi detector (VD) performs maximum-likelihood detection in an efficient fashion and its operation is based on dynamic programming [3]. Viterbi detection is possible for optical and magnetic storage channels because these types of channels have a finite memory

length M . For example, a memory length M of 2 bits means that the channel output depends on the current bit and the previous 2 bits. The memory length can be reduced by using an equalizer before detection. So in general, at the time k , the detector input signal can be written as

$$y_k(\mathbf{s}_{k-1}, \mathbf{s}_k) = \sum_{i=0}^M p_i a_{k-i} + n_k, \quad (1.2)$$

where p_k is the impulse response function at the detector input, $a_k \in \{-1, 1\}$ are the bits, n_k is a noise term and $\mathbf{s}_{k-1} = [a_{k-M}, \dots, a_{k-1}]$ is the state describing the memory of the channel and \mathbf{s}_k is the current state.

At the heart of the VD is a so-called trellis that is determined by the memory length of the channel and the modulation code. In Fig. 1.12 an example of such a trellis is shown for uncoded data and a memory length $M = 2$. The nodes of the trellis at time instant k represent all possible states \mathbf{s}_k . In total there are $S = 2^M$ possible states. In the trellis, each state \mathbf{s}_k has two possible predecessors \mathbf{s}_{k-1} . The transitions between states are called branches. Every branch uniquely describes a noiseless detector input $d_k = \sum_{i=0}^M p_i a_{k-i}$ (denoted as the reference value) where a_k are the bits described by the two states that are connected by the branch and where p_k is assumed to be known by the detector. Maximum-likelihood sequence detection boils down to finding the admissible sequence d_k that is closest to the detector input sequence y_k in the Euclidean sense. This can be accomplished by first calculating a metric for every branch in the trellis according to

$$\beta(\mathbf{s}_{k-1}, \mathbf{s}_k) = (y_k - d_k)^2. \quad (1.3)$$

Subsequently the path through a sequence of states in the trellis has to be found that has the least sum of branch metrics. This path reflects the best fit of the detector input signal y_k to the calculated reference values d_k . The goal of the VD is to find the path through the entire trellis that has the smallest path metric. This goal is achieved by following a stage-wise approach. At each stage k and for every state \mathbf{s}_k it finds the path with the smallest metric. This path is denoted as a survivor path, so at stage k , there are $S = 2^M$ survivor paths, one for each state. The survivor paths can be constructed by a recursive procedure. At each time instant k , for all states the path with the smallest metric is calculated. The smallest metric is obtained by comparing path metrics for its two predecessor states and subsequently selecting the smallest path metric. The path metric for going from state \mathbf{s}_{k-1} to state \mathbf{s}_k can be calculated as

$$\lambda_k^{\{\mathbf{s}_{k-1}, \mathbf{s}_k\}} = \hat{\lambda}_{k-1}^{\mathbf{s}_{k-1}} + \beta(\mathbf{s}_{k-1}, \mathbf{s}_k), \quad (1.4)$$

where $\hat{\lambda}_{k-1}^{\mathbf{s}_{k-1}}$ was the smallest path metric leading to state \mathbf{s}_{k-1} at time instant $k-1$. As a result the smallest path metric leading to state \mathbf{s}_k is calculated as

$$\hat{\lambda}_k^{\mathbf{s}_k} = \min_{\mathbf{s}_{k-1}} \{\lambda_k^{\{\mathbf{s}_{k-1}, \mathbf{s}_k\}}\}. \quad (1.5)$$

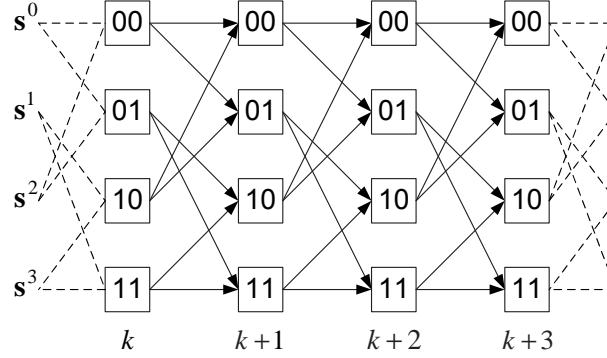


Figure 1.12: Example of a trellis for a channel with memory length 2.

The selection of the path with the smallest path metric is done by a so-called Add-Compare-Select unit. Because of the finite memory length of the channel, each of the survivor paths will, after some amount of back-tracking, pass through the same node in the trellis. This node uniquely defines the detected bit that is used as output of the detector.

1.4.4 Adaptation

In Section 1.4.2 it was mentioned that equalization is an effective mechanism to shape the signal suitable for detection, by mitigating the effects of ISI and noise. Adaptation techniques are used in the receiver to compensate, in real-time, for variations in the optical channel. The choice of the adaptation speed, i.e. the speed at which we react to variations in the optical channel, is a key parameter in the design of adaptation loops. On the one hand, we would like to make the adaptation as fast as possible to adequately track all variations. On the other hand, fast adaptation invariably leads to worse steady state performance due to the increased amount of gradient noise, which is noise in the adaptation loop due to noise sources present in the system. As a result it is desirable to adapt not faster than strictly needed to track the bulk of the variations. For the same reason it is desirable to adapt no more parameters than strictly needed. Based on this reasoning, it is appropriate to slowly adapt the large number of equalizer coefficients to account for the overall channel response variation and to adapt a few important and fast-varying parameters with some dedicated fast adaptation loops. Usually in storage applications these parameters are sampling phase, gain and DC level [4].

Another key factor that plays a role in the determination of the choice of adaptation speed is the overall latency in the adaptation loop. In general, latencies cause the tracking capabilities of the loop to deteriorate with respect to a loop without any latency. Detection latency is the main source of latency in an adaptation loop and as a result the latency of the detector should be taken into account in the design of

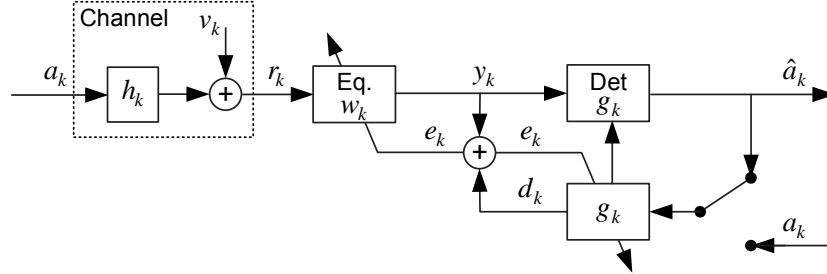


Figure 1.13: Schematic overview of an adaptive PRML system.

all adaptation loops. In 2-D systems, detection latencies are even larger and as a result the design of adaptation loops for 2-D systems are of particular importance to guarantee acceptable system performance.

In this subsection the adaptation loops used for the different parameters are discussed. First the adaptation of the equalizer and the target response is discussed.

Equalizer and Target Response

In Fig. 1.13 a schematic overview of an adaptive PRML system is depicted, where both the equalizer impulse response w_k and the target response g_k are adapted. The channel has an unknown impulse response h_k while the detector is designed for the target response g_k . The equalizer serves, roughly speaking, to transform h_k into g_k . To this end, the equalizer impulse response is adapted based on the error signal e_k . This signal is a measure of the difference between the actual detector input signal y_k and the desired input signal $d_k = (\mathbf{a} * \mathbf{g})_k$, where $*$ denotes convolution. A problem is that the bit sequence a_k is generally not known in the receiver. However, an estimate \hat{a}_k may be used to calculate d_k . This will work well in case erroneous bit decisions are rare. At the startup of the system, however, many bit errors may occur when the initial equalizer settings are far off. To avoid problems at startup, in many systems initial acquisition is achieved by using a known data sequence a_k , like a preamble. In Fig. 1.13, a switch is present at the detector input that selects the detected bit-sequence \hat{a}_k in case an arbitrary bit sequence is read from the disc, or the known sequence a_k if a preamble is read. This switch enables a reliable initial acquisition: during start-up, adaptation is enabled only during read-out of a preamble.

In many systems, the target response is optimized before system production and as a result it can be considered to be fixed. However an alternative configuration could be to also make the target response adaptive (this configuration is shown in Fig. 1.13). Both w_k and g_k are adapted based on the error $e_k = y_k - d_k$. To avoid interaction between the adaptation of w_k and g_k , a constraint needs to be imposed on the equalizer impulse response and/or on the target response. In many systems a monic constraint is imposed on the target response, i.e. $g_0 = 1$. This constraint causes the equalizer

to approximate a canonical whitened matched filter, which is the optimum prefilter for the VD [106]. However this optimum filter is only approximated if the target response has sufficient length, which causes the full-fledged VD to have a rather high complexity. The VD complexity can be reduced by either using sequence feedback or by constraining the length of the target response (causing again a loss with respect to the whitened matched filter) [3].

The adaptation of the equalizer coefficients can be based on the LMS technique or on the Zero-Forcing (ZF) technique. The LMS technique is based on the Minimum Mean Square Error (MMSE) criterion, where the objective is to minimize $E[e_k^2]$. This results in the following update rule for equalizer

$$w_k(i) = w_{k-1}(i) + 2\mu e_k r_{k-i} \quad \text{for } i \in \{-W, -W+1, \dots, W\}, \quad (1.6)$$

where $2W+1$ is the overall equalizer length and μ is an adaptation constant which determines the bandwidth of the adaptation loop. The choice of μ is always a trade-off between tracking speed and gradient noise, where gradient noise is defined as variations around the steady-state coefficient value caused by the channel noise u_k . The LMS technique will converge to the MMSE solution, yielding a compromise between minimizing RISI and noise enhancement at the detector input.

In the ZF technique, RISI is forced to be zero at the detector input. This means that the detector input will be a noisy version of d_k (the ideal detector input). In practice, there are infinitely many ISI components (related to past, present and future bits) and as a result not all RISI components will be zero but the overall amount of RISI will be minimized across a predefined span. The update rule for the ZF adaptation of the equalizer is

$$w_k(i) = w_{k-1}(i) + 2\mu e_k d_{k-i} \quad \text{for } i \in \{-W, -W+1, \dots, W\}. \quad (1.7)$$

This update rule has the disadvantage that the loop will become ill-conditioned in case the channel has frequencies with zero transfer, i.e. spectral nulls, while the target response is non-zero at one or more of these frequencies. In an attempt to force the RISI to zero for these frequencies, the equalizer coefficients will blow up to large values, resulting in a lot of noise enhancement. A possible solution for this is the introduction of tap leakage. The update rule is changed to

$$w_k(i) = (1 - \alpha)w_{k-1}(i) + 2\mu e_k d_{k-i} \quad \text{for } i \in \{-W, -W+1, \dots, W\}, \quad (1.8)$$

where α is the so-called leakage factor. When the error is zero, the equalizer coefficients slowly leak to zero. This means that for frequencies where the channel has spectral nulls the equalizer transfer will not blow up to large values.

For adaptation of the target response, the LMS and the ZF criteria are equivalent and yield the following update rule

$$g_k(i) = g_{k-1}(i) + 2\mu e_k \hat{a}_{k-i} \quad \text{for } i \in \{0, \dots, G-1\}, \quad (1.9)$$

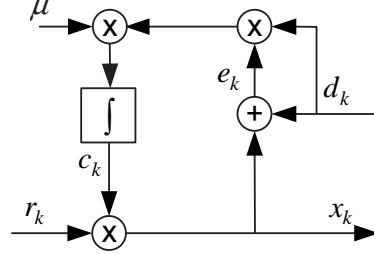


Figure 1.14: Zero-forcing AGC loop.

where G is the target length and μ is again an appropriate adaptation constant. As already stated, to avoid interaction between the adaptation of w_k and g_k a constraint needs to be imposed on the target response: the monic constraint ($g_0 = 1$) [67], the unit energy constraint ($\sum_k g_k = 1$) [65] or some other constraint [92].

Note that the LMS and ZF techniques try to minimize an objective that is not directly linked to the Bit-Error Rate (BER) of the system. Recently, adaptation methods have been proposed that try to directly optimize the BER by using the sequence amplitude modulation criterion. This allows a further performance improvement compared to the LMS or ZF methods. These new techniques have been developed for equalizer adaptation [107, 108] and for target response adaptation [109].

Automatic Gain Control

In digital storage systems, received signal levels may vary widely. These variations can be due to gain variations in the analog front-end of the system or due to variations in the optical path (e.g. dust, fingerprints). Although the equalizer accomplishes gain control, a dedicated adaptation loop is commonly used that accomplishes Automatic Gain Control (AGC) [3]. The most important reason why gain variations are handled by a separate adaptation loop is that normally the ADC is just behind the AGC and in this way the ADC needs far fewer bits. Another reason why gain variations are handled by a separate adaptation loop is that in general fast gain-variations cannot be handled by the equalizer adaptation because it is only adapted slowly (due to the large number of coefficients). The AGC loop can be fast because only a single parameter has to be adapted, namely the gain parameter c .

The AGC loop can be based on the ZF or the LMS criterion. Here the focus will be on the ZF criterion. The LMS loop is similar [3]. Fig. 1.14 depicts a ZF AGC loop. The AGC input r_k is multiplied by the gain parameter c_k to produce the output x_k . The crosscorrelation of e_k and d_k serves as a basis for the adaptation of the gain parameter c_k . The update rule is given by

$$c_k = c_{k-1} + \mu e_k d_k. \quad (1.10)$$

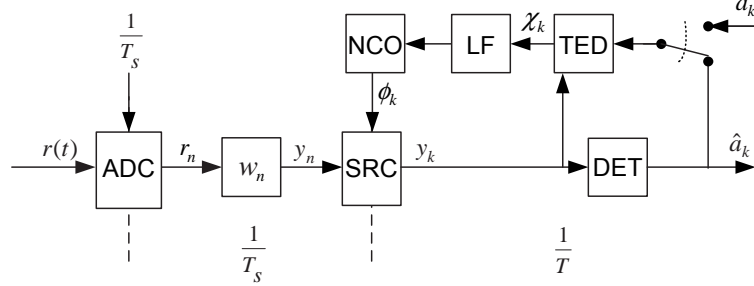


Figure 1.15: Receiver model with an inductive data-aided timing recovery circuit.

An analysis of the dynamic behavior of this adaptation loop can be found in Section 6.3 and will not be given in this section.

To avoid interaction between the equalizer adaptation and the AGC, commonly a constraint is imposed on the equalizer coefficients: limit the overall energy of the equalizer coefficients or fix the central coefficient to a fixed value (e.g. $w_0 = 1$).

Timing Recovery

Up to this point, we have assumed that the receiver operates at exactly the baud rate $1/T$ (see Fig. 1.10). In practice this will not be true and a local resynchronisation in terms of frequency and phase must be established in the receiver. The problem of timing recovery is concerned with the determination of the ideal time instants at which the samples should be taken. Clearly, timing recovery is very important since errors in the choice of the sampling instants will directly be translated in to poor detection performance [110]. This poor performance is because the appropriate ISI structure (according to the target response) at the detector input is only guaranteed at the correct sampling instants. In literature many different types of timing recovery schemes have been proposed [3]. One possible way to accomplish timing recovery in digital storage systems is to use a digital Phase Locked Loop (PLL) [3]. In this work a digital PLL will be used that employs an inductive data-aided timing recovery circuit based on a ZF adaptation criterion [47]. Inductive means that the timing information is extracted behind the resynchronisation, data-aided refers to the knowledge of the transmitted bits and ZF refers to the way the timing information is extracted.

In Fig. 1.15 a receiver model is depicted in which a digital PLL based on inductive timing recovery is used to accomplish timing recovery. The analog received signal $r(t)$ is sampled by the Analog to Digital Converter (ADC) which is controlled by a free-running clock signal with frequency $1/T_s$ which is close to baud rate $1/T$. The digital signal $r_n = r(nT_s)$ is used as input of the asynchronous equalizer (a discussion about the operation and adaptation of an asynchronous equalizer can be found in [111]), which shapes the channel ISI into an ISI structure defined by the target

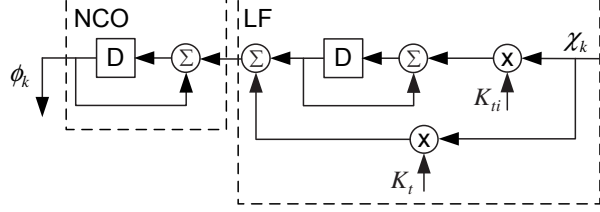


Figure 1.16: Phase-domain model of the loop filter (LF) and the numerically controlled oscillator (NCO).

response. Subsequently, a Sample Rate Converter (SRC) is used to resample the equalized signal y_n at the baud rate $1/T$. The sampling phase of the SRC is determined by a phase estimate ϕ_k . The SRC together with the Timing Error Detector (TED), the Loop Filter (LF) and the Numerically Controlled Oscillator (NCO) form the PLL that tracks the clock phase and frequency of the binary data. The TED serves to generate a timing error estimate χ_k which is an indication of the sampling phase error at the detector input. The LF serves to filter the timing error estimate to produce an instantaneous frequency value. This value is used by the NCO to produce the sampling phase ϕ_k which is used by the SRC. The NCO is implemented as an integrator.

As already stated above, the TED is of the data-aided type, i.e. it uses the data sequence a_k (either as a known preamble sequence or as bit decisions \hat{a}_k). To ensure proper acquisition, the preamble is used as input of the TED. When the end of the preamble is reached the PLL has sufficiently converged such that reliable bit decisions \hat{a}_k are produced. Beyond that point these bit decisions are used as input of the TED. Besides the data sequence also the detector input is used in the TED. The output of the TED can be expressed as

$$\chi_k = e_k(\mathbf{q} * \mathbf{a})_k = (y_k - d_k)(\mathbf{q} * \mathbf{a})_k, \quad (1.11)$$

where e_k is the error signal, d_k is the desired detector input signal (see Fig. 1.13) and \mathbf{q} is the sampled derivative of the target response \mathbf{g} . The signal $(\mathbf{q} * \mathbf{a})_k$ can easily be approximated by convolving d_k with the $(1 - D^2)$ operator [112]. By using this operator, also e_k needs to be delayed by one bit interval.

Fig. 1.16 depicts a phase-domain model of the LF and the NCO. In a phase-domain model the operation of the different blocks is described in terms of the sampling phase. The input of the LF is the phase error estimate χ_k generated by the TED and the output of the model is the sampling phase estimate ϕ_k that is used in the SRC. Because the timing recovery circuit needs to be able to handle frequency offsets, the LF is a second-order filter. The LF has a proportional branch with weighing factor K_t that ensures an average steady-state phase error equal to zero, and an integrating branch with weighing factor K_{ti} , which lets the loop converge towards the correct

frequency value. The dynamic behavior of the loop can be expressed as a function of the ideal sampling phase θ_k and the phase ϕ_k produced by the timing recovery loop. The transfer function of the complete loop can be expressed as

$$G_\phi(z) = \frac{\Phi(z)}{\Theta(z)} = \frac{K_t(z-1) + K_{ti}}{(z-1)^2 + K_t(z-1) + K_{ti}}. \quad (1.12)$$

This is a second-order transfer function which may be characterized by a normalized natural frequency $\omega_n T = \sqrt{K_{ti}}$ and a damping factor $\zeta = K_t/(2\sqrt{K_{ti}})$ [3].

Because at low SNRs accurate timing recovery is of crucial importance for system reliability, many improvements to the timing recovery circuit described above have been proposed in recent years. One of these improvements is a modification made to the TED to account for the signal-dependent nature of media noise [113–115]. Furthermore, the joint estimation of data and ideal sampling instants has received a lot of attention: either a direct optimization [116] or in the form of a per-survivor timing recovery [117–120].

1.5 Motivation and Content of this Thesis

In Chapter 2 the TwoDOS concept will be explained in detail. The main purpose of the chapter is to explain the basic techniques that are needed for the realization of a 2-D storage system. The characteristics and the characterization of the TwoDOS system will be discussed in Chapter 3. There the strengths and weaknesses of the TwoDOS system will be gauged based on experimental results. The conclusions that are drawn from that chapter will serve as a basis of motivation for the subsequent chapters.

A first conclusion is the fact that media noise (i.e. variations in bit size and bit position) is the dominant source of noise in the experimental TwoDOS system. In Chapter 4 a novel method for the accurate characterization of media noise is presented. The knowledge of the magnitude of the different noise types in a digital storage system is prerequisite for taking measures against the negative impact of the noise types on system performance. In Chapter 4 a real-time characterization algorithm is proposed that is able to efficiently decompose the different noise sources based on the information available within the system. The proposed algorithm makes use of the data-dependency of the media noise to distinguish between the different noise sources. The algorithm is simple and as a result only a very limited amount of complexity is required to implement it in storage systems as an easy “add-on” to read channels ICs.

A second conclusion of Chapter 3 concerns the existence of large amounts of ISI in the experimental TwoDOS system. The ISI can be divided into two categories: nonlinear ISI caused by bits within a small span of the current bit and a substantial amount of linear ISI caused by bits that are located further away. The nonlinear

ISI can be taken into account in the detector by using a pattern-dependent reference value. If also the linear ISI would have to be taken into account by the detector this would result in a very high complexity. A simple yet highly efficient alternative solution is the cancellation of this linear ISI. In Chapter 5, ISI cancellation in 2-D storage systems is discussed. In this chapter, a theory is developed for the error rate of receivers that use tentative decisions to cancel ISI. Furthermore, precise conditions are formulated under which such ISI cancellation can be applied effectively. Experimental results for the TwoDOS system show that a substantial performance improvement may be obtained by applying ISI cancellation.

A third conclusion of Chapter 3 is the fact that some channel parameters exhibit rapid variations. Because system performance is very sensitive to variations of storage channel parameters, accurate tracking of rapid variations is of major importance in the TwoDOS receiver. Tracking capabilities are limited by latencies in the adaptation loops. These latencies are largely governed by delays of the bit-detector. In 2-D storage systems data are packaged in a group of adjacent tracks, and for some of the tracks the detection delays can increase dramatically with respect to 1-D systems. As a result the effective latencies in the adaptation loops preclude the tracking of rapid variations and really limit the performance of the system. In Chapter 6 a scheme is proposed that overcomes this problem and that can be used for timing recovery, automatic gain control and other adaptive circuits. Rapid variations for all the tracks are tracked using control information from tracks for which detector latency is smallest. This works properly if rapid variations are common across the tracks as is the case, for example, for the TwoDOS system. Experimental results for TwoDOS confirm that the scheme yields improved performance with respect to conventional adaptation schemes. Finally in Chapter 7 the conclusions are given together with recommendations for future research.

1.6 List of publications and Patents

This section contains a list of publications related to the contents of this thesis.

1.6.1 Papers

- [P-1] S. Van Beneden, J. Riani, J.W.M. Bergmans, A.H.J. Immink, "Minimum Latency Tracking of Rapid Variations in Two-Dimensional Storage Systems," *IEEE Trans. Magn.*, vol. 43, nr. 1, Part 1, pp. 67 - 78, Jan. 2007.
- [P-2] S. Van Beneden, J. Riani, J.W.M. Bergmans, "Adaptive Decomposition of Noise Sources in Digital Recording Systems with Media Noise," *IEEE Trans. Magn.*, vol. 43, nr. 2, Part 1, pp. 561 - 571, Feb. 2007.

-
- [P-3] S. Van Beneden, J.W.M. Bergmans, J. Riani, A.H.J. Immink, "Cancellation of Linear Intersymbol Interference for Two-Dimensional Storage Systems," *IEEE Trans. Magn.*, vol. 42, nr. 8, pp. 2096 - 2106, Aug. 2006.
 - [P-4] S. Van Beneden, J.W.M. Bergmans, J. Riani, M. Ciacci, A. Nowbakht, A.H.J. Immink, "Minimum Loop-Delay Adaptation and Timing Recovery for Two-Dimensional Optical Storage," *Proc. 25-th Symp. Inf. Th. Benelux*, ISBN 90-71048-20-9, pp. 257-264, 2004.
 - [P-5] S. Van Beneden, J. Riani, J.W.M. Bergmans, "Cancellation of Linear Intersymbol Interference for Two-Dimensional Storage Systems," 2006 IEEE International Conference on Communications (ICC 2006), Istanbul, Turkey, Vol. 6, pp. 3173-3178, 11-15 June 2006.
 - [P-6] S. Van Beneden, J. Riani, J.W.M. Bergmans, "Characterization of Two-Dimensional Storage Systems," *Acoustics, Speech, and Signal Processing*, 2006. *ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, Toulouse, France, vol. 4, pp. IV-801 - IV804, 14-19 May, 2006.
 - [P-7] S. Van Beneden, J. Riani, J.W.M. Bergmans, "Cross-Talk and ISI cancellation for Two-Dimensional Storage Systems", *Proceedings of SPS-DARTS 2006*, 2006, pp.61-64.
 - [P-8] S. Van Beneden, J. Riani, J.W.M. Bergmans, "Data-Dependent Noise Estimation in Digital Recording Systems," *Global Telecommunications Conference, 2006. GLOBECOM '06. IEEE*, San Francisco, USA, GEN04-5, pp. 1-5, 27 Nov. - 1 Dec., 2006.
 - [P-9] D.M. Bruls, A.H.J. Immink, A.M. van der Lee, W.M.J. Coene, J. Riani, S. Van Beneden, M. Ciacci, J.W.M. Bergmans, M. Furuki, "Two-Dimensional Optical Storage: High-Speed Read-Out of a 50 GByte Single-Layer Optical Disc with a 2D Format Using $\lambda=405$ nm and NA=0.85," *Jpn. J. Appl. Phys.*, Vol. 44, No. 5B, pp. 3547-3553, 2005.
 - [P-10] J. Riani, S. Van Beneden, J.W.M. Bergmans, A.H.J. Immink, "Near Minimum BER Equalizer Adaptation for PRML systems," *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE*, Saint Louis, USA, vol. 4, pp. 2123-2128, 28 Nov. - 2 Dec., 2005.
 - [P-11] J. Riani, A.H.J. Immink, J.W.M. Bergmans, S. Van Beneden, "Near Minimum BER, All Adaptive Partial Response Equalization for High Density Recording Systems," *Global Telecommunications Conference, 2006. GLOBECOM '06. IEEE*, San Francisco, USA, GEN04-6, pp. 1-6, 27 Nov. - 1 Dec., 2006.

- [P-12] J. Riani, S. Van Beneden, J.W.M. Bergmans, A.H.J. Immink, "Near Minimum Bit-Error Rate Equalizer Adaptation for PRML Systems," *Submitted for publication to IEEE Trans. Magn.*, 2005.
- [P-13] J. Riani, S. Van Beneden, J.W.M. Bergmans, A.H.J. Immink, "Data-Aided Timing Recovery for Recording Channels with Data-Dependent Noise," *IEEE Trans. Magn.*, vol. 42, nr. 11, pp. 3752 - 3759, Nov. 2006.
- [P-14] J. Riani, S. Van Beneden, J.W.M. Bergmans, "Data-Aided Timing Recovery for Recording Channels with Data-Dependent Noise," *European Signal Processing Conference, Eusipco*, 4 Sept.-8 Sept. 2005.
- [P-15] A.H.J. Immink, J.Riani, S. Van Beneden, J.W.M. Bergmans, M. Ciacci, A. Nowbakht Irani, W.M.J. Coene, A.M. van der Lee, D. Bruls, "Adaptation and Timing Recovery for Two-Dimensional Optical Storage," *Optical Data Storage, Proceedings of SPIE*, vol. 5380, pp. 90-104, 2004.
- [P-16] A. Nowbakht Irani, J.W.M. Bergmans, S. Van Beneden, W.M.J. Coene, M. Ciacci, A.H.J. Immink, J. Riani, "Improved Correlation Receiver for Frame Synchronization," *Proceedings of the 25-th Symp. Inf. Th. Benelux. Ed. R. Pellikaan*, pp. 65-72, 2004.
- [P-17] A.H.J. Immink, W.M.J. Coene, A.M. van der Lee, C. Busch, A.P. Hekstra, J.W.M. Bergmans, J. Riani, S. Van Beneden, T. Conway, "Signal Processing and Coding for Two-Dimensional Optical Storage," *Global Telecommunications Conference, GLOBECOM, IEEE*, vol.7, pp. 3904-3908, 2003.
- [P-18] J. Riani, J.W.M. Bergmans, S. Van Beneden, W.M.J. Coene, A.H.J. Immink, "Equalization and Target Response Optimization for High Density Two-Dimensional Optical Storage," *Proc. 24-th Symp. Inf. Th. Benelux, ed. L. Tolhuizen, ISBN 90-71048-18-7*, pp. 141-148, 2003.
- [P-19] W.M.J. Coene, D.M. Bruls, A.H.J. Immink, A.M. van der Lee, A.P. Hekstra, J. Riani, S. Van Beneden, M. Ciacci, J.W.M. Bergmans, M. Furuki, "Two-Dimensional Optical Storage," *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on* Vol. 5, 18-23 March 2005, pp. 749 - 752.
- [P-20] J.Riani, S. Van Beneden, J.W.M. Bergmans, "Equalizing Sampling Rate Convertors for Storage Systems," *Acoustics, Speech, and Signal Processing, 2006. ICASSP 2006 Proceedings. IEEE International Conference on* , Toulouse, France, vol. 3, pp. III392 - III395, 14-19 May, 2006.

-
- [P-21] D.M. Bruls, A.H.J. Immink, A.M. van der Lee, W.M.J. Coene, J. Riani, S. Van Beneden, M. Ciacci, J.W.M. Bergmans, M. Furuki, “Two-dimensional Optical Storage: High-speed read-out of a 50 GByte single-layer optical disc with a 2D format using $\lambda = 405\text{nm}$ and $\text{NA}=0.85$ ”, *International Symposium on Optical Memory 2004*, 13 Oct. 2004.

1.6.2 Patents

- [PA-1] A.H.J. Immink, S.J.L. Van Beneden, J.W.M. Bergmans, J. Riani, *Feedback control loop for bit detection in an N-dimensional data block*, PHNL 040117, filing date: 20-01-2005.
- [PA-2] J. Riani, S.J.L. Van Beneden, J.W.M. Bergmans, A.H.J. Immink, *Data Dependent Timing Recovery*, PHNL 000669, filing date: 13-7-2005.
- [PA-3] J. Riani, S. Van Beneden, J.W.M. Bergmans, A.H.J. Immink, *Near-minimum bit-error rate equalizer adaptation*, PH 002710 EPP, filing date: 18-11-2005.

Chapter 2

Two-Dimensional Optical Data Storage

“The further growth of today’s information society will require larger storage capacities and faster access to ever growing data bases and digital content. The classical roadmap for optical storage seems to come to an end, both in capacity and in data rate. We propose to develop a new and challenging concept for optical storage, in which the information written on the disc fundamentally has a two-dimensional character. We call this new concept TwoDOS (Two-Dimensional Optical Storage). The aim is to generate key technologies that realize an increase over the 3rd generation of optical storage with a factor of two in data density and a factor of 10 in data rate for an optical read-only system, based on innovative two-dimensional channel coding and advanced signal processing, in combination with a read channel consisting of a multi-spot light path realizing a parallel read-out. TwoDOS will achieve a capacity of at least 50 GB for a 12 cm disc, with a data rate of at least 300 Mb/s”

This is an extract of the project definition of the TwoDOS project [121]. In this chapter we will give an introduction to the TwoDOS system. Section 2.1 describes the format of the disc. Section 2.2 explains how these discs can be read using a linear array of laser spots. Section 2.3 gives a detailed analysis of the two-dimensional (2-D) channel and introduces models that describe the complete 2-D system. Subsequently, the basic principles to recover the 2-D bit patterns from the replay signals are discussed in Section 2.4. These principles will be expanded further in Section 2.5, where the data receiver is introduced. In this section the signal processing path from Photo Detector IC (PDIC) up until the bit-detector will be discussed. The bit-detector itself will be discussed in Section 2.6.

2.1 Two-Dimensional Disc Format

The continually increasing demand for data rate and capacity has urged the development of new storage techniques. The use of a two-dimensional lattice to store the information on a disc is one of these techniques. One could argue that all disc-based storage is actually based on a 2-D format. Conventionally, the data is considered to be 1-D in the sense that the information is stored as a 1-D sequence of bits that is

arranged along a spiral track on the disc. Successive rotations of the spiral are considered as different tracks on the disc that evolve in the tangential direction of the disc. The distance between adjacent tracks is chosen such that during read-out the different tracks simply do not interfere. Furthermore there is no physical relationship between the data on the adjacent tracks.

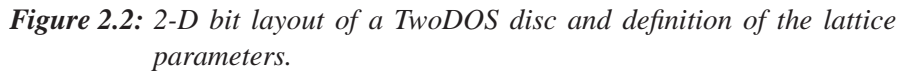
In a 2-D storage system this is different: the adjacent tracks are aligned in phase with respect to each other and the intertrack distance is reduced significantly. Furthermore the data of the different tracks belong to a single entity that should be read simultaneously. The reduction in intertrack distance and the simultaneous read-out increases respectively the capacity and the data rate of the system. In this section the 2-D disc format of TwoDOS will be discussed. First the lattice characteristics will be discussed, followed by the modulation code and the test format.

2.1.1 2-D Lattice Characteristics

In the TwoDOS system, the bits are stored in a metaspiral consisting of a number of bit-tracks stacked upon each other with a fixed phase relation in the radial direction. In this way the bits form a 2-D lattice. A 2-D close packed hexagonal ordering was chosen because it has a 15% higher packing density than the square lattice as used in [122–125]. On the other hand, in a hexagonal lattice 6 direct neighbors will cause the bulk of the Intersymbol Interference (ISI), while in a square lattice only 4 direct neighbors are present.

In Fig. 2.1 the broad spiral of the TwoDOS system is shown. The broad spiral contains L bit-tracks (in the figure, $L = 7$), stacked upon each other to form a hexagonal structure. Adjacent rotations of the broad spiral are separated by a so-called guard band with a width of one bit-track. This guard band serves the following purposes:

- A primary function of the guard band is to introduce a discontinuity in the phase relationship between adjacent rotations of the broad spiral to ensure a constant areal density at different radial positions on the disc.
- The guard band can serve as a starting point for a 2-D detection process. The a-priori known content of the guard band (i.e. an empty bit-track) can be used as side information in a first detection step (see Section 2.6 for more information about 2-D detection).
- The guard band can be used as a basis for radial tracking [18]. A tracking signal for a 2-D system can be generated by comparing the averaged amplitude signals of the two boundary tracks of the broad spiral. If a radial offset is present, one of the spots of the boundary tracks will be scanning an empty guard band, resulting in a lower average amplitude for that track.


$$\mathbf{G} = \frac{a_H}{2} \begin{bmatrix} 2 & 1 \\ 0 & \sqrt{3} \end{bmatrix}. \quad (2.1)$$

- the matrix \mathbf{A}_k contains all bits defined by the following coordinates (u, v) with $v \in \{0, L-1\}$ and $u \in \{k-I, \dots, k+I\}$ where I is a predefined window length;
- the set \mathbf{B}_k^l contains all bits that are within the first shell of the bit with coordinates (k, l) ;
- the set \mathbf{C}_k^l contains all bits of \mathbf{A}_k that are not within the first shell of a_k^l , i.e. that are not part of \mathbf{B}_k^l .

In a TwoDOS disc the data are stored as lands or pits in the hexagonal lattice. The circular pit-holes are embossed into the disc by means of a so-called stamper in a mass production process. The stamper can be manufactured using one of the following techniques: Laser Beam Recording (LBR), Liquid Immersion Mastering (LIM) or Electron-Beam Recording (EBR). A stamper is created by illuminating a photo-sensitive resist layer, which is spin-coated on top of a glass substrate, using a focused laser beam (LBR and LIM) or an electron-beam (EBR). The photo resist is etched away completely at the positions at which it was illuminated. Subsequently a nickel alloy is deposited on the top of the resist layer by means of a sputtering process. The metallized glass master is electroplated to form a thick nickel, so-called father stamper, which can be removed from the glass substrate.

The father stamper is used to replicate discs. Several replication methods exist to produce many identical reproductions of the father stamper. In the TwoDOS project, the glass-2P (Photo-Polymerization) process is used. In this process, a liquid lacquer is deposited on top of the glass substrate on which the stamper is pressed against. The lacquer is cured by ultra-violet (UV) light through the transparente substrate. After curing the newly formed disc can be released from the stamper and an aluminum and additional protective layer can be deposited on top of the lacquer.

Laser Beam Recording

Generally the spot in a mastering machine should be smaller than the spot in the read-out system. Because the spot size is proportional to $\phi_{Airy} = \frac{\lambda}{2NA}$, it is not possible for LBR mastering to achieve the very small pit size that is necessary for a 35 GB or a 50 GB disc capacity. For this reason, experiments are performed with a scaled NA equal to 0.57. For example, for the hexagonal lattice with $a_H = 248$ nm the disc capacity amounts only to 15.5 GB. However when read out with a lens with $NA = 0.57$, this is equivalent to a capacity of 35 GB at $NA = 0.85$.

Liquid Immersion Mastering

As already stated above, a conventional far-field optical system with Blu-Ray parameters ($\lambda = 405$ nm and $NA = 0.85$) is not sufficient to record a TwoDOS disc. As a result a system with a shorter wavelength or a higher NA should be used. A possible solution to achieve this is to use Near-Field mastering (see Section 1.2.3). Another solution is called liquid immersion mastering (LIM) [127, 128]. Here a far-field objective lens is used ($NA=0.9$) and an immersion liquid is applied between the lens and the rotating disc. The immersion liquid, which has a refractive index considerably higher than 1, allows light waves at angles above the critical angle to pass through the liquid film without the problem of total internal reflection. This gives a diffraction limited spot corresponding to a NA of approximately 1.2.

For the LIM discs, a run-length-limited (RLL) modulation code is used similar to the 17PP (parity preserving) code that is used in the Blu-ray Disc format. The shortest mark on the disc has a length of 2 channel bits, while the longest mark is equal to 8 channel bits. Fig. 2.3 depicts a Scanning Electron Microscope (SEM)

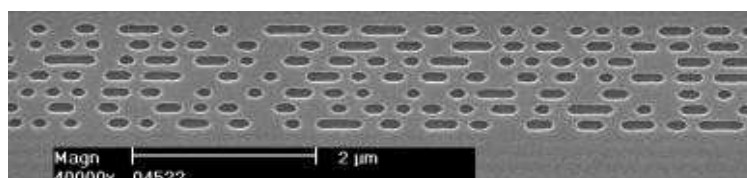


Figure 2.3: SEM image of a LIM disc.

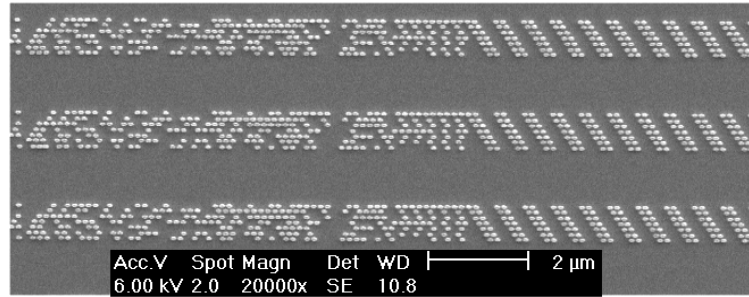


Figure 2.4: SEM image of an EBR disc.

image of a LIM disc. In the TwoDOS project, several LIM discs have been produced with capacities ranging from 37.5 GB to 70 GB.

Electron-Beam Recording

For the hexagonal format even smaller pits are required and also LIM may not be sufficient anymore. A disruptive technology to master even smaller features is to use an electron-beam recorder (EBR) [129]. In the EBR process, an electron beam with a very small wavelength can be obtained by accelerating electrons with a very high voltage. The effective spot diameter is however not directly determined by the wavelength. It is rather limited by spherical aberrations of the objective lens in the electron-optical setup, and due to the finite energy spread of the high-energy electrons. Within the resist the effective spot size is further increased by forward and backward scattering of the electrons in the resist. Therefore, in regions with a large number of pits there is a considerable amount of background illumination and the pits tend to be somewhat larger. This effect is known as the proximity effect and it causes a pattern-dependent variation in the signal level (see Section 2.3.2). In this thesis, a disc with lattice characteristics $a_H = 138$ nm (50 GB) and $b_H = 102$ nm will be mainly used to produce experimental results. In Fig. 2.4, a SEM image of this type of EBR disc is shown. In this image three rotations of the broad spiral are shown which are separated by approximately 10 empty bit-tracks.

2.1.3 Test Format

The data that is written on the disc is in principle not known at the receiving end of the system (i.e. the data receiver). In a data receiver (see Section 1.4 for 1-D receivers and Section 2.5 for the TwoDOS receiver), several adaptation loops are present that use the knowledge of the data to base their operation on. For evaluation purposes, it is convenient to use the receiver in a data-aided (DA) operation mode, which means the actual data is assumed to be known and is used to generate the error signals for the adaptation loops. The final receiver, however, should operate in a decision-directed

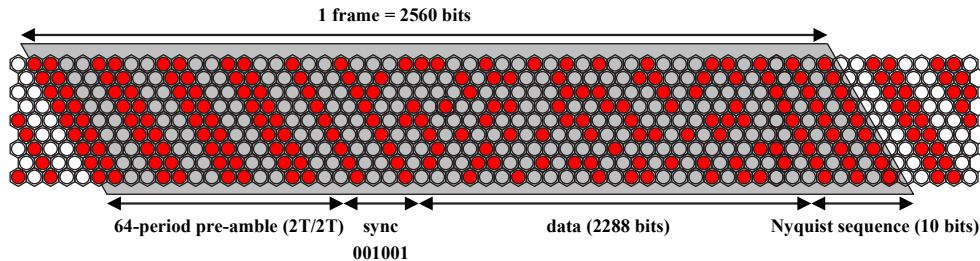


Figure 2.5: Schematic representation of the frame-based test format of the TwoDOS discs: preamble structure, synchronization pattern, data pattern and Nyquist sequence.

(DD) mode, which means that for the generation of the error signals the output of the bit-detector should be used. In DD mode an acquisition problem occurs: the adaptation loops need correct bit decisions from the bit-detector in order to converge properly, while the bit-detector needs the adaptation loops to be converged in order to produce reliable bit decisions. To overcome this deadlock, the data on the disc is organized in frames. Fig. 2.5 depicts a schematic representation of the frame-based format of the TwoDOS disc.

In every frame the user data is preceded by a sequence of a-priori known data symbols called the preamble. The preamble sequence is chosen such that it contains the adequate amount of information for the critical adaptation loops: for the TwoDOS system the critical adaptation loop is the timing recovery circuit. A periodic pattern of two pits and two lands (a so-called $2T$ -pattern) is chosen as preamble sequence for the timing recovery loop, the DC adaptation loop and the gain adaptation loop. This $2T$ -pattern is repeated 64 times such that the adaptation loops have sufficient time to converge. At that point the bit-detector is able to produce reliable bit decisions. As a result after the preamble has ended the receiver can operate reliably in a DD mode. After the preamble a sync-pattern is inserted that indicates the beginning of the user data in the frame. The sync-pattern is chosen such that it has a large detection distance with respect to the periodic pattern of the preamble, i.e. that it is sufficiently different from the preamble [130]. The total length of the frame is 2560 bits.

2.2 Read Out of a TwoDOS disc

In the previous section, the format of the TwoDOS disc is discussed. In this section we will describe the optical system that is used to read out the TwoDOS discs. To achieve an increase in data rate with respect to Blu-Ray disc, the different tracks of the broad spiral should be read out simultaneously (i.e. in parallel). The simultaneous read-out is accomplished by the use of an array of L laser spots arranged such that each spot is centered on one of the bit-tracks within the broad spiral.

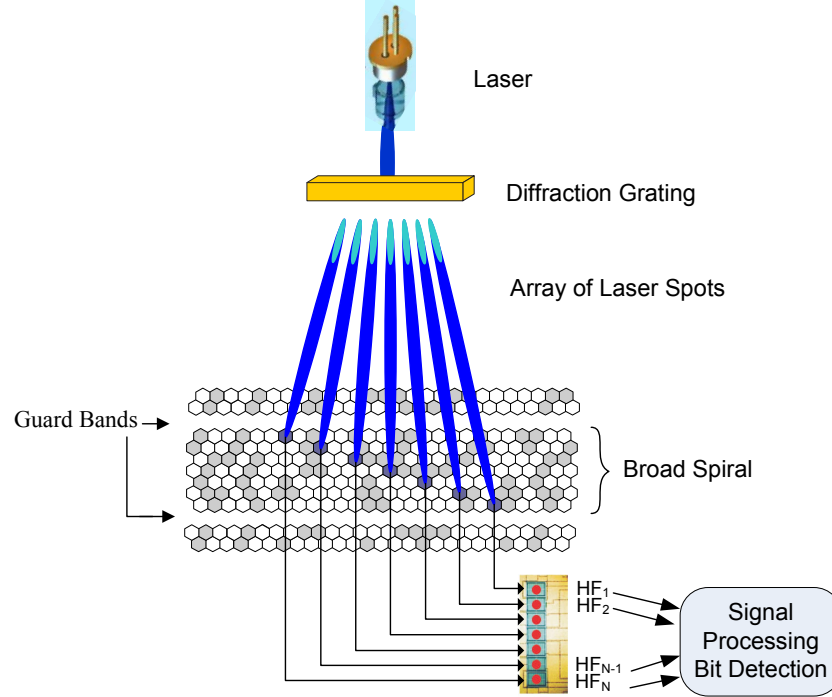


Figure 2.6: Simplified representation of the read-out principle of a 2D disc.

A schematic representation of a solution that achieves this parallel read-out is depicted in Fig. 2.6. In this solution, the array of laser spots is generated from a single laser diode by using a diffraction grating. Such a phase grating consists of a periodic pattern [18]. The advantage of using a diffraction grating is that only a single laser diode is needed. This argument reveals also the main disadvantage: multi-spot writing is not possible. After the diffraction grating the array of laser beams passes to some optical components (not shown in the figure): a collimator, a beam shaper, a telescope and an objective lens. At this point every laser spot is focussed on a different track of the broad spiral on the disc. Subsequently, the reflected light is focussed on a photo detector IC (PDIC). A multi-spot PDIC is used to generate a so-called high-frequency (HF) signal for every bit-track. The analog HF signals are used as input of the data receiver (which consists of signal processing blocks and a bit-detector, see Section 2.5 for further explanation).

Due to the fact that the different spots originate from the same source and due to the coherent nature of the laser light, interference will occur in areas where the different spots overlap. The different spots in the array are diffraction limited and have the Airy intensity profile [17, 131]. To limit the amount of interference, the minimal distance between two consecutive laser spots on the disc has been chosen such that the local maximum of the second Airy rings of adjacent spots are aligned, as is

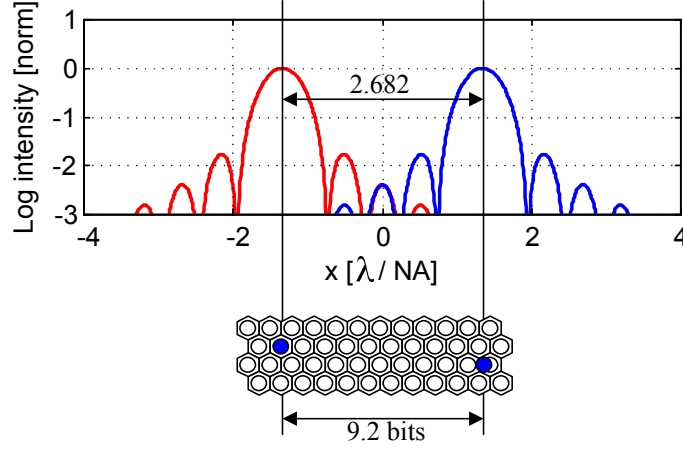


Figure 2.7: Configuration of adjacent optical spots. The airy intensity profiles are shown for two adjacent spots. As an example, a delay of 9.2 bits is shown as it applies for the discs with $a_H = 138$ nm).

shown in Fig. 2.7. The minimum distance between two spots can be calculated to be $2.682 \lambda / \text{NA}$. For the 50 GB disc ($a_H = 138$ nm), the time delay between two successive tracks can be calculated to be 9.2 bit periods [18]. A dedicated signal processing block is present in the data receiver that compensates for this delay (namely the delay compensation block, see Section 2.5).

2.3 Optical Storage Channel Models

The optical detection technique used in the TwoDOS read-out process is the so-called central aperture detection as is described in [17]: all photons that are reflected or diffracted from the optical disc into directions that are comprised within the (so-called central) aperture at the exit pupil of the objective lens, constitute the physical signal which is detected by the PDIC. The resulting HF signal is further processed in the data receiver. To analyse the operation of the data receiver it is important to have a model of the optical channel (see Section 1.4). In this section several channel models will be discussed. In general, the HF signal can be divided into two parts: first, a part that arises from the bits on the disc, more precisely the desired signal component and the linear and nonlinear ISI, and second, a random part, namely the noise in the system.

2.3.1 Intersymbol Interference Model

Increases in capacity and in data rate with respect to Blu-Ray disc are the key benefits of 2-D optical storage. These increases are enabled in part by applying enhanced

signal processing techniques while using the same optical system as in 1-D (i.e. the same laser wavelength and the same objective lens). The main difference with a 1-D system is the reduced intertrack distance, while the optical spot is basically identical. As a result, from a physical point of view, there is no difference between the 1-D and the 2-D optical channel: both have a 2-D optical laser spot and hence a 2-D impulse response. For 1-D systems, the 2-D optical channel is simply replaced by its 1-D equivalent because the intertrack distance is so large that only the track under consideration causes ISI. In this section, several 2-D channel models are discussed, starting from the simple linear model to the more sophisticated (and more accurate) scalar diffraction model.

Linear Braat-Hopkins Model

Simulation tools for signal waveforms in 1-D optical storage (CD, DVD and BD, the Blu-Ray Disc format) often use a linear model for the optical channel. This linear model is characterized by the modulation transfer function (MTF) as derived in the Braat-Hopkins formalism for the read-out of optical discs [17]. The 1-D Braat-Hopkins MTF can easily be extended to its 2-D equivalent. This 2-D MTF, denoted by $H_{2-D}(f, \theta)$, is given by a circularly symmetric function

$$H_{2-D}(f, \theta) = \begin{cases} \frac{\sin(\pi f T)}{\pi f T} \left(\arccos \left| \frac{f}{f_c} \right| - \frac{f}{f_c} \sqrt{1 - \left(\frac{f}{f_c} \right)^2} \right), & |f| < f_c, \\ 0, & |f| \geq f_c, \end{cases} \quad (2.2)$$

with f the spatial frequency, θ the azimuth in the 2-D spatial frequency plane and f_c the cut-off frequency of the (central-aperture) optical channel. This cut-off frequency is given by $2NA a_H / \lambda$ with a_H the lattice constant, λ the wavelength of the laser light and NA the numerical aperture of the (objective) lens.

The 2-D MTF is characterized by an almost linear roll-off from DC up to f_c and is shown in Fig. 2.8(a) for the optical parameters λ and NA , and the lattice constant a_H used in the TwoDOS project. The MTF has a characteristic shape, reminiscent of a Chinese hat. Note that this simple model does not account for channel non-linearities, which are expected to be quite relevant in TwoDOS.

The channel output signal r_k^l (at time instant k and for track l) is given by

$$r_k^l = \sum_{i,j} h_{2-D}(i, j) a_{k-i}^{l-j} \quad (2.3)$$

where $a_k^l \in \{0, 1\}$ (1 represents a pit and 0 represents a land) and h_{2-D} is the 2-D impulse response function obtained by taking the inverse Fourier transform of H_{2-D} (see Fig. 2.8(b)).

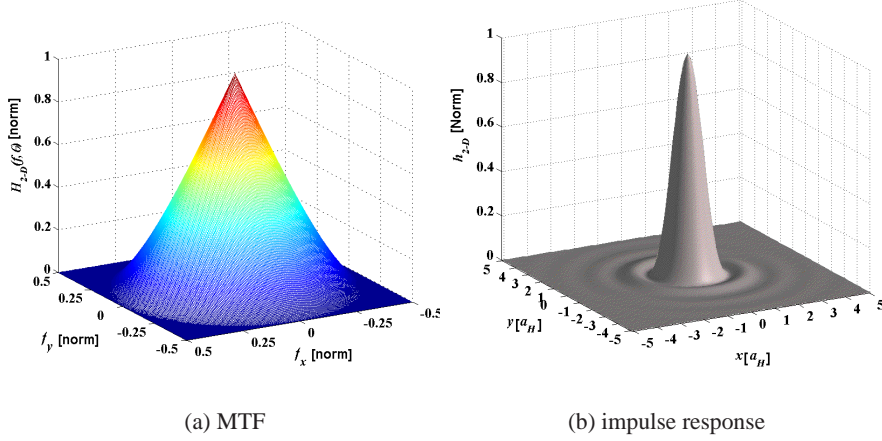


Figure 2.8: Nominal MTF and impulse response function based on the linear Braat-Hopkins model for $\lambda = 405$ nm, $\text{NA} = 0.85$ and $a_H = 138$ nm.

Scalar Diffraction Model

A more sophisticated way of computing the channel output is the approach based on scalar diffraction theory [17, 131]. Fig. 2.9 depicts the principle of the scalar diffraction approach to compute the HF signal. In this approach, the HF signal is computed in a couple of steps.

- First the probe function $P(\mathbf{p})$ is computed. This represents the (complex-valued) spot profile that is focused on the disc at a position determined by the 2-D position vector \mathbf{p} . It is obtained by taking the inverse 2-D Fourier transform of the waveform $P(f)$ at the entrance pupil of the objective lens.
- The disc reflection function $D(\mathbf{p})$ has to be computed next. This function of the disc accounts for the relative phase difference between the light reflected from the mirror at land level and the light reflected from the mirror at the bottom of the pit-area. For TwoDOS, a pit-bit ($a_i = 1$) is mastered as a pit-hole centered in the hexagonal cell that is available for each bit; a land-bit ($a_i = 0$) is characterized by the absence of a pit-hole. The disc reflection function can be written as

$$D(\mathbf{p}) = 1 + \sum_i b_i W(\mathbf{p} - \mathbf{p}_i), \quad (2.4)$$

where $W(\mathbf{p} - \mathbf{p}_i)$ is the pit-window function for the bit at position \mathbf{p}_i which is unity inside the pit-hole and vanishes outside. Furthermore $b_i = a_i(e^{j\phi_p} - 1)$ where ϕ_p is the double phase depth of the pit (i.e. from the top of the pit to the

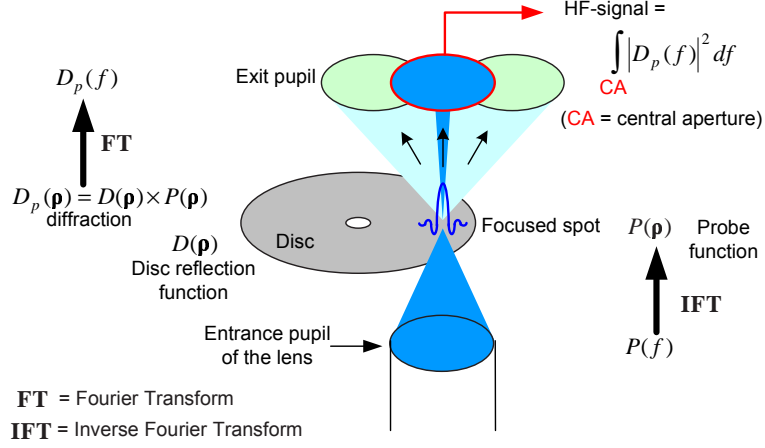


Figure 2.9: Principle of the scalar diffraction approach.

bottom and back) which can be expressed as $\phi_p = 2\pi \frac{2d}{\lambda}$, where d is the depth of the pit.

- The diffraction of the laser spot caused by the pit-structures on the disc is mathematically described by the product of the disc reflection function with the probe function of the laser spot: this product, denoted by $D_p(\mathbf{p})$, yields the reflected and diffracted wavefronts.
- Fourier transformation of $D_p(\mathbf{p})$ from the disc plane towards the plane of the exit pupil of the lens yields the wavefront $D_p(f)$ in the plane of the exit pupil where the physical detection takes place. The HF signal can easily be computed as the integration of the power $|D_p(f)|^2$ over the frequency region of the central aperture that is used for detection.

The approach based on scalar diffraction described above has two main disadvantages: its computational complexity (e.g. two 2-D Fourier transforms) and the fact it is like a black-box, i.e. it does not offer additional insight into the linear and nonlinear characteristics of the optical channel. For these reasons, in [132] a new channel model is proposed which describes the optical channel via linear and bi-linear dependencies on the binary channel bits written on the disc. These dependencies are derived based on the scalar diffraction approach. The complex-valued waveform $D_p(f)$ at the pupil plane will be denoted as $|\psi\rangle$ and can easily be rephrased as

$$|\psi\rangle = |\psi_L\rangle + \sum_i a_i |\psi_i\rangle, \quad (2.5)$$

with a first term $|\psi_L\rangle = \text{FT}[P(\mathbf{p} - \mathbf{p}_p)]$, denoting the waveform in case only land-pits were present (\mathbf{p}_p is the current position on the disc and FT denotes Fourier transform)

and with a second term, yielding the contributions from all pit-bits that are within the area of the laser-spot. In the latter term, a single pit-bit at position \mathbf{p}_i with a pit-area defined by $W(\mathbf{p} - \mathbf{p}_i)$ causes a contribution $|\psi_i\rangle = \text{FT}[P(\mathbf{p} - \mathbf{p}_p)W(\mathbf{p} - \mathbf{p}_i)]$.

The channel output signal when the laser spot is focused at position \mathbf{p}_p , can be written as

$$r(\mathbf{p}_p) = \langle \psi_L | \psi_L \rangle + \sum_i b_i \langle \psi_L | \psi_i \rangle + \sum_i b_i^* \langle \psi_i | \psi_L \rangle + \sum_{i,j} b_j^* b_i \langle \psi_j | \psi_i \rangle, \quad (2.6)$$

where the mathematical operation $\langle \phi | \phi \rangle = \int_{CA} \phi^*(f) \phi(f) df$ is the integration over the central aperture (CA) of the pupil. This channel output signal is dependent on the bits a_i through the coefficients b_i . The first term is a DC-term, the next two terms represent that part of the signal that is linearly dependent on the bits a_i , while the last term is the bi-linear term, which corresponds to the interference between two pit-bits at positions \mathbf{p}_i and \mathbf{p}_j , contributing to the HF signal with the laser spot focused at \mathbf{p}_p . The channel output signal can be written in terms of the bits a_i as

$$r(\mathbf{p}_p) = 1 - \sum_i c_i a_i + \sum_{i \neq j} d_{i,j} a_i a_j, \quad (2.7)$$

where the coefficients c_i represent the linear ISI contributions and the coefficients $d_{i,j}$ represent the bi-linear ISI contributions. They can be calculated as

$$\begin{aligned} c_i &= 2(1 - \sin(\phi_p))(\langle \psi_L | \psi_i \rangle - \langle \psi_i | \psi_L \rangle), \\ d_{i,j} &= 2(1 - \sin(\phi_p))\text{Re}(\langle \psi_i | \psi_j \rangle). \end{aligned} \quad (2.8)$$

The coefficients of the linear and bi-linear kernels have to be computed only once, at the initialization stage of the computations. Only in that stage, a limited number of time-consuming fast Fourier transforms (FFTs) have to be carried out. Once the coefficients of the kernels have been set-up, the actual signal generation boils down to a small number of additions per HF sample to be computed: consequently, a large number of HF-samples can be computed in a short computation time. The values of the kernels for different lattice constants a_H and pit-hole diameters b can be found in [132] and in [18].

2.3.2 Noise Model

In this section a short description of noise models will be given. In an optical storage system several noise sources might be present, each contributing to the overall amount of noise at the receiver input. In this section the focus will be on the physical origin of these noise sources and less on the mathematical description of them (see Section 3.3 for mathematical noise models). The different noise sources are described in the order in which they appear in the physical channel.

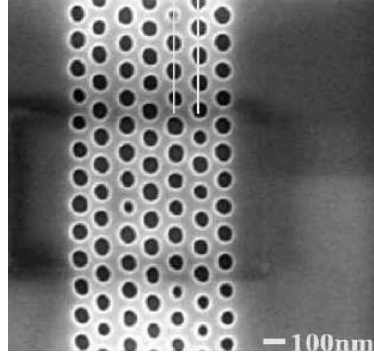


Figure 2.10: Detailed image of an EBR disc containing a substantial amount of pit-size noise. This disc is only used to illustrate the existence of pit-size noise but is not used for obtaining experimental results.

Media Noise In general, media noise can be described as the noise that is induced by the write process. In optical storage, media noise originates from small deviations in the pit-shape from its ideal form. For example, random inaccuracies in the pit-position will cause media noise. Another inaccuracy might be the variation of the pit-size from one pit to the other. This effect is clearly observed in some worst case EBR-mastered discs. One reason for this effect is that only a limited number of electrons is available to master a single pit. A photograph of a clear, worst-case example is shown in Fig. 2.10.

In general, media noise can be considered as a random variation of the pits that are recorded in the medium. During the read-out process, however, it is filtered by the MTF. As a consequence the media noise has a spectrum corresponding to the transfer function of the optical channel. Due to the fact that only pits show random variations, the amount of media noise that is present at the channel output will highly depend on the type of pattern that was stored on the disc. This implies that pit-noise will be data-dependent. This knowledge can be used to derive some quantitative estimates on the amount of media noise during experiments. This estimation will be the subject of Chapter 4.

Laser Noise: Relative Intensity Noise The light intensity of semiconductor lasers tends to fluctuate due to a couple of reasons: spontaneous emission at low laser power levels, and fluctuations in the current that drives the laser. These fluctuations cause a Gaussian, uncorrelated variation at the PDIC output but with only limited effect on the overall noise power [18].

Shot Noise Shot noise arises due to the fact the photons arrive at the PDIC input following a Poisson process. Therefore, even at a constant average incident power on the PDIC, the number of photons arriving within a certain limited amount of time is fluctuating leading to a non-zero variance of the detected photo-current [133]. For the TwoDOS system it was found that this type of noise can be neglected with respect to other types of noise.

Electronics Noise Electronics noise is an accumulation of all the different noise sources that are present in the integrated circuits of the photo detector and in the current-amplifier. This noise can roughly be divided into two categories: current noise and voltage noise. The current noise can be considered to have a flat power spectral density, while the voltage noise increases with increasing frequency. As a result, the increase in data rate in 1-D systems by spinning the disc faster will make the detection less reliable as more voltage noise arises. In a 2-D system, however, a lower rotation speed yields the same data rate and hence a lower voltage noise. In the TwoDOS system, electronics noise at high frequencies together with media noise at low frequencies were identified to be the main sources of noise [18].

Quantization Noise The analog output of the PDIC is transformed into the digital domain by an ADC. In the digital domain only a limited number of amplitude levels can be represented. The number of levels depends on the number of bits at the ADC output. The mapping of the signal into a specific level causes some random quantization noise. In the TwoDOS project an 8-bit ADC is used and only a limited amount of quantization noise will occur.

2.4 Signal Processing Principles

Up to this point, the optical storage system up to the ADC has been elaborated on. The digital replay signals that are produced by the ADC will be processed by the data receiver in order to detect the bits that were written on the disc. In this section the basic signal processing principles will be introduced that can be used for the 2-D data receiver. Especially the differences with the 1-D data receiver, see Section 1.4, will be discussed in detail.

2.4.1 Modulation Code

As has been seen in the previous section, the MTF of the optical channel is low-pass with a hard cut-off frequency f_c . As a result all frequency content above f_c is not transmitted across the optical channel. For this reason modulation codes are used to shape the user data such that it has limited or no content above f_c . In 2-D systems

modulation codes can have a more sophisticated design because they are not limited to use the input data sequence of a single track but they can operate across different tracks.

In literature several 2-D low-pass codes have been proposed [134–136]. Multi-track (d, k) constrained binary codes can be defined in a couple of alternative ways depending on whether the k -constraint and/or the d -constraint is defined only along the track or also across the different tracks [46]. A low-pass coding technique for 2-D systems was proposed in [137]. The technique is based on defining constraints on the bits in the lattice where each point has 6 direct neighbors. A first constraint is the number of bits in the direct neighborhood with the same sign and a second constraint specifies the minimum number of azimuthally-contiguous nearest neighbors (specifying the minimum mark size). Additional constraints are required for the boundaries of the broad spiral, where not all of the 6 neighbors are present. In [137] the bits were considered to be stored on a rectangular lattice. In [18], the technique was rephrased for a hexagonal lattice.

Low-pass coding schemes generally remove patterns that cause large amounts of ISI. Another criterion to remove patterns from the data stream is to remove those patterns that cause small detection distances in maximum likelihood detectors. The detection distances are defined as the difference in total Euclidean distance between admissible waveforms received in the absence of noise. For a single bit error this distance can be formulated as $d^2 = \sum_{\forall(i,j)} h_{2-D}(i,j)^2$. For error patterns that consist of multiple bit errors, ISI has to be taken into account and smaller values of d^2 are possible due to the partial cancellation of impulse response coefficients due to neighboring bits in the error pattern.

For 2-D channels, due to the large amount of possible error patterns, finding error patterns with small detection distances is more difficult than for 1-D channels and some procedures of doing so are reported in [138]. For the TwoDOS optical channel, a brute-force search of error patterns shows that closed-rings of alternating +1 and -1 all result in low Euclidean distances. These rings will be referred to as Nyquist rings. In Fig. 2.11(a) two possible Nyquist rings are shown. The left pattern with 6 bit errors has the smallest Euclidean distance and its spectrum is shown in Fig. 2.11(b). By comparing this spectrum with the transfer function of the optical channel shown in Fig. 2.8, one can see why these patterns are the worst-case patterns: the error patterns have mainly high-frequent content while the channel itself passes only low-frequent content. As a result these error patterns have a low power spectrum at the channel output and they result in small detection distances.

The elimination of these critical patterns is similar to the effect achieved by the MTR-constraint in 1-D systems (see Section 1.4.1). In a 2-D system this type of modulation code is constructed based upon two types of building blocks. The first type of building block is constructed by dividing the meta-spiral into a small number

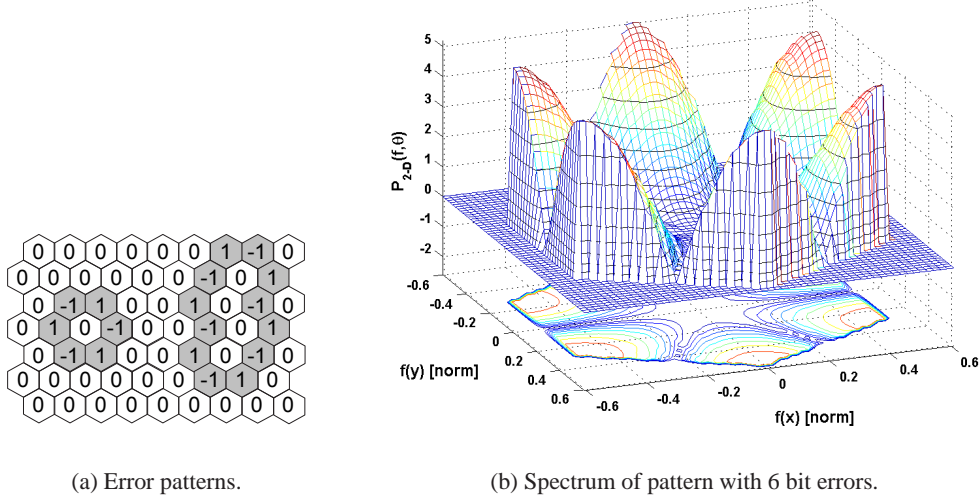


Figure 2.11: Worst-case error patterns for the TwoDOS channel.

of strips, each containing a limited amount of tracks (for example 3 tracks). Within such a strip, encoding is done by applying a 1-D encoding scheme by considering M -ary symbols (where M denotes the number of possible symbol combinations within the strip, $M = 2^3 = 8$ in the example). Strips are concatenated in the radial direction. The second type of building block is a single bit-track that is 1-D modulated and is inserted between consecutive strips such that the track serves as merging bit-track effectively gluing consecutive strips together. More details about the procedure to encode and decode this 2-D modulation code can be found in [139] and will not be repeated here.

2.4.2 Receiver principles

The receiver principles for 1-D systems, as explained in Section 1.4.2, can easily be extended to 2-D systems. In literature several types of 2-D receivers have been proposed. In this section, we will give a brief overview of these types together with their main advantages and disadvantages.

A 2-D storage channel can essentially be considered as a Multiple-Input Multiple-Output (MIMO) channel. Given a MIMO channel, different types of receiver structures can be realized, including those based on a Decision-Feedback Equalizer (DFE) and Partial Response Maximum Likelihood (PRML) detection. In the MIMO receiver the equalizer essentially becomes a 2-D filter that transforms the 2-D ISI, induced by the channel, into a desired ISI structure at the detector input, denoted as the target response. Exactly at this point, an extra degree of design freedom enters:

namely this ISI structure is not restricted to be 1-D but can also be 2-D. For example, a possible choice for the target response can be the full-response target response, i.e. an ISI structure with a central ‘1’ surrounded by all ‘0’s. As is well-known, in this case the optimum 2-D receiver structure consists of a matched-filter front-ends followed by tapped-delay line equalizers, of dimension equal to the number of tracks [140, 141]. The optimal performance of this receiver structure can be achieved in the limit of an infinite number of tracks by a linear MMSE equalizer [42]. In this case a symbol-by-symbol detector is used. Also in the DFE receiver, this type of detector is exploited. For 2-D systems different types of DFEs can be conceived of: a number of 1-D DFEs preceded by an equalizer that minimizes the cross-track ISI [41], a DFE that eliminates all cross-track ISI of tracks that not have been detected [142] or a full 2-D DFE that operates on a column by column basis [143]. These types of DFEs can also be constructed to operate based on a different type of detector, like for example the Fixed-Delay Tree Search (FDTS).

Besides DFEs also PRML receivers can be considered. In general, a PRML receiver using a Viterbi detector (VD) can be constructed using two alternative realizations: one with a 1-D VD, in which a 1-D target response along the track should be chosen and one with a 2-D VD that is the most general. The latter realization has the disadvantage of the greatly increased detector complexity, as discussed in more detail in Section 2.6. The choice of the 2-D target response is a crucial factor for achieving acceptable system performance [144]. In the TwoDOS project, a PRML receiver structure has been chosen. In the next section the structure and the building blocks of the data receiver used in the TwoDOS project will be explained in more detail.

2.5 Data Receiver

In the TowDOS project, the operation of the data receiver is based on a PRML detection technique employing a 2-D Viterbi detector. In this section the different parts of the PRML receiver will be discussed in detail, except for the 2-D Viterbi detector, which will be the topic of Section 2.6. First, the general receiver architecture is briefly discussed.

2.5.1 Basic operation

A top-level block diagram of the receiver architecture under consideration is shown in Fig. 2.12. The analog replay signals $\mathbf{r}(t) = [r^0(t), r^1(t), \dots, r^{L-1}(t)]$, where L denotes the number of tracks on the disc, coming from the multi-spot PDIC are sampled at a fixed frequency $f_s = 1/T_s$, asynchronous with respect to the baud rate, by a multi-channel ADC. To avoid any aliasing from high-frequent noise components back to

the baseband of the signal, the analog signal is filtered with an analog low-pass filter (not shown in the figure). On the one hand, f_s should be chosen as low as possible for practical reasons. On the other hand it should be chosen high enough such that data components do not cause aliasing, i.e. it should obey $f_s > 2f_c$ where f_c is the channel cut-off frequency (see Section 2.3.1). When $f_s = 2f_c$, the analog low-pass filter should have a pass-band with amplitude 1 in the frequency region $[0, f_c]$, an infinitely steep transition band at f_c , and a stop-band with infinite attenuation for $f > f_c$. In practice such a filter cannot be realized. For this reason, the sampling frequency is increased to relax the requirements for the analog low-pass filter. In the experimental TwoDOS system, the sampling frequency f_s is chosen to be 2.5 times the baud rate, i.e. $f_s = 2.5/T$ [145].

$$d_k^l = \sum_{i,j} g^l(i,j) \hat{a}_{k-i}^{l-j}. \quad (2.9)$$

The signal processing blocks in the receiver are a delay-compensation block, a

2-D equalizer, a DC and gain compensation block and finally a sample-rate converter. The first block in the signal processing path is the delay compensation block that eliminates relative track delays caused by the slanted orientation of the linear array of read-out spots with respect to the tangential direction of the broad spiral (see Section 2.2). Also here a high sampling frequency f_s helps to limit the complexity of this block to an acceptable level [146]. At the output of the delay compensation block, the signal is decreased in sample-rate by a factor of 2, i.e. the signal \mathbf{r}_n has a sample-rate of $1/2T_s$. The required anti-aliasing filtering before this sample-rate reduction is combined in an efficient way in the delay compensation block [18]. The reduction of the sample-rate is done to lower the processing rate and the complexity of the subsequent blocks in the signal processing path. After the delay compensation block, a preamble detection block is used to identify the start and the end of the preamble in the data sequence. This block is needed as the delay compensation block estimates the relative delays between the different tracks only during a preamble [18]. Also other adaptation loops (see below) use the output of the preamble detector to switch between different operation modes: acquisition and tracking.

After the delay-compensation block, a 2-D adaptive equalizer is used to transform the ISI induced by the optical channel into the ISI as defined by the target response. The coefficients of the equalizers are updated via an adaptation loop to account for channel variations. The signals \mathbf{r}_n are the inputs of the equalizer and \mathbf{x}_n denotes the equalized output signals. In Section 2.5.2 more details about the 2-D equalizer can be found.

After the 2-D equalizer, the DC and gain compensation block accounts for offset and amplitude variations that might occur in the total system (e.g. drift in the analog amplifiers). The DC and gain estimates are produced by adaptation loops. The delay in these loops is minimized by placing the DC and gain compensation block as close as possible to the detector input (only the SRC is positioned in between). This allows fast response of these adaptation loops.

The bit-detector is designed to operate on samples synchronous to the baud rate. As a result a sample-rate converter is needed to convert the output signals \mathbf{y}_n of the DC and Gain compensation (at a frequency $\frac{1}{2T_s}$) to synchronous input samples for the bit-detector (at a frequency $\frac{1}{T}$). The Sample Rate Converter (SRC) is implemented as a polyphase filter structure in combination with linear interpolation. The SRC forms part of the Phase Locked Loop (PLL) which accurately controls the frequency and phase of the samples at the output of the SRC.

Each block in the receiver accounts for a particular effect of the channel on the received signal $\mathbf{r}(t)$. Because the channel might vary over time, the blocks are required to compensate for these variations. In other words, adaptivity is needed in the receiver to compensate for channel variations. This adaptivity is achieved by utilizing adaptation loops that are able to adjust parameters of the blocks. These adaptation

loops serve to minimize the mismatch between \mathbf{y}_k and \mathbf{d}_k . To achieve this, they use the error signal $\mathbf{e}_k = \mathbf{y}_k - \mathbf{d}_k$. In practice the delay of the bit-detector needs to be compensated for when deriving the error signal \mathbf{e}_k . This is achieved by delaying \mathbf{y}_k (not shown in the figure). In the following subsections the most important blocks in the receiver will be discussed together with their associated adaptation loops.

2.5.2 Equalization

The 2-D equalizer transforms the 2-D ISI induced by the optical channel into a 2-D ISI structure defined by the target response \mathbf{g} . The equalizer is implemented as a set of 2-D hexagonal filters, one filter for every track:

$$x_n^l = \sum_{i,j} w^l(i,j) r_{n-i}^{l-j}, \quad (2.10)$$

where x_n^l is the equalizer output signal, r_n^l is the input signal and $w^l(i,j)$ denotes the filter coefficients for track l , with an extent determined by $i \in \{-W, W\}$ and $j \in \{0, L-1\}$, W defines the length of the filters and their height is limited to the number of tracks within the broad spiral. It should be noted that in the TwoDOS receiver, the number of equalizer input signals is equal to the number of equalizer output signals. In 2-D systems this is not generally true. For example, for the systems shown in Fig. 1.8 (magnetic systems with L tracks that are read out by N heads), the equalizer would have N input signals and L output signals. Another example is given in [147] for 2-D optical systems. Here a tribinary target response is used, such that the equalizer has L inputs and produces $L+1$ outputs. In the TwoDOS receiver, the size of the 2-D equalizer should be chosen carefully. The bulk of the ISI power at the channel output should be covered by the span of the equalizer. By using the scalar diffraction model, it can be computed that at least 5 shells surrounding the central bit should be covered [18].

Because of the nonlinear nature of the optical channel, non-linear equalizers might yield an improved performance [83, 148]. The added implementation cost of these types of equalizers is likely to be significant and for this reason, a linear equalizer is used in the TwoDOS receiver. The channel nonlinearities are taken into account in the detection process, see Section 2.6.

The equalizer coefficients are continuously adapted based on either the Least Mean Square (LMS) or the Zero-Forcing (ZF) adaptation technique. Both techniques use the error signal \mathbf{e}_k to update the equalizer coefficients. The LMS technique tries to minimize the power of the error while the ZF technique forces residual ISI (RISI) to be zero [106], where RISI is defined as the difference between the actual ISI and the ISI as defined by the target response. In the TwoDOS receiver the choice has been made to use an asynchronous equalizer. This choice complicates the adaptation considerably because an inverse SRC is needed to transfer the synchronous error signal

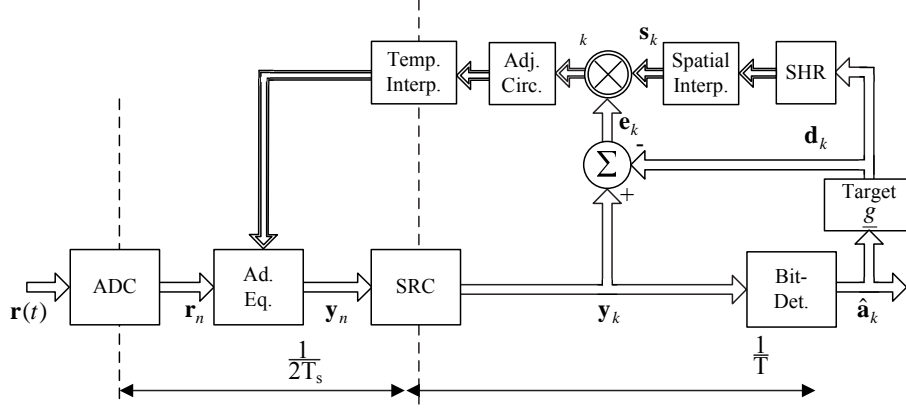


Figure 2.13: Zero-forcing equalizer adaptation loop in the overall block diagram.

to the asynchronous domain. Here, only the asynchronous ZF adaptation technique will be discussed. For asynchronous LMS adaptation we refer to [149] where the 1-D situation is explained.

In Fig. 2.13 the ZF adaptation loop is depicted. This loop forces RISI to be zero or, equivalently, it causes the detector input signal \mathbf{y}_k to be a noisy version of the reference signal \mathbf{d}_k . A synchronous ZF adaptation loop for track number l would accomplish this by forcing for every equalizer coefficient $w^l(i, j)$ the correlation between the error signal e_k^l and the reference signal d_{k-i}^j to zero. In the asynchronous case, the equalizer coefficients are spaced by the double sampling period $2T_s$ rather than the bit period T . This means that the correlation between e_k^l and $s_k(n - i, j)$ should be forced to zero (where n is the index associated with the rate $1/2T_s$), where \mathbf{s}_k is a matrix with columns containing spatially interpolated versions of \mathbf{d}_k (roughly for $k \in \{-WT_s/T, \dots, WT_s/T\}$ to cover the entire equalizer span). To accomplish this interpolation, the samples \mathbf{d}_k are stored in a shift register (SHR). The output of the SHR is a matrix of desired signals, one vector for each track (matrix signals are denoted by a double arrow in the block diagram). These vectors are then interpolated along the track such that the spacing is correct. Across the track no interpolation is needed as the samples are spaced according to the inter-track distance. Summarizing, the matrix \mathbf{s}_k at time instant k , contains for all tracks l a vector of samples spaced at a rate $2T_s$.

The correlation signals $\Delta_k^l(i, j)$ are obtained (for $i \in \{-W, W\}$, $j \in \{0, L-1\}$ and $l \in \{0, L-1\}$) as

$$\Delta_k^l(i, j) = e(k, l) s_k(n - i, l - j). \quad (2.11)$$

These signals are used as inputs to a set of adjustment circuits (denoted as “Adj. Circ.” in the figure), one for every equalizer coefficient that is adapted. The adjustment circuit consists of a scaling factor, i.e. the adaptation constant μ and an ideal integrator.

The outputs of the adjustment circuit are the values of equalizer coefficients given in the synchronous time domain. As the equalizer operates in the asynchronous domain, some kind of conversion is needed to transform the synchronous coefficient values to the asynchronous domain. This conversion is accomplished by a temporal interpolation. As the equalizer coefficients are slowly varying in time, this interpolation can be realized by a set of simple zero-order hold functions with negligible performance degradation [111].

The coefficient update can be written as

$$w_k^l(i, j) = w_k^l(i, j) + 2\mu\Delta_k^l(i, j), \quad (2.12)$$

where μ is a suitable adaptation constant. Note that this update is not fundamentally different from the 1-D case, see (1.7). Just as in the 1-D case, the introduction of tap leakage is essential to limit the sensitivity of the coefficient values to spectral nulls in the channel. The leaky coefficient-update given in (1.8) for the 1-D equalizer, can easily be extended to the 2-D case and is omitted here.

2.5.3 Timing Recovery

The delay compensation block, right after the ADC, already achieves a crude alignment of the different tracks (this block is not discussed separately but the interested reader is referred to [18]). The fine alignment and the transformation into the synchronous domain is done by the timing recovery circuit. Timing recovery for 2-D systems is not very different from the timing recovery for 1-D systems. One could just apply an independent circuit for every track. However there are some differences with respect to the 1-D case.

The first difference arises from the fact that by using 1-D Sample Rate Convertors (SRCs) a rectangular fundamental domain is obtained that is smaller in the temporal (tangential) direction than the hexagonal fundamental domain originating from the hexagonal bit lattice (see [18] for more explanation about the difference in fundamental domains). Due to the smaller fundamental domain in the temporal direction, aliasing might occur in this direction. This type of aliasing might be avoided by using a 2-D SRC for every track. In the TwoDOS project, however, the choice has been made to use a 1-D SRC. Due to the shape of the MTF (see Fig. 2.8), the effect of aliasing is minimal and the performance gain by using a 2-D SRC does not justify the complexity.

The second difference lies in the fact that the frequency (and in principle also the phase) of the different tracks are equal. As a result this information can be used to derive a common frequency error for all the tracks simultaneously by averaging the timing errors of all these tracks. This common derivation of the frequency error makes the timing recovery circuit more robust. As the error is averaged over the

different tracks, the amount of gradient noise in the loop is reduced and as a result the bandwidth of the loop filter might be increased.

Although a separate SRC is used for every track, there is still some relation between the SRCs, namely they need to output new samples at the same time. A third difference with 1-D timing recovery is the fact that due to small phase differences between different tracks at the SRC input, the SRCs of the different tracks do not need to skip or add samples simultaneously to be able to output at the same time. As a result in the implementation of the SRCs some buffering needs to be incorporated to account for this. For the actual implementation of the SRCs we refer to [18].

2.5.4 DC and gain control

Other time-varying parameters in the system are the DC offsets and the gains of the channel. Several artifacts in the system might cause variations of DC offset and/of gain:

- Variations in the analog part of the data receiver cause both DC offset and gain fluctuations. Examples are the DC offset and the gain variations in the PDIC and in the ADC due to temperature variations, voltage variations or ageing.
- Baseline wander has an impact on the DC offset at the detector input. Baseline wander appears in case the signal is AC coupled, which means no DC content is transmitted by the channel. As a result if the data is not DC-free, a slowly time-varying DC offset will be present at the detector input.
- Artifacts in the optical part of the system. For example variations in disc reflection due to scratches, fingerprints or inhomogeneous material may cause rapid variations of both DC and gain values. Furthermore gain variations between different tracks appear in the TwoDOS system due to variations in diffraction efficiency of the grating [18].

This list of possible artifacts suggests that many DC and gain variations are common for the different tracks. For example a scratch on the disc will affect a large 2-D area. Especially the rapid variations appear to be common across the tracks, while static or slowly varying artifacts cause track-dependent variations. These observations reveal a worthwhile adaptation strategy: rapidly varying parameters are identified as much as possible and controlled commonly in fast adaptation loops (common timing parameters, a common DC offset and a common gain factor). The remaining response after compensation of the rapidly-varying components is mainly subject to slow variations and is further shaped towards the target response by slow adaptation loops. These slow loops have in general a large number of parameters to adapt. This adaptation strategy is the subject of Chapter 6 where the theoretical issues and the practical implementation is described in more detail.

Unless mentioned otherwise, individual adaptation loops for every track will be used. As their functionality and their implementation are not fundamentally different from the 1-D case, the description is omitted and the interested reader is referred to Section 1.4.4.

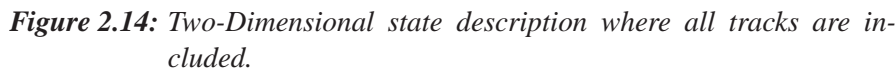
2.5.5 Interaction between adaptation loops

The different adaptation loops in the TwoDOS receiver each have a dedicated function. For example, the gain loop controls the signal amplitude at the detector input and the timing recovery loop controls the sampling phase. It must be noted that there can, in general, be interaction between the different loops. For example, also the equalizer is able to adjust the gain of the system. If the equalizer coefficients are all multiplied by a common factor, the gain value produced by the gain adaptation-loop will be divided by the same factor. This means that both loops might drift in an opposite direction without the error generation circuit of the loops being able to detect this. This interaction is undesirable because in a fixed-point implementation it leads to a loss in dynamic range and causes quantization errors. Another example is the interaction between the equalizer and the timing recovery loop. Both are able to adjust the phase at the detector input.

To avoid interaction between adaptation loops, a constraint may be imposed on one of the loops. A common constraint to avoid interaction between the equalizer adaptation loop and the gain adaptation loop, is to fix the central coefficient of the equalizer to 1. In this way the gain factor of the equalizer is bounded and the gain variations in the optical system are handled by the gain adaptation loop. The interaction between the equalizer and the timing recovery loop is also reduced but not avoided by this constraint [111]. In the TwoDOS system, however, this type of interaction does not cause the equalizer adaptation loop and the timing recovery loop to deviate a lot from their normal range of operation, i.e. the largest coefficient values are still located around the central coefficient and the sampling phase obtained by the timing recovery circuit does not deviate more than 1 symbol period from its central value. As a result no extra measures need to be taken in the TwoDOS receiver to further reduce the interaction.

2.6 Bit Detection Techniques

The bit-detector that is used in the TwoDOS system is a 2-D Viterbi detector. The aim of this detector is to jointly detect the bits of all tracks in the broad spiral. A straightforward detection approach is to generalize the 1-D VD to multi-track systems by grouping the tracks. This approach results in a 2-D VD where a normal Viterbi algorithm is applied that uses M -ary symbols instead of binary symbols, with



2-D Stripewise Viterbi detection The SWVD is based on the MVA. In the MVA, joint detection is performed on a number of $L_s < L$ tracks. Such a subset of tracks will be referred to as a stripe. To cover all different tracks in the broad spiral, $L_n = L - L_s + 1$ stripes are needed, if the different stripes are shifted by a single bit-track. The stripe-wise approach of the MVA-algorithm is based on a concatenation scheme of interconnected Viterbi detectors in which the output of previous stripes is passed as side-information for the branch metric calculations of subsequent stripes.

In Fig. 2.15 a SWVD is shown with stripes containing 3 bit-tracks. In total 5

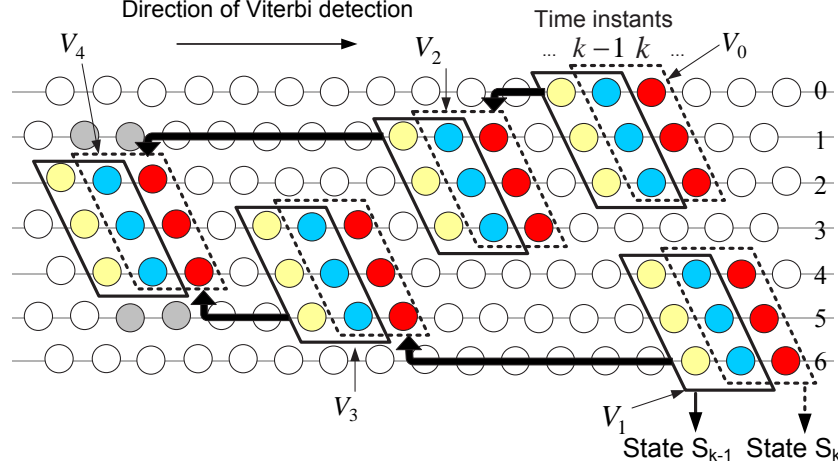


Figure 2.15: Configuration of the stripe-wise VD for $L = 7$, five stripe VDs are organized in a $<-$ shape.

stripe VDs are needed to cover the complete broad spiral. The different stripe VDs are organized in a “ $<-$ ”-shape. This shape is utilized because it results in the lowest latency between input and output, namely three times the latency of a single stripe VD. If the stripe VDs were organized sequentially from the upper to the lower track, the latency would be 5 times the stripe VD latency. The stripes adjacent to the guard band are used as starting points because the guard band contains known bits (namely zeros). These known bits can be used as side-information for the branch metric calculation for the first stripe VDs. As there are two guard bands, the stripe VDs V_0 and V_1 are the first to launch. The stripe VD V_0 outputs a bit decision for the upper track $l = 0$, while V_1 outputs a bit decision for the lower track $l = L - 1$. These outputs are passed as side-information to the next stripe VDs, respectively V_2 and V_3 . The last stripe VD outputs bit decisions for the inner 3 bit-tracks.

The branch metric used in V_0 is the sum of the Euclidean distances of the metrics of the different tracks belonging to that stripe;

$$\beta_{V_0}(\mathbf{S}_k, \mathbf{S}_{k-1}) = \sum_{l=0}^2 (y_{k-1}^l - d_{k-1}^l)^2, \quad (2.13)$$

where d_k^l is the ideal detector input signal for track l according to the target response \mathbf{g}^l (this signal will be also referred to as the reference signal). The ideal detector input signal is determined by the bits that are defined by the matrix \mathbf{B}_{k-1}^l , namely the current bit and the bits in the first shell. Note that for the two outer tracks in the stripe VD (for example, for V_4 these tracks are track number 2 and track number 4) not all the bits in \mathbf{B}_k^l are defined by the two states \mathbf{S}_k and \mathbf{S}_{k-1} . For V_4 the bits that are not defined, are colored grey in Fig. 2.15. As these bits are required to calculate

the branch metric, the values are taken from an array $\hat{\mathbf{a}}$ of most-recent bit decisions that are available. In the figure, the thick black arrows indicate where bit decisions of other stripe VDs are used as side information in the branch metric calculation of other stripe VDs. For stripe VD V_4 the needed bits are all already detected by other stripe VDs, however for V_0 no bit decisions are readily available. As a starting point, $\hat{\mathbf{a}}$ is filled with zeros or with some best guesses of the bit values as obtained from a preliminary bit-detector like a threshold detector.

As these preliminary decisions are not very reliable, the performance of the stripe VD might suffer from this lack of prior knowledge. A possible way to reduce this effect, is by introducing weights in the calculation of the branch metric (for stripe VD V_0) as

$$\beta(\mathbf{S}_k, \mathbf{S}_{k-1}) = \sum_{l=0}^2 \omega^l (y_{k-1}^l - d_{k-1}^l)^2, \quad (2.14)$$

where ω^l are some suitable weights. For example, for stripe VD V_0 these weights are $\omega_0 = \omega_1 > \omega_2$. For the stripe VDs in the lower part of the broad spiral, the weights of the upper track should be smaller than the other weights because for these stripe VDs no prior information is available for the tracks above the stripe.

As already stated above, side-information is taken from $\hat{\mathbf{a}}$. It seems natural that the reliability of this prior information influences the reliability of the final decisions of the stripe VD. As a result it would be preferable to have reliable prior information for all the stripe VDs. A way to achieve this, is to do multiple iterations of the different stripe VDs, as shown in Fig. 2.15. As a result it would not be needed during the second iteration to use the initial guesses contained in $\hat{\mathbf{a}}$. This leads us to the final implementation of the SWVD as it is used in the TwoDOS receiver for bit-detection. This implementation is shown in Fig. 2.16. The SWVD consists of two detection iterations, where every iteration consists of stripe VDs which are organized in a “<”-shape. The first iteration is performed by stripe VDs (V_{00} up to V_{05} in the figure) each covering two bit-tracks, which results in a stripe VD with 16 states. The second iteration consists of stripe VDs (V_{10} up to V_{14} in the figure) each covering three bit-tracks which results in a stripe VD with 64 states. In every iteration the binary output from a first stripe VD is passed to a next stripe VD to be used as side information in the branch metric calculations [152]. The outputs of the first iteration are used as side information in the second iteration.

2.7 Conclusions

In this chapter, a global overview of the TwoDOS system is given. In the TwoDOS system, bits are stored on a hexagonal lattice. In contrast with conventional optical storage systems (CD, DVD and BD), where the bits are stored in a single spiral (a 1-D sequence of bits), in TwoDOS the bits are organized in a so-called broad spiral.

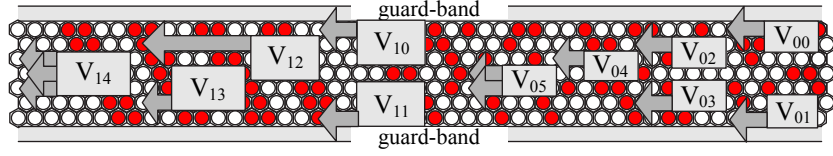


Figure 2.16: The configuration of stripe-wise Viterbi detection: two iterations of stripe detectors are used to detect the bits of the broad spiral.

The broad spiral contains a number L of bit-tracks, stacked upon each other to form a hexagonal structure. Adjacent rotations of the broad spiral are separated by a guard band consisting of a bit-track without any pits. The data is read out with an array of L laser spots arranged such that each spot is centered on one of the bit-tracks within the broad spiral. A multi-spot photo detector integrated circuit is used to generate a so-called high-frequency (HF) signal for every bit-track.

A PRML receiver has been built for TwoDOS. It consists of a bit-detector preceded by an adaptive equalizer, an adaptive DC compensator, an AGC and a timing recovery loop. A 2-D VD performs joint bit-detection on all bit-tracks. To reduce the complexity of a full-fledged 2-D VD, the VD is divided into smaller processing units (called stripe VD). Each stripe VD covers a limited number of bit-tracks (so-called stripes with a typical height of 2 or 3 bit-tracks). This detection configuration is called a Stripe-Wise Viterbi Detector (SWVD). The SWVD consists of two detection iterations, where every iteration consists of stripe VDs which are organized in a “>”-shape. The first iteration is performed by stripe VDs each covering two bit-tracks, which results in a stripe VD with 16 states. The second iteration consists of stripe VDs each covering three bit-tracks which results in a stripe VD with 64 states. In every iteration the binary output from a first stripe VD is passed to a next stripe VD to be used as side information in the branch metric calculations. The outputs of the first iteration are used as side information in the second iteration.

Chapter 3

Characterization of Experimental TwoDOS PRML System

This chapter is concerned with the detailed characterization of the TwoDOS system. A system model is proposed that accurately describes both the intersymbol interference structure and the noise structure. The intersymbol interference is modeled via a Look-Up Table (LUT). The noise structure is modelled as the output of a 2-D data-dependent autoregressive filter. The proposed system model is simple, accurate and it is able to track channel variations. Throughout this chapter experimental signal waveforms coming from the TwoDOS system will be used to illustrate the characterization in terms of the proposed model. The results of this experimental characterization provides important information about the quality of the mastered discs and about possible modifications to the TwoDOS PRML receiver to improve its performance.

3.1 Introduction

In general, characterization techniques are used to analyze the signal waveforms of an experimental system. Characterization results allow researchers to identify various distortions in the experimental storage channel, and are hence fundamental for compensating or counteracting these distortions. In optical storage systems, characterization serves three main purposes: 1) to monitor the quality of the mastered media (e.g. pit imperfections), 2) to interpret the characterization results in order to fine-tune the experimental read-out system and 3) to provide information for possible improvements in several signal processing algorithms utilized in the modulation code, the mastering, the read-out and the receiver.

The characterization of an experimental system involves two steps: the selection of a system model with a proper structure, and the accurate estimation of the model parameters based on experimental signal waveforms. The first step is crucial: if the structure of the selected model is not matched to the experimental system under consideration, the estimated parameter values are meaningless and no relevant conclusions can be drawn from the characterization results. Therefore in this chapter

the focus will be on the development of a system model matched to the experimental TwoDOS system.

Modelling high density storage effects has been the subject of intensive research interest in recent years [153–155]. The typical tradeoff in modelling is accuracy for tracking speed, with detailed but computationally intensive models [156] on one end of the spectrum and fast parametric models [157] on the other end. Ideally, a model must be computationally efficient and therefore simple parametric models are commonly used. Parametric models with one and more parameters have been proposed for 1-D magnetic systems [154, 155, 158, 159] and for 1-D optical systems [3, 160]. In this chapter we will develop a simple parametric model suitable for 2-D systems by extending these 1-D models to their 2-D equivalents.

In general, a channel model consists of two distinct parts: a intersymbol interference (ISI) model that is a deterministic function of the recorded data and a noise model that describes the indeterministic or random portion. The ISI model describes the portion of the signal waveform that is uniquely dependent on the information (bits) written on the disc, i.e. it covers the linear and nonlinear ISI originating from the bits. The second part describes the portion of the signal waveform that is not uniquely dependent on the written bits, i.e. this portion will not be the same for different occurrences of the same bit pattern. As it is subject to random processes, this part will be denoted as the noise model.

In this chapter we will propose a system model for 2-D digital data storage systems. The ISI model will be developed in Section 3.2 and it consists of two parts: a Look-Up Table (LUT) to cover all nonlinear and linear ISI originating from bits with a limited distance from the current bit, and a linear filter that covers all linear ISI originating from bits with a larger distance. In Section 3.3, the noise model will be developed. In general the noise can be modelled as the output of a 2-D data-dependent autoregressive filter. The proposed noise model is the 2-D equivalent of the 1-D model described in [157]. It matches the observed waveform up to the second-order statistics. When the noise is Gaussian, the first- and the second-order statistics uniquely define the probability distribution.

The development of the system model will be accompanied by results for the experimental TwoDOS system to illustrate the agreement of the model with an experimental system. The PRML receiver that has been built for TwoDOS will be used as basis for the experimental characterization [145]. The receiver consists of a bit-detector preceded by an adaptive equalizer, an adaptive DC compensator, an AGC and a timing recovery loop (see Section 2.5). To improve the signal processing algorithms applied in the receiver, the characterization is performed on the detector input signals because the quality of these signals are directly related to the performance of the receiver. The characterization results (i.e. the values of the estimated parameters) yield important information about the ISI structure and the physical noise sources in

the TwoDOS system.

Throughout this chapter, a data-aided adaptive scheme will be used for the estimation of model parameters. This scheme has the advantage that variations in model parameters can be monitored easily. These variations provide important information about time-varying artifacts in the experimental system. Time variations of different parameters will be discussed in more detail in Section 3.4.

3.2 Intersymbol interference Characterization

In Fig. 3.1 a model of a 2-D storage system is given. In this model the bit sequence \mathbf{a}_k is input of the continuous-time channel which is characterized by a (possibly nonlinear) 2-D function $\mathbf{h}(\cdot)$ and the noise signal $\mathbf{v}(t)$. The resulting replay signal $\mathbf{r}(t)$ has a continuous-time nature and is the input of the data receiver. These replay signals are subsequently sampled at the baud rate $1/T$. Hence the resulting discrete-time signal \mathbf{r}_k is synchronous with respect to the input bit sequence. In order to do reliable bit-detection the 2-D channel impulse response is transformed by the equalizer into a predefined target response. This equalizer is characterized by a 2-D impulse response \mathbf{w}_k . The resulting signal \mathbf{y}_k is hence a synchronized signal characterized by a predefined impulse response and is used as input of the bit-detector (denoted by DET in the figure).

In this section the focus is on the development of a discrete-time ISI model suitable for the characterization of the detector input signal \mathbf{y}_k of the experimental TwoDOS system. In other words we want to have a discrete-time model of the detector input signal \mathbf{y}_k as a function of the input sequence \mathbf{a}_k . In the remainder of this chapter the part of the signal due to ISI will be described by the function $\mathbf{p}(\cdot)$. This function describes the chain of operations of the continuous-time channel, the sampler and the equalizer. These operations will be grouped and will be denoted as the equivalent channel or just as the channel in the remainder of this chapter, not to be confused with the time-continuous channel from Fig. 3.1. This ISI model will be part of the overall system model for TwoDOS.

The configuration that is used for ISI characterization, is shown in Fig. 3.2. In response to the bit sequence $\mathbf{a}_k = [a_k^0, a_k^1, \dots, a_k^{L-1}]^T$ (with $a_k^l \in \{0, 1\}$ and where L is

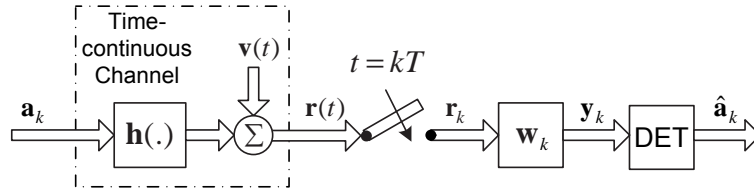


Figure 3.1: Block diagram of a storage system model.

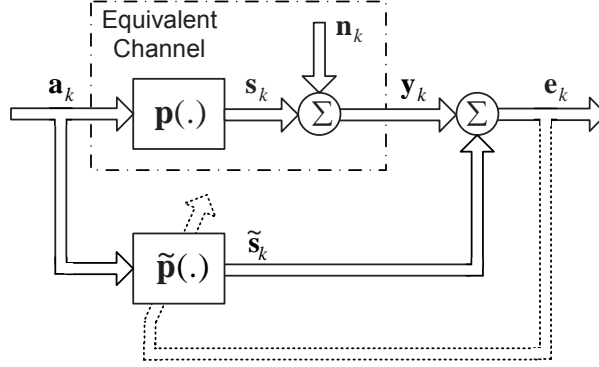


Figure 3.2: Basic configuration used for ISI characterization.

the total number of tracks on the disc and T denotes matrix transposition) which is recorded on the disc, an equivalent channel produces an output signal $\mathbf{y}_k = [y_k^0, y_k^1, \dots, y_k^{L-1}]^T$. The channel output signal y_k^l at time k for track l is given by

$$y_k^l = s_k^l + n_k^l, \quad (3.1)$$

where s_k^l is the ISI signal, containing all ISI and the signal due to the current bit a_k^l , and n_k^l is the noise signal (which is data-dependent and colored, see Section 3.3). As already stated in the previous section, the signal s_k^l is a deterministic function of the bits written on the disc, denoted $s_k^l = p^l(\mathbf{a}_k)$. For every track l , the function $p^l(\cdot)$ defines the relationship between the bits on the disc and the received signal and each of these functions might be nonlinear and time-dependent. To characterize the signal, the bits \mathbf{a}_k are applied to a set of filters with input-output function $\tilde{p}^l(\cdot)$. The output signals \tilde{s}_k of these filters are subtracted from the channel output signals \mathbf{y}_k to produce an error signal for every track. The error signal of track l at time k is given by

$$e_k^l = p^l(\mathbf{a}_k) - \tilde{p}^l(\mathbf{a}_k) + n_k^l. \quad (3.2)$$

For every track l the power of e_k^l is minimized when $\tilde{p}^l(\cdot)$ is selected to equal $p^l(\cdot)$. This suggests the use of an independent adaptation algorithm for every track l that minimizes e_k^l by adjusting the function $\tilde{p}^l(\cdot)$. When the power of e_k^l reaches its minimum value for every track l , $\tilde{\mathbf{p}}(\cdot)$ specifies the unknown function $\mathbf{p}(\cdot)$ and \mathbf{e}_k describes \mathbf{n}_k . In practice, one is most interested in the precise knowledge of the function for the middle track, as in general the central track is the most difficult one to do reliable detection on. For this reason in the remainder of this chapter characterization results will be shown for the middle track only.

A proper choice for the function $\tilde{\mathbf{p}}(\cdot)$ is crucial for accurate ISI characterization. In general, in low-density storage applications linear functions are often used

because they provide sufficient accuracy [3]. However at increasing densities, non-linear ISI becomes more pronounced and more complicated functions are needed to fully describe the signal waveform. For optical storage systems this nonlinear ISI can be modelled by adding bilinear ISI coefficients to the linear ISI coefficients [132]. These bilinear coefficients represent the interaction of pairs of bits. More generally a Volterra kernel can be used to model storage channels [161]. The resulting ISI model will be more accurate but it will also have many more parameters that need to be estimated. Another approach to model nonlinear channels with a finite memory length is the usage of a look-up table (LUT) [148]. This approach has the advantage that it completely describes the deterministic channel operation. Disadvantages of the LUT table are the fact that the table entries do not reveal much information about the operation of the channel and that a large number of parameters are needed [162].

To characterize the experimental TwoDOS signals, different function types will be used. In Section 3.2.1 a linear function is used for the characterization. The experimental results for this function type clearly suggest that besides linear ISI also nonlinear ISI is present. In Section 3.2.2 bilinear coefficients are added to the linear function to account for the main nonlinearities induced by the optical channel. Inspired by detection methods suited for nonlinear channels, in Section 3.2.3 a LUT will be used for the ISI characterization. To limit the total number of required entries in the LUT, in Section 3.2.4 the LUT is combined with a linear filter.

3.2.1 Linear ISI Model

In digital storage systems, the channel output y_k^l can often be approximated as a linear function of the input bits according to

$$y_k^l = \mathbf{A}_k^T \mathbf{P}_k^l + n_k^l, \quad (3.3)$$

where

$$\mathbf{A}_k = \begin{bmatrix} a_{k-I}^0 & \dots & a_k^0 & \dots & a_{k+I}^0 \\ a_{k-I}^1 & \dots & a_k^1 & \dots & a_{k+I}^1 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ a_{k-I}^{L-1} & \dots & a_k^{L-1} & \dots & a_{k+I}^{L-1} \end{bmatrix}^T \quad (3.4)$$

is the matrix containing input bits of all tracks with a span of I bits around the current bit ($a_k^l \in \{0, 1\}$). Furthermore,

$$\mathbf{P}_k^l = \begin{bmatrix} p_k^l(-I, 0) & \dots & p_k^l(0, 0) & \dots & p_k^l(I, 0) \\ p_k^l(-I, 1) & \dots & p_k^l(0, 1) & \dots & p_k^l(I, 1) \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ p_k^l(-I, L-1) & \dots & p_k^l(0, L-1) & \dots & p_k^l(I, L-1) \end{bmatrix}^T \quad (3.5)$$

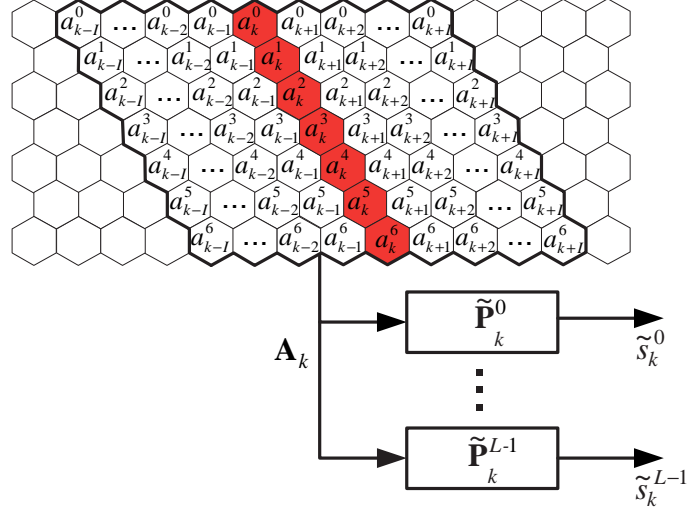


Figure 3.3: Configuration of linear ISI characterization technique for TwoDOS experimental system. The matrix \mathbf{A}_k with the bit values is used as input of a set of linear filters with impulse responses $\tilde{\mathbf{P}}_k^l$ to produce for every track l an ISI signal \tilde{s}_k^l .

is the channel impulse response matrix for track l at time instant k and n_k^l is an additive noise signal. To emphasize the time dependence of the channel impulse response a subscript k is added to \mathbf{P}^l .

The functions $\tilde{\mathbf{P}}(\cdot)$ which are used in the ISI characterization, will also be linear with the same structure as the one defined in (3.5). The output signal \tilde{s}_k^l of the characterization technique can be written

$$\tilde{s}_k^l = \mathbf{A}_k^T \tilde{\mathbf{P}}_k^l. \quad (3.6)$$

All the coefficients $\tilde{p}_k^l(n, m)$ of $\tilde{\mathbf{P}}_k^l$ ($n \in [-I, I]$, $m \in [0, L-1]$ and $l \in [0, L-1]$) are adapted based on the LMS adaptation algorithm [153, 163]

$$\tilde{p}_k^l(n, m) = \tilde{p}_{k-1}^l(n, m) - \mu_h e_k^l a_{k+n}^m, \quad (3.7)$$

where μ_h is an adaptation constant which enables a tradeoff between speed of convergence and steady-state excess mean-square error. The behavior of this algorithm has been studied extensively [164–166].

A schematic overview of this linear ISI characterization technique applied to the experimental TwoDOS system, is shown in Fig. 3.3. For every track l , a 2-D adaptive filter with impulse response $\tilde{\mathbf{P}}^l$ produces, based on \mathbf{A}_k which contains all relevant bit values a_i^j for $i \in \{k-I, \dots, k+I\}$ and $j \in \{0, \dots, L-1\}$, a signal \tilde{s}_k^l that is an estimate

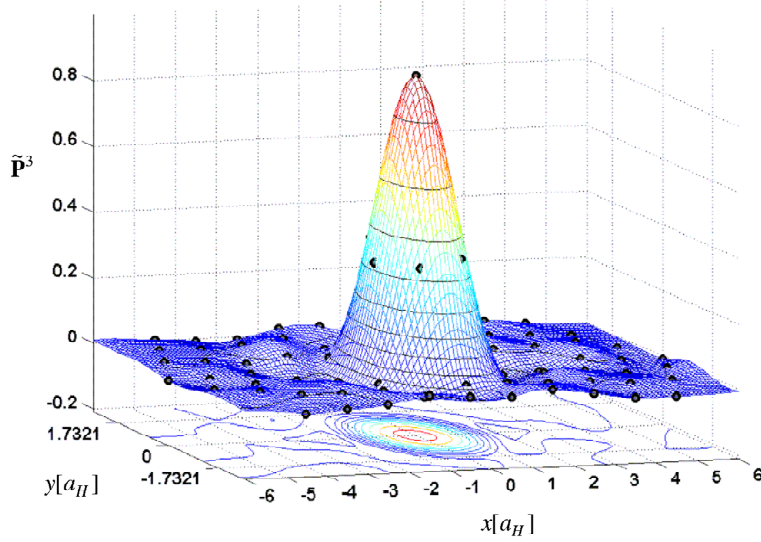


Figure 3.4: Estimated linear coefficients of the detector impulse response of the central track versus spatial coordinates x and y (with units given in $a_{bit} = 138\text{nm}$). The coefficients are centered around the central coefficient $\tilde{p}^3(0, 3)$.

of the ISI part of the experimental TwoDOS signal y_k^l , which is, in this case, the detector input signal. Experimental TwoDOS discs are used with 7 parallel bit-tracks ($L = 7$) and a capacity of 50 GB. These discs are read out under normal conditions by the optical player and experimental replay signals are produced which are digitized and used in the TwoDOS receiver. Adaptive signal processing blocks in the receiver (equalizer, timing recovery, DC/gain control) transform these digitized replay signals into a synchronous detector input signal with a predefined target impulse response, more precisely a one-shell target with a central coefficient equal to 1, i.e. $g_0 = 1$ and $g_1 = 0.26$ (and as explained in Section 2.5.4, the coefficients are subsequently normalized to have a total power equal to 1).

The estimated impulse response $\tilde{\mathbf{P}}$ for the central track is shown in Fig. 3.4. Clearly the bulk of the signal energy is contained within the first shell (central 7 bits). These bits are defined by the set \mathbf{B}_k^l for track l , see Section 2.1.1. Only a limited amount of ISI originates from bits outside this first shell for these experimental signals.

To judge the quality of the ISI characterization technique, the actual (experimental) signal values and expected signal values (according to the linear model) should be compared. As the signal values mainly depend on the bits defined by the set \mathbf{B}_k^l , they are defined by all possible combinations of bits within the first shell ($2^7 = 128$

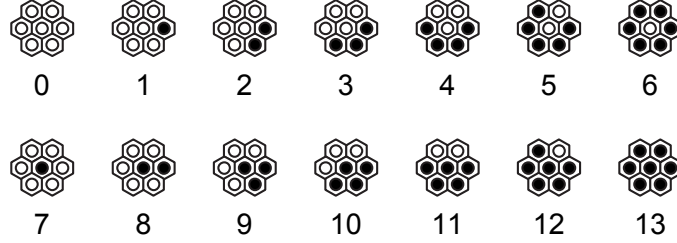


Figure 3.5: Overview of 7-bit cluster types for linear characterization (14 different types).

combinations). All these possible combinations can be grouped further in to clusters with the same amount of pits ($a_k^l = 1$) in the first shell. This grouping is described by the cluster number which is defined as $7a_k^l + \sum_{(i,j) \in \mathbf{B}_k^l} a_i^j$. In Fig. 3.5 an overview of all cluster numbers (with a specific example for every cluster number) is given.

At the detector input, it is expected that the estimated impulse response $\tilde{\mathbf{P}}$ is equal to the target response which is linear and circularly symmetric. As a result the calculated ISI signal value \tilde{s}_k^l is expected to depend only on the associated cluster number. To judge whether the experimental system results in a channel impulse response that is really linear and circularly symmetric, the experimental signal values should be compared with the expected ISI signal values for all possible bit combinations.

In Fig. 3.6 the mismatch between the experimental signal values and the calcu-

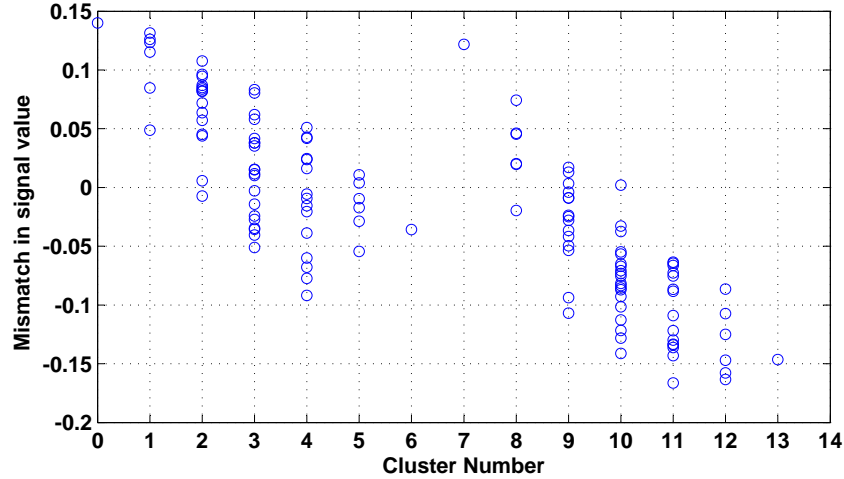


Figure 3.6: Mismatch between experimental signal values (averaged over time) with respect to expected ISI signal values according to the linear model versus cluster number for the 128 different clusters.

lated ISI signal values is plotted versus cluster number for all possible different bit clusters. The experimental signal values are obtained by averaging the experimental values y_k^l over time for a specific cluster number to minimize the effect of noise. If the linear model would be an appropriate model for the experimental system, these mismatches would be zero for all clusters. Clearly these mismatches are not zero for numerous clusters. As a result we can conclude that the linear model does not fully describe the experimental signal values and a more sophisticated model should be developed. Channel nonlinearity seems the most likely cause for the offsets in signal values. In the next section bilinear coefficients are added to the linear ISI characterization technique to account for this nonlinear ISI.

3.2.2 Bilinear ISI Model

Many storage applications are severely hampered by nonlinear channel effects. To be able to describe these nonlinearities several models have been introduced [82]. For optical storage, it is shown that not all the coefficients of the Volterra decomposition are needed [161]. The nonlinear operation is sufficiently described by adding bilinear combinations to the linear system model [132]. This means that only combinations of two pit-positions need to be considered. A further simplification can be realized by explicitly imposing rotational symmetry, by neglecting all the interferences beyond the nearest-neighbor pits and by neglecting the cross-interference between nearest-neighbor pits that are separated by a distance larger than a_{bit} [18]. For this simplification, the system model needs two additional parameters to account for the remaining bilinear interference: α for the interaction between the central pit and the pits of the first shell and β for the interaction between two pits of the first shell that are neighbors. These coefficients are updated according to the standard LMS adaptation technique (see Section 3.4.1 for more details about adaptive parameter estimation). The signal \tilde{s}_k^l is given by

$$\tilde{s}_k^l = \mathbf{A}_k^T \mathbf{P}_k^l + \alpha a_k^l p_{1,0}(\mathbf{B}_k) + \beta p_{1,1}(\mathbf{B}_k), \quad (3.8)$$

where $p_{1,0}(\mathbf{B}_k) = -a_k^l + \sum_{(i,j) \in \mathbf{B}_k} a_i^j$ and $p_{1,1}(\mathbf{B}_k)$ is the number of pit-pairs within the first shell that are nearest neighbors of each other. The bilinear coefficients α and β represent respectively the interaction between a central pit and pits in the first shell, and the interaction between neighboring pits within the first shell.

Characterization results for the 50 GB disc show that the mean power of \mathbf{e}_k decreases from -18.45 dB (in case the linear model is used) to -18.75 dB (in case the bilinear terms are added to the linear model). This improvement indicates that bilinear ISI is present in the experimental TwoDOS signal. More bilinear coefficients can be added to see whether these results could be improved. We can capture all (not only the bilinear) nonlinear effects originating from pits within the first shell by the use of a look-up table (LUT), which is the topic of the next section.

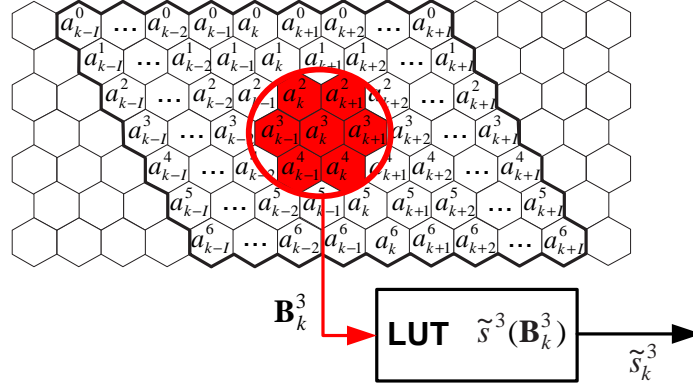


Figure 3.7: LUT configuration for ISI characterization technique for track $l = 3$.

3.2.3 Look-Up Table Model

To include nonlinearities, \tilde{s}_k^l can be constructed as a LUT (which can be implemented as a RAM) rather than as a linear filter as discussed in Section 3.2.1. However, if we want to cover the entire matrix \mathbf{A}_k , this table will have a very large number of entries ($2^{L(2l+1)}$). Because the bulk of the nonlinear ISI is induced by bits close to the central bit a_k^l , the dependence window for the LUT is centered around this bit. This window of the LUT is described by the subset \mathbf{B}_k^l . It covers all bits that are adjacent to the current bit, e.g. $\mathbf{B}_k^2 = [a_k^1, a_{k-1}^1, a_{k+1}^2, a_k^2, a_{k-1}^2, a_{k+1}^3, a_k^3]$. As a result, for every track l , the LUT has $2^7 = 128$ entries. The entries in the LUT can be computed beforehand by averaging or can be updated adaptively based on the LMS adaptation technique (see Section 3.4.1).

This dependence window will be used to characterize the TwoDOS system and the configuration is shown schematically in Fig. 3.7.

The reference values $\tilde{s}^l(\mathbf{B}_k^l)$ (128 values in total) are plotted versus cluster number in Fig. 3.8 for the central track $l = 3$, where cluster number is again defined as $7a_k^l + \sum_{j,i \in \mathbf{B}_k} a_i^j$. The plot shows two branches: the left branch corresponds to clusters with $a_k^l = 0$ and the right branch corresponds to clusters with $a_k^l = 1$. The mean power of \mathbf{e}_k^l for the LUT characterization technique amounts to -21.84 dB for the central track ($l = 3$), which is an improvement of 3.4 dB with respect to the linear model and nearly 3.1 dB with respect to the bilinear model. This shows that higher-order nonlinearities are truly significant in the TwoDOS system and may not be neglected.

The LUT model covers all linear and nonlinear ISI originating from bits within the first shell. However, from Fig. 3.4 we can see that although the bulk of the energy is coming from linear ISI from within the first shell, there is still linear ISI from outside the first shell. The magnitude of this ISI for different experimental conditions

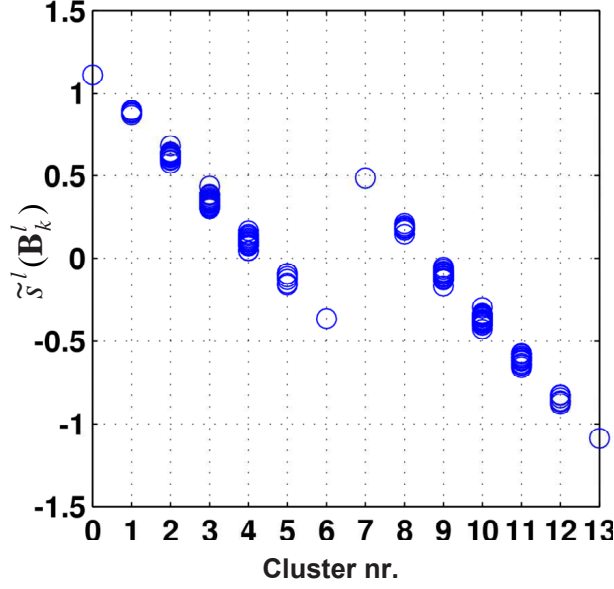


Figure 3.8: LUT values for the central track $l = 3$.

is studied in the next section.

3.2.4 Residual ISI Model

The signal \tilde{s}_k^l should depend on all the channel input bits in \mathbf{A}_k . In the LUT-model however, the signal-dependence window of the LUT covers only a subset of \mathbf{A}_k , namely \mathbf{B}_k^l and therefore not all ISI is modelled by the LUT. To also include linear contributions from bits outside the first shell, we propose to use a combination of a LUT that encapsulates the main nonlinearities (which originate from \mathbf{B}_k^l) and a linear 2-D convolution that generates the signal due to residual intersymbol interference (RISI) originating from the remaining input bits. The signal output \tilde{s}_k^l can then be expressed as the sum of the output of the LUT and the RISI signal which is the convolution output:

$$\tilde{s}_k^l = \tilde{s}(\mathbf{B}_k^l) + \mathbf{C}_k^{lT} \tilde{\mathbf{F}}_k^l. \quad (3.9)$$

The RISI signal is computed by convolving the bits which are in \mathbf{A}_k but not in \mathbf{B}_k^l (these bits are described by the subset \mathbf{C}_k^l , which has zero coefficients for the 7 central bits) with the RISI impulse response $\tilde{\mathbf{F}}^l$. Hence, the impulse response $\tilde{\mathbf{F}}^l$ has $L(2I - 1) - 7$ nonzero coefficients. Summarizing for every track l , a LUT and a 2-D impulse response $\tilde{\mathbf{F}}^l$ is used to characterize the noiseless signal output \tilde{s}_k^l .

This configuration with two parts is shown in Fig. 3.9 for the TwoDOS system.

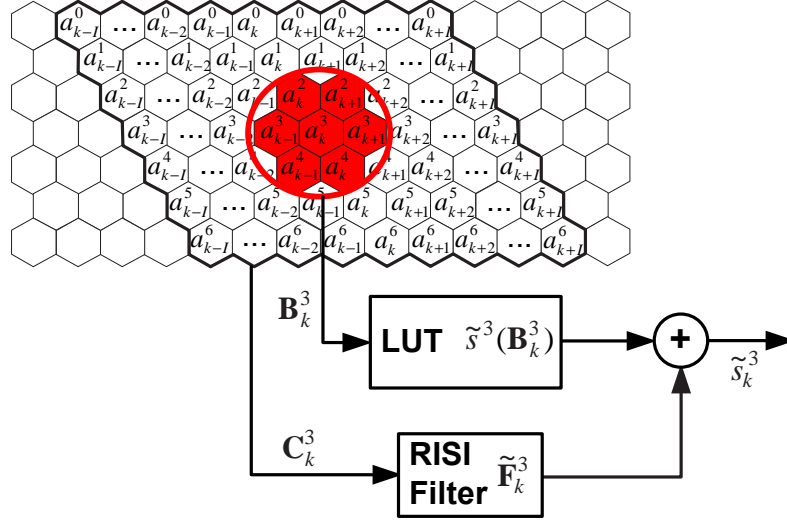


Figure 3.9: Configuration for ISI characterization technique for a combination of a LUT and a linear RISI Filter for track $l = 3$.

The RISI impulse response $\tilde{\mathbf{F}}^l$ obtained from the experimental TwoDOS characterization is shown in Fig. 3.10 for $l = 3$ in case there is an angle between the disc and the laser beam (also referred to as a tilt angle) of -1.0° in the radial direction. The RISI components are limited in amplitude. Even so, RISI originating from symbols with limited temporal separation from the symbols of the target response is non-negligible. This RISI hampers the performance of the SWVD considerably, suggesting that some extra measures against RISI should be taken. One of these measures is the cancellation of RISI which is treated in Chapter 5.

3.2.5 Experimental Results

A PRML receiver has been built for TwoDOS [145]. It consists of a bit-detector preceded by an adaptive equalizer, an adaptive DC compensator, an AGC and a timing recovery loop. The purpose of the adaptive equalizer is to shape the ISI structure induced by the channel into a predefined linear ISI structure, characterized by the target response \mathbf{g}_k .

Because at the detector input the bulk of the ISI energy is concentrated within the span defined by the target response (typically this span covers all bits within the first shell, see Fig. 2.1), the signal LUT is implemented to incorporate this energy, i.e. the LUT of $\tilde{s}^l(\mathbf{B}_k^l)$ has $2^7 = 128$ entries. If due to misequalization, not all the ISI is incorporated into the target span, some RISI is left outside the span of the target. This RISI is characterized by the impulse response $\tilde{\mathbf{F}}^l$. A 2-D VD performs joint bit-

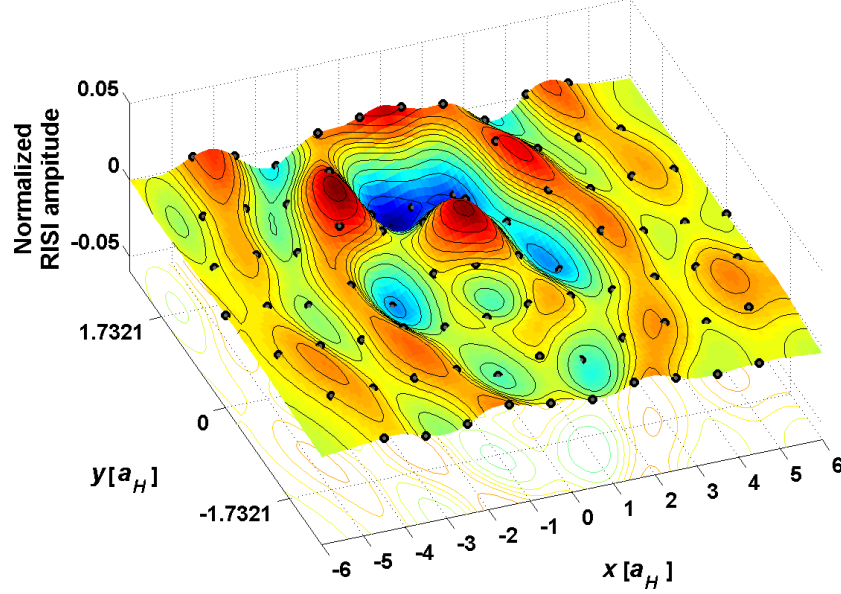


Figure 3.10: Estimated amplitudes (normalized with respect to $E[|s^l(\mathbf{B}_k^l)|])$ of the RISI impulse response $\tilde{\mathbf{F}}^l$ at the detector input for $l = 3$ ($\tilde{\mathbf{F}}^l$ is centered around track $l = 3$). Both axes are scaled in terms of a_H , where a_H is the distance between two bits measured on the disc ($a_H = 138$ nm for 50 GB disc).

detection on all bit-tracks based on the values in the LUT entries $s^l(\mathbf{B}_k^l)$. To reduce the complexity of a full-fledged 2-D VD, the VD is divided into smaller processing units (called stripe VD). Each stripe VD covers a limited number of bit-tracks (so-called stripes with a typical height of 2 or 3 bit-tracks). This detection configuration is called a stripe-wise viterbi detector (SWVD) [145]. More details about the SWVD are given in Section 2.6.

Characterization of the experimental TwoDOS system is based on the read-out of a 50 GB single layer TwoDOS disc with $L = 7$ bit-tracks within the broad spiral. The actual characterization is performed on the input signals of the SWVD. The signal-dependence window used in the characterization is defined by $I = 7$, i.e. the matrix \mathbf{A}_k contains $L(2I + 1) = 105$ elements.

To judge which characterization model is the most accurate and which ISI structure is present, different BER experiments are conducted for different conditions. In Fig. 3.11, the BER of the TwoDOS system is shown for different tilt angles. In the detector two different target responses are used: a linear target response and a LUT target response. Clearly the second target response performs better. The difference in performance can be explained by the fact that the VD is highly sensitive to residual linear and nonlinear ISI [167]. Because the nonlinear ISI originating from bits within



Figure 3.11: BER of TwoDOS PRML receiver with two types of target response: linear target response and LUT target response.

Table 3.1: MSE in dB for different configurations of TwoDOS PRML receiver

MSE	Radial Tilt angle		
Configuration	-1.0°	0.0°	1.0°
Linear ISI	17.12	-18.75	-18.04
Linear and Bilinear ISI	-17.77	-19.45	-18.70
Reference LUT	-19.72	-21.84	-20.83
Reference LUT and RISI filter	-21.48	-23.49	-22.47

the first shell is included in the LUT, the performance is improved significantly by including the information of the LUT in the branch metric computation of the VD. Also the remaining linear ISI originating from bits outside the first shell, causes some performance degradation. The treatment of this RISI is the topic of Chapter 5.

As mentioned in Section 3.2, the power of the error signal is also a good measure to quantify the characterization accuracy. In Table 3.1 the MSE is given for different characterization techniques and for different tilt angles. The characterization technique with the LUT and the RISI linear filter yields an improvement of approximately 4.5 dB with respect to the linear-model technique. Further improvements with respect to the LUT and RISI technique might be accomplished by adding for example

bilinear ISI components at the edge of the first and second shell of bits. These additions are found to yield only minor further improvements in MSE and for this reason in this chapter the technique with the LUT and the RISI cancellation will be used to characterize the experimental TwoDOS signals. In the next section different noise characterization techniques will be proposed and experimental TwoDOS results will be shown to validate these techniques.

3.3 Noise Characterization

The input signal of the detector at time k for track l may be represented as

$$y_k^l = s_k^l + n_k^l, \quad (3.10)$$

where s_k^l is the noiseless ISI signal and n_k^l is an additive noise signal. The ISI characterization provides an estimate \hat{s}_k^l of the ISI signal. The error signal $e_k^l = y_k^l - \hat{s}_k^l$ (see Fig 3.2) contains all relevant information about n_k^l . For this reason the development of an accurate noise model is based on the analysis of the error vector $\mathbf{e}_k = [e_k^0, \dots, e_k^{L-1}]$. A crucial condition for the accuracy of the noise model is that the ISI characterization should be highly accurate. This condition is essential because if RISI overwhelms the noise then a false indication of noise is obtained and as a result noise characterization becomes pointless. For the experimental TwoDOS characterization this condition is guaranteed by the elaborate research that has been conducted to develop the ISI model (see Section 3.2). Adding linear and bilinear RISI components on top of the combination of LUT and linear RISI filter was shown to not yield additional improvements in MSE. For this reason it is guaranteed that \mathbf{e}_k does not contain any ISI components and is primarily determined by noise.

A first and basic noise characterization technique is to measure the power of \mathbf{e}_k for every track l . This technique is used to compute the predetection SNR for every track l according to

$$\text{SNR}^l = \frac{E_k [\hat{s}_k^{l2}]}{E_k [e_k^{l2}]}, \quad (3.11)$$

where E_k denotes the expectation over all time indices k . For Viterbi detection techniques it is well known that the performance is uniquely determined by the predetection SNR given by (3.11) if \mathbf{e}_k is Gaussian and white [151]. However in many practical systems the noise at the detector input is not white and not Gaussian. Many different noise models might be considered: going from simple models with a few parameters to complicated models with many parameters. In general the simple models rely on the physical nature of the noise source (see Section 2.3.2), and as a result they are attractive because they provide direct indications for improvements in the system. In Chapter 4 a characterization technique based on such models is treated. In

this chapter the choice has been made to use more general noise models that can be applied to all 2-D storage systems and many types of noise sources. In Section 3.3.1 the correlation of \mathbf{e}_k is examined. In Section 3.3.2 a data-dependent auto-regressive noise model is proposed that can be used to accurately characterize the noise in 2-D systems.

3.3.1 Correlated Gaussian Noise Model

The stochastics of \mathbf{e}_k reveal insightful information about the noise processes in the TwoDOS system. To this end, the 2-D autocorrelation function of e_k^l is calculated. This 2-D autocorrelation function, denoted as R_e , is calculated over all tracks and over all time instants and is defined as

$$R_e(i, j) = E_{k,l} \left[e_k^l e_{k-i}^{l-j} \right], \quad (3.12)$$

where $E_{k,l}$ denotes the expectation over all time indices k and over all tracks l .

In Fig. 3.12 this 2-D autocorrelation function is plotted versus the tangential and radial directions. In this figure R_e is normalized with respect to $R_e(0,0)$. The central coefficient of the function obviously has a value of 1. If the noise would be white, all other coefficients should be zero. This is not the case: the coefficients in the first shell all have a value bigger than zero (around 0.15). This indicates that \mathbf{e}_k has a low-pass nature both in the radial and in the tangential direction. Furthermore the behavior of R_e outside the first shell shows some high-frequency oscillations. These high-frequency components in the error signal \mathbf{e}_k can be explained by the fact that the adaptive equalizer tries to minimize the mismatch between the actual and the desired (target) response especially at low frequencies (as the optical channel is a low-pass channel) and does not explicitly minimize the error at high frequencies. As a result some noise enhancement at high frequencies arises at the detector input.

By taking the 2-D hexagonal Fourier transform of the autocorrelation function, the Power Spectral Density (PSD) of the noise is obtained [168, 169]. This PSD, denoted as $S_e(f_x, f_y)$, is shown in Fig. 3.13 and it exhibits rather large variations (a difference of 7 dB exists between the minimum and the maximum). These variations will deteriorate the performance of the VD as it is designed for white noise [72]. In the next section a noise model is presented that is suitable for a wide variety of 2-D storage systems: the 2-D data-dependent Auto-Regressive (AR) noise model.

3.3.2 Data-Dependent Auto-Regressive Noise Model

In the previous section, the noise autocorrelation function and its Fourier transform were used to provide information about the noise at the detector input. Additional information can be obtained by performing a Karhunen-Loeve decomposition on

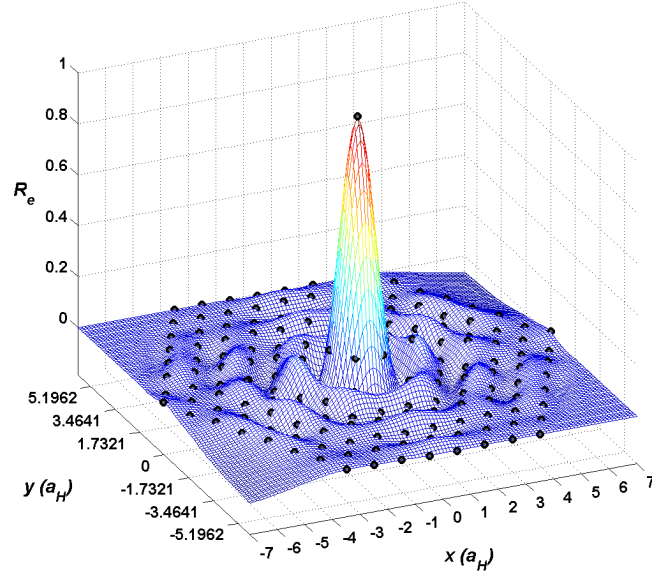


Figure 3.12: Normalized autocorrelation function R_e versus x and y , the tangential and radial directions respectively (both axis are scaled in terms of a_H).

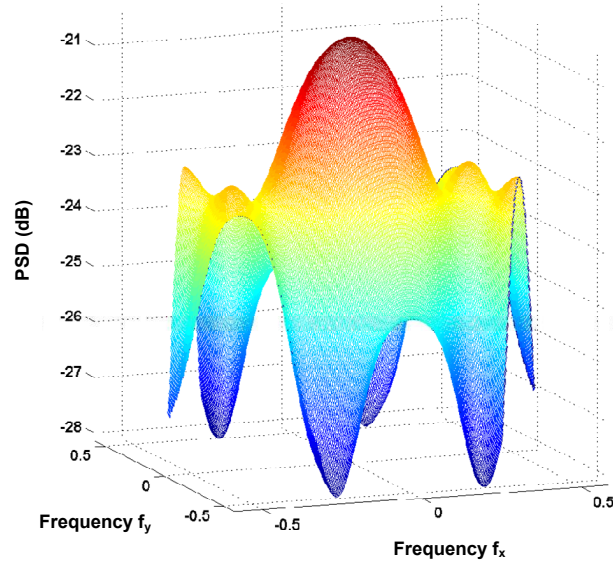


Figure 3.13: 2-D amplitude representation of the PSD of the error signal.

the covariance matrix of the noise, from which for example pulse jitter and amplitude variations can be identified [157]. However, Karhunen-Loeve decompositions are static matrix decompositions and hence are not very well suited for real-time characterization. For real-time characterization, where characterization results are generated on the fly, the more suitable matrix decomposition is the Cholesky decomposition of the covariance matrix inverse, which actually results in an AR noise model [157]. This AR noise model can be improved even further by taking into account the data-dependent nature of the media noise (see Section 2.3.2). The resulting model is a data-dependent AR noise model. In this section a 2-D version of this model will be introduced and used to characterize the TwoDOS system.

The 2-D data-dependent AR noise model of the central track ($l = 2$) is depicted in Fig. 3.14 for a rectangular lattice for $L = 5$ tracks together with the applied ISI model (see Section 3.2.4). Together they form the system model. This system model generates, besides an ISI signal output s_k^l (which has the same configuration as the one described in Section 3.2.4), also a noise output n_k^l for every track l at every time

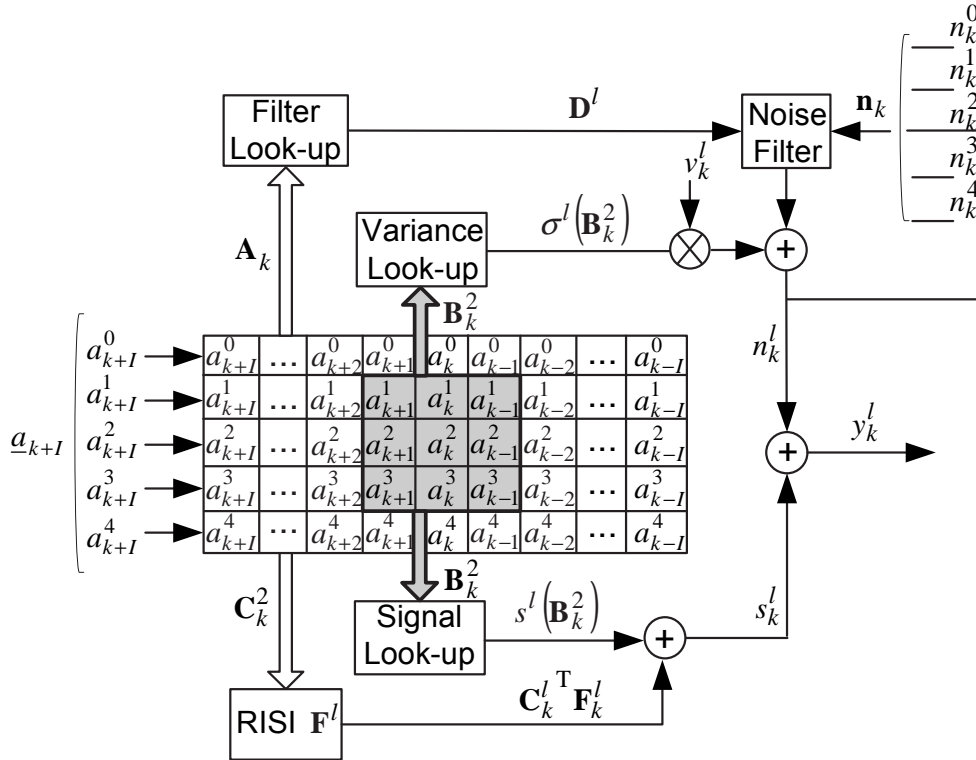


Figure 3.14: Two-dimensional AR system model for $L = 5$ tracks, shown for the central track $l = 2$.

index k . This noise output is modelled as the output of a causal 2-D AR filter (denoted noise filter in Fig. 3.14) with an impulse response \mathbf{D}^l , where every coefficient $d_{i,j}^l$ ($i \in [0, L-1]$ and $j \in [1, R]$, where R is the noise memory length) is data-dependent. To limit the complexity of the noise model, these coefficients are considered to be dependent only on a subset of \mathbf{A}_k . Here the choice is made to make the coefficient $d_{i,j}^l$ depend on the window described by \mathbf{B}_{k-j}^i . Although this window does not cover all bits, in most cases it captures sufficient data-dependency. The noise output n_k^l can be written as the output of a 2-D AR filter with at its input a white Gaussian noise sequence $\{v_k^l\}$ (with zero mean and unit variance):

$$n_k^l = \sum_{i=0}^{L-1} \sum_{j=1}^R d_{i,j}^l(\mathbf{B}_{k-j}^i) n_{k-j}^i + \sigma^l(\mathbf{B}_k^l) v_k^l, \quad (3.13)$$

where the coefficients of the noise filter, i.e. the standard deviation $\sigma^l(\mathbf{B}_k^l)$ and all coefficients $d_{i,j}^l$ (with $i \in [0, L-1]$ and $j \in [1, R]$), are data-dependent. This noise structure makes the noise sequence n_k^l both signal-dependent and correlated (where the correlation is also signal-dependent). The estimation of the different parameters is treated in more detail in Section 3.4.1.

The estimated noise filters $\tilde{\mathbf{D}}^l$ are used to produce a whitened error signal

$$\mathcal{N}_k^l = e_k^l - \sum_{i=0}^{L-1} \sum_{j=1}^R \tilde{d}_{i,j}^l(\mathbf{B}_{k-j}^i) e_{k-j}^i. \quad (3.14)$$

In Fig. 3.15 the PSD of this whitened error signal is shown for experimental TwoDOS replay signals. By comparing this PSD with the PSD of the original error signal (see Fig. 3.13), it is clear that the spectrum has become much flatter (variations of only 2dB between the minimum and the maximum). Still some residual variations are left in the whitened error signal. These variations can be explained by the data-dependent nature of the media noise. The PSD is obtained by averaging over all possible data sequences without making distinction between data sequences with a low or a high amount of media noise. Hence a further noise characterization step involves measuring data-dependent noise variances.

Based on the whitened error signal the data-dependent variances $\tilde{\sigma}^l(\mathbf{B}_k^l)^2$ can easily be calculated. In Fig. 3.16 the signal-dependent variances $\tilde{\sigma}^l(\mathbf{B}_k^l)^2$ are plotted versus the cluster number for track $l = 3$ for different angles of radial tilt. The observation that the right branch has higher variances than the left branch, indicates that pit size variations are the dominant sources of media noise in the optical recording process, see Section 3.3.4. A noise-predictive maximum likelihood detector can be used to improve the overall system performance. This detector uses the estimated noise filters $\tilde{\mathbf{D}}^l$ to whiten the noise in the detector [170]. The estimated signal-dependent noise variances $\tilde{\sigma}^l$ can be used in the computation of a modified branch metric to improve the performance even further [170].

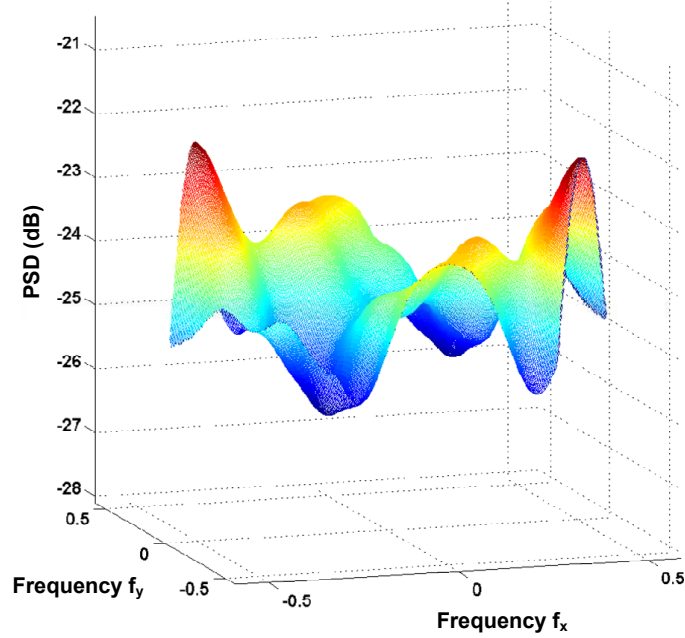


Figure 3.15: 2-D amplitude representation of the PSD of the whitened error signal.

3.3.3 Experimental Results

AR Model Validation

A strong test of the proposed channel model is the bit-error rate (BER) comparison presented in Fig. 3.17. HF signals coming from the experimental set-up are applied to the TwoDOS receiver to produce equalized synchronous inputs of the SWVD. These inputs were used to estimate the parameters of the proposed channel model. The error rates in Fig. 3.17 present the performance of the SWVD for the real noisy experimental data, the data generated by applying the proposed channel model and the data generated by applying a channel model that has the same ISI structure but generates only white noise (with $\sigma^2 = E[(e_k^l)^2]$) [171]. The first bars compare the overall BER, while the other bars compare the BER due to error events with a specific length. The data-dependent AR channel model is clearly a more accurate model than the white noise model. Because the white noise model does not take the noise coloration into account, it produces BERs that are too optimistic.

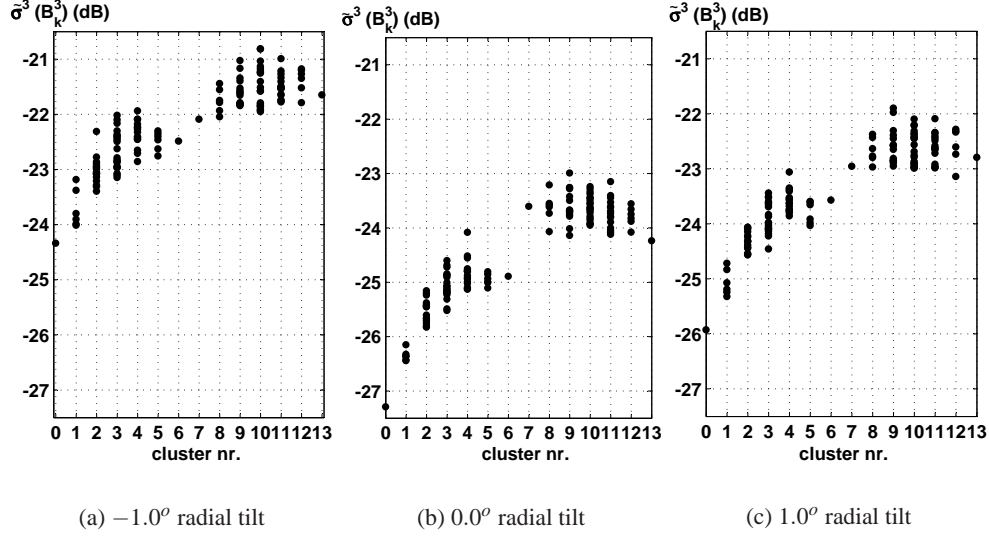


Figure 3.16: Estimated noise variances for the central track $l = 3$ for different values of radial tilt.

MSE Performance of TwoDOS system

As mentioned in Section 3.2, the power of the error signal is also a good measure to quantify the fit between the actual and the estimated channels. In Table 3.1 the MSE is given for different characterization configurations and for different tilt angles. To evaluate the data-dependent AR noise model, the whitened error signal is used to

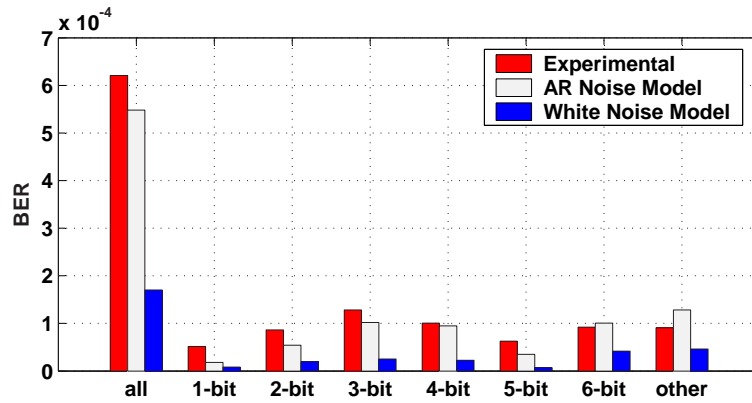


Figure 3.17: Error rate comparison between experiment, data-dependent AR channel model and white noise channel model.

Table 3.2: *MSE for different configurations of TwoDOS PRML receiver*

Configuration	-1.0°	0.0°	1.0°
Reference LUT and RISI cancellation	-21.48	-23.49	-22.47
Data-Dependent AR	-21.84	-24.21	-22.92

compute the MSE. These results are shown in Table 3.2. The improvement in MSE is due to the whitening filters which act as noise prediction filters. The error signal which is left after ISI characterization still contains a considerable amount of correlation (see Fig. 3.13). This correlation is predictable and as a result can be removed by decorrelating the error signals over time and over the different tracks. This decorrelation is accomplished by the whitening filters which produce the whitened error signal \mathcal{N}_k^l . In average the MSE has improved 0.5 dB by applying these whitening filters.

3.3.4 Media Noise

In Section 3.3.2 noise characterization according to a data-dependent AR model was discussed. This noise model can be applied to a wide variety of storage systems because it describes the noise structure in a very generic way by using many different parameters. A disadvantage of this generality is that the characterization results, i.e. the estimated model parameters, are not linked to the physical origin of the noise. Every noise source has very specific characteristics which are not readily visible in the estimated parameters. Therefore it would be very interesting to link the estimated parameters to specific noise sources because information about specific noise sources (which can be related to a physical cause) would provide important information about possible improvements in the system.

As stated in Section 2.3.2, media noise is the dominant noise source in optical storage channels [18]. In this section we will try to link the estimated parameters of the noise model to a specific media noise source. The observation that in Fig. 3.16 the right branch has higher variances than the left branch, suggests that pit-size variations, i.e. variations in the diameter of the pit-hole (see Section 2.3.2), are the most likely sources of media noise. But the estimated parameter values do not reveal any information about the actual amount of pit-size variations, i.e. they do not directly give the variance of the pit sizes as outcome. In this section a quantitative measure for the amount of pit-size noise will be derived based on the estimated parameter values.

This quantitative measure will be derived in a couple of steps. First, for different amounts of pit-size noise theoretical variances for every cluster type will be computed based on a realistic channel. Second, the estimated values will be plotted for every cluster type together with the theoretical values to have a visual identification of the amount of pit-size noise. Finally, the most likely pit-size variance can be computed based on the theoretical and estimated variances by averaging over all cluster types.

To obtain an accurate estimation of the pit-size variation, the selection of an ac-

curate channel model is crucial. For this reason we select the scalar diffraction model to base the computations on. In the scalar diffraction model the detected signal for a optical spot centered at the position \mathbf{p}_p is given by

$$r(\mathbf{p}_p) = \langle \psi | \psi \rangle, \quad (3.15)$$

where the complex-valued waveform at the pupil plane is denoted by $|\psi\rangle$. Furthermore the inner product is defined as the integration in the frequency domain over the central aperture of the objective lens:

$$\langle \theta | \varphi \rangle = \int_{CA} \theta^*(\mathbf{v}_s) \varphi(\mathbf{v}_s) d\mathbf{v}_s. \quad (3.16)$$

The complex-valued waveform $|\psi\rangle$ is given by

$$|\psi\rangle = \mathcal{F}_{\mathbf{p} \rightarrow \mathbf{v}_s} [P(\mathbf{p} - \mathbf{p}_p) D(\mathbf{p})], \quad (3.17)$$

where $P(\mathbf{p})$ is the spot profile and $D(\mathbf{p})$ is the disc reflection function. Note that the bracket notation is used as introduced by Dirac [172] in the field of quantum mechanics.

To incorporate the effect of pit-position and pit-size noise the disc reflection function of Eq. (2.4) can be rewritten as:

$$D(\mathbf{p}_p) = 1 + \sum_i a_i W_i(\mathbf{p} - \mathbf{p}_i - \delta\mathbf{p}_i), \quad (3.18)$$

where $W(\mathbf{p} - \mathbf{p}_i)$ defines the area of the pit on the disc which is centered at position \mathbf{p}_i . It is equal to 1 inside the pit area and 0 elsewhere. Pit-position noise is modelled as a random displacement vector \mathbf{p}_i and pit-size noise is considered with respect to the nominal pit-size window W as:

$$W_i = W + dW_i. \quad (3.19)$$

Pit-size noise is quantized through a parameter σ_{dW} with dimension of an area (m^2).

To get some insight into the influence of a certain amount of pit-size noise on the variance per cluster level, simulations are performed based on the scalar diffraction model. Pit sizes are distributed normally with standard deviation σ_{dW} . Average signal levels are derived and also the variances for each of the clusters are calculated. These variances uniquely define the effect of pit-size noise and as a result can be used to compare with the experimentally obtained variances.

For a fair comparison, no variance due to differences in ISI signal values is allowed. Therefore, the set of clusters that give unique ISI signal values in a circularly symmetric channel model has 26 entries. Any arbitrary cluster (1 of 128) can be derived from these 26 basic types by rotation or point-inversion with respect to the

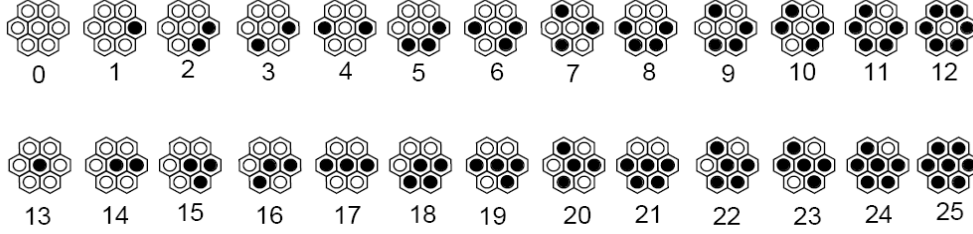


Figure 3.18: Overview of 7-bit cluster types for noise characterization (26 different types).

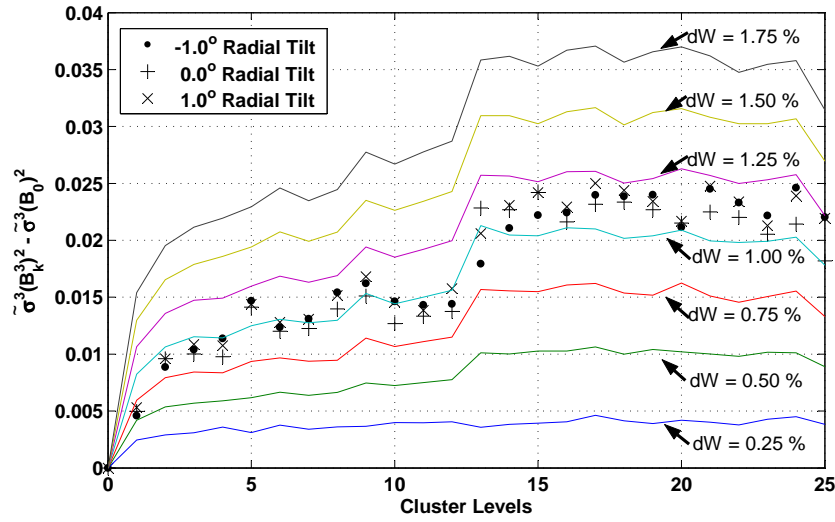


Figure 3.19: Variances for experimental signals (based on the AR channel model) and simulated signals (based on the scalar diffraction model) versus cluster levels.

center bit. The 26 clusters are listed in Fig. 3.18. Note that the relative ordering of the bits of the first shell does matter, but not cyclic permutations.

In Fig. 3.19 the theoretical variances obtained from scalar diffraction simulations for different standard deviations σ_{dW} are shown together with the variances obtained from the experimental TwoDOS signals (for different tilt angles: -1.0° , 0.0° and 1.0°). The experimentally obtained variances are obtained by using the AR channel model derived in Section 3.3.2. For the experimental signals the variance of the all-land cluster (no pits) is taken as a baseline. The pit-size noise of the scalar diffraction model is denoted relative to the area of a hexagonal fundamental domain.

By visual inspection of Fig. 3.19, it is clear that the experimental disc has a pit-size variation of about 1%. In Chapter 4 an alternative way to characterize pit-size

noise in optical storage system is given. There a simple noise model is used to characterize the media noise directly based on the replay signals.

3.4 Time Variations

So far we have developed a system model suitable for the characterization of 2-D storage systems. Characterization results have been shown for the experimental TwoDOS system. These characterization results are the estimated parameter values of the proposed system model. Up to now we have only shown static estimated parameter values, more precisely values averaged over an entire revolution of the disc. However these estimated values may vary over time due to time-varying artifacts in the TwoDOS system. In Section 3.4.1 adaptive parameter estimation is discussed for the data-dependent AR system model. Also in this section, based on this adaptive parameter estimation, time-varying effects in the characterization results of the TwoDOS system are discussed. Furthermore, there are also other time-varying artifacts in the TwoDOS system which are not readily visible in the characterization results because they are counteracted by the adaptive signal processing blocks in the TwoDOS receiver (and in the signal processing chain these blocks appear before the channel characterization block). In Section 3.4.2 these other time-varying artifacts in the TwoDOS system are discussed.

3.4.1 Adaptive Data-Aided Parameter Estimation for the Channel Characterization

In this section, the channel model parameters are estimated given the model size pair (I, R) (the data-dependence window length I and the noise correlation length R), the known data sequences $\mathbf{a}_0 \dots \mathbf{a}_N$ and their associated waveforms $\mathbf{y}_0 \dots \mathbf{y}_N$, with $N \gg 2^{I+1}$. The waveforms are experimental replay signals that are assumed to be synchronous with respect to the baud rate. The classical approach to estimate model parameters is to compute means and covariance matrices and to calculate the model parameters $s^l(\mathbf{B}_k^l)$, \mathbf{F}^l , $\sigma^l(\mathbf{B}_k^l)$ and $d_{i,j}^l(\mathbf{B}_{k-j}^i)$ (for $i \in [0, L-1]$ and $j \in [1, R]$) based on these matrices, involving complex matrix operations (known as the Yule-Walker equations [157]). Because of these complex operations, the classical approach is numerically not very appealing. For this reason a method to adaptively track the model parameters directly on the experimental signals, is proposed. Moreover, tracking these parameters adaptively is preferable in many applications because the noise may be non-stationary and the channel response may change over time.

The error signal e_k^l of track l at time k is defined as

$$e_k^l = y_k^l - \hat{s}_k^l = y_k^l - \left(\hat{s}^l(\mathbf{B}_k^l) + \mathbf{C}_k^{lT} \tilde{\mathbf{F}}_k^l \right). \quad (3.20)$$

At every clock cycle the signal LUT is easily updated according to

$$\hat{s}^l(\mathbf{B}_k^l)^{(new)} = \hat{s}^l(\mathbf{B}_k^l)^{(old)} + \mu e_k^l, \quad (3.21)$$

where μ denotes the adaptation constant. The RISI impulse response coefficients $\tilde{f}_{i,j}^l$ (for $i \in \{0, L-1\}$, $j \in \{-I, +I\}$ and only if $a_j^i \in \mathbf{C}_k^l$) are updated according to

$$\tilde{f}_{i,j}^l{}^{(new)} = \tilde{f}_{i,j}^l{}^{(old)} + \mu e_k^l a_j^i. \quad (3.22)$$

Equations (3.21) and (3.22) together define the updates for the ISI model parameters. For the updates of the parameters of the AR noise model a different error signal needs to be defined, namely the whitened error signal \mathcal{N}_k^l , already introduced by Eq. (3.14), given by

$$\mathcal{N}_k^l = e_k^l - \sum_{i=0}^{L-1} \sum_{j=1}^R \tilde{d}_{i,j}^l(\mathbf{B}_{k-j}^i) e_{k-j}^i. \quad (3.23)$$

A scheme to estimate and to track the filter coefficients $\tilde{\mathbf{D}}^l$ can be based on minimizing the cost function $J^l = \mathbb{E}[(\mathcal{N}_k^l)^2]$. The LMS solution for this minimization is to adjust the filter coefficient $\tilde{d}_{i,j}^l$ (omitting the data-dependency of the filter coefficients) as follows:

$$\tilde{d}_{i,j}^l{}^{(new)} = \tilde{d}_{i,j}^l{}^{(old)} + \mu \frac{\delta \mathcal{N}_k^{l2}}{\delta \tilde{d}_{i,j}^l}. \quad (3.24)$$

The LMS solution leads to the following update rule for the noise filter coefficients

$$\tilde{d}_{i,j}^l(\mathbf{B}_{k-j}^l)^{(new)} = \tilde{d}_{i,j}^l(\mathbf{B}_{k-j}^l)^{(old)} + \mu \mathcal{N}_k^l e_{k-j}^i. \quad (3.25)$$

These rules are applied at every clock cycle, for every track l and for all coefficients defined by $j \in [0, L-1]$ and $i \in [1, R]$. Subsequently the estimation of the data-dependent variances $\tilde{\sigma}^l(\mathbf{B}_k^l)^2$ is achieved by estimating $(\mathcal{N}_k^l)^2$. The update rule for the variance value $\tilde{\sigma}^l(\mathbf{B}_k^l)^2$ is given by

$$\tilde{\sigma}^l(\mathbf{B}_k^l)^2{}^{(new)} = \tilde{\sigma}^l(\mathbf{B}_k^l)^2{}^{(old)} + \mu \left(\mathcal{N}_k^{l2} - \tilde{\sigma}^l(\mathbf{B}_k^l)^2{}^{(old)} \right). \quad (3.26)$$

The equations (3.21), (3.22), (3.25) and (3.26) together define the updating rules that are applied for the adaptive estimation of the different model parameters. By using these updating rules time-varying effects in the experimental TwoDOS systems can be monitored and if needed countermeasures can be taken.

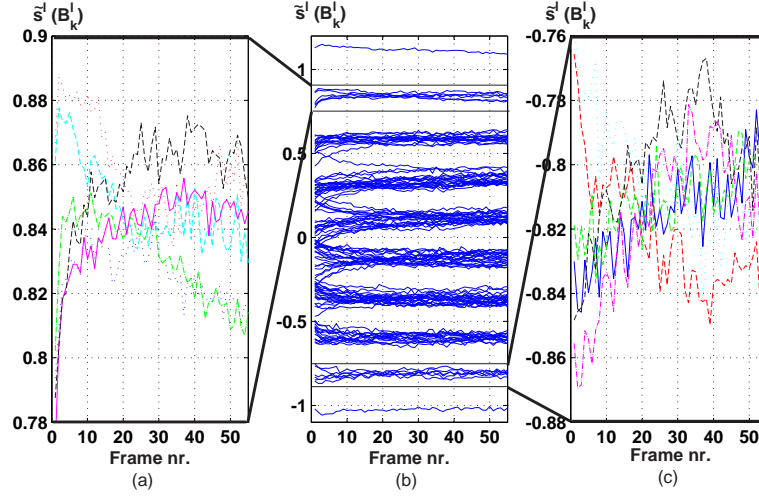


Figure 3.20: Time evolution of the estimated ISI signal values for different clusters: (a) the clusters with a central land and one pit in the first shell (6 possibilities), (b) all 128 different clusters, (c) the clusters with a central pit and 5 pits in the first shell (6 possibilities).

ISI Characterization In Fig. 3.20 the estimated ISI signal values $\tilde{s}(\mathbf{B}_k^l)$ are shown versus frame number. The values are obtained for experimental TwoDOS replay signals. In the TwoDOS disc format every frame consists of 2560 bit intervals. The first 10 frames are needed for acquisition of the parameter values, while from that point in time one can assume that the estimated parameter values equal the actual values. In the adaptive estimation a suitable adaptation constant is used: $\mu = 0.01$. This adaptation constant is a trade-off between acquisition speed (or equivalently tracking speed) and gradient noise.

In Fig. 3.20 (b) the evolution for all 128 different clusters is shown. The average signal levels are consistent with the levels shown in Fig. 3.8. In Fig. 3.20 (a) and (c) detailed snapshots for specific cluster types are shown, respectively for clusters with a central land and a single pit in the first shell (cluster number 1 in the convention of Fig. 3.5) and for clusters with a central pit and a single land in the first shell (cluster number 12). For each of these cluster types 6 different clusters exist. After initial acquisition substantial amplitude variations (up to ± 0.05 in value) can be observed. These variations are real fluctuations of the ISI signal values and are not due to gradient noise as the variance of the gradient noise can be calculated to be 0.00005 (see Eq. (6.3): $\sigma_\Delta^2 \simeq K_t \sigma_v^2 / 2$ where σ_Δ^2 is the variance of the gradient noise, $K_t = \mu = 0.01$ and $\sigma_v^2 = E[e(k, l)^2] < 0.01$ represents the noise in the estimation loop). If these variations are not counteracted or not taken into account during detection, the BER of the

system could severely deteriorate as they can be regarded as RISI at the detector input (see Chapter 5 for the impact of RISI on BER). For this reason the continuously adapted LUT is used as target response of the Viterbi detector (see Section 3.2.5, more precisely Fig. 3.11).

Noise Characterization In Fig. 3.21 (a) estimated noise variances for different clusters are shown versus the frame number (every frame consists of 2560 bit intervals). Experimental replay signals are used where 1.0° of radial tilt was present during the read-out. Four different clusters are taken into consideration: the all-land cluster (number 0), the all-pit cluster (number 127) and two randomly chosen clusters (numbers 42 and 70). The cluster numbers are obtained by convolving the bits in \mathbf{B}_k^l with the mask shown in Fig. 3.21 (b). By averaging the depicted variances over time,

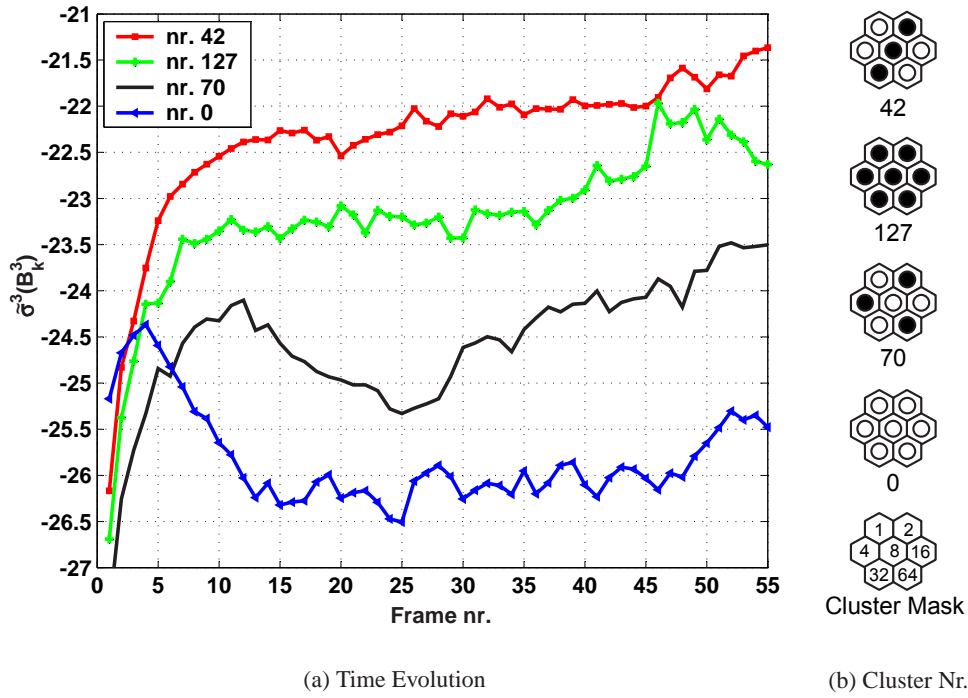


Figure 3.21: (a) Time evolution of estimated noise variances of the central track ($l = 3$) for different clusters. A 50 GB TwoDOS disc is read out with 1.0° of radial tilt. (b) The different clusters for which the time evolution is shown. Here cluster number is defined as the output of the convolution of \mathbf{B}_k^3 with the cluster mask.

the variances shown in Fig. 3.16 (c) are obtained.

For these clusters the variations amount up to 2 dB. For these estimation loops it is more difficult to calculate which part of the variations are due to real variations of the noise sources or are due to statistical artifacts in the estimation loop. As a small adaptation constant μ is used, variations within a dB can be caused by statistical artifacts but larger variations most likely cannot be. Variations of more than 1 dB have a severe impact on the overall BER of the system. Hence it would be beneficial to discover the source of these variations such that countermeasures can be taken. As already pointed out in Section 3.3.2 the data-dependent AR noise model does not intuitively reveal which portion of the variance is due to which noise source. In Chapter 4 a noise characterization model is developed that is directly related to physical noise sources.

3.4.2 Time-varying Channel Artifacts

In the previous section, several time-varying results of the channel characterization have been shown. Because the channel characterization is performed on the detector input signal, several time-varying artifacts have already been accounted for by the adaptive signal processing blocks in the TwoDOS receiver. As a result, to envision all the time-varying artifacts, the control signals of the adaptive blocks should be monitored and interpreted.

In Fig. 3.22 a model of the TwoDOS system is given, with a detailed view of the receiver. The bit sequence \mathbf{a}_k is the input of the system. The physical channel (denoted as channel in the figure) transforms the input sequence into a continuous replay signal $\mathbf{r}(t)$. This physical channel consists of the disc, the optical path and the Photo-Detector Integrated Circuit (PDIC) and is characterized by the channel response $\mathbf{h}(\cdot, \theta)$, which can be nonlinear and time-varying. Several channel parameters, denoted by the symbol θ , determine the overall channel response. Examples of channel parameters are the pit size on the disc, the rotation speed of the disc, the tilt angle of the disc with respect to the laser beam,... The continuous replay signal $\mathbf{r}(t)$ is digitized at a sampling frequency $1/T_s$, which is approximately 1.25 times the baud rate $1/T$, by the Analog to Digital Convertor (ADC). The resulting digitized signal \mathbf{r}_n is input of the receiver part of the system. The receiver performs detection on synchronous replay signals using a PRML scheme in the form of a Viterbi detector. A 2-D target response \mathbf{g}_k of limited size is specified which is the desired impulse response at the detector input. In order to obtain bit-synchronous replay signals \mathbf{y}_k that comply with \mathbf{g}_k , the receiver is composed of several signal processing blocks, each with a specific function. First, the adaptive equalizer transforms the channel response into the target response. Second, DC and gain control blocks normalize the DC and amplitude levels as well as possible. Finally the Sampling Rate Convertor (SRC) resamples the signals to the correct frequency and phase. The resulting signal \mathbf{y}_k is

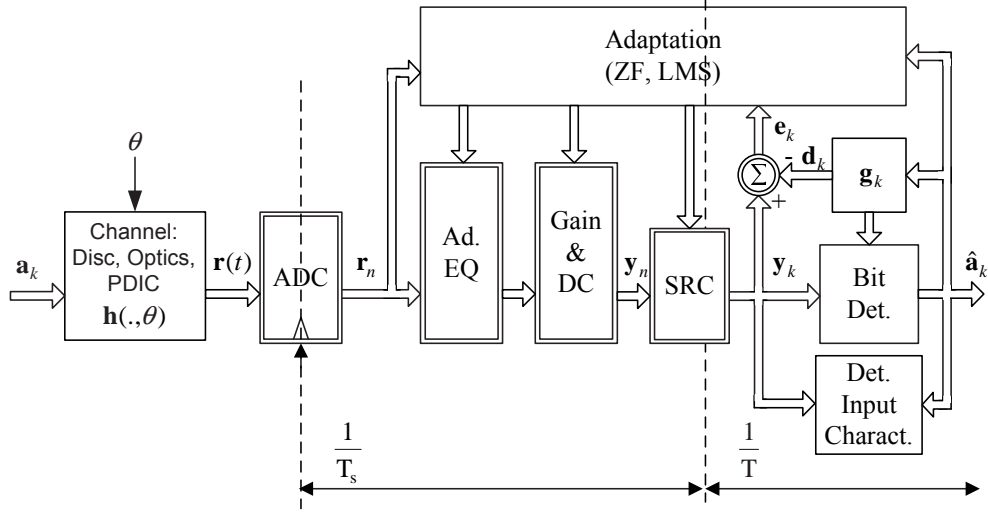


Figure 3.22: Model of the TwoDOS system with a detailed view of the receiver.

the input of the detector and is also the input of the system characterization block (denoted as detector input characterization in the figure) which is described earlier in the chapter.

In order to be able to compensate for varying channel parameters the signal processing blocks are made adaptive. The adaptation is based on the Zero-Forcing (ZF) or on the Least Mean Square (LMS) criterion; they both minimize the error signal \mathbf{e}_k which is the difference between the actual detector input \mathbf{y}_k and the desired detector input \mathbf{d}_k , where $\mathbf{d}_k^l = (\mathbf{a} * \mathbf{g}^l)_k$.

In the scheme described above, slowly varying channel parameters can be successfully tracked and compensated by the adaptive equalizer. Although this compensation can allow reasonable bit-error rates, it is desirable to derive real-time information about some relevant channel parameters, such as the direction and amount of disc tilt, focus conditions and timing misadjustments between different tracks. Because of the adaptive nature of the receiver, variations of channel parameters θ are not visible in the detector input characterization results but they appear as variations in control information in the adaptation block. For example, cover-layer thickness variations would appear in the receiver mainly as DC variations.

DC Variations The DC adaptation loops serve to counteract time-varying DC-offsets [145]. Here DC-offset estimates ϕ_O^l for every track l are generated by separate adaptation loops where the gain values K_r (see Section 2.5.4) are chosen such that ϕ_O^l is able to track fast variations of the ideal DC-offset values θ_O^l (in the experimental

estimates $K_t = 0.06$ was found to be a proper value). Because $G_\phi(z)$ has unit amplitude up to the normalized loop cut-off frequency ($\Omega_c \approx 0.01$), the spectral content of ϕ_O^l resembles the spectral content of θ_O^l up to Ω_c (if noise is neglected). The spectral content of the inner-tracks DC-offset estimate ϕ_O is shown in Fig. 6.6. This estimate ϕ_O is composed of different components which are also shown in the figure. The common DC-offset component ϕ_{Oc} is calculated by averaging the DC-offsets over all tracks. The track-dependent DC-offset component ϕ_{Or} is obtained by subtracting ϕ_{Oc} from ϕ_O .

For low frequencies the track-dependent component ϕ_{Or} is the most important component of ϕ_O . At higher frequencies ($\Omega = [7e^{-4}, 2e^{-3}]$) the common component ϕ_{Oc} determines ϕ_O . For even higher frequencies the offset estimate is determined by the noise in the system (denoted as v in the figure). The common offset component ϕ_{Oc} can be explained by the fact that certain channel parameters (e.g. the amount of defocus and the cover-layer thickness) are common across the adjacent tracks. For the TwoDOS system, the cover-layer thickness exhibits variations that extend over a limited amount of bits (100-1000 bits). As a result these variations result in high-frequency common offset variations. Other reasons for fast common channel parameter variations are: dust, fingerprints, scratches on the disc, dropouts... [173, 174].

Phase Variations The sampling phase used in the SRC can be used to estimate lattice distortion that is present due to a time varying relative phase between the tracks. Such a phase variation is likely because of the multiple-pass mastering process as discussed in Section 2.1.3. An estimation of the peak-to-peak phase variations can be derived by plotting the difference in sampling phase between two adjacent tracks as function of time. This is done in Fig. 3.24 for experimental TwoDOS replay signals obtained from the LBR disc 12 (with an equivalent density of 50 GB). This figure shows that the peak-to-peak delay variation between the tracks is in the order of 10% of a bit interval.

Equalizer-Based Parameter Estimation In the receiver, slowly varying channel parameters can be tracked and compensated to a certain extent via adaptive equalization. In [175] a general scheme is proposed that, based on a processing of the equalizer-coefficient values, permits simultaneous estimation of a given set of channel parameters. This scheme utilizes a mask \mathbf{B} such that projecting the equalizer coefficients onto \mathbf{B} leads to zero in absence of channel parameter variations, and does provide a non-zero estimate in presence of a non-zero parameter value.

The design of the mask \mathbf{B} suited for different channel parameters is described in [176]. For a channel parameter θ (e.g. radial tilt), the procedure to design the corresponding mask \mathbf{B}^θ can be summarized as follows. For $\theta = 0$ the equalizer \mathbf{w} is

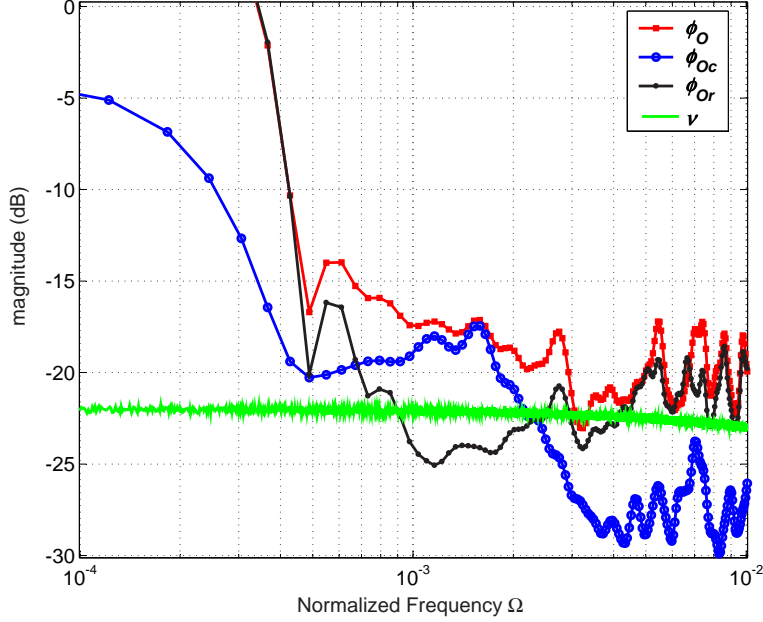


Figure 3.23: Experimentally estimated spectral content of DC-offsets in TwoDOS system: the DC-offset estimate ϕ_O of the inner track and its different components: ϕ_{Oc} and ϕ_{Or} . Also the spectrum P_v of the noise v is shown.

given by the nominal equalizer \mathbf{w}_n . For small values of θ around zero the first-order Taylor approximation yields

$$\mathbf{w}^\theta = \mathbf{w}_n + \theta \mathbf{B}_0, \quad (3.27)$$

meaning that the equalizer impulse response in the presence of a value θ can be approximated as the nominal one, plus a correction term that is proportional to the parameter value. A tilt estimate $\xi(\theta)$ can be produced by projecting the equalizer coefficients \mathbf{w}^θ onto \mathbf{B}_0 according to

$$\tilde{\theta} = \langle \mathbf{w}^\theta | \mathbf{B}_0 \rangle = \langle \mathbf{w}_n | \mathbf{B}_0 \rangle + \theta \|\mathbf{B}_0\|^2, \quad (3.28)$$

where the first term is constant and the second is proportional to the amount of tilt. In principle this projection can be carried out over any \mathbf{B} non-orthogonal to \mathbf{B}_0 . This design freedom can be exploited to make the estimation of one channel parameter orthogonal (or in other words, insensitive) to the estimation of other channel parameters. More details about the orthogonalization with respect to other channel parameters can be found in [175].

Here, the described design procedure was used to generate a mask for the estimation of radial and tangential tilt. The resulting masks were used to estimate tilt

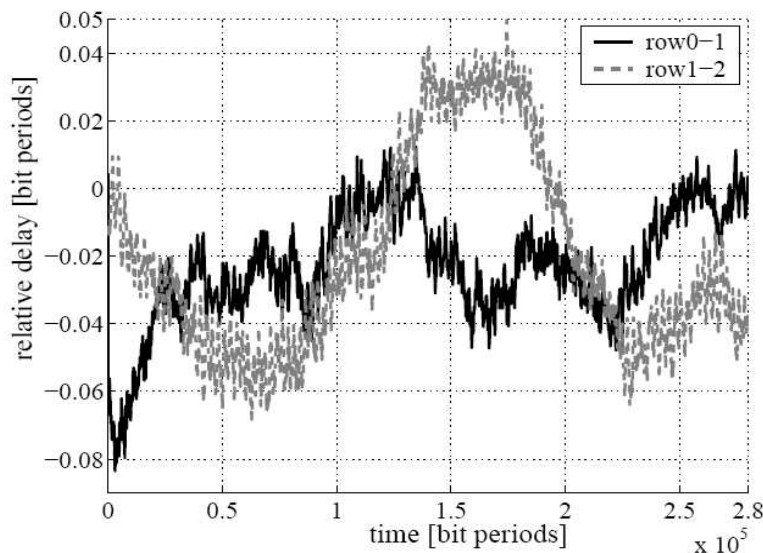


Figure 3.24: Variation in relative delay between adjacent tracks in the broad spiral.

variations in the experimental TwoDOS system. In Fig. 3.25 the time evolution of the estimates of radial tilt is shown for different experimental signals. The different experimental signals are obtained by mechanically invoking different tilt angles in the optical player. For all these signals, tilt variations of 0.3° peak-to-peak can be observed. As is well known (see Fig. 3.11) tilt variations have a severe impact on system performance and as a result some additional measures against tilt variations might have a positive impact on system performance.

3.5 Conclusions

In this chapter a signal-dependent model for multi-track storage systems was presented with (nonlinear) intersymbol interference and signal-dependent and correlated media noise. Several attractive features of this model were demonstrated: simplicity, straightforward parameter estimation and a direct relationship with receiver optimization. The experimental TwoDOS system was used to gauge the accuracy of the model.

The model has several distinct features. 1) It is conceptually simple and computationally efficient. 2) Estimation of the model parameters can be based on a simple data-aided adaptive scheme which achieves a high accuracy and moreover is able to track channel variations. 3) Agreement with experimental data is very good not only for second-order statistics but also for bit-error rates.

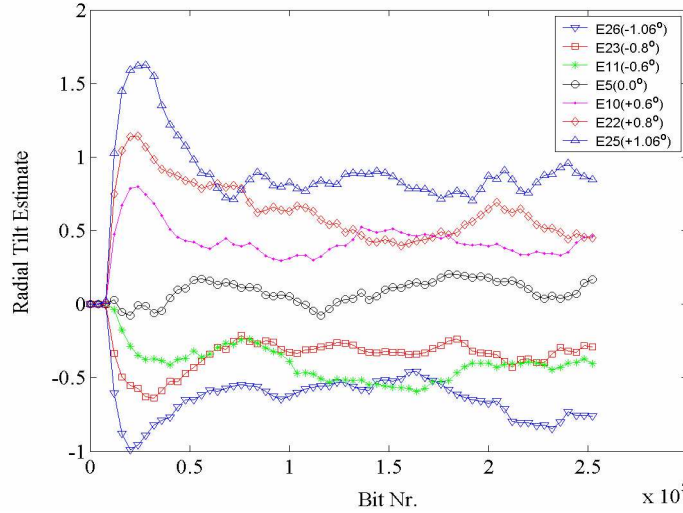


Figure 3.25: Radial tilt estimates versus bit number for different experimental conditions. Different radial tilt angles were mechanically invoked in the optical player.

The model was developed for the experimental TwoDOS system. Based on the characterization results, three main observations can be made.

- *Media noise.* In the TwoDOS system many different noise sources are present. Based upon the noise characterization, it was clearly demonstrated that pit-size noise is the dominant noise source. In general, media noise is highly data-dependent and can severely deteriorate system performance. As a result an accurate characterization of media noise (composition, time variations...) would be beneficial to improve the performance of the TwoDOS system. This media noise characterization is treated in Chapter 4.
- *Residual ISI.* An adaptive equalizer is used in the TwoDOS receiver to transform the ISI structure induced by the channel into an ISI structure with a limited span as defined by the target response. This limited span is needed to limit the detector complexity. However based upon the ISI characterization results, it was shown that for the TwoDOS system still considerable amounts of residual ISI (RISI) originate from bits outside this span. This RISI is highly undesired as it is not accounted for by the detector. As a result a proper handling of RISI would be very beneficial for the TwoDOS performance. This proper handling is the topic of Chapter 5.
- *Time Variations.* Throughout this chapter, a data-aided adaptation scheme was

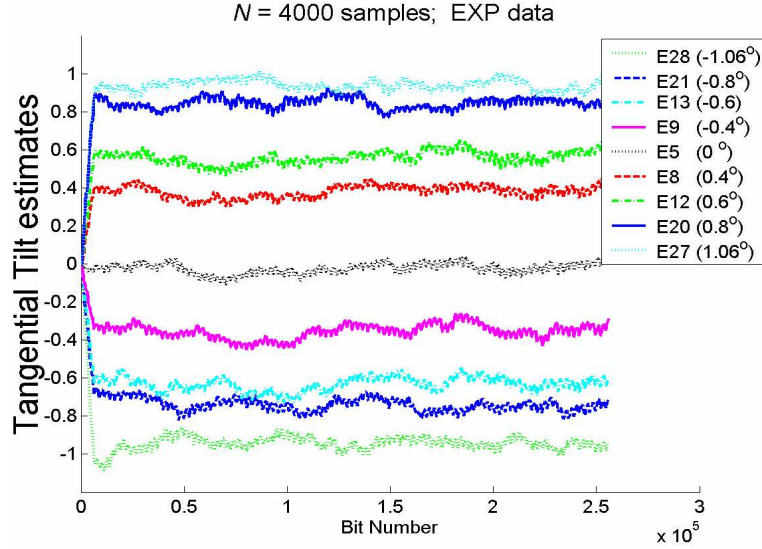


Figure 3.26: Tangential tilt estimates versus bit number for different experimental conditions. Different tangential tilt angles were mechanically invoked in the optical player.

used for the estimation of model parameters. The adaptive nature of the estimation makes it possible to easily monitor variations in the parameter values. These variations provide important information about time-varying artifacts in the experimental system. In the TwoDOS system many different time variations are present: going from slowly-varying disc tilt to rapidly-varying DC offsets. Especially rapidly-varying parameters might cause problems in the TwoDOS receiver. A key reason for these problems is that the adaptation loops are not able to track fast variations because of the latency in the loops caused by the 2-D bit-detector. In Chapter 6 a solution is proposed that is based on efficiently using minimum-latency control in the loops.

Chapter 4

Adaptive Decomposition of Noise Sources in Digital Storage Systems with Media Noise.

In digital recording systems the total amount of data-dependent media noise increases considerably as recording densities increase. A proper noise characterization is crucial for the design of receivers for high-density storage systems. This characterization involves the selection of a proper noise model and subsequently the accurate estimation of the parameters of the selected model. The estimation algorithm proposed in this chapter, jointly estimates the parameters of both media and additive noise with a high accuracy. The proposed algorithm makes use of the data-dependency of the media noise to distinguish between the different noise sources. The algorithm is simple and as a result only very limited amount of complexity is required to implement it in recording systems as an easy "add-on" to read channels ICs. Based on the simulation results and the analytical derivation of the estimation algorithm, one can clearly indicate which data patterns yield near-optimal estimation performance. These patterns are the ideal test patterns in experimental systems: examples of test patterns for magnetic and optical storage systems are proposed and discussed.

4.1 Introduction

In many storage applications, the noise present in the read-out signal can be modelled as a combination of media noise and electronics noise (here referred to as additive noise). Media noise is typically associated with the write process whereas additive noise is associated with the read-out process [177]. Because the write process is determined by the input data, media noise is strongly data-dependent. It is well known that the presence of data-dependent noise significantly deteriorates the error-rate performance of partial-response maximum-likelihood (PRML) data detection [178–181]. Various codes and detectors have been proposed to improve the performance based on specific media-noise models [170, 179, 182]. In an actual

system, the performance improvement is only guaranteed if the noise is characterized properly. A proper characterization involves the selection of a model with a proper structure. Furthermore it involves an accurate estimation of the model parameters based on experimental read-out signals. Ideally, the models must be computationally efficient and therefore simple parametric models are commonly used. Parametric models with one and more parameters have been proposed for magnetic systems [154, 155, 158, 159] and for optical systems [3, 160]. These models are based on the physical nature of the media noise. Hence the knowledge of the parameter values provides important information for the development and evaluation of recording media as well as detectors for next generation disk drives. All these parametric models can be considered to be special cases of the signal-dependent autoregressive (AR) channel model [157, 183]. In this model the media noise is the output of an AR system driven by white noise and whose parameters are signal-dependent (see Section 3.3.2). The signal-dependent AR model can be transformed by using complex matrix operations into the parametric models that relate well with the physical nature of the media noise [184]. In this chapter parametric models will be used for two reasons: only a few parameters have to be estimated (which results in a limited overall complexity), and the parameter values immediately reveal insightful information about properties of the physical channel. In practice it is highly desirable to be able to characterize the media noise parameters simultaneously with those of the additive noise, i.e. to decompose the noise components of the replay signals into media and additive noise components.

In general, the estimation of parameter values is based on fitting the selected models to read-out signals. In experimental systems one has the freedom to design test patterns and the ability to base the estimation of the parameters on different reads and writes of the test patterns. As a result a lot of different read-out signals are available to base the characterization on, and there is a large flexibility in the design of a characterization procedure. Many existing characterization procedures make use of this flexibility by tailoring the procedure to the specific experimental storage application. For example, to achieve a proper decomposition of the noise sources the write noise is extracted from the read-out signals by appropriately averaging over many reads and writes of a periodic pattern [177].

The topic of this chapter is the design of a highly general characterization procedure which is applicable to real-time systems. In real-time systems the constraints are very different than in experimental systems. The large flexibility available in experimental systems (multiple reads and writes of predefined test patterns) is in great contrast with the single read-out of a-priori unknown patterns available in real-time systems. Hence in real-time systems much less flexibility is available to design a proper decomposition procedure. Moreover computational complexity weighs more heavily. In this chapter a decomposition procedure is proposed that operates on a

single read-out of an arbitrary pattern and is able to properly decompose the noise sources present in a real-time storage systems. The procedure is able to jointly estimate the parameter values of media and additive noise sources by exploiting the data-dependent nature of the media noise. It is computationally simple and very suitable for incorporation as an “add-on” to existing read-channel integrated circuits (ICs).

In any characterization procedure, estimation of the parameter values can be based on spectral or on temporal techniques or on a combination of both [185]. The first class of techniques, often used in experimental environments, involves measuring noise-amplitude spectra for specific test patterns (for magnetic applications, DC and block wave patterns are commonly used) [185, 186]. An advantage of these techniques is the fact that synchronization is not needed. As a result the analog read-out signal can be used as input of a spectrum analyzer and little or no additional signal processing is needed. Furthermore parameter estimates are obtained by fitting the model to approximate the measured noise spectra. Fitting techniques are often based on the minimization of a suitable criterion or on exhaustive search algorithms, both resulting in a high estimation accuracy [187]. Spectral techniques often rely on highly-oversampled read-out signals obtained around isolated bit transitions or pairs of transitions [154, 158].

In modern read channels, the analog read-out signals are sampled, equalized and synchronized to the baud-rate. As a result it is convenient to base the characterization on the equalized, synchronized, detector input signal [177, 188]. The second class of estimation techniques (temporal techniques) can use this synchronized signal. In these techniques the noise correlation matrix is measured based on the synchronized signal and subsequently key portions of this matrix are used to estimate different noise parameters [189, 190]. Temporal techniques are sometimes used in experimental environments where equalized, synchronized signals are available because measuring and fitting the noise correlation matrix are operations with little complexity and still a high estimation accuracy is achieved. In principle temporal techniques can be applied in real-time environments.

The decomposition procedure proposed in this chapter is based on such a temporal technique. More precisely, accurate parameter estimation is achieved based on the open-loop estimation of the ensemble variances of noise samples as a function of the data pattern under consideration. The estimates of the parameter values are obtained by solving a matrix equation with the estimated ensemble variances as input. As opposed to the solution for periodic data patterns presented in [190], it is applicable to arbitrary patterns and therefore it is very well suited for real-time systems. Moreover, our solution is also applicable to experimental systems and we will see that the estimation accuracy is similar to the one achieved by spectral techniques.

The complexity required to calculate the open-loop solution, is quite significant

and as a consequence it is a disadvantage of the open-loop estimation technique. Hence a simplification is desired for implementation in real-time environments. A closed-loop version of the procedure provides such a simplification. This closed-loop procedure results in an adaptation algorithm which estimates parameter values on a sample-by-sample basis and requires only a very limited amount of complexity. Besides this limited complexity, another advantage of the adaptation algorithm is the ability to track variations of media noise parameters. This tracking ability provides information about time-varying artifacts in the storage system and hence gives insight in possible improvements in the recording media and/or read channel.

The adaptation algorithm can be utilized to characterize various types of media noise. In this chapter a model for pit size variations in optical storage systems is used to illustrate the design of the algorithm. In optical storage systems, media noise is due to imperfections in the optical layer of the optical medium, e.g. variations in the dimensions of the pits [18, 191]. This media noise passes through the read-out by the optical spots. Hence we can say that this media noise originates at the channel input. For example, in CD-RW systems binary input symbols are stored as pits (the laser current is switched on such that a phase change is realized in the recording layer) or lands (no current), only the pits are affected by media noise and not the lands. As a consequence for such systems media noise can be modelled as noisy pits and noise-free lands. Hence the media noise is data-dependent. In this chapter we will assume that the media noise samples originate from a white Gaussian stochastic process and that the additive noise samples originate from a correlated Gaussian stochastic process. Exploiting the data-dependency of the media noise, the adaptation algorithm decomposes the different noise sources and provides an estimate of the different parameters, i.e. it performs a joint estimation of the parameters of the media noise and of the additive noise. The analysis and results for this simple optical media noise model can be found in Section 4.2. Because the adaptation algorithm is based on the data-dependency of the media noise and not on any assumption about the media-noise correlation, the algorithm can be extended to cover also correlated media noise.

In Section 4.3, an analysis similar to the one presented in Section 4.2 is given for the characterization of media noise in magnetic storage systems. In magnetic storage, position-jitter is considered to be the dominant media noise effect [70], yet also other types of media noise arise at high densities, like nonlinear transition shifts and width-variations [192]. All these types of media noise sources are data-dependent and as a result the proposed decomposition procedure can also be applied to magnetic storage systems. A theoretical analysis and simulation results are provided to demonstrate the decomposition capabilities and the estimation accuracy of the proposed adaptation algorithm in magnetic storage applications. Based on the simulation results and the analytical derivation of the adaptation algorithms, one can clearly indicate which data

patterns yield near-optimal estimation performance. As a result these patterns are the ideal test patterns to use in experimental environments. In Section 4.4 the design of these test patterns for magnetic and optical systems is discussed. Conclusions are provided in Section 4.5.

4.2 Media Noise in Optical Storage

A system model of an equalized, synchronized optical storage channel is shown in Fig. 4.1. The bits $a_k \in \{0, 1\}$ are inputs of the channel, where k represents the discrete time index, $k = 1 \dots K$ (K is the total number of transmitted bits) and $a_k = 0$ for $k \notin [1, K]$. These bits are distorted by multiplicative noise to produce the channel input signal $x_k = a_k(1 + u_k)$, where u_k is a sequence of white Gaussian noise samples with zero mean and standard deviation σ_u . As a result, the sequence $a_k u_k$ represents the data-dependent media noise: pits ($a_k = 1$) are noisy while lands ($a_k = 0$) are non-noisy. This is a reasonable first-order approximation of media noise in read-write optical media [18]. In this chapter this specific type of data-dependency is used to illustrate the decomposition of two noise sources, but the decomposition technique can be applied to other data-dependencies and more noise sources. The distorted input signal x_k is used as input of the channel with impulse response (IR) $\mathbf{h} = [h_0, h_1, h_2 \dots h_M]$, where M defines the memory span of the channel. Finally a stationary additive noise component v_k is present at the output of the channel. This component represents the electronics noise in the storage system and is characterized by the autocorrelation function

$$R_v(n) = E[v_{k-n}v_k], \quad (4.1)$$

for $n = 0, \dots, N$, where N defines the memory length of the additive noise process. Summarizing, the channel output y_k can be expressed as

$$\begin{aligned} y_k &= (h * a)_k + (h * au)_k + v_k \\ &= \sum_j h_j a_{k-j} + \sum_j h_j a_{k-j} u_{k-j} + v_k, \end{aligned} \quad (4.2)$$

where ‘ $*$ ’ denotes linear convolution. The first component is the output due to the input bits a_k , the second component represents the noise at the channel output due to the media noise $a_k u_k$ and the third component is the additive noise at the channel output. The characterization of the noise sources present in this system model, concerns the joint estimation of the standard deviation σ_u , or equivalently the variance σ_u^2 of u_k , and the autocorrelation function $R_v(n)$ of v_k (Section 4.2.1). A key assumption in the estimation is the knowledge of the model output y_k , the channel IR \mathbf{h} and the bits a_k . In a practical system this knowledge is present at the receiver side: the detector input is readily available and an identification scheme is able to accurately estimate the channel IR (see Section 3.2). Finally also the bits are known: either as a specific preamble pattern which is used to acquire parameter values in acquisition mode of

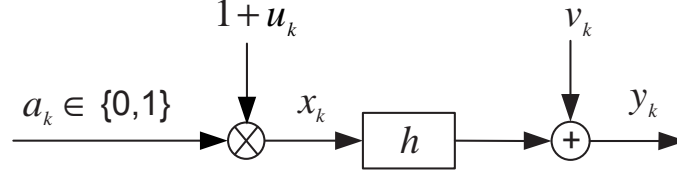


Figure 4.1: Model of an equalized, synchronized, optical channel with media noise and additive noise.

operation (referred to as data-aided, DA, mode) or as the outputs of a bit-detector in tracking mode of operation (referred to as decision-directed, DD, mode). Generally the bit-detector provides fairly accurate bit decisions. However occasional bit decision errors will affect the overall estimation accuracy. In Section 4.2.3 we will show that even for a significant Bit-Error Rate (BER, up to 10^{-2}) the estimation accuracy is just slightly influenced.

4.2.1 Data-Dependent Media Noise Characterization

In this section an algorithm is derived to estimate the parameters σ_u^2 and $R_v(n)$ based on the observed signal y_k and the knowledge of the transmitted data a_k . To distinguish between the two noise sources u_k and v_k , we have to make use of the data-dependent nature of the media noise. An obvious way to distinguish is to first transmit a long sequence of zero bits followed by random data. In the first part, at the channel output only additive noise is present and as a result $R_v(n)$ can easily be estimated. Subsequently during the random data part the media noise variance σ_u^2 can be estimated. Such a procedure is, however, not preferable because the transmission of zero bits decreases the overall information throughput or in some cases it is simply not possible to transmit a sequence of zero bits. For this reason a procedure that is fully based on arbitrary data patterns is desirable. The procedure developed next applies to arbitrary data patterns. As a result it is applicable in case random data is present but also short test patterns can be designed to improve the estimation accuracy.

The error signal e_k is defined as follows

$$e_k = y_k - (h * a)_k = (h * au)_k + v_k \quad (4.3)$$

for $k = 1 \dots K$. This error signal consists of a data-independent part, i.e. the additive noise v_k , and a data-dependent part, namely the media noise at the channel output. Based on a data-dependent averaging procedure the data-independent and the data-dependent parts can be characterized jointly. Because the channel IR has a finite memory length M , i.e. $h_k = 0$ for $k \notin [0, M]$, the data-dependency has also a finite memory length. Now define \mathbf{a}_k^M as the vector $[a_k, a_{k-1}, \dots, a_{k-M}]$. The data-dependent

autocorrelation function of the error signal can be written as

$$\begin{aligned} R_e(n, \mathbf{a}_k^M) &\triangleq \mathbb{E}[e_k e_{k-n} | \mathbf{a}_k^M] \\ &= \mathbb{E}[v_k v_{k-n}] + \sum_{j=0}^M h_j h_{j-n} a_{k-j}^2 \mathbb{E}[u_{k-j}^2] \\ &= R_v(n) + \sigma_u^2 H(n, \mathbf{a}_k^M), \end{aligned} \quad (4.4)$$

where $H(n, \mathbf{a}_k^M) = \sum_{j=0}^M h_j h_{j-n} a_{k-j}^2$. This autocorrelation function can be calculated for $n = 0, \dots, N$ and for all possible data patterns \mathbf{a}_k^M which are part of a set \mathcal{S} . This set \mathcal{S} is defined by the modulation code and consists of all permissible data patterns. As a result a least-squares solution can be formulated for the estimation of $R_v(n)$ (for $n = 0, \dots, N$) and σ_u^2 . This solution is derived in Appendix A and is given by the solution of the following Yule-Walker equations (in matrix formulation):

$$\begin{bmatrix} \mathcal{H} & H(0) & \cdots & H(N) \\ H(0) & 2^M & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ H(N) & 0 & \cdots & 2^M \end{bmatrix} \begin{bmatrix} \sigma_u^2 \\ R_v(0) \\ \vdots \\ R_v(N) \end{bmatrix} = \begin{bmatrix} \mathcal{G} \\ G(0) \\ \vdots \\ G(N) \end{bmatrix}, \quad (4.5)$$

where

$$H(n) = \sum_{\mathbf{a}_k^M \in \mathcal{S}} H(n, \mathbf{a}_k^M), \quad \mathcal{H} = \sum_{\mathbf{a}_k^M \in \mathcal{S}} \sum_{n=0}^N H(n, \mathbf{a}_k^M)^2, \quad (4.6)$$

$$G(n) = \sum_{\mathbf{a}_k^M \in \mathcal{S}} R_e(n, \mathbf{a}_k^M) \text{ and } \mathcal{G} = \sum_{\mathbf{a}_k^M \in \mathcal{S}} \sum_{n=0}^N R_e(n, \mathbf{a}_k^M) H(n, \mathbf{a}_k^M). \quad (4.7)$$

All coefficients $H(n, \mathbf{a}_k^M)$ are only a function of the channel IR \mathbf{h} and the possible bit patterns (within the memory span of the channel), whereas the coefficients $G(n, \mathbf{a}_k^M)$ depend on the data-dependent autocorrelation function of the error signal.

The accuracy of the estimates is determined by the accuracy of the computed error correlation coefficients $R_e(n, \mathbf{a}_k^M)$. As there are $(N+1) \times 2^M$ different coefficients for uncoded data, a long averaging period is required to guarantee good estimation performance. The proposed algorithm achieves a one-shot estimation and therefore it is not able to deal with time-varying noise conditions. Moreover to solve equation (4.5), complex matrix operations have to be performed resulting in a computationally intensive procedure. In the next section a closed-loop adaptive estimation scheme for the noise parameters is proposed which overcomes these disadvantages.

4.2.2 Adaptive Estimation Scheme

The adaptive estimation scheme is based on a stochastic gradient search. The noise parameters are estimated adaptively based on arbitrary data patterns. This approach has the advantage that noise parameter variations can be tracked. The gradient search

aims to minimize the cost function defined by (4.19) (see Appendix B). At every time instant k , a new estimate of the different noise parameters is computed based on the previous estimate and an update term which is the gradient of the cost function with respect to that specific parameter. The adaptation rules are (see Appendix B for the derivation)

$$\begin{aligned}\tilde{\sigma}_u^2(k) &= \tilde{\sigma}_u^2(k-1) - 2\mu_u \left[\sum_{n=0}^N H(n, \mathbf{a}_k^M) \Lambda_k(n) \right], \\ \tilde{R}_v(n)^{(k)} &= \tilde{R}_v(n)^{(k-1)} - 2\mu_v \Lambda_k(n),\end{aligned}\quad (4.8)$$

for $n = 0, \dots, N$, where $\Lambda_k(n) = H(n, \mathbf{a}_k^M) \tilde{\sigma}_u^2(k-1) + \tilde{R}_v(n)^{(k-1)} - e_k e_{k-n}$ is the instantaneous estimation error of the n^{th} correlation lag. Here μ_u and μ_v are the adaptation constants of respectively the media noise update and the update of the additive noise correlation function. A block diagram of the adaptive estimation scheme is shown in Fig. 4.2 for $N = 0$ (white additive noise). The bits a_k and the error signal e_k are the only inputs and as they are readily known in read channels, the adaptive estimation scheme can be used as an “add-on” to existing read channel ICs. The variance estimates $\tilde{\sigma}_u^2(k)$ and $\tilde{R}_v(n)^{(k)}$ of respectively the media and the additive noise are the outputs. All the coefficients $H(n, \mathbf{a}_k^M)$ for $n = 0, \dots, N$ and for all possible \mathbf{a}_k^M , are computed beforehand and stored in a look-up table (LUT).

Steady-State Behavior. The steady-state value of the estimates can be derived easily. For every possible data pattern \mathbf{a}_k^M , the steady state values of the estimates can be found by solving the equations obtained by putting the expectation of the partial derivatives (4.20) equal to zero. By applying (4.4) it can easily be found that $E[\tilde{\sigma}_u^2] = \sigma_u^2$ and $E[\tilde{R}_v(n)] = R_v(n)$ for all data patterns in S .

Furthermore also the variance of the estimates can be calculated for every data pattern. The derivation is given in Appendix C for the case the additive noise originates from a white Gaussian noise source, i.e. $R_v(0) = \sigma_v^2$ and $R_v(n) = 0$ for $n > 0$. The results are:

$$\begin{aligned}\gamma_{\tilde{\sigma}_u^2}^2 &\triangleq E[\tilde{\sigma}_{u(k)}^2] - E[\tilde{\sigma}_{u(k)}^2]^2 \approx 2\mu_u (HR_u(0) + R_v(0))^2, \\ \gamma_{\tilde{\sigma}_v^2}^2 &\triangleq E[\tilde{\sigma}_{v(k)}^2] - E[\tilde{\sigma}_{v(k)}^2]^2 \approx 2\mu_v (HR_u(0) + R_v(0))^2.\end{aligned}\quad (4.9)$$

where $H = H(0, \mathbf{a}_k^M)$ is used to simplify the notation. Based on these expressions the accuracy of the adaptation scheme can be characterized. It should be noted that these expression roughly approximate the actual variances for a number of reasons: first by neglecting all higher-order terms in (4.28) and second by the fact that the adaptation scheme is data-dependent (at every time instant a different coefficient $H(0, \mathbf{a}_k^M)$ should be used). This data-dependency makes the scheme very difficult to analyse accurately. However in Section 4.2.3 we will show that the variances given by (4.9)

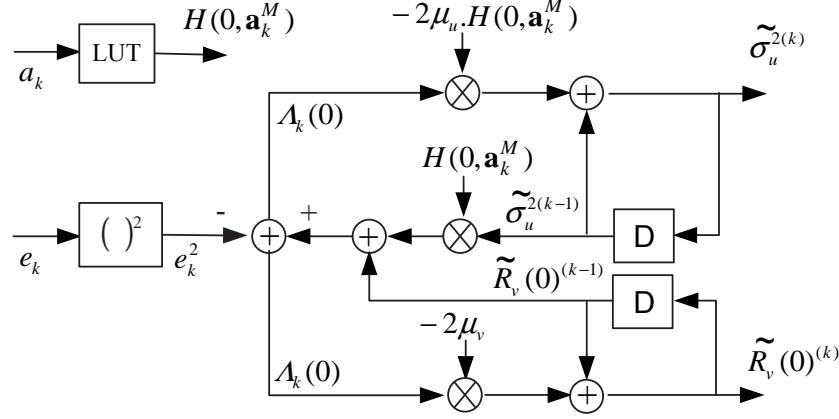


Figure 4.2: Model of adaptive noise variance estimation scheme in case media and additive noise originate from independent white AWGN sources. The first input signal a_k is used to generate the coefficients $H(0, \mathbf{a}_k^M)$. The second input signal e_k together with these coefficients are used to calculate noise variance estimates $\tilde{\sigma}_u^{2(k)}$ and $\tilde{R}_v(0)^{(k)}$.

are reasonably accurate and can be used as a initial design rule for the choice of the adaptation constants μ_u and μ_v .

Example. Assume that the additive noise component v_k originates from a white noise process, i.e. $R_v(0) = \sigma_v^2$ and $R_v(n) = 0$ for $n > 0$. In this case the adaptation rules simplify to

$$\begin{aligned}\tilde{\sigma}_u^{2(k)} &= \tilde{\sigma}_u^{2(k-1)} - 2\mu_u H(0, \mathbf{a}_k^M) [H(0, \mathbf{a}_k^M) \tilde{\sigma}_u^{2(k-1)} + \tilde{R}_v(0)^{(k-1)} - e_k^2], \\ \tilde{R}_v(0)^{(k)} &= \tilde{R}_v(0)^{(k-1)} - 2\mu_v [H(0, \mathbf{a}_k^M) \tilde{\sigma}_u^{2(k-1)} + \tilde{R}_v(0)^{(k-1)} - e_k^2].\end{aligned}\quad (4.10)$$

If a sequence of zero bits is transmitted, the media noise variance is not updated and the additive noise variance is changed according to:

$$\tilde{R}_v(0)^{(k)} = \tilde{R}_v(0)^{(k-1)} - 2\mu_v (\tilde{R}_v(0)^{(k-1)} - e_k^2). \quad (4.11)$$

The latter update rule is the expected one as the error signal e_k does not have a component due to the media noise and is only determined by the additive noise.

Now if random data is transmitted, the coefficient $H(0, \mathbf{a}_k^M)$ is, in general, nonzero. As a result the update of the media noise variance is enabled and the update represents that part of the error power that is not due to additive noise (addition of $\tilde{R}_v(0)^{(k)}$) and not yet present in the previous estimate $\tilde{\sigma}_u^{2(k-1)}$. In the update of the additive noise

variance, the term $H(0, \mathbf{a}_k^M) \tilde{\sigma}_u^{2(k)}$ represents that part of the error power e_k^2 that is due to the media noise.

Extension to multiple media noise sources. In most cases, especially at high densities, more media noise sources are simultaneously present in the read-out signal. The adaptive decomposition procedure defined in this subsection can easily be extended to the case of multiple data-dependent noise sources. For every parameter an individual estimation loop has to be used and the data-dependent coefficients of the loop should be tuned to that specific parameter. If the data-dependencies of the media noise sources are different, the adaptive decomposition procedure will work and no offsets are present in the variance estimates. The variances of the estimates, given by (4.9) for the case a single media noise source is present, will contain a term for every noise source (with the corresponding coefficient $H(0, \mathbf{a}_k^M)$) and as a result the overall variance will increase for increasing number of noise sources.

4.2.3 Simulation results

The functionality of the adaptive estimation scheme is illustrated by simulating the model of Fig. 4.1. An ideally-equalized optical channel is assumed with equalized channel response $h = [0.17 \ 0.5 \ 0.67 \ 0.5 \ 0.17]$, and channel memory length $M = 4$. An uncorrelated input sequence \mathbf{a} is used with $K = 4 \times 10^5$. Furthermore the media noise sequence \mathbf{u} is assumed to be white and Gaussian. The signal-to-noise ratio (SNR) is defined as $\text{SNR} = \sum_j h_j^2 / \sigma_z^2$, where σ_z^2 is the total noise power: $\sigma_z^2 = \sigma_u^2 + \sigma_v^2$. We define media noise percentage (MNP) as the ratio of the media noise power to the total noise power [193]

$$\text{MNP} = \frac{\sigma_u^2}{\sigma_z^2} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}. \quad (4.12)$$

In case the additive noise sequence \mathbf{v} is correlated, the memory length N is in practical situations not known. For this reason N should be chosen large enough to cover all non-zero elements of the autocorrelation function $R_v(n)$. We consider two different cases: $N = 0$ and $N = 2$.

White Gaussian Additive Noise

Here $R_v(n) = 0$ for $n > 0$ and the memory length N is chosen to be zero ($N = 0$). In Fig. 4.3, the estimated parameters $\tilde{\sigma}_u^2$ (upper plot) and $\tilde{\sigma}_v^2 = \tilde{R}_v(0)$ (lower plot) are plotted in dB versus the sample number for $\sigma_u^2 = 0.1$ and $\sigma_v^2 = R_v(0) = 0.1$ (indicated by the dashed lines in the figure). The adaptation constants are tuned such that the time constant τ_e of both estimation loops are equal and a slow but steady convergence is achieved. To this end $\tau_e = 2 \cdot 10^4$ is used. Because on average the media noise variance estimate is updated only half of the time, the adaptation constant μ_u should be a factor of 2 higher than μ_v . The different lines in the plot are the results

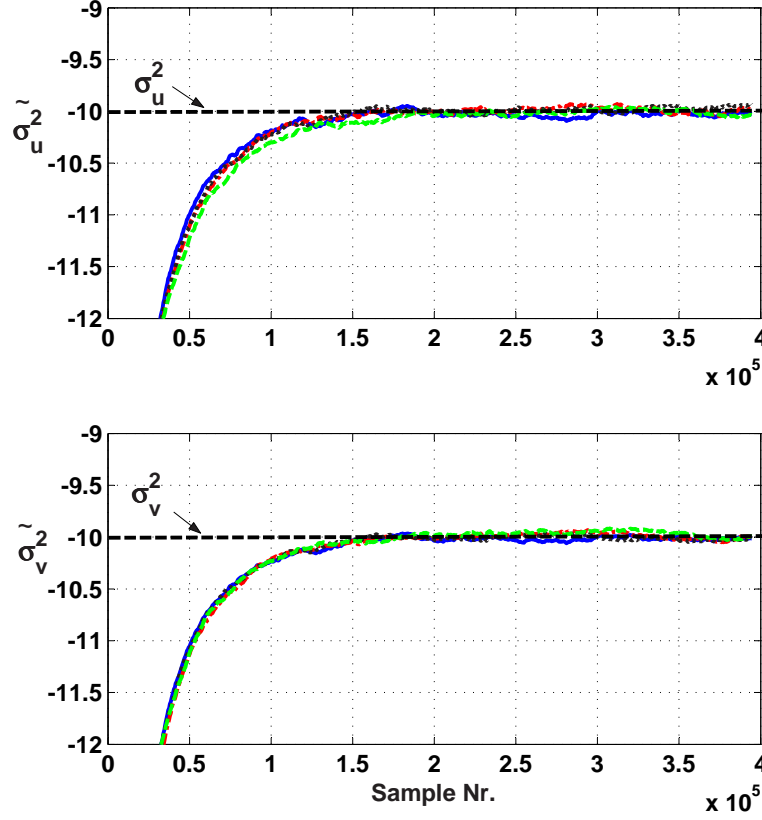


Figure 4.3: Estimated parameter values $\tilde{\sigma}_u^2$ and $\tilde{\sigma}_v^2$ (given in dB) versus the sample number. In the simulation the following values are used: $\sigma_u^2 = 0.1$ and $\sigma_v^2 = R_v(0) = 0.1$, $\mu_u = 1e^{-4}$ and $\mu_v = 5e^{-5}$.

for different noise realizations. Because the control information in the two loops is not completely orthogonal, there is interaction in the estimation of the media noise parameters and the additive noise parameters. This interaction does not introduce a bias in the estimates but influences the convergence to their steady-state values. In steady-state operation, the estimated parameter values do not deteriorate more than 0.02 dB from their actual values for the given time constant.

In practical situations it will be desirable to have an estimate of the overall SNR and an estimate of the MNP, respectively denoted as $\tilde{\sigma}_z^2$ and $\tilde{\text{MNP}}$. These estimates can easily be estimated based on (4.8): $\tilde{\sigma}_z^{(k)} = \tilde{\sigma}_u^{(k)} + \tilde{R}_v(0)^{(k)}$ and $\tilde{\text{MNP}}^{(k)} = \tilde{\sigma}_u^{(k)} / \tilde{\sigma}_z^{(k)}$.

To judge the accuracy of the adaptive estimation scheme, we introduce the Normalized Estimation Accuracy (NEA). The NEA is defined as the deviation in es-

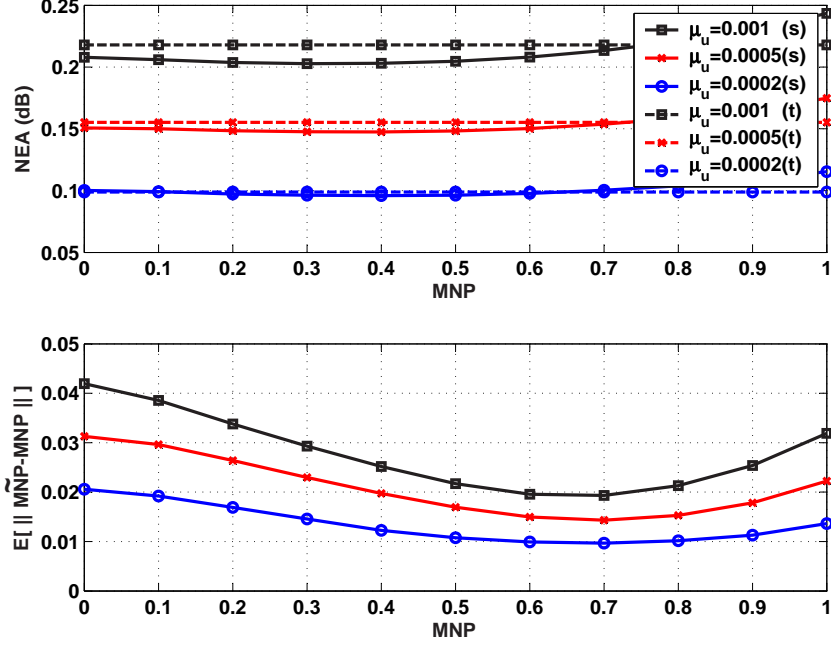


Figure 4.4: a) Theoretical (t) and simulated (s) NEA values (given in dB) versus MNP. b) Simulated MNP estimation accuracies $E[||\tilde{\text{MNP}} - \text{MNP}||]$ versus MNP. SNR = 25dB. Three different adaptation constants are used in the simulations: $\mu_u = 0.001$, $\mu_u = 0.0005$ and $\mu_u = 0.0002$; and $\mu_v = \mu_u/2$.

estimated SNR (in dB) in case the estimate deviates one standard deviation from its mean value, irrespective of the actual SNR value. More precisely the NEA can be expressed as:

$$\text{NEA} = 10 \log_{10} \left(1 + \frac{\gamma_{\tilde{\sigma}_z^2}}{\sigma_z^2} \right), \quad (4.13)$$

where $\gamma_{\tilde{\sigma}_z^2}$ is the standard deviation of the SNR estimate. If we assume that the estimates $\tilde{\sigma}_u^2$ and $\tilde{\sigma}_v^2$ have independent normal distributions (in reality this is not completely true but it provides a good approximation), this standard deviation can be calculated based on $\gamma_{\tilde{\sigma}_z^2}^2 = \gamma_{\tilde{\sigma}_u^2}^2 + \gamma_{\tilde{\sigma}_v^2}^2$. The NEA value indicates the accuracy of the estimation with respect to the overall noise power.

In the upper part of Fig. 4.4, simulated and theoretical NEA values are plotted versus MNP for different values of the adaptation constants μ_v (again $\mu_u = 2\mu_v$) while SNR is fixed at 25 dB. The simulated values are obtained by calculating its expected value over all samples after the estimation loops have converged. The theoretical

values are calculated according to (4.13) and (4.9), where in the latter the maximum value of $H(0, \mathbf{a}_k^M)$ is used to account for worst case noise conditions.

From (4.9) it should be clear that the NEA is mainly determined by the square root of the adaptation constant. If the time constant is increased by a factor of 2, the NEA will decrease by a factor of $\sqrt{2}$ and visa versa. The NEA indicates what is the accuracy of the estimation. For example, the NEA for $\mu_u = 0.0002$ is approximately 0.1, i.e. the estimated SNR level has a standard deviation of 0.1 dB from its actual value. This means that you have a probability of 95% that the estimated SNR is within 0.2 dB (equal to 2 standard deviations) from its actual value.

In the lower part of Fig. 4.4, the standard deviation of the estimated MNP is plotted versus MNP for different values of μ_u . For example for $\mu_u = 0.0002$, this standard deviation is smaller than 0.02. This means that you have a probability of 95% that the estimated MNP is within 0.04 of its actual value. From this figure, it should be clear that the standard deviation of the MNP estimate is sensitive to the media noise vs. additive noise mix in the read-out signal. The best accuracy is achieved in case the magnitude of the noise sources are comparable.

As a conclusion from this plot, it can be stated that sufficient estimation accuracy is achieved in case the time constant τ_e is chosen sufficiently large.

Correlated Additive Noise

we consider additive noise with an autocorrelation function $R_v(1) = 0.47 \times R_v(0)$ and $R_v(2) = 0.19 \times R_v(0)$, together with a media noise source. The memory length N is chosen to cover the memory length of the autocorrelation function, i.e. $N = 2$. In Fig. 4.5 the estimated noise parameters (the full lines) are plotted versus the sample number. Also their actual values are indicated in the figure by the dashed lines. In this example $\tau_e = 5.10^3$, which induces a lot more gradient noise compared to the case where a higher time constant is chosen (see Fig. 4.3).

For the given function $R_v(n)$ the spectrum of the additive noise has a low-pass characteristic, whereas the channel IR is also low-pass. Because of the data-dependent update, the two different noise sources can be separated in the estimation process despite the fact they both have a low-pass nature.

As concluding remark it can be stated that the noise parameter estimates are fairly accurate if τ_e is chosen sufficiently high, i.e. if the adaptation loops exhibit slow convergence.

Decision-directed Estimation

in the simulation results we have always assumed that during the estimation we have perfect knowledge of the bit sequence \mathbf{a} . In DD operation mode however the bit-detector occasionally makes wrong decision and as a result the accuracy of the esti-

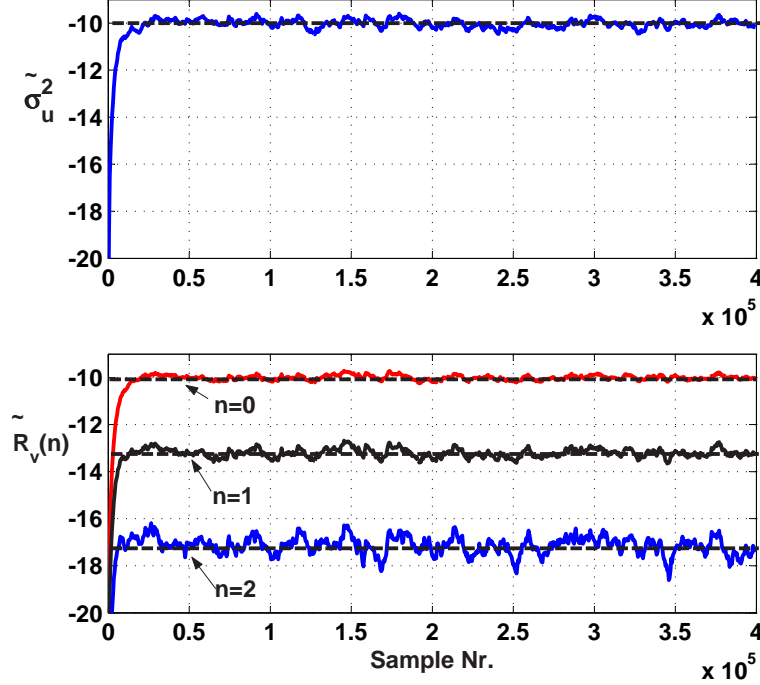


Figure 4.5: Estimated parameter values $\tilde{\sigma}_u^2$ and $\tilde{R}_v(n)$ for $(n = 0, 1, 2)$ versus the sample number. In the simulation the following parameter values are used: $\sigma_u^2 = -10$ dB, $R_v(0) = -10$ dB, $R_v(1) = -13.2$ dB and $R_v(2) = -17.2$ dB. In the simulation the following adaptation constants are used: $\mu_u = 0.0002$ and $\mu_v = 0.0001$.

mation algorithm worsens. To assess how the accuracy is influenced, the adaptation procedure depicted in Fig. 4.1 is simulated. But now instead of taking the actual bit sequence \mathbf{a} as input, an estimate $\hat{\mathbf{a}}$ of the bit sequence is used which is produced by a Viterbi Detector (VD) based on the detector input signal \mathbf{y} . The VD is matched to the target response \mathbf{h} (in case $M = 4$ the VD has 16 states), see Section 1.4.3 for a detailed explanation about the operation of the VD. Furthermore it must be noted that the VD operates optimally in case the white additive noise is present at its input. However in our case media noise is also present and as a result for MNP values different from 0, this VD is not optimal in terms of detection reliability. In our simulations it is used anyway for all MNP values because we are only interested in the effect of bit errors on the estimation accuracy and not in the optimal detection performance for different MNP values.

In Fig. 4.6, the offsets of the estimated SNR and estimated MNP from their actual values are plotted versus the actual MNP for different BERs. From this figure it is clear that in case erroneous bit decisions are used as input of the estimation

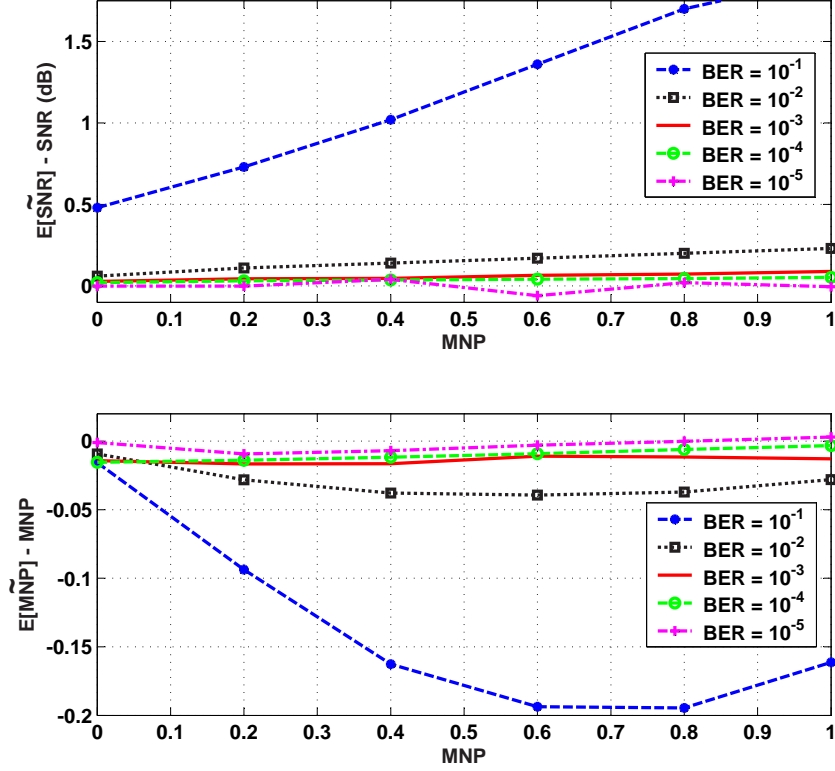


Figure 4.6: Offsets of $E[\tilde{\text{SNR}}]$ and $E[\tilde{\text{MNP}}]$ from their actual values for different BERs.

scheme, the estimated SNR will be higher than the actual SNR and the estimated MNP will be lower than the actual MNP. In practical systems however the BER can be assumed to be reasonably small ($\text{BER} < 10^{-2}$) and as a result the estimated SNR and the estimated MNP will exhibit just a small offset from their actual values. These results show that in systems with a BER smaller than 10^{-2} the estimation scheme can be used in a DD operation mode.

Sensitivity to channel IR

in the simulation results we have always assumed that we have perfect knowledge of the channel IR. In practical systems however the channel IR is estimated using a characterization scheme. Errors in the estimation of the channel IR will cause a bias in the variance estimate because the coefficients used in the adaptive estimation loop do not correspond to the ones of the actual storage channel. A possible error in the channel IR is a gain mismatch between the actual IR \mathbf{h} and the estimated IR $\hat{\mathbf{h}}$. As we

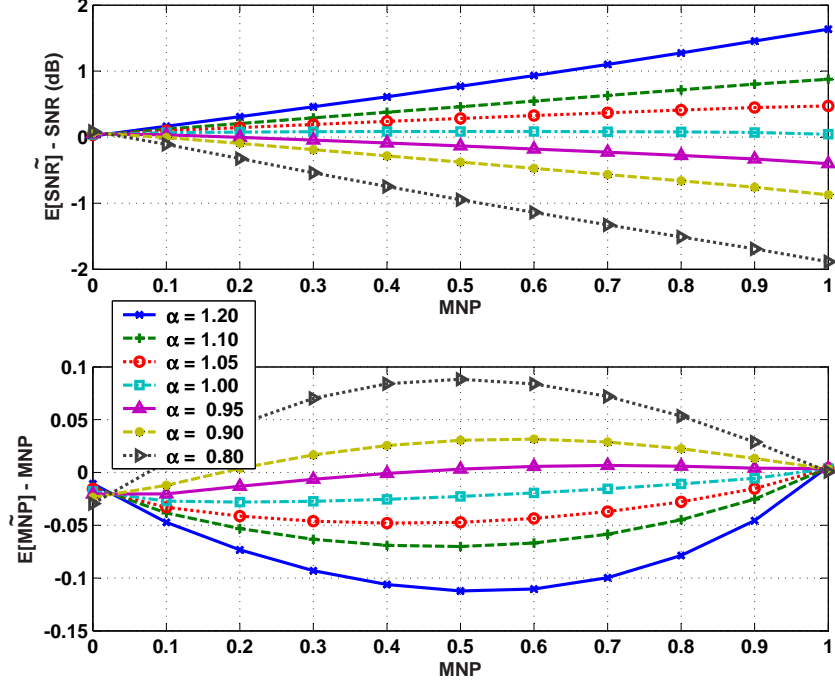


Figure 4.7: Offsets of $E[\tilde{\text{SNR}}]$ and $E[\tilde{\text{MNP}}]$ from their actual values for different gain factors α .

are estimating powers in the adaptive decomposition procedure, gain mismatch can be considered to be the most harmful for a correct operation of the procedure.

In Fig. 4.7, the offsets of the estimated SNR and estimated MNP from their actual values are plotted versus the actual MNP for different gain factors α , where $\tilde{h}_k = \alpha h_k$ for all k . From this figure it is clear that only relatively large gain mismatches ($> 10\%$) cause a substantial bias in the variance estimates ($E[\tilde{\text{MNP}}] - \text{MNP} > 0.05$ and $E[\tilde{\text{SNR}}] - \text{SNR} > 1\text{dB}$). In practical systems however the estimation of coefficients of the channel IR can be assumed to be reasonably accurate if the characterization loop is designed well. As a result in practical system only a very small bias is to be expected and the channel IR estimation will not affect the adaptive noise decomposition procedure.

4.3 Magnetic Storage

In magnetic storage applications, position-jitter is considered to be the dominant media noise effect [70], yet also other types of media noise arise at high densities, like nonlinear transition shifts and width-variations [192]. Furthermore at the output of

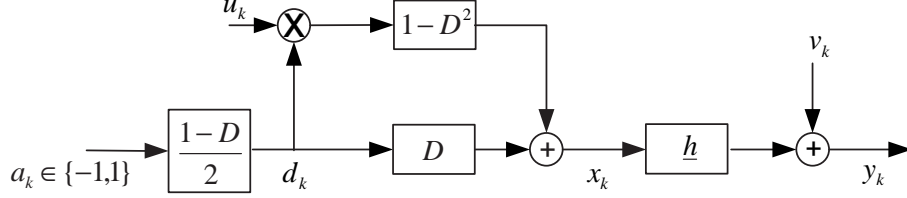


Figure 4.8: Simplified magnetic channel model with jitter noise u_k and additive noise v_k .

the channel, electronics noise is added to the signal. In general this additive noise is correlated. As a result, in storage systems two independent noise sources are commonly present. Because the media noise is data-dependent and the additive noise is data-independent, they have a different impact on the system and they should be treated separately in a characterization of the storage system. The decomposition procedure used in Section 4.2 for optical storage applications, can also be applied to magnetic storage applications. First, in Section 4.3.1, a simplified model for position-jitter is proposed on which the rest of this section is based. Subsequently, in Section 4.3.2, an adaptation algorithm for the decomposition of media and additive noise sources in magnetic storage systems is proposed. The proposed algorithm achieves a high estimation accuracy at the expense of limited computational complexity.

4.3.1 Media Noise Model

In this chapter, a simplified first-order approximation of position-jitter noise is used as media noise model. Furthermore additive noise is also present. The magnetic storage channel model is shown in Fig. 4.8. The discrete-time output of the model is given by

$$\begin{aligned} y_k &= \sum_j h_j x_{k-j} + v_k \\ &= \sum_j h_j d_{k-1-j} + \sum_j h_j (u_{k-j} d_{k-j} - u_{k-2-j} d_{k-2-j}) + v_k, \end{aligned} \quad (4.14)$$

where $\{x_k\}$ is the channel input sequence, \mathbf{h} is the equalized transition response, $\{u_k\}$ is a random noise sequence, $\{d_k u_k\}$ is the media noise sequence reflecting the amount of position jitter and $\{v_k\}$ is the additive noise sequence. The channel input sequence x_k is determined by the transition sequence $d_{k-1} (\in \{-1, 0, 1\})$ and a random part $u_k d_k - u_{k-2} d_{k-2}$ due to the transition jitter. This random part is computed somewhat different than in commonly used models [155]. In commonly used models, the effect of transition jitter at the channel output is modelled as the convolution of the media noise sequence with the first derivative of the transition response with respect to time [194]. In our model, this derivative is approximated by the simple $1 - D^2$ operation which is reasonable because in general \mathbf{h} is low-pass and $1 - D^2$ approximates a

differentiator at low frequencies. The transition sequence is obtained from the binary data sequence $a_k \in \{-1, 1\}$ ($d_k = (a_k - a_{k-1})/2$). All patterns in the sequence a_k are part of the set \mathcal{S} which is defined by the modulation code.

In this section, the random variables u_k and v_k are assumed to originate from independent white Gaussian noise sources with variances respectively denoted σ_u^2 and σ_v^2 . The proposed algorithm can be expanded to cover colored noise sources but for notational simplicity we limit ourselves to the white noise case.

The signal-to-noise ratio (SNR) is defined as $\text{SNR} = \sum_j f_j^2 / \sigma_z^2$, where $f_j = (h_j - h_{j-1})/2$ and σ_z^2 is the total noise power: $\sigma_z^2 = \sigma_u^2 + \sigma_v^2$. We define the media noise percentage (NMP) as the ratio of the media noise power to the total noise power [193]

$$\text{MNP} = \frac{\sigma_u^2}{\sigma_z^2} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}. \quad (4.15)$$

From Fig. 4.8, we can see that the overall noise consists of two different types of Gaussian noise. One of the types, namely the media noise, depends also on the transmitted signal. Therefore a solution similar to the one presented in section 4.2.2, can be designed to adaptively decompose the noise components based on the channel output signal.

4.3.2 Adaptive Estimation Scheme

In this section an algorithm is derived to estimate the stochastic parameters σ_u^2 and σ_v^2 based on the observed signal y_k and the knowledge of the transmitted data a_k . The algorithm requires a single read-out of arbitrary data patterns and as a result it can be used in real-time systems. The data-dependent nature of the media noise is used to adaptively decompose the different noise components at the channel output.

An error signal e_k is defined as follows

$$e_k = y_k - \sum_j h_j d_{k-1-j} = \sum_j h_j (u_{k-j} d_{k-j} - u_{k-2-j} d_{k-2-j}) + v_k. \quad (4.16)$$

This error signal consists of a data-independent part (the additive noise v_k) and a data-dependent part, namely the media noise at the channel output due to the transition jitter sequence u_k . In most practical data-receivers, an equalizer is used to shorten the transition-response memory length to an acceptable duration, such that maximum likelihood sequence detection (MLSD) can be performed with a reasonable complexity [4, 60]. As a result the equalized transition response \mathbf{h} can be considered to have a finite memory length M , and consequently also the data-dependency of the media noise has a finite memory length. Now define \mathbf{a}_k^M as the vector $[a_k, a_{k-1}, \dots, a_{k-M}]$.

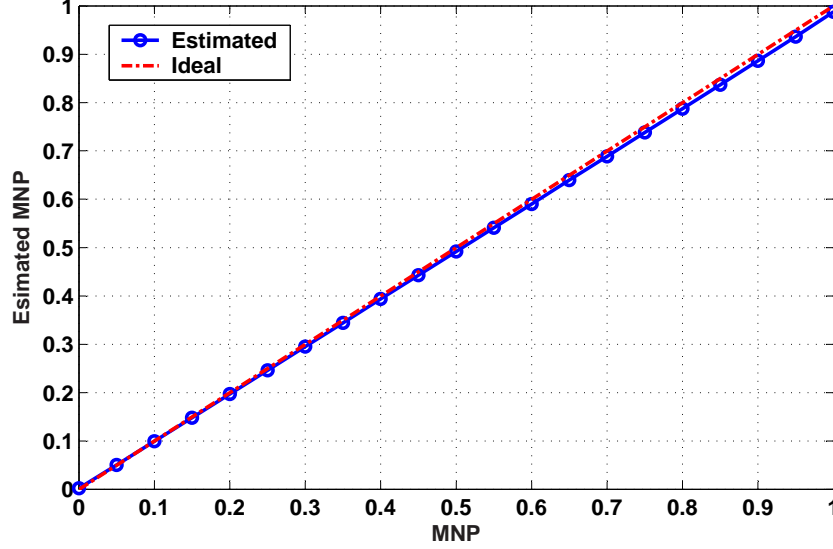


Figure 4.9: Estimated MNP: $E[\hat{\text{MNP}}] = E\left[\frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2}\right]$ for $\mu_u = \mu_v = 0.0001$ and $\text{SNR} = 20$ dB.

The power function of the error signal for a specific pattern \mathbf{a}_k^M can be written as

$$\begin{aligned} E[e_k^2 | \mathbf{a}_k^M] &= \sigma_v^2 + 2\sigma_u^2 \sum_{j=0}^M (h_j^2 (2 - a_{k-j}a_{k-j-1} - a_{k-j-2}a_{k-j-3}) \\ &\quad + 2h_j h_{j+2} (a_{k-j-2}a_{k-j-3} - 1)) \\ &= \sigma_v^2 + \sigma_u^2 H(0, \mathbf{a}_k^M), \end{aligned} \quad (4.17)$$

with $H(0, \mathbf{a}_k^M) = 2\sum_{j=0}^M (h_j^2 (2 - a_{k-j}a_{k-j-1} - a_{k-j-2}a_{k-j-3}) + 2h_j h_{j+2} (a_{k-j-2}a_{k-j-3} - 1))$. Based on this power function an adaptive estimation scheme can be derived following the procedure defined in Appendix B. The resulting adaptation rules are the same as obtained in the optical storage example

$$\begin{aligned} \tilde{\sigma}_u^{2(k)} &= \tilde{\sigma}_u^{2(k-1)} - 2\mu_u H(0, \mathbf{a}_k^M) [\tilde{\sigma}_u^{2(k-1)} + \tilde{\sigma}_v^{2(k-1)} - e_k e_k], \\ \tilde{\sigma}_v^{2(k)} &= \tilde{\sigma}_v^{2(k-1)} - 2\mu_v [H(0, \mathbf{a}_k^M) \tilde{\sigma}_u^{2(k-1)} + \tilde{\sigma}_v^{2(k-1)} - e_k e_k]. \end{aligned} \quad (4.18)$$

The only difference between the optical and magnetic storage example is in the data-dependent coefficients $H(0, \mathbf{a}_k^M)$. More generally for the estimation of different types of media noise the data-dependent coefficients will also be different.

4.3.3 Simulation Results

The simplified magnetic channel model of Fig. 4.8 is used to judge the accuracy of the proposed decomposition algorithm. The SNR is set to 20 dB, σ_u^2 and σ_v^2 are varied according to a given MNP. The ME²PRML channel $(5 + 4D - 3D^2 - 4D^3 - 2D^4)$ is

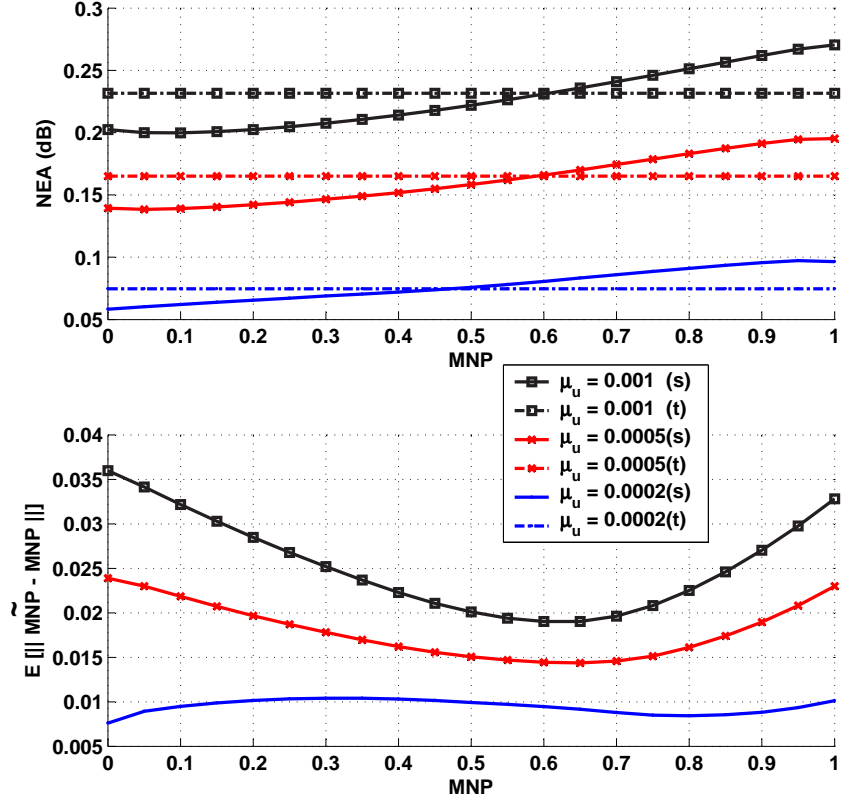


Figure 4.10: Simulated (s) and theoretical (t) values of NEA and $E[\|\tilde{\mathbf{MNP}} - \mathbf{MNP}\|]$ versus MNP for different time constants $\mu_u = \mu_v = 0.0002$, $\mu_u = \mu_v = 0.0005$ and $\mu_u = \mu_v = 0.001$, furthermore SNR = 25 dB.

used in the simulations [101], which is shown to be a good match to magnetic storage channels at high densities [104]. The transition response for the ME²PRML channel is found to be: $\mathbf{h} = 5 + 4D + 2D^2$. In Fig. 4.9, the estimated MNP is shown versus the actual MNP for $\mu_u = \mu_v = 0.0001$. The estimated MNP is almost identical to the actual MNP. To judge the achieved accuracy of the adaptation algorithm, the simulated and theoretical NEA values are plotted in the upper part of Fig.4.10 for 3 different time constants $\mu_u = \mu_v = 0.0002$, $\mu_u = \mu_v = 0.0005$ and $\mu_u = \mu_v = 0.001$ and again SNR is set to 25 dB. The theoretical derivation of the estimation variance presented in Appendix C is also valid for the magnetic storage system and the approximated theoretical curves are also shown. The simulation results are in line with theory. This leads to the same conclusions as in Section 4.2.3, namely that the estimation accuracy is mainly determined by the time constant of the adaptation loops, more precisely by

the square root of μ_u and μ_v .

In the lower part of Fig. 4.10 the standard deviation of the estimated MNP is plotted versus MNP for different values of μ_u .

4.4 Test Pattern Design

The adaptive estimation procedure developed here, can be applied to real-time and experimental systems. In the simulation results of Section 4.2 and Section 4.3 the procedure is applied in a real-time setting, i.e. for a single read-out of arbitrary, a-priori unknown data patterns. In experimental systems however one has the freedom to use a sequence of predefined test patterns instead of a single arbitrary pattern. Based on the theoretical analysis of the adaptation algorithm, we will design test patterns for experimental systems (both magnetic and optical storage systems). These test patterns yield an improved estimation accuracy with respect to the accuracy obtained in real-time systems.

The adaptive estimation scheme given by (4.8) updates the estimated values for every possible data pattern. However, for most of the data patterns a mixture of media noise and additive noise is present and as a result the adaptation of the media noise and of the additive noise interact. This interaction degrades the overall estimation performance. Consequently the algorithm can be changed to update only on data patterns for which one noise source is dominant and as a result also the interaction will be limited.

For the optical channel model presented in Section 4.2, the two opposite data patterns (the all-land pattern 00000 and the all-pit pattern 11111) are of particular interest. The all-land pattern does not induce media noise and therefore the additive noise characteristics should be estimated based on only this pattern. The all-pit pattern however induces the maximum amount of media noise information with respect to a fixed amount of additive noise and as a result the media noise estimate should be based only on this all-pit pattern. As a result the optimal test pattern for optical systems consists of two parts: first an all-land pattern to estimate the additive noise followed by an all-pit pattern to estimate the pit size noise.

In the magnetic channel model presented in Section 4.3, transition jitter is the media noise source under consideration. As transition jitter only occurs if transitions are present in the recorded data sequence, the optimal pattern to estimate the additive noise parameter is the pattern where no transitions occur, namely the all-zero 00000 or the all-one pattern 11111. Consequently the pattern that should be used to estimate the media noise parameters, should have as many transitions as possible, i.e. the Nyquist-pattern 01010. However in state-of-the-art magnetic storage systems, the channel can be considered to have little or no transfer at the Nyquist frequency and as a result the Nyquist-pattern is not very suitable to base the estimation on.

Consequently another pattern with a lot of transitions should be chosen that still has substantial transfer across the channel. A suitable test pattern is the periodic pattern with a period of 2 bits, i.e. 00110011.

Besides the specific test pattern, the accuracy of the estimation algorithm also depends on the knowledge of the channel IR \mathbf{h} . In practical systems \mathbf{h} is also estimated by an identification scheme. This identification scheme estimates the amplitude of the equalized channel IR. As the estimation of the noise parameters essentially boils down to the estimation of the power of the noise sources, the IR identification scheme is not influenced by the noise estimation. Therefore a conventional minimum mean square error estimation can be used where the noise in the system results in gradient noise in the estimation of the channel IR. The acquisition of the channel IR in such estimation is guaranteed by the use of a pseudo-noise random sequence. As a result this sequence needs to be part of the overall test pattern. Furthermore the adaptive estimation scheme is sensitive to residual even-order nonlinearities (DC-offset, quadratic, 4th order). They have to be carefully compensated for before using the adaptive estimation scheme.

4.5 Conclusions

In digital storage systems the total amount of data-dependent media noise increases considerably as storage densities increase. The estimation algorithm proposed in this chapter, jointly estimates the parameters of both media and additive noise with a high accuracy. The proposed algorithm makes use of this data-dependency to distinguish between the different noise sources. The algorithm is simple and as a result only very limited amount of complexity is required to implement it in storage systems as an easy “add-on” to read channels ICs. The resulting estimates of the noise parameters provide important diagnostic information about the functionality of the system and possible improvements in the storage system. Simulation results for an idealized optical channel with data-dependent noise show that the estimation of a specific additive or media noise parameter can be very accurate in case its magnitude is not much smaller than the magnitude of the other parameter.

Appendix 4.A: Derivation of multivariate least squares solution

The multivariate least squares solution is based on the minimization of the total misadjustment power. The misadjustment power is defined as

$$J = \sum_{\mathbf{a}_k^M \in \mathcal{S}} \sum_{n=0}^N \left(R_e(n, \mathbf{a}_k^M) - \sigma_u^2 H(n, \mathbf{a}_k^M) - R_v(n) \right)^2. \quad (4.19)$$

To minimize this cost function, the partial derivatives with respect to the unknown noise parameters are taken and are put to zero.

$$\begin{aligned} \frac{\partial J}{\partial \sigma_u^2} &= 2 \sum_{\mathbf{a}_k^M \in \mathcal{S}} \sum_{n=0}^N \left(\sigma_u^2 H(n, \mathbf{a}_k^M)^2 + \right. \\ &\quad \left. R_v(n) H(n, \mathbf{a}_k^M) - R_e(n, \mathbf{a}_k^M) H(n, \mathbf{a}_k^M) \right) = 0 \\ \frac{\partial J}{\partial R_v(0)} &= 2 \sum_{\mathbf{a}_k^M \in \mathcal{S}} \left(R_v(0) + \right. \\ &\quad \left. \sigma_u^2 H(0, \mathbf{a}_k^M) - R_e(0, \mathbf{a}_k^M) \right) = 0 \\ &\vdots \\ \frac{\partial J}{\partial R_v(N)} &= 2 \sum_{\mathbf{a}_k^M \in \mathcal{S}} \left(R_v(N) + \right. \\ &\quad \left. \sigma_u^2 H(N, \mathbf{a}_k^M) - R_e(N, \mathbf{a}_k^M) \right) = 0 \end{aligned} \quad (4.20)$$

By defining

$$H(n) = \sum_{\mathbf{a}_k^M \in \mathcal{S}} H(n, \mathbf{a}_k^M), \quad \mathcal{H} = \sum_{\mathbf{a}_k^M \in \mathcal{S}} \sum_{n=0}^N H(n, \mathbf{a}_k^M)^2, \quad (4.21)$$

$$G(n) = \sum_{\mathbf{a}_k^M \in \mathcal{S}} R_e(n, \mathbf{a}_k^M) \text{ and } \mathcal{G} = \sum_{\mathbf{a}_k^M \in \mathcal{S}} \sum_{n=0}^N R_e(n, \mathbf{a}_k^M) H(n, \mathbf{a}_k^M), \quad (4.22)$$

the set of equations can be written in matrix form as

$$\begin{bmatrix} \mathcal{H} & H(0) & \cdots & H(N) \\ H(0) & 2^M & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ H(N) & 0 & \cdots & 2^M \end{bmatrix} \begin{bmatrix} \sigma_u^2 \\ R_v(0) \\ \vdots \\ R_v(N) \end{bmatrix} = \begin{bmatrix} \mathcal{G} \\ G(0) \\ \vdots \\ G(N) \end{bmatrix}. \quad (4.23)$$

Appendix 4.B: Derivation of adaptive least mean squares solution

The cost function is given by (4.19). New estimates of the different noise parameters can be computed as

$$\begin{aligned}\tilde{\sigma}_u^{2(k)} &= \tilde{\sigma}_u^{2(k-1)} - \mu_u \frac{\partial J}{\partial \sigma_u^2}, \\ \tilde{R}_v(0)^{(k)} &= \tilde{R}_v(0)^{(k-1)} - \mu_v \frac{\partial J}{\partial \tilde{R}_v(0)}, \\ &\vdots \\ \tilde{R}_v(N)^{(k)} &= \tilde{R}_v(N)^{(k-1)} - \mu_v \frac{\partial J}{\partial \tilde{R}_v(N)},\end{aligned}\tag{4.24}$$

where μ_u and μ_v are adaptation constants. The partial derivatives with respect to the unknown noise parameters are given by (4.20). To be able to use these partial derivatives we replace the expectation values of the error cross correlation by their instantaneous values. Using these partial derivatives in (4.24) establishes the update rules for the estimation of the noise characteristics:

$$\begin{aligned}\tilde{\sigma}_u^{2(k)} &= \tilde{\sigma}_u^{2(k-1)} - 2\mu_u \left(\sum_{n=0}^N H(n, \mathbf{a}_k^M) \right. \\ &\quad \left. (H(n, \mathbf{a}_k^M) \tilde{\sigma}_u^{2(k-1)} + \tilde{R}_v(n)^{(k-1)} - e_k e_{k-n}) \right), \\ \tilde{R}_v(n)^{(k)} &= \tilde{R}_v(n)^{(k-1)} - 2\mu_v \\ &\quad (H(n, \mathbf{a}_k^M) \tilde{\sigma}_u^{2(k)} + \tilde{R}_v(n)^{(k-1)} - e_k e_{k-n}),\end{aligned}\tag{4.25}$$

for $n = 0, \dots, N$.

Appendix 4.C: Steady-State Behavior Adaptation Loops

The steady-state behavior of the adaptation loops presented in Fig. 4.2. Here the media and the noise additive noise are modelled as AWGN processes. The steady-state value of the estimates can be derived easily. For every possible data pattern \mathbf{a}_k^M , the steady state values of the estimates can be found by solving the equations obtained by making the expectation of the partial derivatives given by (4.20) equal to zero. By applying (4.4) it can easily be found that $E[\tilde{\sigma}_u^2] = \sigma_u^2$ and $\tilde{R}_v(0) = R_v(0)$ for all data patterns in S .

Furthermore also the variances of the estimated values $\tilde{\sigma}_u^2$ can be calculated. Based on the adaptation rules (4.8), the variances are given by the following ex-

pressions.

$$\begin{aligned}
\gamma_{\tilde{\sigma}_u^2}^2 &\triangleq E[\tilde{\sigma}_{u(k)}^2] - E[\tilde{\sigma}_{u(k)}^2]^2 \\
&= (1 - 4\mu_u H^2 + 4\mu_u^2 H^4) E[\tilde{\sigma}_{u(k-1)}^2] + 4\mu_u^2 H^2 E[\tilde{\sigma}_{v(k-1)}^2] + 4\mu_u^2 H^2 E[e_{(k)}^4] \\
&\quad + (8\mu_u^2 H^3 - 4\mu_u H) E[\tilde{\sigma}_{u(k-1)}^2 \tilde{\sigma}_{v(k-1)}^2] - (8\mu_u^2 H^3 - 4\mu_u H) E[\tilde{\sigma}_{u(k-1)}^2 e_{(k)}^2] \\
&\quad - 8\mu_u^2 H^2 E[\tilde{\sigma}_{v(k-1)}^2 e_{(k)}^2] - E[\tilde{\sigma}_{u(k)}^2]^2, \\
\gamma_{\tilde{\sigma}_v^2}^2 &\triangleq E[\tilde{\sigma}_{v(k)}^2] - E[\tilde{\sigma}_{v(k)}^2]^2 \\
&= (1 - 4\mu_v + 4\mu_v^2) E[\tilde{\sigma}_{v(k-1)}^2] + 4\mu_v^2 H^2 E[\tilde{\sigma}_{u(k-1)}^2] + 4\mu_v^2 E[e_{(k)}^4] \\
&\quad + (8\mu_v^2 H - 4\mu_v H) E[\tilde{\sigma}_{u(k-1)}^2 \tilde{\sigma}_{v(k-1)}^2] - (8\mu_v^2 - 4\mu_v) E[\tilde{\sigma}_{v(k-1)}^2 e_{(k)}^2] \\
&\quad - 8\mu_v^2 H E[\tilde{\sigma}_{u(k-1)}^2 e_{(k)}^2] - E[\tilde{\sigma}_{v(k)}^2]^2,
\end{aligned} \tag{4.26}$$

where $H = H(0, \mathbf{a}_k^M)$. To simplify these expressions, the expectation values of different signals need to be calculated.

$$\begin{aligned}
E[\tilde{\sigma}_{u(k-1)}^2] &= \gamma_{\tilde{\sigma}_u^2}^2 + E[\tilde{\sigma}_{u(k)}^2]^2, \\
E[\tilde{\sigma}_{v(k-1)}^2] &= \gamma_{\tilde{\sigma}_v^2}^2 + E[\tilde{\sigma}_{v(k)}^2]^2, \\
E[\tilde{\sigma}_{u(k-1)}^2 \tilde{\sigma}_{v(k-1)}^2] &= R_u(0)R_v(0), \\
E[\tilde{\sigma}_{u(k-1)}^2 e_{(k)}^2] &= R_u(0)E[e_{(k)}^2] = HR_u(0)^2 + R_u(0)R_v(0), \\
E[\tilde{\sigma}_{v(k-1)}^2 e_{(k)}^2] &= R_v(0)E[e_{(k)}^2] = HR_v(0)^2 + R_v(0)R_u(0), \\
E[e_{(k)}^4] &= 3E[e_{(k)}^2]^2 = 3H^2R_u(0)^2 + 6HR_u(0)R_v(0) + 3R_v(0)^2.
\end{aligned} \tag{4.27}$$

The latter equation is derived by applying (4.4) and the formula of Isserlis [195]. By applying (4.27), the expressions given by (4.26) can be simplified.

$$\begin{aligned}
\gamma_{\tilde{\sigma}_u^2}^2 &= \frac{1}{1-\mu_u H^2} \left[2\mu_u H^2 R_u(0)^2 + 4\mu_u H R_u(0)R_v(0) + 2\mu_u R_v(0)^2 + \mu_u \gamma_{\tilde{\sigma}_v^2}^2 \right], \\
\gamma_{\tilde{\sigma}_v^2}^2 &= \frac{1}{1-\mu_v} \left[2\mu_v R_v(0)^2 + 4\mu_v H R_u(0)R_v(0) + 2\mu_v H^2 R_u(0)^2 + \mu_v H^2 \gamma_{\tilde{\sigma}_u^2}^2 \right].
\end{aligned} \tag{4.28}$$

In normal conditions the adaptation constants are chosen such that: $1 > \sigma_u^2 \gg \mu_u > 0$ and $1 > R_v(0) \gg \mu_v > 0$. Based on these considerations, the expressions can be approximated by

$$\begin{aligned}
\gamma_{\tilde{\sigma}_u^2}^2 &\approx 2\mu_u (HR_u(0) + R_v(0))^2, \\
\gamma_{\tilde{\sigma}_v^2}^2 &\approx 2\mu_v (HR_u(0) + R_v(0))^2.
\end{aligned} \tag{4.29}$$

Chapter 5

Cancellation of Linear Intersymbol Interference for Two-Dimensional Storage Systems

In this chapter the cancellation of linear Intersymbol Interference (ISI) in two-dimensional systems is discussed. A theory is developed for the error rate of receivers that use tentative decisions to cancel ISI. Furthermore, precise conditions are formulated under which such ISI cancellation can be applied effectively. For many two-dimensional systems these conditions are easily met and therefore the application of ISI cancellation is of significant interest for these systems. The theory and the conditions are validated by simulation results for a two-dimensional channel model. Furthermore, results for an experimental two-dimensional optical storage system show that for a single layer disc with a capacity of 50 GB a substantial performance improvement may be obtained by applying ISI cancellation.

5.1 Introduction

Steadily increasing storage densities are a clear trend in storage systems. Increasing amounts of Intersymbol Interference (ISI) are a consequence of this trend. A possible technique to deal with this ISI is ISI cancellation. In this technique tentative decisions are used as input to an interference canceller, which attempts to remove those linear or nonlinear ISI components that are not expected by the main bit-detector. These unexpected ISI components are denoted in the remainder as Residual ISI (RISI) components. In general RISI components originate from the fact that the equalizer is not able to perfectly shape the ISI structure induced by the storage channel into the ISI structure expected by the bit-detector (defined by a so-called target response). The part of the RISI components that originates from symbols subsequent to the current one (i.e. from future symbols), is denoted as precursive ISI.

The general structure of an ISI cancellation scheme is depicted in Fig. 5.1 for a one-dimensional (1-D) system. The equalized signal y_k is a distorted and noisy ver-

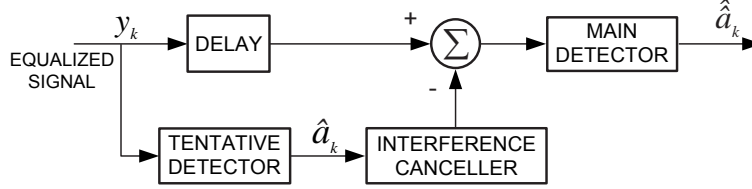


Figure 5.1: General structure of an ISI cancellation scheme.

sion of the recorded bits a_k . Based on tentative decisions \hat{a}_k with respect to these recorded bits a_k , the interference canceller generates an estimate of the RISI contained in the equalized signal y_k . To make the scheme causal, y_k is delayed until tentative decisions are available for all symbols that induce precursive RISI. The RISI estimate is subtracted from this delayed version of y_k to produce a signal that is ideally free from RISI. This signal is used as input of the main detector. This main detector produces the final decisions $\hat{\hat{a}}_k$. The benefits of the ISI cancellation technique are: simplicity, ability to handle both linear as nonlinear RISI components and absence of loops (i.e. ISI cancellation can be pipelined).

In general ISI cancellation works effectively if two conditions are fulfilled. First, all the cancelled RISI components should have a relatively small magnitude (with respect to the magnitude of the target-response components). Second, errors that affect the main and the tentative decisions measured at the same instant should be statistically independent. In practice the latter condition is fulfilled if the cancelled RISI components originate from symbols that have sufficient “temporal separation” from the current symbols [196]. In a 1-D system the number of small RISI components tends to be limited and for this reason the performance gain of ISI cancellation is usually small. As a result ISI cancellation is seldom used in practical systems because the substantially increased complexity is not justified by the marginally increased performance.

Besides the trend of increasing storage densities, there is also a general trend of increasing data rates. The development of two-dimensional (2-D) storage systems fits with this trend and permits exploitation of parallelism. The parallelism is achieved by packaging data in a group of adjacent tracks or rows and by parallel processing of these tracks [197]. The physical proximity of the tracks causes 2-D ISI during read-out. The main topic of this chapter is the use of linear ISI cancellation in 2-D systems to deal with 2-D RISI components. The 2-D ISI cancellation technique presented in this chapter is general and can be applied to a variety of 2-D systems: MIMO, holographic storage, page-oriented optical memories, patterned magnetic media and 2-D optical storage.

In this chapter we argue that the application of ISI cancellation is of significant interest for 2-D systems. A first argument is the fact that the number of small RISI

components is increased considerably with respect to the 1-D case. As a result the performance gain by applying the cancellation technique will also increase. A second argument is the fact that only little additional complexity may be required for the application of ISI cancellation techniques. Because 2-D detection is often accomplished by several iterations of smaller detection units to avoid the complexity of a full 2-D Viterbi detector [152, 198], decisions of one of these smaller detection units can be used as tentative decisions by the canceller. As a result no additional bit-detectors need to be implemented to produce these tentative decisions. Summarizing these arguments, with a limited additional complexity (only the interference canceller needs to be added, not an additional detector) ISI cancellation in a 2-D system may improve performance significantly.

In this chapter the attention is limited to the cancellation of linear ISI. This limitation is reasonable as in general the linear components account for the bulk of the total RISI and moreover in [196] the cancellation of nonlinear ISI components is shown to be ineffective.

An experimental 2-D optical storage system, called TwoDOS, is used to illustrate the performance improvement. For TwoDOS a Partial-Response Maximum-Likelihood (PRML) receiver with a Stripe-Wise Viterbi Detector (SWVD) was developed [145]. This SWVD performs two consecutive detection iterations. As a result the outputs of the first iteration can be used as tentative decisions in an ISI cancellation scheme. The application of linear 2-D ISI cancellation in the PRML receiver improves the performance of the system significantly at very limited additional complexity.

In Section 5.2 a historical overview of ISI cancellation schemes for 1-D systems is given. Also in Section 5.2 existing reception techniques for 2-D systems are discussed. In Section 5.3 the 2-D ISI cancellation scheme is proposed and analyzed. Finally in Section 5.4 experimental results of linear 2-D ISI cancellation are presented for the TwoDOS system. These experimental results show a substantial performance improvement by applying linear 2-D ISI cancellation.

5.2 Overview of ISI Cancellation

In 1-D systems, the adaptive canceller usually employed in echo cancellation was first applied to adaptive equalization by Gersho and Lim [165]. Their work was the extension of early work by Proakis [199]. They developed a cancellation structure that achieves the optimal performance, i.e. the isolated-pulse, matched filter reception. For this to happen, cancellation should be based on the actual data, i.e. on data without decisions errors. Because of the noncausal nature of the cancellation structure, it is necessary to use tentative decisions to synthesize the precursive RISI components. This cancellation structure is sometimes referred to as a two-stage equalizer. The

first stage produces tentative decisions which are used by the cancellation filter while the second stage produces final decisions based on the equalizer output signal after cancellation of the RISI components. Gersho and Lim suggest the use of a linear equalizer (LE) as the first-stage equalizer. Significant SNR gains are observed using the canceller compared to an LE. A theory has been worked out on data-aided equalization techniques including linear, decision-feedback and canceller based equalizers [200]. Both in [165] and in [200] the analysis was based on the mean-square error criterion and ideal (i.e. correct) decisions are assumed. Wesolowski [166,201] showed that the error performance of the canceller critically depends on the performance of the first-stage equalizer. Therefore a Decision-Feedback Equalizer (DFE) was proposed as the first-stage equalizer. The replacement of the LE by DFE yields a moderate improvement in performance. Based on the assumption that the final decisions will be better than the tentative decisions, it has been argued that replacing tentative decisions by final decisions to synthesize the postcursive RISI contribution (RISI originating from past symbols), will improve performance [202].

The combination of a linear ISI canceller combined with an error reducing circuitry, called quantized logical equalizer, has been proposed in [203]. In [204], through a simulation study, the performance of the canceller was shown to lag well behind the Viterbi Detector and to provide little improvement over a DFE at high storage densities. However for specific low-pass channels, the performance of the canceller was shown to approach that of a maximum-likelihood detector if the input data is constrained using a run-length-limited code [205]. The use of reliabilities produced by the error-correcting decoder in the ISI cancellation mitigates the error propagation [206]. This approach can be generalized to an iterative scheme of alternately cancelling RISI components and decoding. Furthermore, the ISI cancellation technique can be applied to nonlinear channels [202,207,208] resulting in SNR gains at the detector input. The corresponding error rate improvements are not shown. In [209] the problem of error propagation is addressed. Error propagation caused by the first stage can degrade the effectiveness of the cancellation technique severely. To mitigate the degradation, nonlinear cancellation combined with trellis coding was proposed. This scheme requires two Viterbi detectors, which leads to an increased complexity.

In general the cancellation technique can be shown to be effective if the ISI that is being cancelled is “small” and if errors affecting final and tentative decisions are statistically independent [196,210]. The latter condition basically means that the cancelled RISI components should originate from symbols that have sufficiently “temporal separation” from the current symbols.

In commercial 1-D storage systems the application of ISI cancellation is limited. Two main reasons can be identified for this limited application: (1) the presence of an auxiliary (Viterbi) detector to produce the tentative decisions will cause a substantial

increase in overall complexity; (2) the number of “small” ISI components that can be cancelled effectively is in general quite low, and as a result the potential performance improvement is small [196].

For 2-D storage systems, receiver structures have been presented based on Linear Equalization (LE) [211,212], Decision-Feedback Equalization (DFE) [213–215] and iterative detection [216,217]. The extension of the 1-D ISI canceller to its 2-D equivalent has however not been reported. In this chapter the application of cancellation techniques is described for 2-D storage systems. The major drawbacks of ISI cancellation in 1-D systems become much smaller for 2-D systems. First, due to the 2-D nature, it should be clear that the number of small ISI components is increased considerably with respect to the 1-D case. As a result the performance gain by applying the cancellation technique will increase. Second, because of complexity issues, bit-detection in 2-D systems is hardly if ever accomplished by a full 2-D Viterbi detector [152, 198]. Instead, detection is often accomplished by several iterations of smaller detection units. As the detection process is divided into smaller subprocesses it is possible to use decisions of one of these smaller subprocesses as tentative decisions in cancellation techniques. As a result, with a very limited additional complexity (only the interference canceller needs to be added), ISI cancellation in a 2-D storage system can yield a significant performance gain.

5.3 Linear ISI Cancellation in 2-D Systems

In this section, linear 2-D ISI cancellation is presented and analyzed. The 2-D ISI cancellation scheme together with the assumed channel model is shown in Fig. 5.2. The signals $\mathbf{y}_k = [y_k^0 y_k^1 \dots y_k^{L-1}]^T$ where L represents the number of tracks, are output of a linear 2-D channel model. Because of the linearity, the overall 2-D channel response of each track l can be separated into two 2-D responses: one response expected by the detector (the target response) and one response which is undesired in the detector (the undesired RISI response). Therefore the signal y_k^l belonging to track l can be expressed as the sum of three terms. The first term is the desired partial response signal value (obtained by convolving the target response with the bits); the second term is the undesired RISI value (obtained by convolving the undesired RISI impulse response with the bits); and finally a noise term.

$$y_k^l = \sum_{p=0}^{G-1} \sum_{q=0}^{L-1} g_0(p, l-q) a_{k-p}^q + \sum_{p=-\gamma}^{\kappa-1} \sum_{q=0}^{L-1} g_1(p, l-q) a_{k-p}^q + n_k^l, \quad (5.1)$$

where $g_0(p, q)$ is the target response expected by the detectors, $g_1(p, q)$ is the RISI impulse response, a_k^l are the channel inputs ($a_k^l \in \{-1, 1\}$), n_k^l are noise samples and

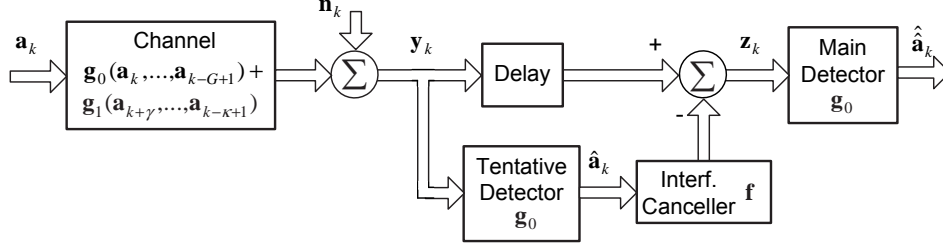


Figure 5.2: System model. The assumed channel model is depicted in combination with the ISI cancellation scheme.

G is the length of the target response. Here we assume all tracks have the same target and RISI response. In total there are $L\gamma$ precursive RISI components and $L\kappa$ postcurusive RISI components. Notice that $g_1(p, q)$ is a noncausal impulse response if $\gamma > 0$ (precursive RISI). In this section the noise samples are assumed to be white and Gaussian with variance σ^2 and the noise samples of different tracks are uncorrelated. For every track l an interference cancellation filter with impulse response $f^l(p, q)$ generates an estimate of the RISI contained in y_k^l based on the tentative decisions \hat{a}_i^j for $i \in [0, L-1]$ and $j \in [k+V-\gamma, k+V+\kappa]$, where V is the delay introduced by the tentative detector. In general to effectively cancel all RISI, the response $f^l(p, q)$ should be equal to $g_1(p, q)$. The RISI estimates of all tracks are subtracted from delayed versions of \mathbf{y}_k (with a delay of $V + \gamma$ symbols such that tentative decisions are available for all precursive RISI components). The resulting signals \mathbf{z}_k are used as inputs of the main detector which produces the final bit decisions $\hat{\mathbf{a}}_k$. In this chapter, we assume for simplicity reasons that the main and the tentative detector operate based on the same target response $g_0(p, q)$. This assumption is however not strictly needed. For example, a configuration is possible where besides $g_1(p, q)$ also part of $g_0(p, q)$ is cancelled and as a result the main detector operates on a truncated version of $g_0(p, q)$ [218].

In Subsection 5.3.1 the symbol error rate of a Viterbi detector is analyzed in case RISI is present at the detector input. Subsequently the symbol error rate of the ISI cancellation scheme of Fig. 5.2 is analyzed in Subsection 5.3.2. The effect of error propagation on the overall receiver performance is discussed in Subsection 5.3.3. Finally Subsection 5.3.4 illustrates the effectiveness of the ISI cancellation scheme for three simplified channel models.

5.3.1 Probability of Error of a Viterbi Detector in the presence of RISI

The probability of symbol error of a 2-D Viterbi Detector (VD) in the presence of 2-D RISI can be derived following the same approach as presented in [196] for a 1-D system. The differences between the 1-D and 2-D case together with the main

conclusions are highlighted here. The probability of symbol error is

$$P_e \leq \sum_{\epsilon \in E} w_H(\epsilon) P(\epsilon), \quad (5.2)$$

where E is the set of all possible 2-D error events ϵ in which the null event (no errors) is excluded, $w_H(\epsilon)$ is the number of symbol errors in the error event ϵ and $P(\epsilon)$ is the probability that error event ϵ occurs. Also

$$P(\epsilon) = P_1(\epsilon) P_2(\epsilon), \quad (5.3)$$

where $P_1(\epsilon)$ is the probability that the VD selects the path corresponding the error event instead of the path corresponding to the actual recorded data sequence, i.e. $P(a_k^l) < P(a_k^l + \epsilon_k^l)$. Furthermore $P_2(\epsilon)$ is the probability of occurrence of a data sequence \mathbf{a}_k that supports the error event ϵ .

Assume that the path associated with ϵ differs from the correct path for $k_0 \leq j \leq k_1$. Define

$$\begin{aligned} s^{\mathbf{a}}(k, l) &= \sum_{p=0}^{G-1} \sum_{q=0}^{L-1} g_0(p, l-q) a_{k-p}^q, \\ i^{\mathbf{a}}(k, l) &= \sum_{p=-\gamma}^{\lambda-1} \sum_{q=0}^{L-1} g_1(p, l-q) a_{k-p}^q, \\ \Delta_0^{\epsilon}(k, l) &= s^{\mathbf{a}^{\epsilon}}(k, l) - s^{\mathbf{a}}(k, l), \\ \Delta_1^{\epsilon}(k, l) &= i^{\mathbf{a}^{\epsilon}}(k, l) - i^{\mathbf{a}}(k, l), \end{aligned} \quad (5.4)$$

where $s^{\mathbf{a}}(k, l)$ is the ideal detector input for track l at time k given a specific data sequence $\mathbf{a} = [\mathbf{a}_{k_0-\kappa+1} \dots \mathbf{a}_{k_1+\gamma}]$, $i^{\mathbf{a}}(k, l)$ is the total RISI for track l at time k and \mathbf{a}^{ϵ} is the data sequence according to the error event ϵ (i.e. $a_k^{\epsilon l} = a_k^l + \epsilon_k^l$).

The VD will select the wrong path if

$$\sum_{j=k_0}^{k_1+G-1} \sum_{i=0}^{L-1} (y_j^i - s^{\mathbf{a}}(j, i))^2 > \sum_{j=k_0}^{k_1+G-1} \sum_{i=0}^{L-1} (y_j^i - s^{\mathbf{a}^{\epsilon}}(j, i))^2, \quad (5.5)$$

or, equivalently, if

$$\begin{aligned} \sum_{j=k_0}^{k_1+G-1} \sum_{i=0}^{L-1} \Delta_0^{\epsilon}(j, i) n_j^i &> \\ \frac{1}{2} \sum_{j=k_0}^{k_1+G-1} \sum_{i=0}^{L-1} \Delta_0^{\epsilon}(j, i)^2 - 2i^{\mathbf{a}}(j, i) \Delta_0^{\epsilon}(j, i). \end{aligned} \quad (5.6)$$

Let K be the total number of symbols transmitted for a single track and \mathcal{R}^K the vector space of K -tuples of real numbers. It is convenient to define the following vectors in \mathcal{R}^K

$$\begin{aligned} \Phi_0^{\epsilon}(l) &= [\Delta_0^{\epsilon}(0, l) \Delta_0^{\epsilon}(1, l) \dots \Delta_0^{\epsilon}(K, l)]^T; \\ \Phi_1^{\epsilon}(l) &= [\Delta_1^{\epsilon}(0, l) \Delta_1^{\epsilon}(1, l) \dots \Delta_1^{\epsilon}(K, l)]^T; \\ \Lambda(l) &= [i^{\mathbf{a}}(0, l) i^{\mathbf{a}}(1, l) \dots i^{\mathbf{a}}(K, l)]^T; \\ \mathbf{n}(l) &= [n(0, l) n(1, l) \dots n(N, l)]^T. \end{aligned} \quad (5.7)$$

Since clearly $\Delta_0^\varepsilon(k, l) = 0$ for $k \geq k_1 + G$ or $k < k_0$, we can express the condition for error event ε to occur as

$$\sum_{l=0}^{L-1} \Phi_0^\varepsilon(l) \mathbf{x}(l)^T > \frac{d(\varepsilon)}{2\sigma}, \quad (5.8)$$

where $\mathbf{x}(l) = \frac{1}{\sigma} \mathbf{n}(l)$ and $d(\varepsilon)$ is the Euclidian weight (denoted as Euclidian distance in the remainder of the text) of a particular error event ε :

$$d(\varepsilon) = \sum_{l=0}^{L-1} \|\Phi_0^\varepsilon(l)\|^2 - 2\Lambda(l)\Phi_0^\varepsilon(l)^T. \quad (5.9)$$

In the absence of any RISI, $\Lambda(l) = [0 \ 0 \dots 0]^T$, and the Euclidian distance reduces to $d_0(\varepsilon) = \sum_{l=0}^{L-1} \|\Phi_0^\varepsilon(l)\|^2$ which is the usual expression for the distance of error event ε .

5.3.2 Probability of Error of the ISI cancellation scheme

It is convenient for the analysis of the error performance of the ISI cancellation scheme to assume that both the main and the tentative detectors of Fig. 5.2 are VDs. Both VDs are matched to the desired component of the channel (i.e the target response g^0). An interference canceller with response $f(p, q) = g^1(p, q)$ for every track is fed with the decisions of the tentative VD. The probability of error for the cancellation scheme can be expressed as

$$P_e \leq \sum_{\varepsilon \in E, \varepsilon' \in E_0} w_H(\varepsilon) P_1(\varepsilon, \varepsilon') P_2(\varepsilon, \varepsilon'), \quad (5.10)$$

where E is the set of all error events without the null event, E_0 is the set of all error events including the null event, $w_H(\varepsilon)$ is the number of symbol errors in the error event $\varepsilon(k, l)$, $P_1(\varepsilon, \varepsilon')$ is the probability that the tentative VD selects the path associated with error event ε' and the main VD selects the path associated with error event ε , and $P_2(\varepsilon, \varepsilon')$ is the probability of the occurrence of a data sequence \mathbf{a} that supports both ε and ε' as possible error events.

The probability $P_1(\varepsilon, \varepsilon')$ can be computed by expressing the conditions for which both VDs make a decision error. The condition for the tentative VD is given by (5.8), whereas the condition for the main VD can be obtained by replacing $\Lambda(l)$ in condition (5.8) by $\Phi_1^{\varepsilon'}(l)$. Therefore, the conditions for ε and ε' are

$$\sum_{l=0}^{L-1} \Phi_0^{\varepsilon'}(l) x > \frac{d(\varepsilon')}{2\sigma} \quad (5.11)$$

and

$$\sum_{l=0}^{L-1} \Phi_0^\varepsilon(l) x > \frac{d(\varepsilon|\varepsilon')}{2\sigma}, \quad (5.12)$$

where

$$d(\epsilon') = \sum_{l=0}^{L-1} \|\Phi_0^{\epsilon'}(l)\|^2 - 2\Lambda(l)\Phi_0^{\epsilon'}(l) \quad (5.13)$$

and

$$d(\epsilon|\epsilon') = \sum_{l=0}^{L-1} \|\Phi_0^{\epsilon}(l)\|^2 + 2\Phi_1^{\epsilon'}(l)\Phi_0^{\epsilon}(l). \quad (5.14)$$

Conditions (5.11) and (5.12) define a region in \mathcal{R}^N delimited by two $N-1$ -dimensional hyperplanes. It is always possible to introduce an orthogonal transformation such that Φ_0^{ϵ} and $\Phi_0^{\epsilon'}$ lie on a 2-D plane. Then the hyperplanes become simple straight lines and the region can be easily visualized. As a result the joint probability of ϵ and ϵ' can be computed by integrating the 2-D Gaussian density with unit variance

$$\mathcal{N}(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 + y^2)}. \quad (5.15)$$

Three cases are of interest:

- **Case I:** The two vectors Φ_0^{ϵ} and $\Phi_0^{\epsilon'}$ are orthogonal. In this case the joint probability of ϵ and ϵ' can be expressed as

$$P_1(\epsilon, \epsilon') = \mathcal{Q}\left(\frac{d(\epsilon')}{2\sigma}\right) \mathcal{Q}\left(\frac{d(\epsilon|\epsilon')}{2\sigma}\right). \quad (5.16)$$

- **Case II:** The same error event occurs in the tentative and the main detectors ($\epsilon = \epsilon'$, and $\Phi_0^{\epsilon} = \Phi_0^{\epsilon'}$). Because the two conditions (5.12) and (5.11) must be satisfied, the joint probability can be expressed as

$$P_1(\epsilon, \epsilon') = \mathcal{Q}\left(\frac{\max(d(\epsilon'), d(\epsilon|\epsilon'))}{2\sigma}\right). \quad (5.17)$$

In general $d(\epsilon') > d(\epsilon|\epsilon')$ and hence the probability of error of the ISI cancellation scheme is basically determined by the probability of error of the tentative detector.

- **Case III:** Vectors Φ_0^{ϵ} and $\Phi_0^{\epsilon'}$ are neither parallel nor orthogonal. In this case the integral does not have a closed-form solution but some tight upper bounds can be computed in many cases of interest.

In the next subsection the effect of error propagation on the performance of the receiver is studied.

5.3.3 Error Propagation in the Receiver using Tentative Decisions for ISI Cancellation

For each ϵ , consider the set E_ϵ of all those events ϵ' that satisfy the condition

$$\sum_{l=0}^{L-1} \Phi_1^T(l, \epsilon') \Phi_0(l, \epsilon) = 0. \quad (5.18)$$

This condition is obviously met in case $\Phi_1(l, \epsilon') = 0$ for $l = [0, L-1]$ (the case of no RISI), but also in case $\Phi_1(l, \epsilon')$ is orthogonal to $\Phi_0(l, \epsilon)$. Therefore, as a result of (5.14), the distance of event ϵ is not affected by the existence of event ϵ' in the tentative detector. The summation of (5.10) can then be split into two terms as follows:

$$P_e \leq \mathcal{P}_1 + \mathcal{P}_2, \quad (5.19)$$

where

$$\mathcal{P}_1 = \sum_{\epsilon \in E, \epsilon' \in E_\epsilon} w_H(\epsilon) P_1(\epsilon, \epsilon') P_2(\epsilon, \epsilon') \quad (5.20)$$

and

$$\mathcal{P}_2 = \sum_{\epsilon \in E, \epsilon' \in \bar{E}_\epsilon} w_H(\epsilon) P_1(\epsilon, \epsilon') P_2(\epsilon, \epsilon'), \quad (5.21)$$

where \bar{E}_ϵ is defined as the complement of E_ϵ with respect to E_0 (as defined in the text following (5.10)). The probability \mathcal{P}_1 represents the error rate of the main VD in case RISI is absent, i.e. the case of ideal cancellation. Following the results presented in Section 5.3.1, this error rate can be expressed as

$$\mathcal{P}_1 = \sum_{\epsilon \in E} w_H(\epsilon) P_1(\epsilon) P_2(\epsilon), \quad (5.22)$$

where

$$P_1(\epsilon) = Q\left(\frac{d_0(\epsilon)}{2\sigma}\right) \quad (5.23)$$

is the probability of event ϵ in the main detector.

The probability \mathcal{P}_2 represents the error propagation effect caused by the errors made by the tentative detector. The cancellation scheme will be effective if $\mathcal{P}_2 < \mathcal{P}_1$, i.e. bit errors due to error propagation will not significantly determine the overall bit-error rate. However if $\mathcal{P}_2 \geq \mathcal{P}_1$, error propagation will mainly determine the overall bit-error rate. Here we analyze individual terms contributing to \mathcal{P}_2 and determine when they will lead to error propagation.

- **Case I:** For small RISI values, $d(\epsilon|\epsilon')$ will not be significantly smaller than $d_0(\epsilon)$, i.e. $\sum_{l=0}^{L-1} \Phi_1^T(l, \epsilon') \Phi_0(l, \epsilon) \ll \sum_{l=0}^{L-1} \|\Phi_0(l, \epsilon)\|^2$ and $\sum_{l=0}^{L-1} \Lambda^T(l) \ll \sum_{l=0}^{L-1} \|\Phi_0(l, \epsilon)\|^2$. Therefore,

$$Q\left(\frac{d(\epsilon')}{2\sigma}\right) Q\left(\frac{d(\epsilon|\epsilon')}{2\sigma}\right) \ll Q\left(\frac{d_0(\epsilon)}{2\sigma}\right) \quad (5.24)$$

and error events of this type will not cause error propagation. If \mathcal{P}_2 is dominated by these events, the cancellation scheme will be effective because P_e is essentially determined by the error rate \mathcal{P}_1 (ideal cancellation). The performance of the tentative detector will not influence the overall performance as the performance of ideal cancellation determines the overall performance, i.e. no additional performance improvement can be obtained by improving the reliability of the tentative detector. For this reason a simple detector (e.g. a symbol-by-symbol detector) may be used as tentative detector. The cancellation scheme only ceases to be effective when RISI values tend to become large.

- **Case II:** The term $\varepsilon = \varepsilon'$ can contribute to \mathcal{P}_2 only if $\sum_{l=0}^{L-1} \Phi_1^T(l, \varepsilon') \Phi_0(l, \varepsilon) \neq 0$. For many channels, events satisfying this condition have large distances and as a result will not significantly contribute to \mathcal{P}_2 . However there are channels where events satisfying this condition do have minimum distance and as a result are the dominating terms. Since usually $d(\varepsilon|\varepsilon') < d(\varepsilon')$ (in words, cancellation with erroneous decisions is worse than no cancellation at all), the error rate of the cancellation scheme is essentially determined by the error rate of the tentative detector. Therefore the cancellation will be ineffective.
- **Case III:** Nonorthogonal error events with minimum or nearly minimum distance will cause error propagation and as a result the cancellation will be ineffective.

To summarize, the main conditions for which ISI cancellation can work effectively are stated:

- the RISI must be small such that the main VD can make relatively reliable decisions even if the tentative detector makes a decision error and such that the tentative detector can make relatively reliable decisions in spite of the RISI,
- errors affecting the main and the tentative detector must be statistically independent.

5.3.4 Examples

In this section a simple example of a 2-D ISI cancellation scheme will be used to study the effect of error propagation. A system with 3 adjacent tracks is considered ($L = 3$), where the data symbols in the different tracks form a hexagonal structure. A full 2-D VD is used as tentative and as main bit-detector. These detectors operate

based on the target response

$$g_0 = \begin{bmatrix} 0 & 0 & 0 & c & c & 0 & 0 & 0 \\ 0 & 0 & 0 & c & 1 & c & 0 & 0 & 0 \\ 0 & 0 & 0 & c & c & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (5.25)$$

where the parameter $c = 0.3$. For every track l , the noise samples n_k^l are white and Gaussian with variance σ^2 . The noise samples of different tracks are uncorrelated. Furthermore, SNR is defined as

$$\text{SNR} = 10 * \log \left(\frac{S_g^2}{\sigma^2} \right). \quad (5.26)$$

where $S_g^2 = 1 + 6c^2$ is the total received energy per transmitted bit. Three different RISI impulse responses g_1 are considered.

Example 1

In this example the RISI impulse response g_1 is defined as (g_1 is aligned with g_0 defined above)

$$g_1 = \begin{bmatrix} 0 & 0 & \eta & 0 & 0 & \eta & 0 & 0 \\ 0 & 0 & \eta & 0 & 0 & 0 & \eta & 0 & 0 \\ 0 & 0 & \eta & 0 & 0 & \eta & 0 & 0 & 0 \end{bmatrix}. \quad (5.27)$$

The different RISI components originate from symbols adjacent to the symbols that have non-zero coefficients in the target response g_0 . The theoretical Bit Error Rates (BERs) of the main and the tentative VD are plotted versus SNR in Fig. 5.3. In the left plot $\eta/S_g = 0.05$ and in the right plot $\eta/S_g = 0.1$. Furthermore the BER in case ideal cancellation is performed (\mathcal{P}_1), is also shown in Fig. 5.3.

These theoretical results are validated by simulation results. For low SNRs the theoretical BER of the tentative VD is not very accurate. This can be explained by the fact that in the simulation (as in reality) only one error event can occur for a single data pattern while the theoretical BER is calculated by summing the errors of all possible error events for a single data pattern. Furthermore, it must be noted that the theoretical and the simulated results for the main VD do not match well for large values of η/S_g . This is due to the fact that all tentative error events ϵ' that lead to case II and case III situations (non-orthogonal error events) are not taken into account because of the computational complexity and especially for large values of η/S_g these error events may have small Euclidian distances. As a result the theoretical BER of the main VD is not a very accurate estimate of the actual BER. The BERs presented in Fig. 5.3 show that the ISI cancellation scheme does not achieve the performance of ideal cancellation even for small values of η/S_g . For this example, the temporal separation between the symbols causing the RISI and the detected symbols is very limited

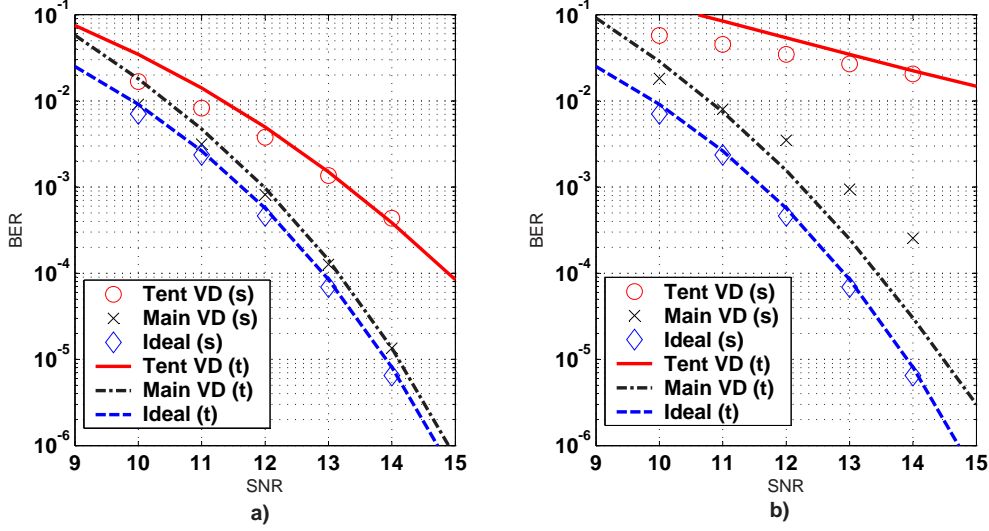


Figure 5.3: Example 1: BERs of the main and the tentative VD are plotted versus SNR together with BERs of ideal cancellation. Both theoretical (t) and simulation (s) results are shown. a) $\eta/S_g = 0.05$. b) $\eta/S_g = 0.10$.

and as a result the ISI cancellation will suffer from error propagation. However, the BER of the main VD is better than the BER of tentative VD. This can be explained by the fact that due to the large number of RISI components, cancellation with a limited amount of erroneous decisions is better than no cancellation at all. As a result impressive gains in BER are observed for both small and large RISI amplitudes despite the error propagation. Concluding, ISI cancellation for this RISI impulse response substantially improves the BER even though error propagation due to non-orthogonal error events prevents the system to achieve ideal cancellation performance.

Example 2

In this example the RISI impulse response g_1 is defined as

$$g_1 = \begin{bmatrix} \eta & 0 & 0 & 0 & 0 & 0 & 0 & \eta \\ \eta & 0 & 0 & 0 & 0 & 0 & 0 & \eta \\ \eta & 0 & 0 & 0 & 0 & 0 & 0 & \eta \end{bmatrix}. \quad (5.28)$$

Theoretical and simulated BERs of the main VD, the tentative VD and ideal cancellation are plotted in Fig. 5.4 versus SNR. For this RISI impulse response the symbols causing RISI have sufficient temporal separation such that the vectors $\Phi_0(\epsilon)$ and $\Phi_0(\epsilon')$ are orthogonal for the error events with minimum distance. As a result error

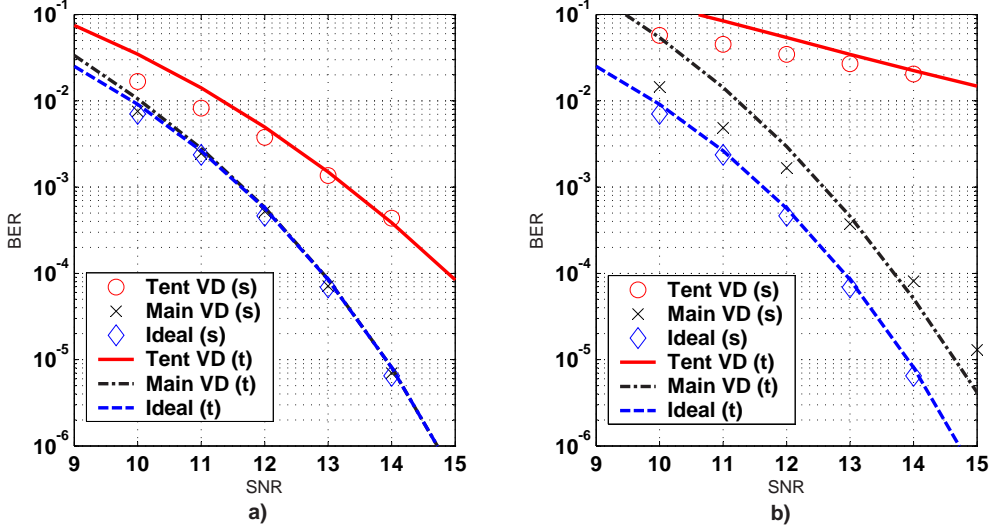


Figure 5.4: Example 2: BERs of the main and the tentative VD are plotted versus SNR together with BERs of ideal cancellation. Both theoretical (t) and simulation (s) results are shown. a) $\eta/S_g = 0.05$. b) $\eta/S_g = 0.10$.

propagation is limited and the BER of the ISI cancellation scheme is almost equal to the BER of ideal cancellation. The ISI cancellation scheme ceases to work efficiently if the value η/S_g becomes too large. In this case also non-minimum distance error events will cause error propagation and as a result the BER deteriorates. But the obtained BER improvement is so impressive (the SNR gain amounts to 5 dB at $\text{BER} = 10^{-2}$) that cancellation of RISI components with large amplitudes is nevertheless very valuable.

Example 3

In the previous examples the RISI impulse response represented amplitude distortions. In this example phase distortion is treated. The RISI impulse response g_1 is defined as

$$g_1 = \begin{bmatrix} 0 & 0 & \eta & 0 & 0 & \eta & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\eta & 0 & 0 & -\eta & 0 & 0 \end{bmatrix}. \quad (5.29)$$

Theoretical and simulated BERs of the main VD, the tentative VD and ideal cancellation are plotted in Fig. 5.5 versus SNR. Despite the fact that the error events of the main and the tentative VD are not always orthogonal, the performance of the ISI cancellation scheme is not severely affected by error propagation. From the results

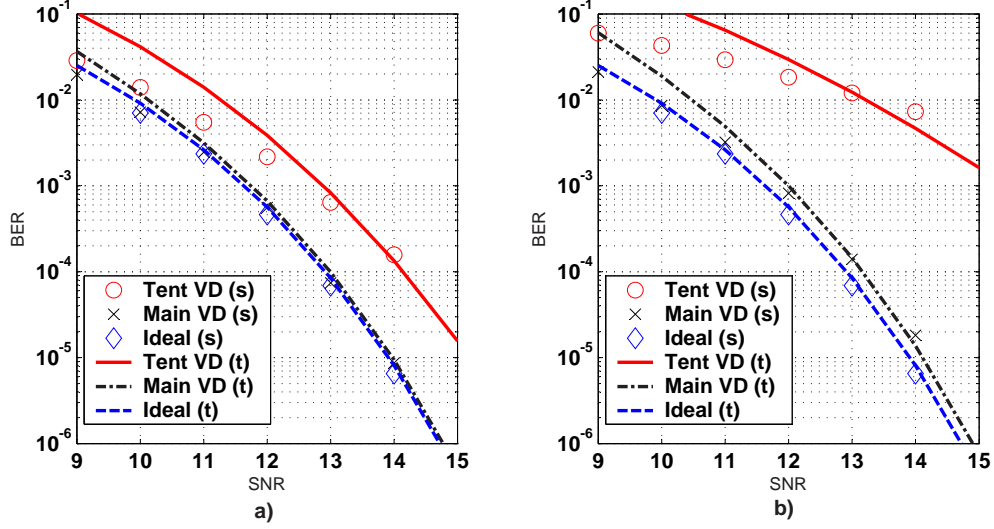


Figure 5.5: Example 3: BERs of the main and the tentative VD are plotted versus SNR together with BERs of ideal cancellation. Both theoretical (t) and simulation (s) results are shown. a) $\eta/S_g = 0.05$. b) $\eta/S_g = 0.10$.

shown in Fig. 5.5, it is clear that the ISI cancellation scheme works efficiently in case RISI is present in the form of phase distortion.

The results presented in these examples clearly show that the conditions for effective cancellations are easily met in 2-D systems. Even if error propagation becomes an issue, the performance improvement is significant in all cases. The examples show that ISI cancellation is effective for amplitude and phase distortion and as a result ISI cancellation will also be effective for a combination of both distortions.

5.4 Experimental Results for TwoDOS

In the TwoDOS system, bits are stored on a hexagonal lattice [1]. In contrast with conventional optical storage (CD, DVD and BD), where the bits are stored in a single spiral (a 1-D sequence of bits), in TwoDOS the bits are organized in a so-called broad spiral. The broad spiral contains a number L of bit-tracks, stacked upon each other to form a hexagonal structure, see Fig. 5.6. Adjacent rotations of the broad spiral are separated by a guard band consisting of a bit-track without any pits. The data is read out with an array of L laser spots arranged such that each spot is centered on one of the bit-tracks within the broad spiral. A multi-spot photo detector integrated circuit is used to generate a so-called high-frequency (HF) signal for every bit-track.

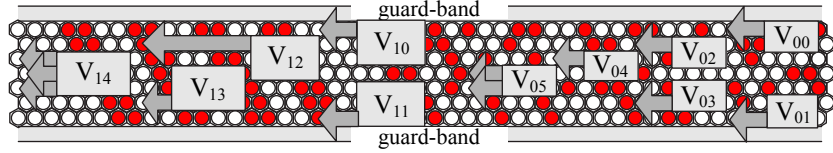


Figure 5.6: Example of hexagonal structure for TwoDOS with $L = 7$. The configuration of stripe-wise Viterbi detection is also shown for this example: two iterations of stripe detectors are used to detect the bits of the broad spiral.

A PRML receiver has been built for TwoDOS [145, 147]. It consists of a bit-detector preceded by an adaptive equalizer, an adaptive DC compensator, an AGC and a timing recovery loop. A 2-D VD performs joint bit-detection on all bit-tracks. To reduce the complexity of a full-fledged 2-D VD, the VD is divided into smaller processing units (called stripe VD). Each stripe VD covers a limited number of bit-tracks (so-called stripes with a typical height of 2 or 3 bit-tracks). This detection configuration is called a Stripe-Wise Viterbi Detector (SWVD) [152] and is shown in Fig. 5.6 together with the hexagonal structure. The SWVD consists of two detection iterations, where every iteration consists of stripe VDs which are organized in a “<”-shape. The first iteration is performed by stripe VDs (V_{00} up to V_{05} in the figure) each covering two bit-tracks, which results in a stripe VD with 16 states. The second iteration consists of stripe VDs (V_{10} up to V_{14} in the figure) each covering three bit-tracks which results in a stripe VD with 64 states. In every iteration the binary output from a first stripe VD is passed to a next stripe VD to be used as side information in the branch metric calculations [152]. The outputs of the first iteration are used as side information in the second iteration. More explanation about the operation of the SWVD can be found in Section 2.6.

Electronic beam recorded discs with a capacity of 50 GB (single layer) are placed in an experimental read-out system to produce experimental replay signals (for these discs $L = 7$). The read-out is conducted under relatively favorable conditions (no scratches, no dropouts, limited amount of dust). The angle of the disc with respect to the laser beam can be varied in a controlled way to identify the performance of the system (BER) for varying angles (denoted as tilt angles). Subsequently the replay signals are digitized and are applied to the TwoDOS receiver. In the TwoDOS receiver an ISI cancellation scheme is implemented that uses the outputs of the first iteration as tentative decisions. The impulse response f_k^l of the interference canceller is estimated using an identification scheme. This identification scheme estimates the RISI impulse response at the detector input in a data-aided way (using a training sequence). This RISI impulse response of the central track ($l = 3$) is shown in Fig. 5.7 for -1.0° of radial tilt. The RISI components are limited in amplitude ($[-0.04, 0.04]$). In this case there are more than 20 RISI components with a significant amplitude (> 0.01).

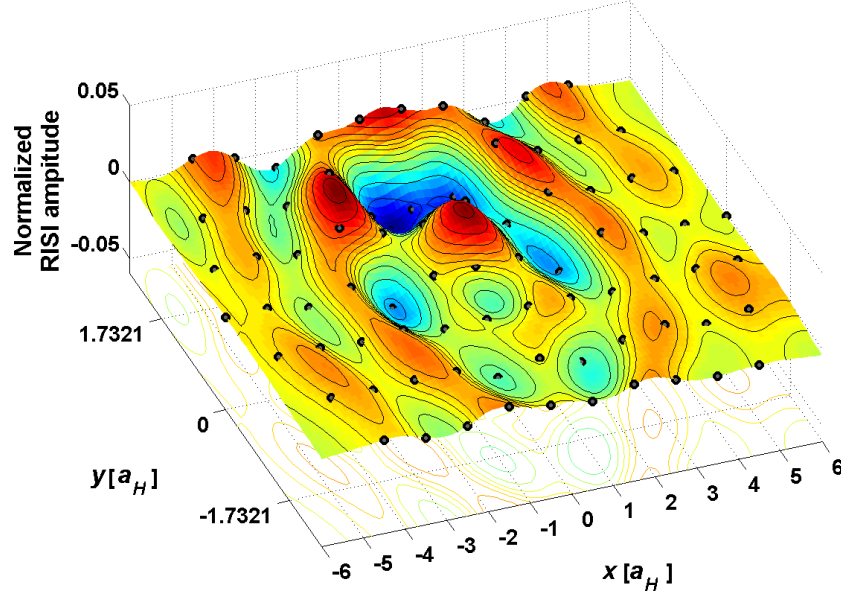


Figure 5.7: Estimated amplitudes (normalized with respect to S_g) of the RISI impulse response at the detector input for $l = 3$. The x -axis is the tangential direction and the y -axis is the radial direction (centered around track $l = 3$). Both axis are scaled in terms of a_H , where a_H is the distance between two bits measured on the disc ($a_H = 138$ nm for 50 GB disc).

Hence application of 2-D ISI cancellation might be very beneficial. In the equivalent 1-D case there would be only 4 or 5 significant RISI components and ISI cancellation would not be beneficial. Furthermore, RISI originating from symbols with limited temporal separation from the symbols of the target response (in the figure dots with indices $x \in \{-1, 0, 1\}$, $y \in \{-0.866, 0, 0.866\}$) is non-negligible. Based on the latter observation, ISI cancellation will suffer from error propagation and as a result the performance of ideal cancellation will not be achieved. In Subsection 5.4.1 the performance of ISI cancellation is discussed in case a SWVD is used that consists of two detection iterations. The reliability of the tentative decisions can be improved by inserting an additional detection iteration. This topology with three detection iterations is discussed in Subsection 5.4.2. Another method to improve the reliability of the tentative decisions is the application of cross-talk cancellation before the first detection iteration. This topology with cross-talk cancellation in combination with two and three detection iterations is discussed in Subsection 5.4.3.

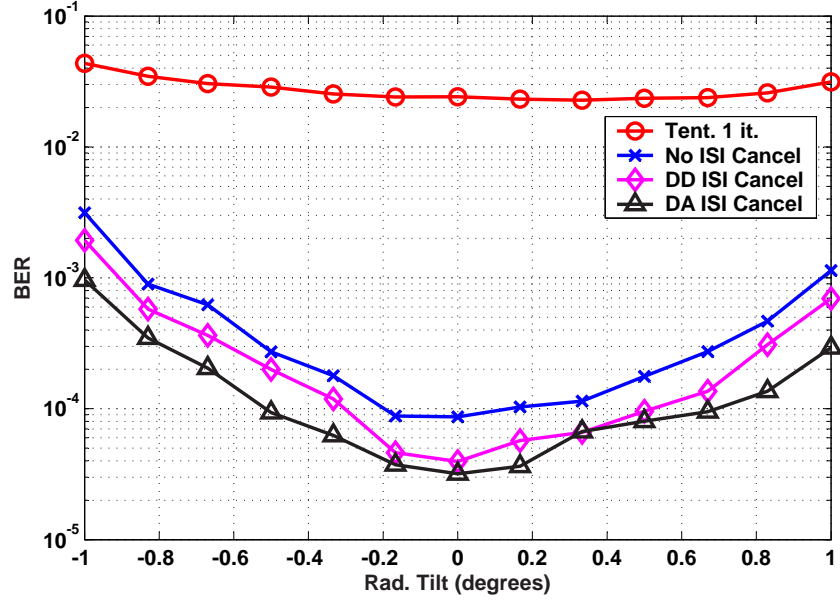


Figure 5.8: BER for different radial tilt angles (in degrees). BER is calculated for different receiver topologies: without ISI cancellation, with ISI cancellation based on the tentative decisions (DD ISI Cancel) and with ISI cancellation based on the actual data (DA ISI cancel).

5.4.1 SWVD with Two Detection Iterations

The results of the experimental system with a SWVD that consists of two detection iterations are shown in Fig. 5.8. In this figure, together with the BER of tentative decisions (outputs of the first iteration), the BER after the second iteration is plotted versus the radial tilt angle for three different topologies: 1) no ISI cancellation, 2) ISI cancellation based on the outputs of the first SWVD iteration (denoted as Decision-Directed, DD cancellation), and 3) ISI cancellation based on the actual bits written on the disc (denoted as Data-Aided, DA cancellation which is clearly not applicable in practical systems but serves as reference for ideal cancellation). The first topology (no ISI cancellation) is the detection topology described in [152]. The performance of this topology will be used as reference to judge the performance of the different topologies.

The application of ISI cancellation is beneficial for this experimental system. The BER at nominal conditions (no radial tilt) is improved from 8.7×10^{-5} to 3.9×10^{-5} . Also the so-called bath tub curve (the BER vs. tilt angles) has broadened, i.e. higher tilt angles can be allowed to achieve the same performance. For example at a given

$\text{BER} = 10^{-4}$, the allowed margins for radial tilt are improved from $[-0.2^\circ, 0.2^\circ]$ to $[-0.3^\circ, 0.5^\circ]$. Hence ISI cancellation nearly doubles the allowed tilt margins for this experimental system. These results show that still a substantial amount of RISI is left at the input of the tentative detector. The comparison of the results of the DA and the DD ISI cancellation shows that although the ISI cancellation scheme considerably improves the performance, it does not reach the performance of ideal cancellation. This performance gap between DA and DD ISI cancellation indicates that error propagation is an issue for this kind of RISI impulse response.

5.4.2 SWVD with Three Detection Iterations

By improving the reliability of the tentative decisions used in the cancellation scheme, error propagation may be lowered and as a result the overall performance may be enhanced. One way to improve this reliability is the insertion of an additional detection iteration with 2-track stripe VDs. This insertion will not substantially increase the overall complexity as an iteration with the 3-track stripe VDs is much more complex than an iteration with 2-track stripe VDs. As a result a topology with three detection iterations is used: first, two iterations of 2-track stripe VDs and finally, one iteration of 3-track stripe VDs. Before the second and before the third iteration interference cancellers are applied to cancel RISI components based on the decisions of respectively the first and the second iteration. The results of the detector with three iterations are presented in Fig. 5.9. These results should be compared with the results of the detector with two iterations (see Fig. 5.8).

By comparing the results of these two detectors, a couple of conclusions can be drawn:

- The insertion of an additional iteration of 2-track stripe VDs improves the reliability of the tentative decisions by a factor of 10.
- Even without any ISI cancellation the additional detection iteration improves the tilt margin from $[-0.2^\circ, 0.2^\circ]$ to $[-0.3^\circ, 0.4^\circ]$.
- The BER of the final decisions by applying DD ISI cancellation is improved by inserting an additional iteration. The tilt margin in this case is improved from $[-0.3^\circ, 0.5^\circ]$ to $[-0.5^\circ, 0.7^\circ]$.
- Even for the configuration with three detection iterations, error propagation is still an important issue. This can be seen by comparing the results of the DA and the DD cancellation schemes. Especially for higher tilt angles, the performance gap can be explained by the fact that the RISI components with limited temporal separation from the target response have a large amplitude such that error propagation is enhanced (see Fig. 5.7). For small tilt angles, the

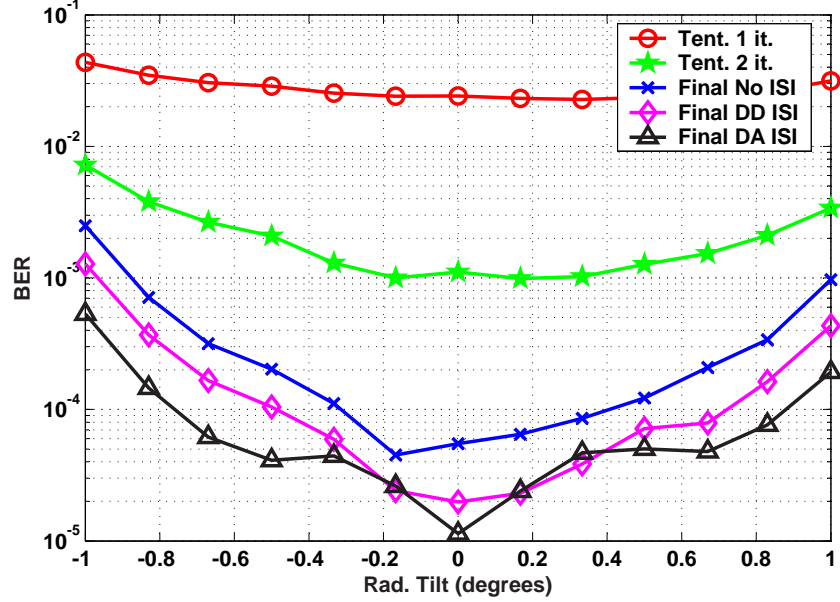


Figure 5.9: BER for different radial tilt angles (in degrees). BER is calculated for different receiver topologies with three detection iterations: without ISI cancellation, with ISI cancellation based on the tentative decisions (DD ISI Cancel) and with ISI cancellation based on the actual data (DA ISI cancel).

amplitude of these RISI components is so small that error propagation is very limited and as a result, the DD ISI cancellation scheme (almost) achieves the performance of ideal cancellation (DA ISI cancellation).

Summarizing, the insertion of an additional iteration of 2-track stripe VDs improves the performance of the ISI cancellation scheme substantially.

5.4.3 Cross-Talk Cancellation

In the detection process not all bits are defined by the branch in the trellis of a stripe VD: some of them lie either within the bit-track immediately above the stripe or within the bit-track immediately below the stripe. These bits are considered as the side-information that is required for the stripe VD. In the SWVD side information is taken from the array of most recent bit decisions. However, in the first iteration no bit decisions are available yet. This means that side information is either not present or has a very poor quality for example when it is generated by simple threshold detection. This lack of side information is reflected in the fact that BER of the first detection iteration is quite low. This can be seen in Fig. 5.9 where a BER improve-

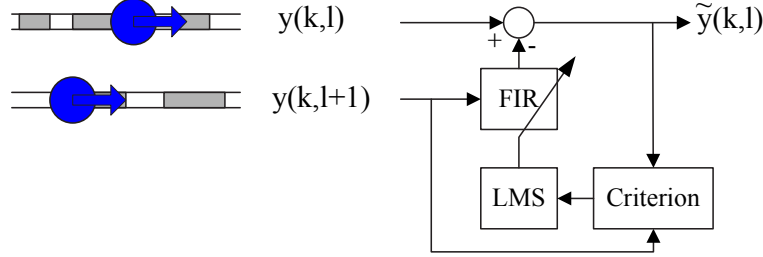


Figure 5.10: Single-sided version of the cross-talk cancellation scheme.

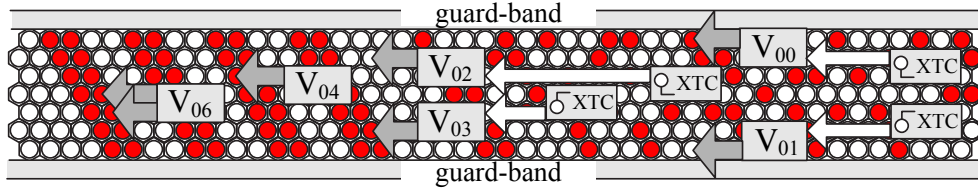


Figure 5.11: Scheme showing the first iteration of the stripe-wise detection including cross-talk cancellation. The white arrows indicate where the compensated signals are used. Note that the compensated signals are only used at the boundaries with low certainty.

ment of a factor of 10 can be observed between an iteration of 2-track stripe VDs without and with side information (the tentative decisions of respectively the first and the second iteration).

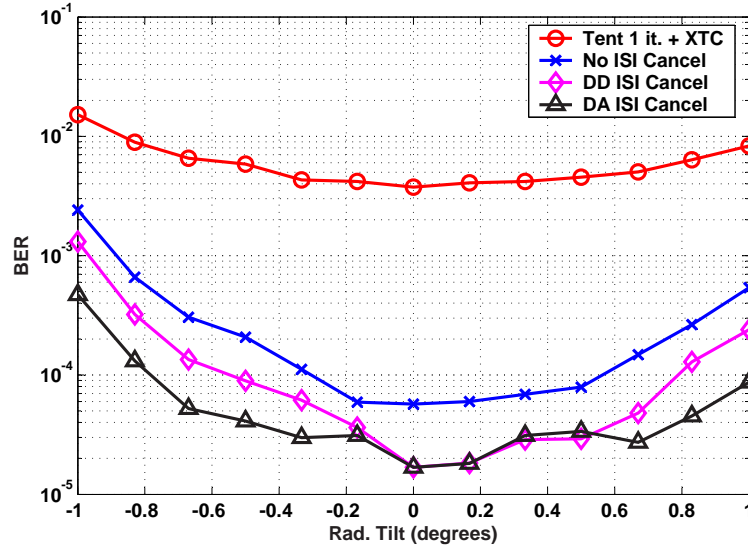
A possible alternative that enables us to avoid using unreliable side information is cross talk cancellation (XTC). Here, we will start the discussion with a single-sided version of the conventional XTC scheme as shown in Fig. 5.10, i.e. only using one side-track for compensation of cross talk.

The filtering is performed using a Finite Impulse Response (FIR) filter. The coefficients of this filter are denoted $v(i, l+1)$, and are adapted using a least mean square (LMS) algorithm based on some suitable criterion. We can write the XTC scheme as:

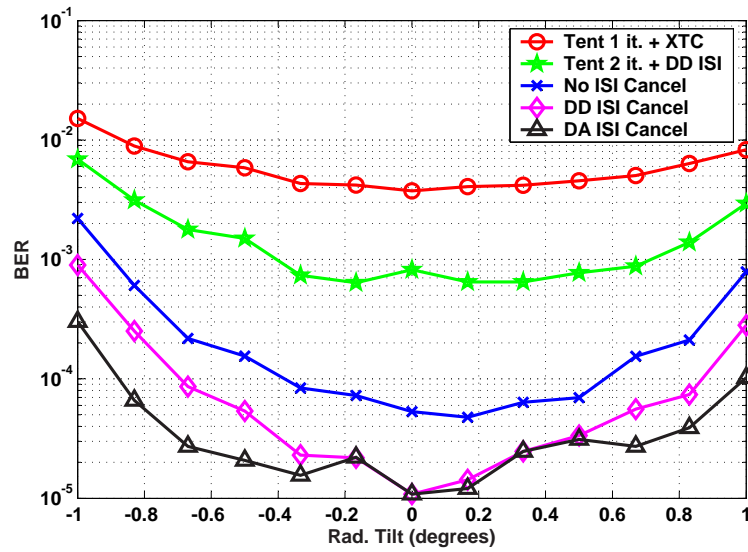
$$\tilde{y}_k^l = y_k^l - \sum_i v(i, l+1) y_{k-i}^{l+1} \quad (5.30)$$

with y_k^l the detector input signal of track l at time instant k . Signal \tilde{y}_k^l is the compensated signal.

Several criteria are possible to update the filter coefficients. For a practical implementation we have chosen to minimize the mean square error, where the error is taken as the difference between the actual result of cross-talk cancellation and the



(a) Two detection iterations



(b) Three detection iterations

Figure 5.12: BER for different radial tilt angles (in degrees). BER is calculated for different receiver topologies with XTC: without ISI cancellation, with ISI cancellation based on the tentative decisions (DD ISI Cancel) and with ISI cancellation based on the actual data (DA ISI cancel).

signal that we would expect after ideal cross-talk cancellation based on the target response. In [18] it is derived that a data aided LMS algorithm to minimize this squared error can be replaced by a non-data aided zero forcing (ZF) algorithm by scaling the resulting filter coefficients and adding a DC-term to the output signal. The big advantage of this scheme is that it does not need any preliminary decisions. The ZF based algorithm is applied in the first iteration of the stripe-wise detection at the low-certainty boundary according to the diagram in Fig. 5.11.

The experimental results where XTC was applied before the first iteration, are shown in Fig. 5.12. In the upper part of the figure, BERs are shown for a SWVD with two detection iterations and in the lower part of the figure a SWVD with three detection iterations was used. Three different topologies were employed: (1) no ISI cancellation, (2) DD ISI cancellation and (3) DA ISI cancellation.

Based on Fig. 5.12 and Fig. 5.9 the following observations can be made.

- The reliability at the output of the first iteration with XTC is not as good as the one obtained after a second iteration of 2-track stripe VDs without XTC: $3.7 \cdot 10^{-3}$ with respect to $1 \cdot 10^{-3}$. This difference can be explained by the fact that the detection with XTC does not take the signal energy of the cancelled track into account during the detection process.
- The performance of the DD ISI cancellation scheme for topology with two detection iterations and XTC is comparable with the performance obtained by applying three detection iterations with in between DD ISI cancellation and without XTC (tilt margin are respectively $[-0.55^\circ, 0.8^\circ]$ and $[-0.5^\circ, 0.7^\circ]$). For this reason the topology with two iterations in combination with XTC is preferable to the topology with three detection iterations, because the latter topology is more complex.
- For the SWVD with both two and three iterations, DD ISI cancellation almost achieves ideal cancellation performance for small tilt angles. As the tilt angle becomes larger the amplitude of the RISI components increases and as a result error propagation is invoked. This error propagation causes the performance of the ISI cancellation scheme to deviate from the performance of ideal cancellation.

By applying XTC, an additional detection iteration of 2-track stripe VDs and ISI cancellation in between every detection iteration the tilt margin is improved from $[-0.2^\circ, 0.2^\circ]$ (see Fig. 5.8) to $[-0.7^\circ, 0.85^\circ]$. Summarizing, in all different receiver topologies (two or three detection iterations, with or without XTC) the application of ISI cancellation consistently improves the performance of the receiver substantially.

5.5 Conclusions

In this chapter we have studied the application of linear ISI cancellation in 2-D systems. A first argument in favor of ISI cancellation is the fact that the number of small RISI components increases considerably with respect to the one-dimensional case. As a result the performance gain by applying the cancellation technique also increases. A second argument is the fact that only little additional complexity may be required for the application of ISI cancellation. Because 2-D detection is often accomplished by several iterations of smaller detection units to avoid the complexity of a full 2-D Viterbi detector, decisions of one of these smaller detection units can be used as tentative decisions by the canceller. As a result no additional bit-detectors need to be implemented to produce these tentative decisions. Summarizing these arguments, with a limited additional complexity (only the interference canceller needs to be added, not an additional detector) ISI cancellation in a two-dimensional system may improve performance significantly. Experimental results based on the read-out of a 50 GB single layer-disc were provided for an experimental two-dimensional optical storage system. These results show that the application of ISI cancellation nearly doubles the allowed tilt margin at a BER of 10^{-4} . Furthermore by applying an additional detection iteration or by applying cross-talk cancellation to improve the reliability of the tentative decisions, the allowed tilt margin increases even further.

Chapter 6

Minimum-Latency Tracking of Rapid Variations in Two-Dimensional Storage Systems.

The trend of increasing storage densities results in growing sensitivity of system performance to variations of storage channel parameters. To counteract these variations more adaptivity is needed in the data receiver. Accurate tracking of rapid variations is limited by latencies in the adaptation loops. These latencies are largely governed by delays of the bit-detector. In two-dimensional storage systems data are packaged in a group of adjacent tracks or rows, and for some of the rows the detection delays can increase dramatically with respect to one-dimensional systems. As a result the effective latencies in the adaptation loops preclude the tracking of rapid variations and really limit the performance of the system. In this chapter, a scheme is proposed that overcomes this problem and that can be used for timing recovery, automatic gain control and other adaptive circuits. Rapid variations for all the rows are tracked using control information from rows for which detector latency is smallest. This works properly if rapid variations are common across the rows as is the case, for example, for the Two-Dimensional Optical Storage (TwoDOS) system.

6.1 Introduction

Steadily increasing storage densities are a clear trend in storage systems [219]. Reduced margins (e.g. margins with respect to disc tilt, Signal-to-Noise Ratio, SNR) and increased sensitivity to piece-wise and temporal variations of physical storage channel parameters are consequences of this trend and necessitate an increasing amount and an increasing accuracy of adaptivity (e.g. timing recovery, automatic gain control and other adaptive loops) in the data receiver [220]. This accuracy is especially hard to accomplish for the tracking of rapid variations, and is limited by latencies in the adaptation loops.

Another consequence of the increasing densities is that SNRs decrease. As a

result the bit-detector that forms part of the data receiver needs to become more complex to maintain detection reliability. This increased complexity inevitably increases the detection delay. Because this delay contributes to the overall latency in the adaptation loops, it will put an increasingly severe limit on their capabilities to track rapid variations.

A widely adopted solution to improve these tracking capabilities is to base adaptation on tentative decisions with a limited detection delay instead of on final bit decisions [4]. This limited delay enables the adaptation loops to track rapid variations. This solution, however, becomes cumbersome as SNRs decrease, as it becomes more difficult to produce tentative decisions with acceptable reliability and delay.

Besides the increasing density, there is also a general trend of increasing data rates [211]. The development of two-dimensional (2-D) storage systems fits with this trend and permits exploitation of parallelism. The parallelism is achieved by packaging data in a group of adjacent tracks or rows and by parallel processing of these tracks. The complexity of 2-D bit-detectors increases dramatically with respect to one-dimensional (1-D) detectors, and also their detection delays can increase dramatically [216, 221].

In many practical systems the 2-D detector is split into several smaller units to limit overall complexity [152, 198, 217]. A couple of schematic models of such detectors reported in literature, are shown in Fig. 6.1. In this figure, seven parallel bit-tracks are shown: two “outer” tracks and five “inner” tracks (tracks positioned near the center of the group of adjacent tracks). The final decisions of the different detector models are indicated in thick black arrows. The different units constituting a detector model are numbered in order of execution. If the output of one unit is used as input of a next unit (indicated by the grey arrows), detection delay increases. Different connections between the different units are possible: a) in a “<”-shape [152] (where the two outer tracks have the smallest detection delay), b) sequentially starting from the top track [198, 212], and c) different iterations of the joint detection [216, 221, 222]. In the latter case (different iterations are performed) the decisions of the outer tracks will be more reliable than the ones of the inner tracks during the first iterations because fewer ISI occurs at the outer tracks (due to the 2-D structure, where inner-track bits have more neighboring bits than outer-track bits) [18]. As a result decisions of the outer track during an early iteration can be used in the adaptation loops limiting the detection delay of the outer track. For the inner track however more iterations are required to achieve an acceptable reliability, resulting in an increased detection delay for the inner tracks.

In all these detectors, the detection delay of inner bit-tracks adds substantial latency in the adaptation loops. In the experimental Two-Dimensional Optical Storage (TwoDOS) system, for example, the delay for the inner tracks is around 100-200 symbol intervals, versus a delay of only 10-20 intervals for the outer tracks [152, 223]. As

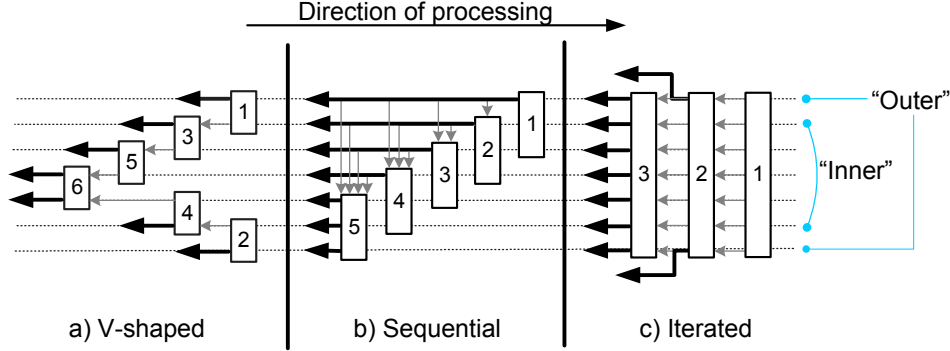


Figure 6.1: Examples of 2-D detectors which have varying detection delays for different tracks.

a result the effective latencies in the adaptation loops for these inner tracks preclude the tracking of rapid variations and really limit the performance of the system.

Benefiting from the fact that in a 2-D system the delays tend to differ per track, in this chapter we propose a scheme that uses control information from tracks for which detector latency is smallest to track rapid variations for all the tracks. The scheme works efficiently if rapid variations are common across the tracks, as is the case, for example, in TwoDOS (see Section 3.4.2). In the proposed scheme these rapid common variations are tracked using control information of the tracks with the minimum latency in the adaptation loop, while the slow track-dependent variations are tracked using the delayed control information of the specific track under consideration. This scheme can be used for timing recovery, automatic gain control and other adaptation loops, and is analyzed and validated experimentally. It shows improved performance with respect to conventional adaptation loops in case substantial loop-delays are present.

The scheme that is proposed in this chapter is general and can be applied to any 2-D storage system. In this chapter a particular example of a 2-D storage system, namely the TwoDOS system, is used to illustrate the design of the adaptation loops and to provide experimental results. In Section 6.2, a general receiver model for 2-D storage systems is discussed and a general parameter-domain model of an adaptation loop is derived from this receiver model. The effect of latencies on the performance of adaptation loops is discussed in Section 6.3. The general scheme for minimum-latency tracking of rapid variations is explained in Section 6.4. In Section 6.5 the design of first-order loops according to the described scheme is explained, analyzed and verified by means of simulations. These first-order loops can be used, for example, for DC compensation and Automatic Gain Control (AGC). Minimum-latency tracking suited for timing recovery is the subject of Section 6.6, where the design of second-order loops is discussed. Finally Section 6.7 presents experimental results for the TwoDOS system. These results show that the new scheme improves the perfor-

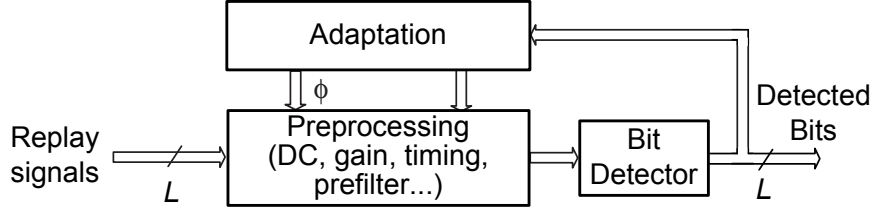


Figure 6.2: Data receiver model for 2-D storages system.

mance of the system with respect to conventional schemes.

6.2 Receiver Model

A data receiver model for 2-D storage systems is shown in Fig. 6.2. Inputs of the model are L digitized replay signals, where L is the number of adjacent bit-tracks. In magnetic storage, for example, these replay signals are generated by read heads. The data receiver contains a bit-detector that relies upon a well-defined relationship between the stored data and the desired detector input signals [4]. This relationship is often characterized by a so-called target response, and is often linear. To approach this relationship as closely as possible, the replay signals are preprocessed by digital signal processing blocks (e.g. timing adjustment, prefilter, DC compensation) before they enter the detector. Because physical parameters of the storage channel (e.g. bandwidth, amplitude, DC offset,...) may vary in time, adaptivity is needed to counteract the parameter variations such that the relationship between the stored data and the detector input signal does not vary in time and is consistent with the target response. To this end the receiver includes a preprocessing block with several adjustable parameters (e.g. an AGC gain, equalizer coefficients,...) that are controlled by dedicated adaptation loops.

For each adjustable parameter a value ϕ is produced by the adaptation block and is subsequently used in the preprocessing blocks to counteract system parameter variations. We denote the ideal value of the adjustable parameter by θ (clearly θ depends on the channel parameters and can hence be time-varying). Ideally ϕ should be equal to θ . A difference Δ between ϕ and θ results in an undesired mismatch between the actual and the desired detector input signal. Accurate tracking of θ (or equivalently, minimizing Δ) will minimize this mismatch and as result will improve receiver performance. An example of an adaptive parameter is the DC-offset for a specific track. The value ϕ is the estimated DC-offset while θ is the ideal DC-offset, i.e. the (possibly time-varying) DC-offset that has to be added to the incoming signal to eliminate any residual DC-offset in the detector input.

The preprocessing block, the bit-detector and the adaptation block form a closed loop which comprises the individual adaptation loops. These loops are of the data-

aided (DA) type. A DA adaptation loop uses the detected bits as side information to facilitate adaptation. As a result the bit-detector forms part of the loop and the detection delays introduce a latency in the loop. This latency will limit the capability of the loop to track fast variations of θ . The effect of latency on the tracking capabilities of an adaptation loop is subject of the next section.

6.3 Effect of latency on loop behavior

In the left part of Fig. 6.3 a discrete-time parameter-domain model of an adaptation loop with latency is given [3]. This model is valid in the tracking mode of operation. An ideal parameter value θ is the first input of the model and this value can be time-varying (the time index k is omitted in the remainder of this chapter for notational simplicity). A noise component v (input-referred noise) is the second input of the loop. The purpose of the loop is to minimize the mismatch Δ between the ideal parameter value θ and its estimate ϕ . To generate this estimate ϕ , a loop filter with transfer characteristic $L(z)$ followed by an ideal integrator is used. In the right part of Fig. 6.3 different types of loop filters are shown. In most cases (e.g. for DC control, automatic gain control, adaptive equalization) a first-order loop is sufficient. In this case the loop filter is just a multiplier: $L(z) = K_t$, where K_t is the total gain of the loop. Timing recovery, however, requires a second-order adaptation loop in order to be able to track frequency variations. To this end the loop filter needs to be extended with an ideal integrator and a second gain K_{ti} , which determines together with K_t the behavior of the loop.

In the model a delay of D symbol intervals is present which mimics the overall latency in the loop. The model has a low-pass frequency characteristic. If θ changes then ϕ will track the slow variations but not the fast ones. In this way also the high-frequency noise components v are rejected.

6.3.1 Loop Behavior

As the dynamic properties of the loop do not depend on the input-referred noise v , we neglect this noise for the time being. The adaptation loop is linear and can be characterized by means of the parameter transfer function $G_\phi(z) = \Phi(z)/\Theta(z)$ and the mismatch transfer function $G_\Delta(z) = \Delta(z)/\Theta(z)$ (the response in the z -domain of respectively ϕ and Δ to the loop excitation θ). Only first-order adaptation loops are considered here. Extension to second-order loops is straightforward. For first-order loops the loop filter has transfer function $L(z) = K_t$, where K_t is the total gain of the loop. The mismatch transfer function $G_\Delta(z)$ is evaluated to be

$$G_\Delta(z) = \frac{z^D(z-1)}{z^D(z-1) + K_t}. \quad (6.1)$$

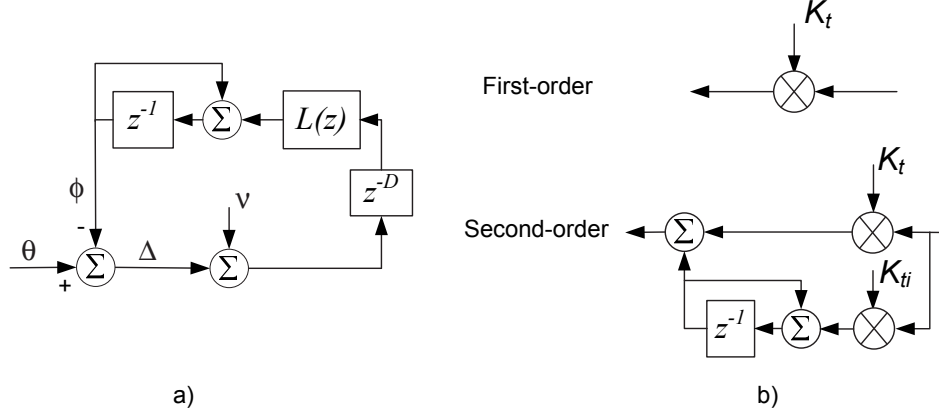


Figure 6.3: a) General parameter-domain model of an adaptation loop with a loop filter characterized by the transfer function $L(z)$. This loop is sampled at the baud rate $1/T$, i.e. z^{-1} corresponds to a delay of T seconds. b) Loop filters of first-order adaptation loop and of second-order high-gain adaptation loop.

Let θ be a unit step function. Then, for $D = 0$, the mismatch can be approximated as an exponentially decaying function $e^{-k/\tau}$, where τ is the time constant of the loop expressed in sampling intervals. For small K_t , the time constant can be expressed as $\tau \approx 1/K_t$.

The mismatch magnitude responses for varying time constants τ and loop-delays D are shown in Fig. 6.4. In the left part of the figure, the response is shown for varying time constants τ and zero delay ($D = 0$). By increasing τ , the equivalent bandwidth of the loop decreases. The equivalent bandwidth is defined by the normalized loop cut-off frequency $\Omega_c = K_t/2\pi$, i.e. the normalized frequency where the amplitude of the transfer function is -3 dB. In the right part of Fig. 6.4, the magnitude response of the mismatch transfer function is shown for a given time constant $\tau = 200$ and for varying loop-delays D . An increasing resonance peak appears near the cut-off frequency Ω_c for increasing loop-delays. If the ideal parameter value θ has spectral content in this frequency region, the total mismatch power will increase due to this resonance peak (in other words the loop enhances rather than suppresses variations around this frequency).

If the delay is increased too much, the loop can become unstable. The edge of the stability region demarcates a relationship between the loop-delay and the time constant of the loop: $\tau \simeq (2D + 1)/\pi$ [224]. This relationship reveals the smallest allowable time constant for a given loop-delay.

The responses shown in Fig. 6.4 indicate that in a practical system the presence of a large loop-delay can influence the choice of a proper time constant considerably. In general, to limit resonance effects or even to avoid instability, a larger time constant

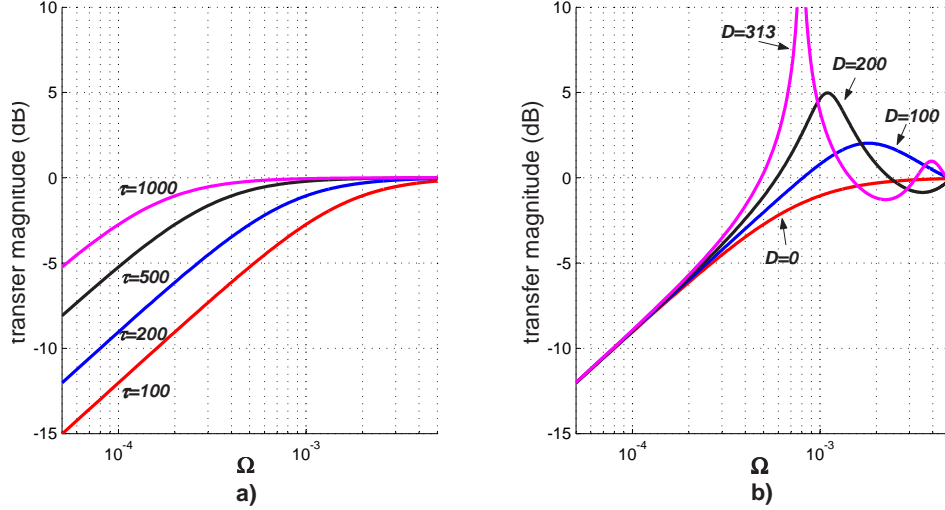


Figure 6.4: Mismatch transfer magnitudes of a first-order adaptation loop as function of the normalized frequency $\Omega = \omega T / (2\pi)$. a) No loop-delay, $D = 0$ (left part). b) Time constant $\tau = 200$ (right part).

with respect to the zero loop-delay case is needed, which will make the loop less capable of tracking fast parameter variations.

6.3.2 Gradient Noise

Gradient noise is defined as the additional mismatch in the adaptation loop due to the input-referred noise v . This gradient noise does not introduce a bias in the estimate ϕ but influences the variance of the mismatch Δ :

$$\sigma_{\Delta}^2 = \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} P_v(e^{j\omega T}) |G_{\phi}(e^{j\omega T})|^2 d\omega, \quad (6.2)$$

where $P_v(e^{j\omega T})$ is the power spectral density of v . If v is assumed to be white and Gaussian, then the mismatch variance can be expressed as $\sigma_{\Delta}^2 = \sigma_v^2 B_l$, where $B_l = \frac{K_t}{2-K_t}$ is the normalized equivalent noise bandwidth (if the loop-delay D is omitted, otherwise the noise bandwidth B_l will be slightly higher due to the resonance effect) and σ_v^2 is the variance of the input-referred noise v . In practice, K_t is much smaller than unity, hence $B_l \simeq K_t/2$. The variance of Δ can then be expressed as

$$\sigma_{\Delta}^2 \simeq \frac{K_t \sigma_v^2}{2}. \quad (6.3)$$

From this equation it is clear that σ_{Δ}^2 is proportional to K_t , which means that the mismatch variance will increase for decreasing time constants.

6.4 Minimum-Latency Adaptation

In a 2-D storage system the detection delay can be especially large for the inner tracks. This large delay results in a large latency in the adaptation loops for these tracks. As described in Section 6.3 this latency makes the loops incapable of tracking rapid parameter variations. In a 2-D system like the TwoDOS system, it is possible to use control information from bit-tracks with smaller latencies to counteract these rapid variations. The principle of using control information with the smallest latency is referred to as the minimum-latency adaptation strategy. As a consequence the aim is to design adaptation loops that make use of the minimum-latency control information to counteract rapid variations in all tracks.

The minimum-latency adaptation strategy is only applicable if rapid variations of system parameters are common for all the tracks. Subject to this basic premise the overall model of the ideal parameter value θ should be: $\theta = \theta_{LF} + \theta_{HF}$, where θ_{LF} is a slowly varying, track-dependent component (with highest frequency Ω_{LF}) and θ_{HF} is a rapidly varying component which is common for all the tracks (highest frequency Ω_{HF} and $\Omega_{HF} > \Omega_{LF}$). The parameters of all tracks must show this behavior: possibly different low-frequency content but the same high-frequency content. The spectral content of the ideal parameter value θ together with the input referred noise v (assumed to be Gaussian and white) is sketched in Fig. 6.5.

The basic premise that rapid variations are common across the tracks can be validated experimentally for the TwoDOS system. By way of illustration we consider the DC adaptation loops [145], which serve to counteract time-varying DC-offsets in the tracks. Here DC-offset estimates ϕ_O for every track are generated by separate adaptation loops where the gain values K_t are chosen such that ϕ_O is able to track fast variations of the ideal DC-offset values θ_O (in the experimental estimates $K_t = 0.06$ was found to be a proper value). Because $G_\phi(z)$ has unit amplitude up to the normalized loop cut-off frequency ($\Omega_c \approx 0.01$), the spectral content of ϕ_O resembles the spectral content of θ_O up to Ω_c (if noise is neglected). The spectral content of the inner-track DC-offset estimate ϕ_O is shown in Fig. 6.6. This estimate ϕ_O is composed of different components which are also shown in the figure. The common DC-offset component ϕ_{Oc} is calculated by averaging the DC-offsets over all tracks. The track-dependent DC-offset component ϕ_{Or} is obtained by subtracting ϕ_{Oc} from ϕ_O . Finally the power spectral density P_v of the input-referred noise v is obtained by taking the Fourier transform of the difference between the ideal and the actual detector input. For low frequencies the track-dependent component ϕ_{Or} is the most important component of ϕ_O . At higher frequencies ($\Omega = [7e^{-4}, 2e^{-3}]$) the common component ϕ_{Oc} determines ϕ_O . For even higher frequencies the offset estimate is determined by the input-referred noise v . The common offset component ϕ_{Oc} can be explained by the fact that certain channel parameters (e.g. the amount of defocus and the cover-layer thickness) are common across the adjacent tracks. For the Two-

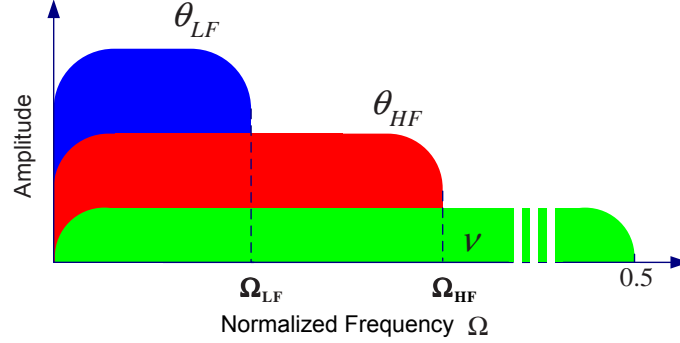


Figure 6.5: Spectral content of different inputs of an adaptation loop. The input parameter value θ is assumed to consist of two components: θ_{LF} and θ_{HF} . Furthermore the spectrum of the noise ν is also shown.

DOS system, the cover-layer thickness exhibits variations that extend over a limited amount of bits (100-1000 bits). As a result these variations result in high-frequency common offset variations. Other reasons for fast common channel parameter variations are: dust, fingerprints, scratches on the disc, dropouts... [173, 174]. These observations lend support to the assumed parameter model, not just for TwoDOS but also for other 2-D storage systems.

A basic assumption of the minimum-latency adaptation strategy is that the common parameter value θ_{HF} is tracked using the minimum-latency control information. In reality, also small and relatively slow variations occur between the tracks. In the minimum-latency adaptation strategy these slow track-dependent components θ_{LF} are handled by using delayed information from the inner tracks.

In Section 6.5 the minimum-latency adaptation strategy for first-order loops is proposed and analyzed (applicable to the AGC loop and the DC adaptation loop). The minimum-latency strategy is applied to second-order loops (as used for timing recovery) in Section 6.6. In Section 6.7, the experimental validation of the minimum-latency adaptation strategy is presented for the TwoDOS system.

6.5 First-Order Minimum-Latency Adaptation Loops

To illustrate the application of the minimum-latency strategy to first-order loops only two tracks (instead of L tracks) are considered for simplicity: an “outer” track with small latency and an “inner” track with large latency. Furthermore every track has a separate adaptation loop. In Fig. 6.7 a parameter-domain model of the minimum-latency first-order adaptation loops is shown together with the assumed parameter

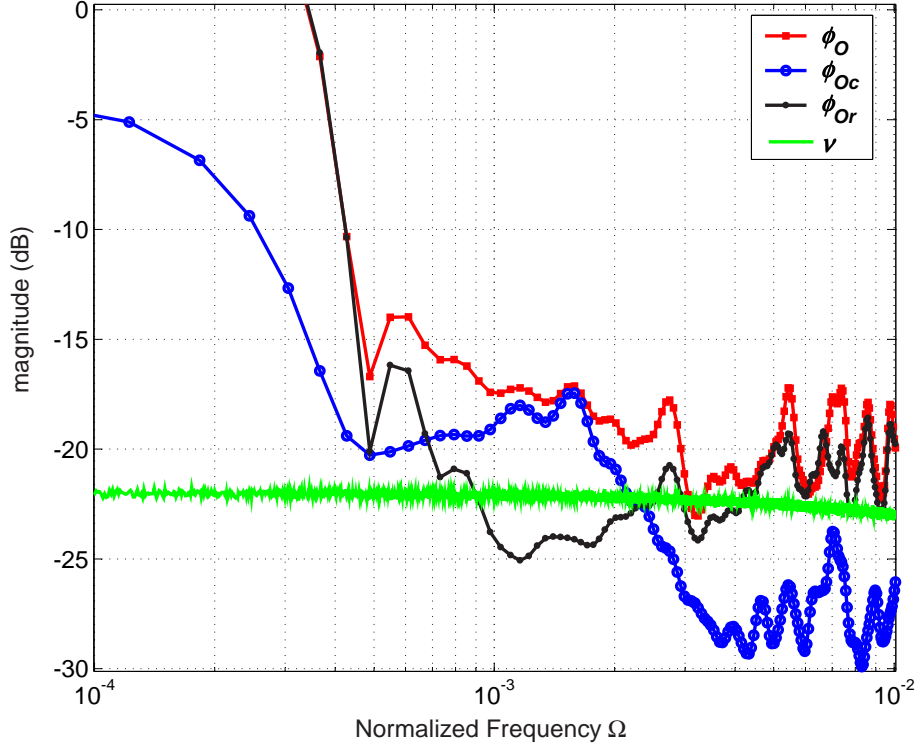


Figure 6.6: Experimentally estimated spectral content of DC-offsets in Two-DOS system: the DC-offset estimate ϕ_O of the inner track and its different components: ϕ_{Oc} and ϕ_{Or} . Also the spectrum P_v of the input noise v is shown.

model¹. The input-referred noises v^i and v^o are inputs of the model and are uncorrelated. Furthermore a delay of D^i bits is present in the inner loop to mimic the large detection delay of the inner track. In reality also a small detection delay ($D^o < D^i$) is present in the outer loop, but this delay is omitted in the model to simplify the analysis. This relatively small delay will not have a major influence on the overall loop behavior. Following the minimum-latency adaptation strategy the outer loop is dimensioned to be fast (large loop gain K_l^o) and the inner loop is dimensioned to be slow (small loop gain K_l^i).

The key innovative feature of the minimum-latency adaptation strategy is the connection between the fast outer loop and the slow inner loop. This connection (thick line in Fig. 6.7) provides the inner loop with control information concerning

¹In the TwoDOS system, there are two outer tracks that can be used to derive the common rapid variations from. The outputs of these outer loops are averaged and the result is used in the inner tracks. This procedure fully exploits all minimum-latency information.

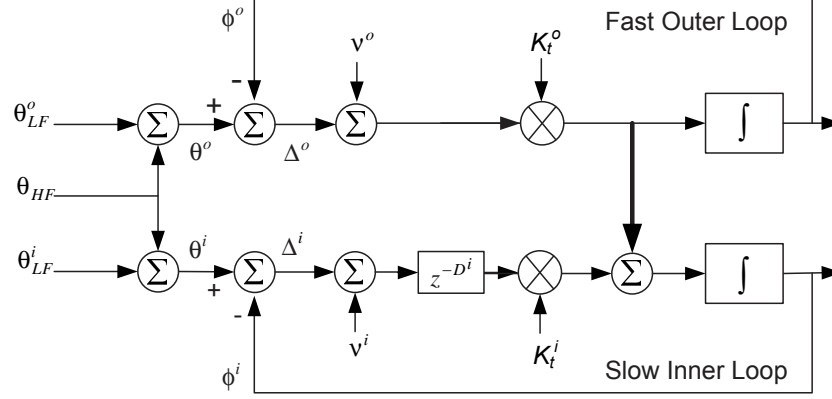


Figure 6.7: Parameter-domain model of the first-order minimum-latency adaptation loops.

the fast common parameter θ_{HF} that is not yet available in the inner loop due to the delay M^i . As a result, ϕ^i is able to track fast variations of θ_{HF} (using control information of the outer track) and slow variations of θ_{LF}^i (using delayed control information of the inner track). However, due to the connection a portion of θ_{LF}^o will inevitably be present in the estimate ϕ^i . In Section 6.5.1, the basic loop behavior is analyzed which will prove that this portion is sufficiently small.

6.5.1 Basic Behavior

Because the inner loop is dimensioned to be slow, the delay D^i of the inner loop will not have strong impact on the loop behavior (see Section 6.3.1). For this reason we initially omit delay in our analysis of the basic behavior, i.e. we set $D^i = 0$. As the dynamic properties of the loops do not depend on the input noises v^i and v^o , these input noises are neglected. By transformation into the z -domain the basic loop behavior can be analyzed. The inner loop will not show first-order behavior anymore but becomes essentially a second-order adaptation loop, whose behavior is determined by the total gains K_t^i and K_t^o . The z -transform of the inner-track estimate ϕ^i can easily be derived to be:

$$\begin{aligned} \Phi^i(z) = & \frac{K_t^i}{z-1+K_t^i} \Theta_{LF}^i(z) \\ & + \left(\frac{K_t^i}{z-1+K_t^i} - \frac{K_t^o(z-1)}{(z-1+K_t^o)(z-1+K_t^i)} \right) \Theta_{HF}(z) \\ & - \frac{K_t^o(z-1)}{(z-1+K_t^o)(z-1+K_t^i)} \Theta_{LF}^o(z), \end{aligned} \quad (6.4)$$

where $\Phi^i(z)$, $\Theta_{LF}^i(z)$, $\Theta_{LF}^o(z)$ and $\Theta_{HF}(z)$ are the z -transform of respectively ϕ^i , θ_{LF}^i , θ_{LF}^o and θ_{HF} .

From (6.4) it is clear that the estimate ϕ^i of the ideal parameter value θ^i is determined by all three ideal parameter components: θ_{LF}^i , θ_{HF} and θ_{LF}^o . In Fig. 6.8 the

mismatch magnitudes due to these components are plotted for two different inner-track time constants τ^i while fixing the outer-loop time constant τ^o at 100 bits. In the left part of the figure, the magnitudes are shown for the case the inner-track adaptation loop is 5 times slower than the outer-track adaptation loop, i.e. $\tau^i = 5\tau^o$. In the right part of the figure $\tau^i = 50\tau^o$. The mismatch magnitudes $|\Delta^i(z)| = |\Theta^i(z) - \Phi^i(z)|$ due to each component are discussed:

- θ_{LF}^i : all spectral content of θ^i (and as a result also of θ_{LF}^i) up to the inner-loop cut-off frequency ($\Omega_c^i = K_t^i/2\pi$) is present in the estimate ϕ^i and is not present in the mismatch Δ^i . As a consequence the inner loop should be designed such that $\Omega_c^i \geq \Omega_{LF}$.
- θ_{HF} : this is the high-frequency component of θ^i and is present in the estimate ϕ^i up to the cut-off frequency of the outer loop ($\Omega_c^o = K_t^o/2\pi$). The key objective of the minimum-latency adaptation strategy is hereby achieved: θ_{HF} is present in the estimate ϕ^i up to the outer-loop cut-off frequency Ω_c^o despite the fact that the inner adaptation loop has a cut-off frequency $\Omega_c^i < \Omega_c^o$. The time constant τ^o should be chosen such that: $\Omega_c^o \geq \Omega_{HF}$.
- θ_{LF}^o : the outer-loop low-frequency component θ_{LF}^o should be rejected as much as possible. This is accomplished if the inner loop is designed properly: $\Omega_c^i > \Omega_{LF}$.

As a conclusion the following criteria for the proper choice of the loop time constants are formulated:

$$\begin{aligned} \text{Outer loop:} \quad & \tau^o \leq \frac{1}{2\pi\Omega_{HF}}; \\ \text{Inner loop:} \quad & \tau^i \leq \frac{1}{2\pi\Omega_{LF}}. \end{aligned} \quad (6.5)$$

6.5.2 Gradient Noise

The outer-loop gradient noise is not influenced by the inner-loop noise v^i and has the same variance as expressed in (6.2). But as expected, the total amount of gradient noise in the inner loop has an extra component due to the minimum-latency adaptation strategy. The variance of the inner-loop mismatch Δ^i is evaluated to be

$$\begin{aligned} \sigma_{\Delta^i}^2 = & \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} P_{v^i}(e^{j\omega T}) |G_{\phi^i}(e^{j\omega T})|^2 d\omega \\ & + \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} P_{v^o}(e^{j\omega T}) |G_{\phi^o}(e^{j\omega T})|^2 d\omega, \end{aligned} \quad (6.6)$$

where $P_{v^i}(e^{j\omega T})$ and $P_{v^o}(e^{j\omega T})$ are the power spectral densities of respectively v^i and v^o . The input-referred noise v^o of the outer loop leaks into the inner loop according to the open-loop transfer function $G_{\phi^o}(z) = \Phi^o(z)/\Theta^o(z)$ which has a bandwidth $B_l = K_t^o/2$ (under the assumption $K_t^i < K_t^o \ll 1$ and $D^i = 0$). In case the input noises v^i and

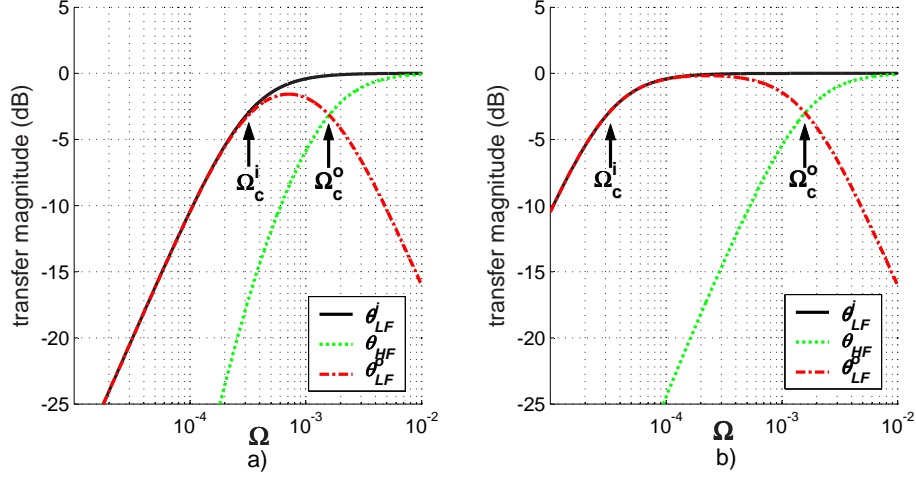


Figure 6.8: Mismatch transfer magnitudes of the inner minimum-latency adaptation loop as a function of the normalized frequency $\Omega = \omega T / (2\pi)$. The outer loop has a time constant $\tau^o = 100$ bits. The inner loop has the following time constants: a) $\tau^i = 5\tau^o$ (left figure). b) $\tau^i = 50\tau^o$ (right figure).

v^o are assumed to be white and Gaussian, the variance of the inner-loop mismatch Δ^i can be approximated as:

$$\sigma_{\Delta^i}^2 \simeq \frac{K_t^i \sigma_{v^i}^2}{2} + \frac{K_t^o \sigma_{v^o}^2}{2}, \quad (6.7)$$

where $\sigma_{v^i}^2$ and $\sigma_{v^o}^2$ are the variances of the input noises v^i and v^o respectively. From this equation it is clear that the gradient noise in the inner loop is proportional to K_t^i and K_t^o , and dominated by the outer-loop noise v^o (because $K_t^o > K_t^i$). Therefore the variances of the inner- and outer-loop mismatches are approximately equal.

6.5.3 Behavior of the Inner Loop with Latency

The insertion of a delay of D^i bit intervals changes the behavior of the inner adaptation loop. The z -transform of the inner-loop mismatch Δ^i is evaluated to be:

$$\begin{aligned} \Delta^i(z) &= \frac{z-1}{z-1+z^{-D^i}K_t^i} \Theta_{LF}^i(z) \\ &+ \left(\frac{z-1}{z-1+z^{-D^i}K_t^i} - \frac{K_t^o(z-1)}{(z-1+K_t^o)(z-1+z^{-D^i}K_t^i)} \right) \Theta_{HF}^i(z) \\ &- \frac{K_t^o(z-1)}{(z-1+K_t^o)(z-1+z^{-D^i}K_t^i)} \Theta_{LF}^o(z). \end{aligned} \quad (6.8)$$

The corresponding transfer magnitudes are shown in Fig. 6.9 for the following conditions: $D^i = 200$ bits and $\tau^o = 100$ bits. The inner adaptation loop is dimensioned to be slow: $\tau^i = 5\tau^o$ in the left part of the figure and $\tau^i = 50\tau^o$ in the right part of

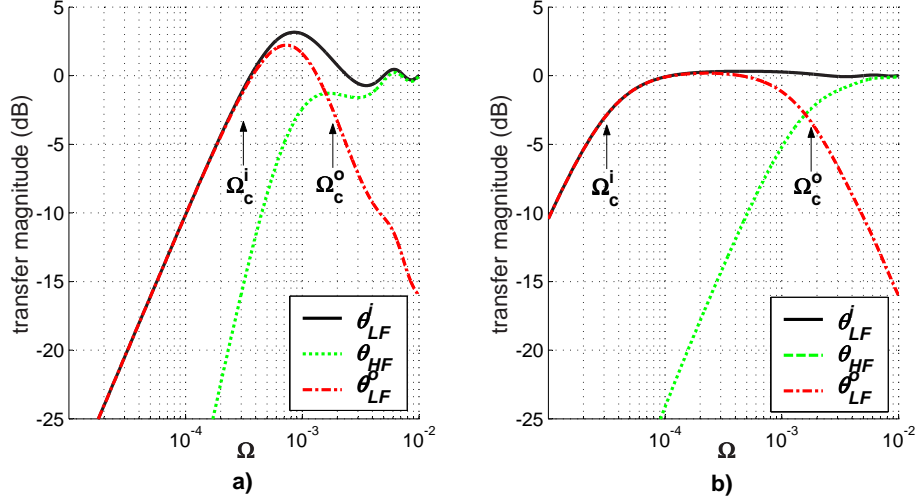


Figure 6.9: Mismatch transfer magnitudes of the inner minimum-latency adaptation loop (first-order) as a function of the normalized frequency $\Omega = \omega T / (2\pi)$. The inner loop has a delay of $D^i = 200$ bits. The outer-loop time constant τ^o is equal to 100 bits. The inner loop has the following time constants: a) $\tau^i = 5\tau^o$ (left figure). b) $\tau^i = 50\tau^o$ (right figure).

the figure. The effect of the insertion of a delay in the inner loop can be evaluated by comparing Fig. 6.9 with Fig. 6.8. Especially the left plot is substantially changed with respect to the situation without delay: a resonance effect clearly appears in the transfer magnitudes at frequencies in the vicinity of the cut-off frequency Ω_c^i of the slow inner loop. Consequently, if the second design criterium ($\Omega_c^i \geq \Omega_{LF}^i$) is obeyed, the correct estimation of the slow parameter variations θ_{LF}^i is guaranteed and also the outer-track parameter θ_{LF}^o is rejected sufficiently in the inner-track estimate ϕ^i . The estimation of the high-frequency component θ_{HF} will deteriorate slightly as the mismatch magnitude is slightly increased near the cut-off frequency Ω_c^o of the outer-track adaptation loop. Consequently if also the first design criterium ($\Omega_c^o \geq \Omega_{HF}$) is obeyed, the tracking capabilities of the minimum-latency adaptation loops with and without a loop-delay will be comparable. The transfer magnitude of a slow inner loop (right part of the figure) is not (or just slightly) influenced by the insertion of the delay. This behavior is expected based on the analysis of Section 6.3.1.

6.5.4 Simulation Results

Simulation results were obtained by simulating the model presented in Fig. 6.10. Replay signals \mathbf{r}_k are produced by passing the data bits \mathbf{a}_k (where $\mathbf{a}_k = [a_k^0 \dots a_k^{L-1}]$) to a simple channel model that has three basic parts. First ideal detector input sig-

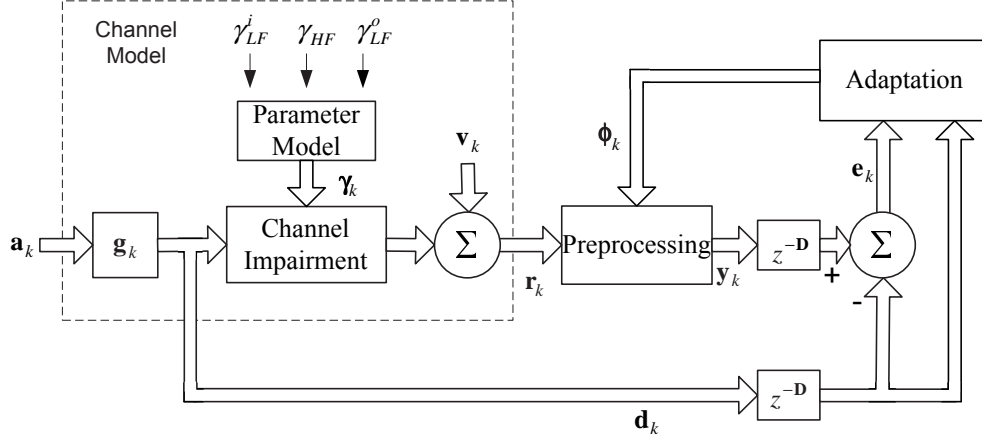


Figure 6.10: *Simulation Model.*

nals \mathbf{d}_k are generated by convolving \mathbf{a}_k with fixed 2-D target responses \mathbf{g}_k (where $\mathbf{g}_k = [\mathbf{g}_k^0 \dots \mathbf{g}_k^{L-1}]$, i.e. each track has its own target response). The second part of the channel model is a channel impairment block which corrupts \mathbf{d}_k according to time-varying system parameter values $\boldsymbol{\gamma}_k = [\gamma_k^0 \dots \gamma_k^{L-1}]$. In this section DC-offsets are considered as channel impairments, i.e. $\boldsymbol{\gamma}_k$ are the DC-offsets induced by the channel. Only one outer-track value ($\gamma_k^0 = \gamma^o$) and one inner-track value ($\gamma_k^{(L-1)/2} = \gamma^i$) are non-zero and they are both generated according to the parameter model of Fig. 6.7: $\gamma^o = \gamma_{LF}^o + \gamma_{HF}$ and $\gamma^i = \gamma_{LF}^i + \gamma_{HF}$, where γ_{LF}^o and γ_{LF}^i are slowly varying components and γ_{HF} is a rapidly varying component. To simplify the simulation only sinusoidally time-varying components are considered, i.e. the DC-offsets induced by the channel vary slowly as a sinusoidal function of time. Moreover only one component has an amplitude equal to one while the other components all have zero amplitude, i.e. the responses to individual parameter components are simulated and not to a combination of different parameter components. Finally, white Gaussian noise components \mathbf{v}_k are added to the channel impairment output signals to produce the replay signals \mathbf{r}_k . The noise components of the different tracks are uncorrelated and have the same variance.

These replay signals \mathbf{r}_k are input of a preprocessing block that tries to undo the corruption induced by the channel impairment block. For this reason this preprocessing block utilizes estimates $\boldsymbol{\phi}_k$ of the ideal values $\boldsymbol{\theta}_k$ (for DC-offsets $\boldsymbol{\theta}_k = -\boldsymbol{\gamma}_k$). To mimic detection delays, the outputs \mathbf{y}_k of the preprocessing block are delayed with delays $\mathbf{D} = [D^0 \dots D^{L-1}]$. To simplify the simulation, the delay of the outer tracks is chosen to be 0 ($D^0 = D^o = 0$) and the delay of the inner tracks is set to D^i bits. Error signals \mathbf{e}_k are calculated by subtracting delayed versions (with delays \mathbf{D}) of \mathbf{y}_k from these delayed versions of \mathbf{y}_k . Finally an adaptation block uses \mathbf{e}_k and delayed versions of \mathbf{y}_k to produce estimates $\boldsymbol{\phi}$ which are used by the preprocessing block. The

adaptation block utilizes the ZF technique [3] to produce the estimates. The variances of the input-referred noises are similar ($\sigma_{v^i}^2 \approx \sigma_{v^o}^2$) provided that the variances of the noise components are similar ($\sigma_{u^i}^2 \approx \sigma_{u^o}^2$), which is the case in our simulations.

The SNR of the system is defined as:

$$\text{SNR} = \frac{\sum g_k^2}{\sigma_{v^i}^2} = \frac{\sum g_k^2}{\sigma_{v^o}^2}. \quad (6.9)$$

The Mean Square Error (MSE) of the inner track is defined as:

$$\text{MSE}^i = E[e_k^i]^2, \quad (6.10)$$

where e_k^i is the error of the inner track and the expectation E is taken over all time instants k . Because MSE quantifies the mismatch between the actual and the desired detector input, it is closely related to bit-error rate. As a result it is a good measure to quantify the performance of the adaptation loops.

The MSE of the inner track is plotted in Fig. 6.11 as a function of the normalized frequency Ω of the different ideal DC-offset components θ_{LF}^i , θ_{HF} and θ_{LF}^o (only one parameter is non-zero at a time). Because the DC-offset estimates ϕ_k are added to \mathbf{r}_k in the preprocessing block, any DC-offset mismatch is part of the error \mathbf{e}_k . As a result the MSE plotted in Fig. 6.11 is directly related to the mismatch transfer magnitudes shown in Fig. 6.8. The time constant τ^o of the outer loop is 100 bits ($\tau^o = 100$ bits), while the time constant τ^i of the inner adaptation loop is 500 bits for the left plot ($\tau^i = 500$ bits) and 5000 bits for the right plot ($\tau^i = 5000$ bits). These simulation results exactly match the theoretical results derived in the previous section.

A validation of the minimum-latency adaptation strategy is shown in Fig. 6.12: a comparison between delayed individual adaptation loops and delayed minimum-latency adaptation loops. The gain in inner-track Minimum MSE (MMSE) is shown as function of the loop-delay D^i for two different SNR values. Here MMSE is defined as the minimum MSE obtained by tuning the time constants τ^i and τ^o . In this simulation sinusoidal DC-offset variations (with unit amplitude) are considered with the following frequencies: $\Omega_{O,LF}^i = \Omega_{O,LF}^o = 10^{-5}$ and $\Omega_{O,HF} = 10^{-4}$. The conditions are representative for severe but not extreme situations (scratches, dropouts, fingerprints...) in an experimental system [173]. Due to the resonance effect the insertion of a delay into the inner loop causes a substantial degradation in MMSE for the individual adaptation loops (which is in agreement with the analysis of Section 6.3.1).

The minimum-latency adaptation strategy however guarantees the tracking of rapid common offset variations $\theta_{O,HF}$ and as a consequence the MMSE degrades less than for individual adaptation if loop-delays are large. In general for small loop-delays the minimum-latency adaptation loops perform worse because outer track noise v^o leaks into the inner-loop mismatch Δ_O^i and at the same time the tracking capabilities of the individual and the minimum-latency loops are comparable. For

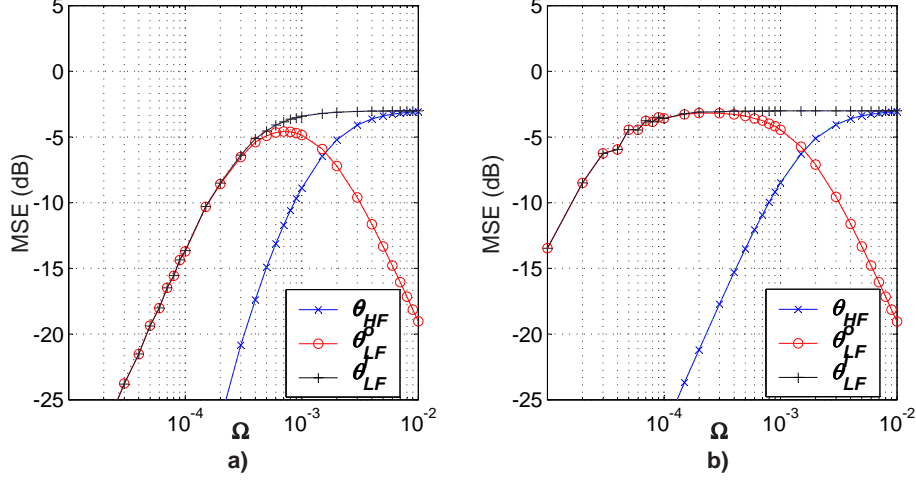


Figure 6.11: MSE of the inner track vs. the normalized frequency Ω of the different DC-offset components. The following settings are used: $D^i = 0$, $\tau^o = 100$ bits, $\tau^i = 500$ (left plot a)) and $\tau^i = 5000$ bits (right plot b)).

larger loop-delays, the minimum-latency adaptation loops outperform the individual adaptation loops because of the strongly improved tracking of rapid common variations θ_{HF} . For low SNR values (in the figure for SNR = 10 dB), the choice of the optimal time constants is mainly dominated by the avoidance of gradient noise and not by the minimization of the mismatch error due to fast parameter variations. For this reason the MMSE for the individual and the minimum-latency adaptation loops will not differ much. For delays larger than 100 bits the minimum-latency adaptation loops outperform the individual loops, even though the MMSE gain is small (0.2 dB for $D^i = 200$ bits). For high SNR values (SNR = 20 dB), the gain in MMSE by using the minimum-latency adaptation loops increases. The time constants can be primarily tuned to improve the tracking capabilities and not to limit the amount of gradient noise in the loop. For a large loop-delay $D^i = 200$ bits the gain in MMSE accumulates up to 1.2 dB. MMSE gains will become even larger if more severe (but still realistic) variations are assumed.

The practical value of these first-order minimum-latency adaptation loops is illustrated in Section 6.7 by means of experimental TwoDOS results. In the next section the minimum-latency adaptation strategy is applied to timing recovery loops.

6.6 Minimum-Latency Timing Recovery

For timing recovery the parameter value θ represents the ideal sampling phase. A second-order loop is generally needed to be able to compensate for frequency errors.

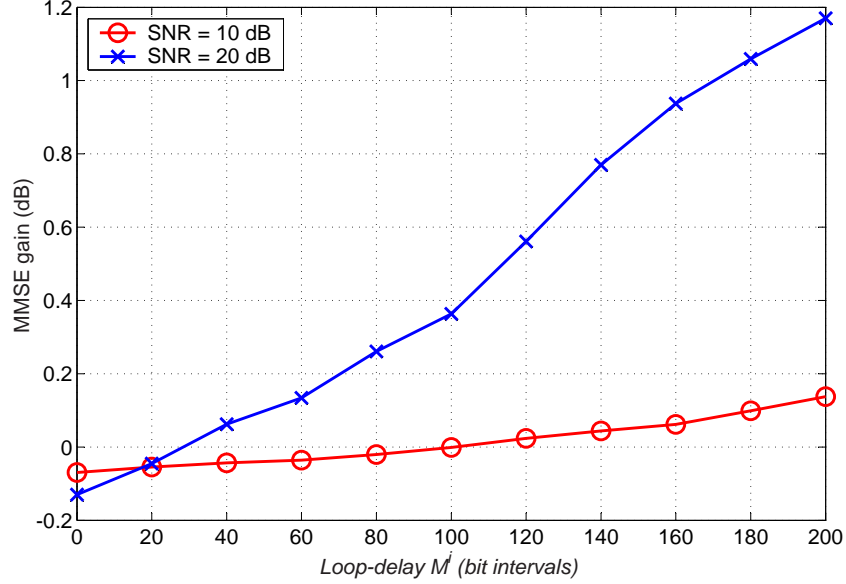


Figure 6.12: Gain in MMSE of the minimum-latency adaptation loops with respect to the individual adaptation loops. The gain in MMSE of the inner track is plotted as a function of the inner-track loop-delay D^l for two different SNR values: 10 dB and 20 dB.

In this case the loop filter $L(z)$ (see Fig. 6.3) has an integrating path (with gain K_{ti}) to track frequency and a proportional path (with gain K_t) to add a phase correction. The normalized natural frequency of a second-order adaptation loop may be defined as $\omega_n T = \sqrt{K_{ti}}$ and the damping factor as $\zeta = K_t / (2\sqrt{K_{ti}})$ [3]. Besides a loop filter the timing recovery loop contains a Numerically Controlled Oscillator (NCO) whose phase-domain model is an integrator, and a phase error detector which can be modelled in the phase-domain by a subtraction of the estimated phase from the ideal phase.

Ideal sampling phases θ in an experimental 2-D system may be common for all tracks (rotation speed variations) but can also be track-dependent [145]. The ideal phase values θ of every track can be written as:

$$\theta = \theta_c + \theta_{LF} = f_p \cdot k + \theta_{HF} + \theta_{LF}. \quad (6.11)$$

where θ_c is a component common for all the tracks and θ_{LF} is a slowly varying ideal phase component specific for the track under consideration. The common component θ_c has again two terms. (a) a frequency deviation term $f_p \cdot k$, where k is the time index and f_p characterizes a small offset between the free-running frequency of the NCO and the frequency of the incoming signal; and (b) a common high-frequency phase component θ_{HF} .

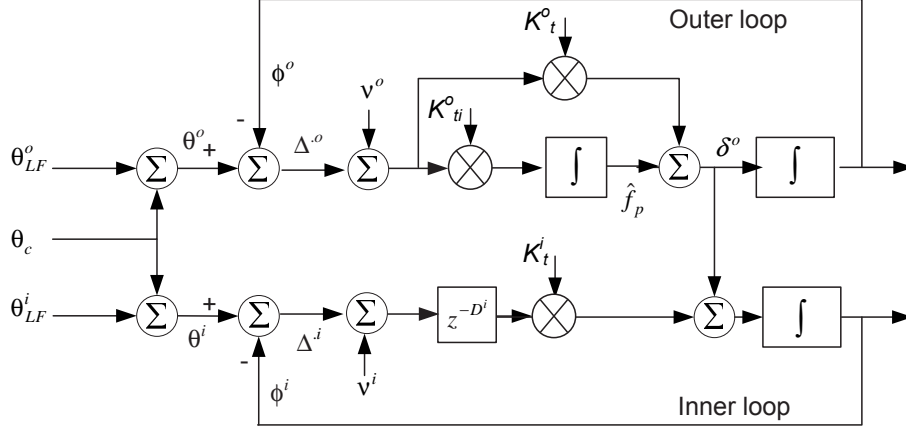


Figure 6.13: Parameter-domain model of the minimum-latency timing recovery loops.

Based on this phase-domain parameter model, a minimum-latency strategy can be developed for the timing recovery loop. In Fig. 6.13 the assumed parameter model together with a parameter-domain model of the minimum-latency timing recovery loops is shown. For simplicity only two tracks are considered: an inner track with small latency and an outer track with large latency. The outer track still has a second-order adaptation loop. A first-order loop filter produces a phase update δ^o which is used as input of the integrator of the outer loop. This phase update δ^o is composed of two components: an estimate of the frequency \hat{f}_p coming from the integrating path (with gain K_{ti}^o) and a phase correction coming from the proportional path (with loop gain K_t^o). Following the minimum-latency adaptation strategy this phase update δ^o is also used in the inner adaptation loop. In our case the frequency errors are common across the tracks and can be tracked based on the outer tracks only. The inner loop is only needed to add a phase correction (accounting for information about θ_{LF}^i) and can therefore remain a first-order loop. Therefore a zeroth-order loop filter is sufficient (with loop gain K_t^i). Furthermore a delay of M^i bit intervals is present in the inner loop to mimic the latency introduced by the detector. Input-referred noises v^i and v^o are also shown in Fig. 6.13 and they are assumed to be uncorrelated.

6.6.1 Basic Behavior

To understand the basic behavior of the minimum-latency timing recovery loops, the delay D^i of the inner track is omitted ($D^i = 0$) and the input-referred noises v^i and v^o are neglected. The behavior of the outer loop is not influenced by the minimum-latency adaptation strategy. As a result the outer loop has a second-order behavior.

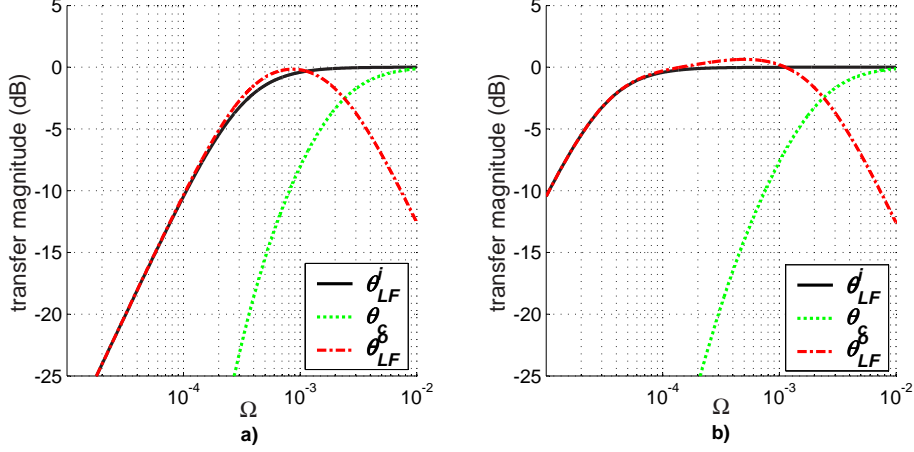


Figure 6.14: Mismatch transfer magnitudes of minimum-latency adaptation loops for timing recovery as a function of the normalized frequency $\Omega = \omega T / (2\pi)$. The outer loop is dimensioned with natural frequency $\omega_n^o T = 0.005$ and damping factor $\zeta^o = 1.5$. The inner loop is dimensioned with the following time constants: a) $\tau^i = 500$ bits (left figure). b) $\tau^i = 5000$ bits (right figure).

The parameter transfer function of the outer loop is evaluated to be:

$$G_{\Phi^o}(z) = \frac{\Phi^o(z)}{\Theta^o(z)} = \frac{K_t^o(z-1) + K_{ti}^o}{(z-1)^2}. \quad (6.12)$$

The parameter transfer function of the inner track is evaluated to be:

$$G_{\Phi^i}(z) = \frac{\Phi^i(z)}{\Theta^i(z)} = \frac{K_t^i}{z-1}. \quad (6.13)$$

The z -transform of the inner-loop mismatch Δ^i is evaluated to be (the inner-track latency D^i is taken into account):

$$\begin{aligned} \Delta^i(z) = & \frac{z-1}{z-1+z^{-D^i}K_t^i} \Theta_{LF}^i(z) \\ & + \left(\frac{z-1}{z-1+z^{-D^i}K_t^i} \right) \Theta_c(z) \\ & - \left(\frac{(K_t^o(z-1) + K_{ti}^o)(z-1)}{((z-1)^2 + (z-1)K_t^o + K_{ti}^o)(z-1+z^{-D^i}K_t^i)} \right) \Theta_c(z) \\ & + \frac{(K_t^o(z-1) + K_{ti}^o)(z-1)}{((z-1)^2 + (z-1)K_t^o + K_{ti}^o)(z-1+z^{-D^i}K_t^i)} \Theta_{LF}^o(z). \end{aligned} \quad (6.14)$$

The mismatch transfer magnitudes of the inner loop are shown in fig. 6.14. These magnitudes are similar to the magnitudes for first-order minimum-latency adaptation loops (see Fig. 6.8). The most important difference is the improved capability of tracking the common component θ_c . For decreasing frequencies the mismatch

magnitude due to the common component θ_c decreases faster with respect to the first-order loops. Moreover tracking of θ_c is guaranteed up to the outer-track cut-off frequency Ω_c^o , while sufficiently suppressing θ_{LF}^o .

The mismatch Δ^i can be shown to go to zero in case of a frequency step (phase ramp) by observing that:

$$\lim_{k \rightarrow \infty} \Delta^i[k] = \lim_{z \rightarrow 1} \Delta^i(z) = 0 \quad (6.15)$$

when for $\Delta^i(z)$ we fill in Eq. 6.14, for $\Theta_c(z)$ we apply a phase ramp ($\Theta_c(z) = 1/(z - 1)^2$) and for the other inputs we assume zero input ($\Theta_{LF}^o(z) = 0$ and $\Theta_{LF}^i(z) = 0$). As a result, no extra measurements have to be taken to guarantee proper convergence of the system.

6.6.2 Gradient noise

The outer-loop gradient noise is not influenced by the inner-loop input-referred noise \mathbf{v}^i . As a result the variance of the mismatch Δ^o is the same as the variance in a normal second-order loop, and can be expressed as:

$$\sigma_{\Delta^o}^2 = \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} P_{\mathbf{v}^o}(e^{j\omega T}) |G_{\phi^o}(e^{j\omega T})|^2 d\omega. \quad (6.16)$$

The open-loop transfer function G_{ϕ^o} has an equivalent bandwidth $B_l^o \geq \tilde{B}_l^o = \omega_n^o T (\zeta^o + 1/4\zeta^o)$ (if $\omega_n^o T$ and ζ^o are chosen within the stability range of the second-order loop and the loop-delay M^i is omitted) [3]. Using this normalized equivalent noise bandwidth \tilde{B}_l^o , the variance $\sigma_{\Delta^o}^2$ can be written as

$$\sigma_{\Delta^o}^2 = \tilde{B}_l^o \sigma_{\mathbf{v}^o}^2, \quad (6.17)$$

if the input-referred noise \mathbf{v}^o is assumed to be white and Gaussian. The gradient noise of the inner loop, however, has an extra component due to the minimum-latency adaptation strategy. The variance of the mismatch Δ^i is evaluated to be

$$\begin{aligned} \sigma_{\Delta^i}^2 &= \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} P_{\mathbf{v}^i}(e^{j\omega T}) |G_{\phi^i}(e^{j\omega T})|^2 d\omega \\ &\quad + \frac{T}{2\pi} \int_{-\frac{\pi}{T}}^{\frac{\pi}{T}} P_{\mathbf{v}^o}(e^{j\omega T}) |G_{\phi^o}(e^{j\omega T})|^2 d\omega. \end{aligned} \quad (6.18)$$

The input-referred noise of the outer loop leaks into the inner loop according to the open-loop transfer function G_{ϕ^o} (with equivalent bandwidth $B_l^o \geq \tilde{B}_l^o = \omega_n^o T (\zeta^o + 1/4\zeta^o)$). Using this normalized equivalent noise bandwidth \tilde{B}_l^o , the variance $\sigma_{\Delta^i}^2$ can be written as

$$\sigma_{\Delta^i}^2 \simeq \frac{K_l^i}{2} \sigma_{\mathbf{v}^i}^2 + \tilde{B}_l^o \sigma_{\mathbf{v}^o}^2 \simeq \frac{K_l^i}{2} \sigma_{\mathbf{v}^i}^2 + \sigma_{\Delta^o}^2, \quad (6.19)$$

if the input noises \mathbf{v}^i and \mathbf{v}^o are assumed to be white and Gaussian. The latter equation indicates that the design of the outer loop (i.e. the choice of values for the natural

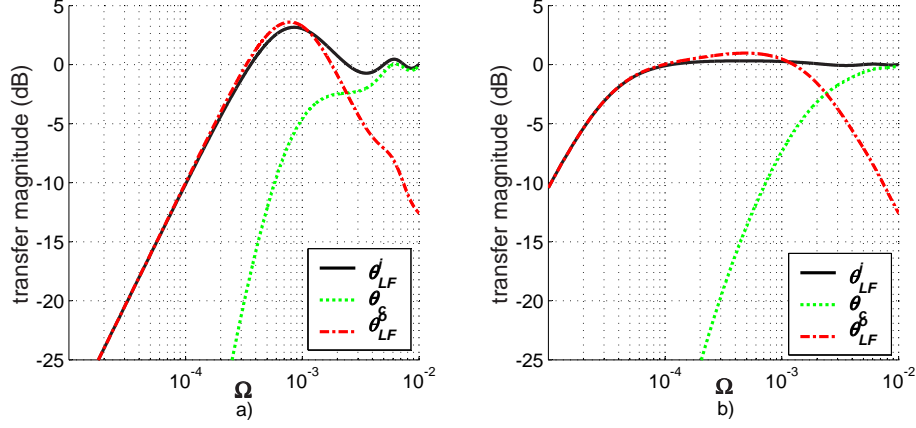


Figure 6.15: Mismatch transfer magnitudes of the inner minimum-latency timing recovery loop as a function of the normalized frequency $\Omega = \omega T / (2\pi)$. The inner loop has a loop delay of $D^i = 200$ bits. The outer loop is dimensioned with a natural frequency $\omega_n^o T = 0.005$ and a damping factor $\zeta^o = 1.5$. The inner loop is dimensioned with the following time constants: a) $\tau^i = 500$ bits (left figure). b) $\tau^i = 5000$ bits (right figure).

frequency ω_n^o and the damping factor ζ^o) considerably influences the amount of gradient noise in the inner loop. But $\sigma_{\Delta^i}^2 \approx \sigma_{\Delta^o}^2$ because $K_t^i \ll \tilde{B}_t^o$, i.e. a negligible extra amount of gradient noise is present in the inner timing recovery loop.

6.6.3 Behavior of Inner Loop with Latency

Inserting a delay in the inner track of D^i bit intervals, causes a degradation in performance of the minimum-latency adaptation loops. The parameter transfer function G_{Φ^i} of the inner track is changed to:

$$G_{\Phi^i} = \frac{\Phi^i(z)}{\Theta^i(z)} = \frac{K_t^i z^{D^i}}{z - 1} \quad (6.20)$$

Consequently also the mismatch of the inner track is changed. In Fig. 6.15 the mismatch magnitudes due to the different components are shown. A similar reasoning as in Section 6.5.3 can be given with the conclusion that the outer loop should be designed slightly faster in order to have MMSE performance.

A resonance peak appears in the transfer magnitudes of the inner loop due to the insertion of the delay D^i . This resonance peak appears near the cut-off frequency Ω_c^i . Consequently, if the second design criterium ($\Omega_c^i > \Omega_{LF}^i$) is obeyed, then the overall mismatch power due to the slowly varying component θ_{LF}^i is guaranteed. Furthermore the outer-track component θ_{LF}^o is rejected sufficiently in the inner-track

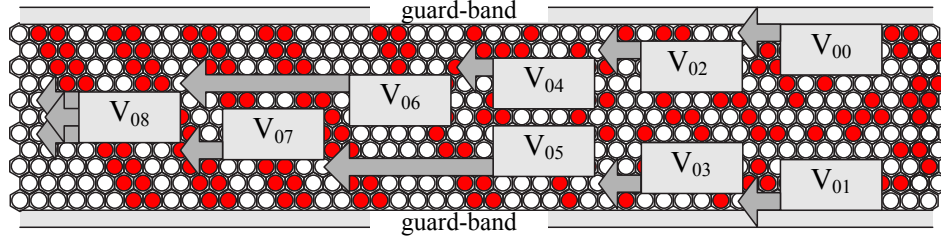


Figure 6.16: Example of hexagonal structure for TwoDOS with $L = 11$. The configuration of stripe-wise Viterbi detection is also shown for this example: in total 9 stripe detectors ($V_{00} \dots V_{08}$) each covering 3 bit-tracks are needed to detect the bits of the broad spiral.

estimate ϕ^i .

6.7 Experimental Results for the TwoDOS system

In the TwoDOS system, bits are stored on a hexagonal lattice [1]. In contrast with conventional optical storage (Compact Disc, CD, Digital Versatile Disc, DVD, and Blu-Ray Disc, BD), where the bits are stored in a single spiral (a 1-D sequence of bits), in TwoDOS the bits are organized in a so-called broad spiral. Within a single rotation of this broad spiral a number L of bit-tracks are stacked upon each other to form a hexagonal structure, see Fig. 6.16. Adjacent rotations of the broad spiral are separated by a guard band consisting of a bit-track without any pits. The data is read out with an array of L laser spots arranged such that each spot is centered on one of the bit-tracks within the broad spiral. A multi-spot photo detector integrated circuit is used to generate a so-called high-frequency (HF) signal for every bit-track.

A Partial-Response Maximum-Likelihood (PRML) receiver has been built for TwoDOS [145, 147]. It consists of a bit-detector preceded by an adaptive equalizer, an adaptive DC compensator, an AGC and a timing recovery loop. A two-dimensional (2-D) Viterbi Detector (VD) performs joint bit-detection on all bit-tracks.

To reduce the complexity of a full-fledged 2-D VD, the VD is divided into smaller processing units (called stripe VD). Each stripe VD covers a limited number of bit-tracks (so-called stripes with a typical height of 2 or 3 bit-tracks). This detection configuration is called a Stripe-Wise Viterbi Detector (SWVD) [152] and is shown in Fig. 6.16 together with the hexagonal structure. The stripe VDs (V_{00} up to V_{08} in the figure) are organized in a “<”-shape. The binary output from a first stripe VD is passed to a next stripe VD to be used as side information in the branch metric calculations [152]. As a result, each next stripe VD adds a delay, which is at least equal to the backtracking depth of the stripe VD [145]. As a result going inwards starting

from the outer bit-tracks the total detection-delay increases considerably. Because all adaptation loops (DC control, AGC, timing recovery,...) use the output of the SWVD as side information, the total latency in the loops for the inner tracks is large and will limit the tracking capabilities severely.

Laser beam recorded discs with a capacity of 35 GB are placed in an experimental read-out system to produce experimental replay signals. The read-out is conducted under relatively favorable conditions (no scratches, no dropouts, limited amount of dust). Subsequently the replay signals are digitized and are applied to the TwoDOS receiver in which the minimum-latency adaptation strategy is utilized for DC control and AGC. In this experiment perfect detection is assumed, i.e. the knowledge of the bits written on the disc is utilized in the receiver. Furthermore the latency M^i induced by the bit-detector can be dimensioned freely to monitor the influence of different detection delays on the overall performance of the system.

The effect of delays on the performance of the first-order DC adaptation loop and the AGC loop is illustrated in Fig. 6.17. The MSE of the inner track after convergence of all adaptation loops, is plotted versus the time constant τ^o . The time constants τ^o of the DC control and the AGC loops are taken equal for three different configurations of the adaptation loops: (1) individual adaptation, $D^i = 0$, $\tau^i = \tau^o$; (2) individual adaptation, $D^i = 150$, $\tau^i = \tau^o$; (3) minimum-latency adaptation, $D^i = 150$, $\tau^i = 5\tau^o$. In this experiment timing recovery is accomplished by individual loops without latency. The natural frequency and damping factor of the second-order loops are defined as: $\omega_n T = 0.01$ and $\zeta = 1.5$. By analyzing the results shown in Fig. 6.17, a couple of conclusions can be drawn.

- **performance of delayed individual adaptation loops:** for small time constants τ^o (fast loops) the performance of the delayed individual adaptation loops is significantly worse with respect to the non-delayed individual loops. The reason for this degradation in MSE is the resonance effect. The spectral content of the system parameter (for the DC-offset in the TwoDOS system see Fig. 6.6) near the cut-off frequency Ω_c^i (where also the resonance peak appears) causes a mismatch error increment. For increasing time constants τ^o (i.e. for decreasing capabilities to track rapid variations), the resonance peak decreases and the performance of the delayed experiments rapidly approaches that of the non-delayed experiments. The insertion of a delay in the individual adaptation loops causes a degradation in MMSE of about 0.35 dB.
- **performance of delayed minimum-latency adaptation loops:** the presence of rapid common variations (see Fig. 6.6) and the presence of latency in the loops causes an improvement in MSE when going from individual to minimum-latency adaptation loops. The minimum-latency adaptation loops achieve an MMSE improvement of about 0.1 dB with respect to the delayed individual adaptation loops.

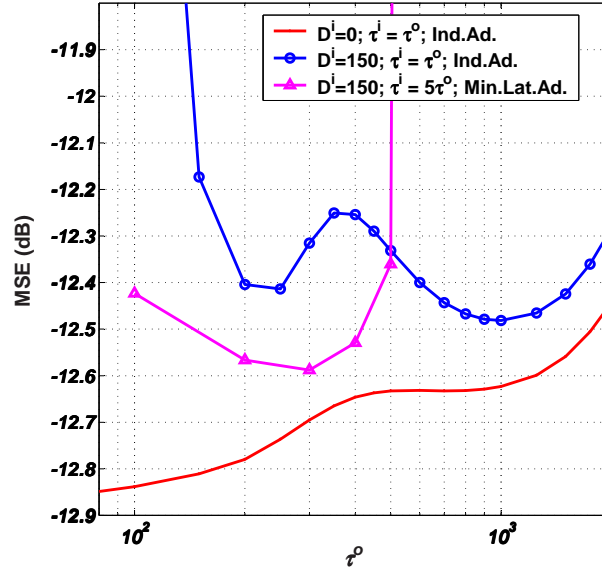


Figure 6.17: Inner-track MSE versus τ^0 . The time constants of the DC adaptation loop and the AGC loop are taken equal.

To assess the strength of the minimum-latency adaptation strategy a last experiment was performed: a detection delay is inserted in all inner-track adaptation loops (DC control, AGC and timing recovery), and the minimum-latency adaptation strategy is also applied to all loops. The results are summarized in the following table.

Type	MMSE
$D^i = 0$, individual adapt.	-12.85
$D^i = 150$, individual adapt.	-12.27
$D^i = 150$, min.-lat. adapt.	-12.71

The insertion of a delay in the adaptation loops causes a loss of 0.58 dB in MMSE performance for individual adaptation loops. The minimum-latency loops make it possible to approach the optimal non-delayed performance up to 0.14 dB, improving upon individual adaptation by 0.44 dB. By comparing the results of the table and Fig. 6.17, one can notice that the MMSE obtained for the third case (minimum-latency adaptation and $M = 150$ for all loops) is better than the MMSE in Fig. 6.17 for the case no latency was present in the timing recovery loop. This MMSE improvement can be explained by the fact that in Fig. 6.17 the normalized natural frequency and the damping factor of the timing recovery loops were fixed and not optimized, whereas in the table they are optimized to obtain the best performance in the different cases.

The gain in bit-error rate is likely to be higher (roughly an order of magnitude for an MSE improvement of 1 dB) but unfortunately insufficient experimental data is available to reliably measure bit-error rates [18]. The results obtained in this section were obtained under relatively favorable conditions as ideal parameter values do not have substantial common high-frequency content. As a result the MMSE gain due to the minimum-latency strategy will be higher if the ideal parameter values will have more common high-frequency content as is the case under extreme situations (severe scratches, dropouts, dust on the disc [173]).

6.8 Conclusion

The presence of delays in adaptation loops introduces resonance effects at frequencies around the cut-off frequency of the loops. These resonance effects cause a degradation of the tracking capabilities of the loop for parameter variations near these frequencies. If these variations are non-negligible, the performance of the loop can degrade considerably. The performance degradation can be limited through an adaptation strategy in which rapid variations are tracked based on control information from tracks with minimum delay. This strategy works well if the rapid parameter variations are common for all the tracks. Experimental results for the TwoDOS system show that even under relatively favorable circumstances (without any scratches or dropouts) the MSE gain is already significant (0.44 dB). In practice more extreme situations (scratches, dust, dropouts,...) can occur and the gain will then definitely be larger.

Chapter 7

Conclusion and Recommendations for Future Work

7.1 Conclusions

In this work, system characterization and reception techniques for 2-D data storage are discussed. A proper and accurate system characterization is of vital importance for the design of a data storage receiver. It is impossible to build a receiver without the knowledge of the system characteristics. After a receiver topology has been chosen and a first receiver implementation is realized, the knowledge of those characteristics that have a big influence on the overall performance, is crucial for improving upon the implemented receiver (e.g. by applying changes to receiver topology or by adding functionalities). This work first discusses the characterization of an experimental 2-D optical storage system, namely the TwoDOS system, and subsequently uses the characterization results to develop reception techniques that efficiently overcome the major bottlenecks in the system.

In Chapter 3, system characterization for two-dimensional storage systems is discussed. A parametric model of the linear and nonlinear ISI structure is used together with an extensive, signal-dependent, autoregressive noise model that is able to accurately describe the noise structure. The parametric ISI model and the noise model together form a 2-D system model that can be used to characterize both 2-D optical and 2-D magnetic storage systems. The proposed 2-D system model has several attractive features: it is conceptually simple and computationally efficient, and estimation of system-model parameters is based on a simple data-aided adaptive scheme which achieves a high accuracy and moreover is able to track channel variations. Agreement with experimental data is very good not only for the second-order statistics but also for the bit-error rates. The results of the characterization (i.e. the values of the estimated parameters) yield important information about the ISI structure and the physical noise sources in the TwoDOS system.

Specifically, the characterization results reveal three characteristics that have a severe impact on system performance and for that reason they have drawn our partic-

ular attention.

- **Media noise:** based upon the noise characterization, it was clearly demonstrated that pit-size noise is the dominant noise source. The characterization results show that in the TwoDOS discs pit-radius variations occur of about 1%. In general, media noise is highly data-dependent and can severely deteriorate system performance. In Chapter 4 a media noise characterization procedure is proposed that is able to properly decompose several types of (media) noise and is able to accurately estimate the key parameters of the different types (e.g. the variance or the autocorrelation).
- **Residual intersymbol interference:** based upon the ISI characterization results, it was shown that for the TwoDOS system still considerable amounts of residual ISI (RISI) originate from bits outside the span of the target response. For severe tilt angles these RISI components have amplitudes in the range $[-0.04, 0.04]$ normalized with respect to the overall signal amplitude. This RISI is highly undesired as it is not accounted for by the detector. As a result a proper handling of this RISI would be very beneficial for the TwoDOS performance. This proper handling is the topic of Chapter 5.
- **Rapid parameter variations:** in the TwoDOS system many different time variations are present: going from slowly-varying disc tilt to rapidly-varying DC offsets. Characterization results show, for example, that offset variations which are common to all the tracks, occur up to a normalized frequency of 0.002. Especially these rapidly-varying parameters might cause problems in the TwoDOS receiver. A key reason for these problems is that the adaptation loops are not able to track fast variations because of the latency in the loops caused by the 2-D bit-detector. In Chapter 6 a solution is proposed that is based on efficiently using minimum-latency control in the loops.

Each of these three characteristics have a clear impact on the performance of the TwoDOS system and for that reason, each characteristic deserves to be treated in more detail in a separate chapter.

In Chapter 4, the characterization of media noise is discussed in more detail. In digital storage systems the total amount of data-dependent media noise increases considerably as storage densities increase. A proper noise characterization is crucial for the design of receivers for high-density storage systems. This characterization involves the selection of a proper noise model and subsequently the accurate estimation of the parameters of the selected model. In Chapter 3 a general noise model (more precisely a data-dependent autoregressive model) was chosen that is suited for the characterization of a wide variety of noise sources. The disadvantage of this general noise model is that it reveals little or no information about which physical noise

sources are the key contributors to the overall noise. In Chapter 4 a general noise-decomposition technique is proposed that is able to jointly estimate the parameters of both media and additive noise sources with a high accuracy. The proposed algorithm makes use of the data-dependency of the media noise to distinguish between the different noise sources. The algorithm is simple and as a result only a very limited amount of complexity is required to implement it in storage systems as an easy “add-on” to read channels ICs. The decomposition technique is explained and elaborated in detail for combinations of additive noise and typical media noise sources in both optical and magnetic storage. Simulation results show that high estimation accuracies are achieved. Based on the simulation results and the analytical derivation of the estimation algorithm, one can clearly indicate which data patterns yield near-optimal estimation performance. These patterns are the ideal test patterns in experimental systems. For optical storage systems a long run of lands followed by a long run of pits is the ideal test pattern and for magnetic storage systems a periodic pattern with a period of 2 bits or more is a suitable test pattern in combination with a DC pattern to characterize the additive noise.

Chapter 5 discusses the cancellation of linear Intersymbol Interference (ISI) in two-dimensional (2-D) systems. It develops a theory for the error rate of receivers that use tentative decisions to cancel residual ISI. Furthermore, it formulates precise conditions under which such ISI cancellation can be applied effectively. For many 2-D systems these conditions are easily met, and therefore the application of ISI cancellation is of significant interest. The theory and the conditions are validated by simulation results for a 2-D channel model. Furthermore, results for the TwoDOS system show that, for a single layer disc with a capacity of 50 GB, the allowed tilt margin at a BER of 10^{-4} is nearly doubled by applying ISI cancellation.

The trend of increasing storage densities results in growing sensitivity of system performance to variations of storage channel parameters. To counteract these variations more adaptivity is needed in the data receiver. Accurate tracking of rapid variations is limited by latencies in the adaptation loops. These latencies are largely governed by delays of the bit-detector. In two-dimensional storage systems, data are packaged in a group of adjacent tracks or rows, and for some of the tracks the detection delays can increase dramatically with respect to one-dimensional systems. As a result the effective latencies in the adaptation loops preclude the tracking of rapid variations and really limit the performance of the system. In Chapter 6, a scheme is proposed that overcomes this problem and that can be used for timing recovery, automatic gain control and other adaptive circuits. Rapid variations for all the tracks are tracked using control information from tracks for which detector latency is smallest. This works properly if rapid variations are common across the tracks as is the case, for example, for the TwoDOS system. Experimental results for TwoDOS confirm that the scheme yields improved performance with respect to conventional adapta-

tion schemes.

7.2 Recommendations for Future Work

Finally, a couple of recommendations for future work can be made based upon the results presented in this work.

The characterization of the TwoDOS system can be expanded further in two directions: the ISI characterization can be improved by adding higher-order nonlinearities between bits that are not both within the first shell, and the noise characterization can be improved by increasing the size of both the data-dependency window and the noise filters. Based on the experimental results, one can state that the chosen sizes yield sufficient accuracy if the system is operating under normal conditions. However in extreme situations other choices might be appropriate. Hence it would be worthwhile to establish a criterion to select appropriate sizes for the data-dependency window and for the noise filters.

In Chapter 4 an adaptive decomposition technique is proposed to analyse noise sources in systems where media noise is the dominant noise source. The accuracy of the decomposition technique is investigated for systems that operate in ideal conditions and for systems that have anomalies, e.g. non-zero bit-error rates, channel IR estimation errors. The results show that the decomposition technique is suitable for a wide range of conditions and only in very severe conditions, very high bit-error rates and large estimation errors, the accuracy of the technique deteriorates.

As a first recommendation for future work, the effect of multiple media noise sources on the estimation accuracy should be studied: is there fundamental interaction between the estimation of different data-dependent noise sources and for which particular situations this does occur? The answer to this question might yield important information about the scope of application of the decomposition technique.

Furthermore, it would be very interesting to judge the quality of the decomposition technique for experimental systems. In experimental systems, many more anomalies might arise and as a result, it would be valuable to see whether the decomposition technique still achieves an acceptable accuracy. Especially a comparison between the results of the decomposition technique and the results of the noise characterization presented in Chapter 3, will yield additional insights about the accuracy of the characterization and the decomposition, and about possible improvements in the data receiver. Furthermore, the use of the decomposition results directly in a data receiver, or more precisely, in a bit-detector, might give additional performance improvements. The knowledge of the relative amounts and the types of media noise might be valuable information based on which an appropriate bit-detector might be chosen. A simple example can be thought of: in case additive noise is dominant, the target response should resemble the channel response, if however media noise is

dominant, a very short target response (in the limit only the central coefficient should be non-zero). This simple example shows the value of the knowledge of the noise composition. More sophisticated techniques could yield important performance improvements.

The adaptive decomposition technique of Chapter 4 is not only of interest for optical storage systems but might also be valuable for other applications, more precisely for applications where multiple noise sources are present and where specific data-dependencies can be attributed to each different noise source. As a result, it would be interesting to see how the decomposition technique can be applied to other applications and what the accuracy of the technique would be.

In Chapter 5, the application of ISI cancellation schemes in 2-D storage systems was proposed. Conditions are phrased for which ISI cancellation is effective. In the chapter, experimental results for the TwoDOS system show that these conditions are met for 2-D optical systems. The experimental results for the TwoDOS system have shown that ISI cancellation is very valuable for systems where a significant amount of ISI is present. As a result the application of ISI cancellation is worth investigating thoroughly beyond the realm of TwoDOS. For this reason it would be very valuable to assess the performance of ISI cancellation when applied to other systems: for example 2-D magnetic storage or holographic storage (which has fundamentally a 3-D nature).

The tracking of rapid parameter variations was studied in Chapter 6. The improvement in MSE was shown for the TwoDOS system. Besides these improvements in MSE, it would be interesting to monitor the improvement in bit-error rate due to the proposed adaptation technique. As already stated in Chapter 6, an improvement in bit-error rate of an order of magnitude is obtained for an MSE improvement of 1 dB. This statement should be verified by experimental results. Just like for the techniques presented in Chapter 4 and in Chapter 5, the applicability of this technique is not limited to the TwoDOS system. The technique of tracking rapidly-varying parameters by utilizing common control information from other tracks can be extended to be used in other 2-D systems. For example in MIMO systems, parallel information streams need to be processed. For these information streams it is likely that parameter variations have a common component. As a result the technique of using common control information for the tracking of parameter variations could be utilized. It would be worthwhile to verify whether these variations are rapidly-varying over time and whether the application of the tracking technique as presented in Chapter 6 would yield performance improvements for MIMO systems.

Bibliography

- [1] W. M. J. Coene, "Two-Dimensional Optical Storage," in *ODS 2003, Techn. Digest*, Vancouver, BC Canada, May 11–14, 2003, pp. 90–92.
- [2] M. Mansuripur, *The Physical Principles of Magneto-Optical Recording*, Cambridge University Press, 1995.
- [3] J. W. M. Bergmans, *Digital Baseband Transmission and Recording*, Kluwer Academic Publishers, 1996.
- [4] C.D. Cideciyan, F. Dolivo, R. Hermann, W. Hirt, and W. Schott, "A PRML System for Digital Magnetic Recording", *IEEE J. Select. Areas Commun.*, vol. 10, no. 1, pp. 38-56, Jan. 1992.
- [5] J. Moon and L.R. Carley, "Performance Comparison of Detection Techniques in Magnetic Recording", vol. 26, no. 6, pp. 3155–3172, Nov. 1990.
- [6] J. Moon, "The Role of SP in Data-Storage Systems", *IEEE Signal Processing Mag.*, vol. 15, no. 4, pp. 54-72, 1998.
- [7] B.H. Schechtman, "Long Term Outlook for Magnetic and Optical Disk Storage Technologies," in *Optical Data Storage Topical Meeting, 1997. ODS. Conference Digest*, Tucson, Arizona, USA, Apr. 7–9, 1997, pp. 3–4.
- [8] T. Narahara, S. Kobayashi, M. Hattori, Y. Shimpuku, G. van den Enden, J. Kahlman, M. van Dijk, and R. van Woudenberg, "Optical Disc System for Digital Video Recording", *Jpn. J. Appl. Phys.*, vol. 39, Pt. 1, no. 2B, pp. 912-919, Feb. 2000.
- [9] A.S. Hoagland, *Digital Magnetic Recording*, John Wiley and Sons Inc, New York, 1991.
- [10] H.J. Richter, A.Y. Dobin, O. Heinonen, K.Z. Gao, R.J.M. v.d. Veerdonk, R.T. Lynch, J. Xue, D. Weller, P. Asselin, M.F. Erden, and R.M. Brockie, "Recording on Bit-Patterned Media at Densities of 1 Tb/in² and Beyond", *IEEE Trans. Magn.*, vol. 42, no. 10, pp. 2255-2260, 2006.

- [11] R.J. Wood, “A Hard Disk Drive Using Magnetic Recording at One Terabit Per Square Inch,” in *NSIC Alternative Storage Technologies Symp.*, Monterey, CA, USA, June 26, 2001.
- [12] K. Compaan and P. Kramer, “The Philips ‘VLP’ System”, *Philips Tech. Rev.*, vol. 33, no. 7, pp. 178-180, 1973.
- [13] J.B.H. Peek, “Communications Aspects of the Compact Disc Digital Audio System”, *IEEE Commun. Mag.*, vol. 23, no. 2, pp. 7-15, Feb. 1985.
- [14] H. van Houten and W. Leibbrandt, “Phase Change Recording”, *Communications on the ACM*, vol. 43, no. 11, pp. 64-71, 2000.
- [15] T. Iwanaga, S. Ogkubo, M. Nakano, M. Kubota, H. Honma, T. Ide, and R. Katayama, “High-Density Recording Systems using Partial Response Maximum-Likelihood with Blue Laser Diode”, *Jpn. J. Appl. Phys.*, vol. 42, no. 2B, part 1, pp. 1042-1043, Feb. 2003.
- [16] A. Padiy, B. Yin, C. Verschuren, J. Lee, R. Vlutters, and T. Jansen, “Signal Processing for 35GB on a Single-Layer Blu-Ray Disc”, *Proc. of SPIE, Optical Data Storage*, vol. 5380, pp. 56-70, 2004.
- [17] G. Bouwhuis, J. Braat, A. Huijser, J. Pasman, G. van Rosmalen, and K. Schouhamer Immink, *Principles of Optical Disc Systems*, Adam Hilger Ltd, Bristol, UK, 1985.
- [18] A. Immink, *System and Receiver Design for Two-Dimensional Optical Storage*, Ph.D. thesis, Univ. of Technology Eindhoven, The Netherlands, Nov. 2005.
- [19] J. Finkelstein and A. Amir, “Methods and Apparatus for High Speed Optical Storage Device”, *U.S. Patent no. 5,627,805*, May 6, 1997.
- [20] A. Amir, “Method and Apparatus for Reading Multiple Tracks of an Optical Disk”, *U.S. Patent no. 6,449,225*, 2002.
- [21] T. Kosoburd, J. Finkelstein, E. Chachamov, I. Katz, M. Naor, and S. Rogers, “Multi-Element Detector and Multi-Channel Signal Conditioner for use Reading Multiple Tracks of Optical Disks Having Diverse Formats”, *U.S. Patent US20030206503*, 2003.
- [22] L.C. Barbosa, “Simultaneous Detection of Readback Signals from Interfering Magnetic Recording Tracks using Array Heads”, vol. 26, no. 5, pp. 2163–2165, Sept. 1990.

-
- [23] M.Z. Ahmed, P.J. Davey, T. Donnelly, and W.W. Clegg, "Track Squeeze using Adaptive Intertrack Interference Equalization", *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2331-2333, Jan. 2002.
 - [24] M.Z. Ahmed, T. Donnelly, P.J. Davey, and W.W. Clegg, "Increases Areal Density Using a 2-Dimensional Approach", *IEEE Trans. Magn.*, vol. 37, no. 4, pp. 1896-1898, Jan. 2001.
 - [25] F. Zijp, M. van der Mark, C. Verschuren, J. Lee, J. van den Eerenbeemd, P. Urbach, and M. van der Aa, "High-Density Near-Field Optical Recording With a Solid Immersion Lens, Conventional Actuator, and a Robust Air Gap Servo", *IEEE Trans. Magn.*, vol. 41, no. 2, pp. 1042-1046, Feb. 2005.
 - [26] K. Saito, T. Ishimoto, T. Kondo, A. Nakaoki, S. Masuhara, M. Furuki, and M. Yamamoto, "Readout Method for Read Only Memory Signal and Air Gap Control Signal in a Near Field Optical Disc System", *Jpn. J. Appl. Phys.*, vol. 41, no. 3B, pp. 1898-1902, March 2002.
 - [27] T.D. Milster, "Near-Field Optics", *Proc. IEEE*, vol. 88, no. 9, pp. 1480-1490, Sept. 2000.
 - [28] C.A. Verschuren, F. Zijp, J.I. Lee, J.M.A. van den Eerenbeemd, D.M. Bruls, K. Saito, and T. Ishimoto, "Cover-Layer Incident Near-Field Recording: Towards 4-Layer Discs using Dynamic Tilt Control," in *ODS 2006, Techn. Digest*, Montreal, Canada, 23-26, 2006, pp. 135-137.
 - [29] T.D. Milster, S.K. Park, and Y. Zhang, "Prospects and Limitations for Large Numbers of Multi-Layers in Optical Data Storage," in *ODS 2006, Techn. Digest*, Montreal, Canada, 23-26, 2006, pp. 179-181.
 - [30] A. Ohta, M. Miyamoto, Y. Kawata, and M. Nakabayashi, "Multilayered Optical Memories for Terabyte Data Storage", vol. 43, no. 2, Part 2, pp. 832-835, Feb. 2007.
 - [31] M. P. O'Neill and T. L. Wong, "Multi-Level Data Storage System Using Phase-Change Optical Discs", *Optical Data Storage, 2000. Conference Digest*, 14-17 May, pp. 170-172, 2000.
 - [32] S. Kobayashi, T. Horigome, and H. Yamatsu, "High-Track-Density Optical Disc by Radial Direction Partial Response", *Jpn. J. Appl. Phys.*, vol. 40, pp. 2301-2307, 2001.
 - [33] J. P. de Kock, S. Kobayashi, T. Ishimoto, H. Yamatsu, and H. Ooki, "Sampled Servo Read Only Memory System Using Single Carrier Independent Pit Edge Recording", *Jpn. J. Appl. Phys.*, vol. 35, pp. 437-442, 1996.

- [34] T. Kobayashi, M. Masuda, and T. Shiratori, "Wall Simulation for the Domain Wall Displacement Detection (I)", *J. Magn. Soc. Jpn.*, vol. 25, no. 3-2, pp. 371-374, 2001.
- [35] S. Kobayashi, H. Yamatsu, and Y. Takemoto, "A Polarization Readout Method Applied to a Groove Baseband Recording Optical Read-Only Memory Discs", *Jpn. J. Appl. Phys.*, vol. 40, pp. 1704-1710, 2001.
- [36] B. V. Johnson, G. A. McDermott, M. P. O'Neill, C. Pietrzyk, T. Shafaat, S. Spielman, D. K. Warland, and T. L. Wong, "Using Pit-Depth Modulation to Increase Capacity and Data Transfer Rate in Optical Discs", *Optical Data Storage Topical Meeting. ODS. Conference Digest*, 7-9 April, pp. 44-45, 1997.
- [37] S. Levy, "The Hard Disk That Changed the World", *Newsweek*, 2006.
- [38] N.L. Koren, "A Comparison of Longitudinal and Perpendicular Recording Based on Optimum Perpendicular Transition", vol. 21, no. 5, pp. 1404-1407, Sept. 1985.
- [39] M. Madden, M. Oberg, Z. Wu, and R. He, "Read Channel for Perpendicular Magnetic Recording", vol. 30, no. 1, pp. 241-246, Jan. 2004.
- [40] R. Cideciyan, E. Eleftheriou, and T. Mittelholzer, "Perpendicular and Longitudinal Recording: a Signal Processing and Coding Perspective", vol. 38, no. 4, pp. 1698-1704, July 2002.
- [41] P.A. Voois and J.M. Cioffi, "Multichannel Signal Processing for Multiple-Head Digital Magnetic Recording", vol. 30, no. 6, pp. 5100-5114, Nov. 1994.
- [42] P.S. Kumar and S. Roy, "Two-Dimensional Equalization: Theory and Applications to High Density Magnetic Recording", *IEEE Trans. Commun.*, vol. 42, no. 1, pp. 386-395, 1994.
- [43] J.M. Cioffi, "Two-Dimensional Data Storage and Retrieval", *IBM Patent Notebook*, June 1984.
- [44] D.J. Costello Jr., J. Hagenauer, H. Imai, and S.B. Wicker, "Applications of Error-Control Coding", *IEEE Trans. Inform. Theory*, vol. 44, no. 6, pp. 2531-2560, Oct. 1998.
- [45] I.S. Reed and G. Solomon, "Polynomial Codes over Certain Finite Fields", *J. Soc. Indust. Appl. Math.*, vol. 8, pp. 300-304, 1960.

-
- [46] K.A.S. Immink, *Codes for Mass Data Storage Systems*, Shannon Foundation Publishers, The Netherlands, 1999.
 - [47] J.J. Wang, *Timing Recovery Techniques for Digital Recording Systems*, Ph.D. thesis, Univ. of Technology Eindhoven, The Netherlands, 2002.
 - [48] B. Bloodworth and P. Siniscalchi and G. De Veirman and A. Jezdic and R. Pier-son and R. Sundararaman, "A 450 Mb/s Analog Front-End for PRML Read Channels," in *Proc. IEEE Intl. Conf. Solid-State Circuits (ISSCC)*, Feb. 1999, pp. 34-35.
 - [49] S.W. McLaughlin, "Shedding Light on the Future of SP for Optical Record-
ing", *IEEE Signal Processing Mag.*, vol. 15, no. 4, pp. 83-94, 1998.
 - [50] Cai Kui, *Design and Analysis of Parity-Check-Code-Based Optical Recording
Systems*, Ph.D. thesis, Univ. of Technology Eindhoven and National University
of Singapore, Mar. 2007.
 - [51] "Standard ECMA-130 Data Interchange on Read-only 120 mm Optical Data
Disks (CD-ROM)," June 1996.
 - [52] R.D. Cideciyan, J.D. Coker, E. Eleftheriou, and R.L. Galbraith, "Noise Pre-
dictive Maximum Likelihood Detection Combined with Parity-Based Post-
Processing", vol. 37, no. 2, pp. 714-720, Mar. 2001.
 - [53] J. Moon and B. Brickner, "Maximum Transition Run Codes for Data Storage
Systems", vol. 32, no. 9, pp. 3992-3994, Sept. 1996.
 - [54] W.M.J. Coene, "Combi-Codes for DC-Free Runlength-Limited Coding",
IEEE Trans. Consumer Electron., vol. 46, pp. 1082-1087, Nov. 2000.
 - [55] P.H. Siegel, "Applications of a peak detection channel", vol. 18, no. 6,
pp. 1250-1252, Nov. 1982.
 - [56] S. Miyanabe, H. Kurabayashi, and K. Yamamoto, "New Equalizer to improve
signal-to-noise ratio", *Jpn. J. Appl. Phys.*, vol. 38, no. 3B, part1, pp. 1715-
1719, Mar. 1999.
 - [57] K.C. Indukumar, "A Novel Post-Processing Scheme for Threshold Detection
on $d = 2$ optical channels," in *Proc. IEEE Intl. Conf. Global Telecommun.
(Globecom)*, San Antonio, USA.
 - [58] P. dling, H.B. Eriksson, and P.O. Brjesson, "Making MLSD Decisions by
Thresholding the Matched Filter Output", *IEEE Trans. Commun.*, vol. 48,
no. 2, pp. 324-332, Feb. 2000.

- [59] A. Cantoni and K. Kwong, "Further Results on the Viterbi Algorithm Equalizer", *IEEE Trans. Inform. Theory*, pp. 764–767, Nov. 1974.
- [60] G.D. Forney, "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference", *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 363-378, May 1972.
- [61] L.C. Barbosa, "Maximum Likelihood Sequence Estimators: A Geometric View", *IEEE Trans. Inform. Theory*, vol. 35, no. 2, pp. 419-427, Mar. 1989.
- [62] H. Kobayashi and D.T. Tang, "Application of Partial-Response Channel Coding to Magnetic Recording Systems", *IBM Journal of Research and Development*, vol. 14, pp. 368-375, 1970.
- [63] D.D. Falconer and F.R.Jr. Magee, "Adaptive Channel Memory Truncation for Maximum Likelihood Sequence Estimation", *Bell Syst. Tech. J.*, vol. 52, pp. 1541-1562, Nov. 1973.
- [64] C.T. Beare, "The Choice of the Desired Impulse Response in Combined Linear-Viterbi Algorithm Equalizers", *IEEE Trans. Commun.*, vol. 26, no. 8, pp. 1301-1307, Aug. 1978.
- [65] I. Lee and J.M. Cioffi, "Equalized Maximum Likelihood Receiver with a Unit Energy Constraint", *IEEE Trans. Magn.*, vol. 33, no. 1, pp. 855-862, Jan. 1997.
- [66] J. Moon and W. Zeng, "Equalization for Maximum Likelihood Detectors", *IEEE Trans. Magn.*, vol. 31, no. 2, pp. 1083-1088, Mar. 1995.
- [67] I. Lee, C. Modlin and J. M. Cioffi, "Equalized Maximum Likelihood Receiver in a Magnetic Recording Channel", *In Proc. Globecom 93*, pp. 1970–1973, Nov. 1993.
- [68] D. G. Messerschmitt, "Design of Finite Impulse Response for The Viterbi Algorithm and Decision Feedback Equalizer", *Proc. ICC*, June 17-19 1974.
- [69] H. Yang and G. Mathew, "Joint Design of Optimum Partial Response Target and Equalizer for Recording Channels with Jitter Noise", *IEEE Trans. Magn.*, vol. 42, pp. 70-77, Jan. 2006.
- [70] N.M. Zayed, J.C. Park, and L.R. Carley, "Detection for Signal-Dependent Correlated Noise in Magnetic Recording," in *Communications, 1999. ICC '99. 1999 IEEE International Conference on*, June 6-10, 1999, vol. 3, pp. 1972–1976.

-
- [71] S.A. Altekhar and J.K. Wolf, "Improvements in Detectors Base Upon Colored Noise", *IEEE Trans. Magn.*, vol. 34, no. 1, pp. 94-97, Jan. 1998.
 - [72] V. Dorfman and J. K. Wolf, "Viterbi Detection for Partial-Response Channels With Colored Noise", vol. 38, no. 5, pp. 2316-2318, Sep. 2002.
 - [73] J. Caroselli, S.A. Altekhar, and P. McEwen J.K. Wolf, "Improved Detection for Magnetic Recording Systems with Media Noise", vol. 33, no. 5, pp. 2779-2781, Sept. 1997.
 - [74] N.M. Zayed and L.R. Carley, "Equalization and Detection in Nonlinear Storage Channels with Signal-Dependent Noise," in *DSP 97*, Santorini, Greece, July 2-4, 1997, pp. 1027-1030.
 - [75] H. Sun, G. Mathew, and B. Farhang-Boroujeny, "Detection Techniques for High-Density Magnetic Recording", *IEEE Trans. Magn.*, vol. 41, pp. 1193-1199, Mar. 2005.
 - [76] N.M. Zayed and L.R. Carley, "Generalized Partial Response Signalling and efficient MLSD using Linear Viterbi Branch Metrics," in *Globecom 1999*, Rio de Janeiro, Brazil, Dec. 5-9, 1999, pp. 949-954.
 - [77] J. Moon and J. Park, "Pattern-Dependent Noise Prediction in Signal-Dependent Noise", *IEEE J. Select. Areas Commun.*, vol. 19, no. 4, pp. 730-743, Apr. 2001.
 - [78] J.D. Coker, E. Eleftheriou, R.L. Galbraith, and W. Hirt, "Noise-Predictive Maximum Likelihood (NPML) Detection", *IEEE Trans. Magn.*, vol. 34, no. 1, pp. 110-117, Jan. 1998.
 - [79] Y. Okamoto, N. Masunari, H. Yamamoto, H. Osawa, H. Saito, H. Muraoka, and Y. Nakamura, "Jitter-Like Noise Cancellation Using AR Model of PR Channel in Perpendicular Magnetic Recording", *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2349-2351, Sept. 2002.
 - [80] H. Matui, "Adaptive Reduced-State Sequence Estimation for Linearly and Nonlinearly Distorted Signals in Magnetic Recording", vol. 29, no. 6, pp. 4024-4026, Nov. 1993.
 - [81] N.M. Zayed and L.R. Carley, "Equalization and Detection for Nonlinear Recording Channels with Correlated Noise", *IEEE Trans. Magn.*, vol. 35, no. 5, pp. 2295-2297, Sept. 1999.
 - [82] L. Agarossi, S. Bellini, F. Bregoli, and P. Migliorati, "An Effective Non Linear Receiver for High Density Optical Disc," in *Global Telecommunications Conference, 1998. GLOBECOM 98. IEEE*, Nov. 1998, vol. 6, pp. 3374-3378.

- [83] A. Ghrayeb, W.E. Ryan, and Nan-Hsiung Yeh, "Performance of Random-Access-Memory-Based Equalizers in Magnetic Recording", *IEEE Trans. Magn.*, vol. 36, no. 6, pp. 4028-4035, Nov. 2000.
- [84] W.U. Lee and F.S. Hill Jr., "A Maximum-Likelihood Sequence Estimator with Decision-Feedback Equalization", *IEEE Trans. Commun.*, vol. 25, no. 9, pp. 971-979, Sept. 1977.
- [85] C.A. Belfiore and Jr. J.H. Park, "Decision Feedback Equalization", *Proc. IEEE*, vol. 67, no. 8, pp. 1143-1156, Aug. 1979.
- [86] A. Kavčić, "Decision Feedback Equalization in Channels with Signal-Dependent Media Noise", *IEEE Trans. Magn.*, vol. 37, no. 4, pp. 1909-1911, July 2001.
- [87] Weining Zeng and J. Moon, "Decision Feedback Equalizer with Pattern Dependent Dynamic Threshold", *IEEE Trans. Magn.*, vol. 32, no. 4, pp. 3266-3273, July 1996.
- [88] D. Yellin, A. Vardy, and O. Amrani, "Joint Equalization and Coding for Inter-symbol Interference Channels", *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 409-425, Mar. 1997.
- [89] J.W.M. Bergmans, J.O. Voorman, and H.W. Wong-Lam, "Dual Decision Feedback Equalizer ", *IEEE Trans. Commun.*, vol. 45, no. 5, pp. 514-518, May 1997.
- [90] K.C. Indukumar, Y.X. Lee, and G. Mathew, "Performance Comparison of Modified Multi-Level Decision Feedback Equalization Detectors", vol. 3, no. 1, Part 2, pp. 594-604, Jan. 1999.
- [91] S.K. Nair, H. Shafiee, and J. Moon, "Equalization and Detection in Storage Channels", *IEEE Trans. Magn.*, vol. 32, no. 5, pp. 5206-5217, Sept. 1996.
- [92] K. Han and R.R. Spencer, "Performance and implementation of Adaptive Partial Response Maximum Likelihood Detection", *IEEE Trans. Magn.*, vol. 34, no. 5, pp. 3806-3815, Sept. 1998.
- [93] R. Koetter, A.C. Singer, and M. Tüchler, "Turbo Equalization", *IEEE Signal Processing Mag.*, vol. 21, pp. 67-80, Jan. 2004.
- [94] T.V. Souvignier, M. Oberg, P.H. Siegel, R.E. Swanson, and J.K. Wolf, "Turbo Decoding for Partial Response Channels", *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1297-1308, Aug. 2000.

-
- [95] W.H. Gerstacker, R.R. Muller, and J.B. Huber, "Iterative Equalization with Adaptive Soft Feedback", *IEEE Trans. Commun.*, vol. 48, no. 9, pp. 1462-1466, Sept. 2000.
 - [96] Sen Jiang and Hong Sunm Chi Sing Leung, "Modified LMMSE Turbo Equalization", *IEEE Commun. Lett.*, vol. 8, no. 3, pp. 174-176, Mar. 2004.
 - [97] L.L. McPheters and S.W. McLaughlin, "Turbo-Coded Optical Recording Channels with DVD Minimum Mark Size", *IEEE Trans. Magn.*, vol. 38, no. 1, pp. 298-302, Jan. 2002.
 - [98] H. Song, B.V.K. Vijaya Kumar, E. Kurtas, and Y. Yuan, "Turbo Decoding for Optical Data Storage," in *IEEE International Conference on Communications 2000*, June 18-22, 2000, vol. 1, pp. 104-108.
 - [99] L.L. McPheters, S.W. McLaughlin, and E.C. Hirsch, "Turbo Codes for PR4 and EPR4 Magnetic Recording," in *Signals, Systems & Computers, 1998.*, Nov. 1-4, 1998, vol. 2, pp. 1778-1782.
 - [100] Y. Nakamura, Y. Okamoto, H. Osawa, H. Saito, H. Muruoka, and Y. Nakamura, "A Study of Turbo Decoding with Embedded AR Channel Model for Perpendicular Recording", *IEEE Trans. Magn.*, vol. 39, no. 5, pp. 2570-2572, Sept. 2003.
 - [101] Z.-N. Wu and J.M. Cioffi, "Low-Complexity Iterative Decoding with Decision-Aided Equalization for Magnetic Recording Channels", *IEEE J. Select. Areas Commun.*, vol. 19, no. 4, pp. 699-707, Apr. 2001.
 - [102] Y. Lin and M.F. Yuan, "Post Compensation for Nonlinear Distortions in PRML Recording Channel", *IEEE Trans. Magn.*, vol. 31, no. 6, pp. 3033-3035, Nov. 1995.
 - [103] H. Sawaguchi, M. Izumita, and S. Mita, "Soft-Output Post-Processing Detection for PRML Channels in the Presence of Data-Dependent Media Noise," in *Globecom '03*, 2003, vol. 1, pp. 3913-3920.
 - [104] H. Sawaguchi, M. Kondou, N. Kobayashi, and S. Mita, "Concatenated Error Correction Coding for High-Order PRML Channels," in *Global Telecommunications Conference, 1998. GLOBECOM '98. IEEE*, Nov. 1998, pp. 2694-2699.
 - [105] A. Viterbi, "Error Bounds for Convolutional Codes and an Asymptotically Optimum Decoding Algorithm", *IEEE Trans. Inform. Theory*, vol. 13, pp. 260-269, Apr. 1967.

- [106] S.A. Fredricsson, "Joint Optimization of Transmitter and Receiver Filters in Digital PAM Systems with a Viterbi Detector", *IEEE Trans. Inform. Theory*, vol. 22, no. 2, pp. 200-210, Mar. 1976.
- [107] T. Okumura, J. Akiyama, S. Maeda, T. Yamaguchi, and A. Takahashi, "New Adaptive Equalization Method for Partial Response Maximum Likelihood System Optimization Error Rate Performance", *Jpn. J. Appl. Phys.*, vol. 43, no. 10, pp. 7097-7101, 2004.
- [108] J. Riani, S. Van Beneden, J.W.M. Bergmans, and A.H.J. Immink, "Near Minimum BER Equalizer Adaptation for PRML Systems," in *Global Telecommunications Conference, 2005. GLOBECOM '05. IEEE*, Saint Louis, USA, 28 Nov. - 2 Dec. 2005, vol. 4, pp. 2123-2128.
- [109] J. Riani, A.H.J. Immink, J.W.M. Bergmans, and S. Van Beneden, "Near Minimum BER, All Adaptive Partial Response Equalization for High Density Recording Systems," in *Global Telecommunications Conference, 2006. GLOBECOM '06. IEEE*, San Francisco, USA, 27 Nov. - 1 Dec. 2006, vol. 4, pp. 2123-2128.
- [110] A.D. Weathers, "Sensitivity of PRML Systems to Timing Offsets", *IEEE Trans. Magn.*, vol. 32, no. 5, pp. 3971-3973, Sept. 1996.
- [111] M.Y. Lin, *Synchronous/Asynchronous Adaptive Signal Processing For Recording Channels*, Ph.D. thesis, National University of Singapore, 2002.
- [112] J.W.M. Bergmans, "Efficiency of Data-Aided Timing Recovery Techniques", *IEEE Trans. Inform. Theory*, vol. 41, no. 5, pp. 1397-1408, Sept. 1995.
- [113] A. Gameiro, "Baud Sampling Bit Synchroniser for Channels with Data Dependent Noise", *Electron. Lett.*, vol. 34, no. 21, pp. 2000-2002, Aug. 28, 1998.
- [114] J. Riani, J.W.M. Bergmans, S. Van Beneden, and A.H.J. Immink, "Data-Aided Timing Recovery for Recording Channels With Data-Dependent Noise", *IEEE Trans. Magn.*, vol. 42, pp. 3752-3759, Nov. 2006.
- [115] J. Armstrong, "Symbol Synchronization Using Baud-Rate Sampling and Data-Sequence-Dependent Signal Processing", *IEEE Trans. Commun.*, vol. 39, no. 1, pp. 127-132, Jan. 1991.
- [116] B.A. Wilson, R. New, J. Campello, and B. Marcus, "Joint Estimation of Data and Timing in the Presence of Inter-Symbol Interference", *IEEE Trans. Magn.*, vol. 39, no. 5, pp. 2582-2584, Sept. 2003.

-
- [117] R. Raheli, A. Polydoros, and C.K. Tzou, "Per-Survivor Processing: A General Approach to MLSE in Uncertain Environments", *IEEE Trans. Commun.*, vol. 43, no. 2/3/4, pp. 354-364, Feb./Mar./Apr. 1995.
 - [118] P. Kovintavewat, J.R. Barry, M. Fatih Erden, and E. Kurtas, "Per-Survivor Timing Recovery for Uncoded Partial Response Channels", pp. 2715-2719, 2004.
 - [119] J.S. Joo, S.C. Hong, and Y.H. Lee, "Adaptive MLSE Receiver: Hybrid of Per-Survivor Processing and Tentative Decision MLSE", *Electronics Letters*, vol. 36, no. 7, pp. 678-680, Mar. 2000.
 - [120] G. Marino R. Raheli and P. Castoldi, "Per-Survivor Processing and Tentative Decisions: What is in Between?", *IEEE Trans. Commun.*, vol. 44, no. 2, pp. 127-129, Feb. 1996.
 - [121] European Project TwoDOS, "Description of Work", *IST-2001-34168*, Feb. 2002.
 - [122] W. Weeks and R.E. Blahut, "The Capacity and Coding Gain of Certain Checkerboard Codes", *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 1193-1203, May 1998.
 - [123] T. Kato, S. Taira, T. Maeda, Y. Katayama, and T. Nishiya, "Two-Dimensional Run-length-limited Code and Partial Response Maximum Likelihood System with Multi-Track Recording", *ISOM/ODS 2002, Joint International Symposium on Optical Memory and Optical Data Storage, Post-Deadline Paper WP.23*, pp. 51-53, 2002.
 - [124] T. Kato, S. Taira, Y. Katayama, T. Nishiya, and T. Maeda, "Two-Dimensional Partial Response Equalization and Detection Method with Multi-Track Recording for Optical Disks", *Technical Report of IEICE*, vol. 2002-03, pp. 65-70, 2002.
 - [125] S. Taira, T. Hoshizawa, T. Kato, Y. Katayama, T. Nishiya, and T. Maeda, "Study of Recording Methods for Advanced Optical Disks", *Technical Report of IEICE*, vol. 2002-03, pp. 57-64, 2002.
 - [126] T. Kalker, "On Multidimensional Sampling", *Digital Signal Processing (SP) Handbook, CRC Press LLC, Chapter 4*, 1998.
 - [127] M. Boamfa and J. H. M. Neijzen, "Two-Dimensional Optical Storage Mastering: Adding a New Dimension to Liquid Immersion Mastering", *Techn. Dig. ISOM 2004*, , no. Tu-B-04, pp. 10-11, 2004.

- [128] H. van Santen and J. H. M. Neijzen, "Deep-UV Liquid Immersion Mastering of High Density Optical Discs", *Jpn. J. Appl. Phys.*, vol. 42, no. 2B, pp. 1110-1112, 2003.
- [129] M. Takeda, M. Furuki, M. Yamamoto, M. Shinoda, K. Saito, Y. Aki, H. Kawase, M. Koizumi, and T. Miyokawa, "Progress in Electron Beam Mastering of 100 Gbit/inch² Density Disc", *Jpn. J. Appl. Phys.*, vol. 43, no. 7B, pp. 5044-5046, 2004.
- [130] A. Nowbakht and J.W.M. Bergmans, "Design of Optimum Sync and Detection Patterns for Frame Synchronisation", *Electronic Letters*, vol. 40, no. 16, pp. 1000-1001, Aug. 2004.
- [131] H.H. Hopkins, "Diffraction Theory of Laser Read-Out Systems for Optical Video Discs", *J. Opt. Soc. Am.*, vol. 69, pp. 4, 1979.
- [132] W.M.J. Coene, "Non-Linear Signal Processing Model for Scalar Diffraction in Optical Recording", *OSA Applied Optics*, vol. 42, no. 25, pp. 6525-6535, Nov. 2003.
- [133] A.M. van der Lee, *Unraveling the Mechanism of Excess Quantum Noise*, Ph.D. thesis, University of Leiden, 2000.
- [134] M.W. Marcellin and H.J. Weber, "Two-Dimensional Modulation Codes", *IEEE J. Select. Areas Commun.*, vol. 10, pp. 254-266, Jan. 1992.
- [135] R.E. Swanson and J.K. Wolf, "New Class of Two-Dimensional RLL Recording Codes", vol. 26, no. 6, pp. 3407-3416, Nov. 1992.
- [136] A. Kato and K. Zeger, "On the Capacity of Two-Dimensional Run-Length Constrained Channels", *IEEE Trans. Inform. Theory*, vol. 45, no. 5, pp. 1527-1540, jul 1999.
- [137] J.J. Ashley and B.H. Marcus, "Two-Dimensional Low-Pass Filtering Codes", *IEEE Trans. Commun.*, vol. 46, no. 6, pp. 724-727, June 1998.
- [138] E. Soljanin and C.N. Georgiades, "Multihead Detection for MultiTrack Recording Channels", *IEEE Trans. Inform. Theory*, vol. 44, no. 1, pp. 2988, Jan. 1998.
- [139] W.M.J. Coene and A.H.J. Immink, "Modulation Coding for a Two-Dimensional Optical Storage Channel", *DIMACS Series in Discrete Mathematics and Theoretical Computer Sciences*, vol. 6, 2004.

-
- [140] M. Kavehrad and J. Salz, "Cross-Polarization Cancellation in Digital Transmission Over Dually Polarized Multipath Fading Channels", *ATT Technical Journal*, vol. 64, no. 10, pp. 2211–2245, Dec. 1985.
 - [141] A. Duel-Hallen, "Equalizers for Multiple Input/Multiple Output Channels and PAM Systems with Cyclostationary Input", *IEEE J. Select. Areas Commun.*, vol. 10, no. 3, pp. 630–639, Apr. 1992.
 - [142] Y.-C. Lee and C.-H. Wei, "A Pseudo-2D Adaptive Decision Feedback Equalizer for High Density Magnetic Recording System," in *Global Telecommunications Conference, 1995. GLOBECOM '95. IEEE*, Nov. 1995, vol. 1, pp. 582–586.
 - [143] Y. Mulero, *Phase Equalization and Detection for multitrack data storage*, Ph.D. thesis, Univ. of Limerick, Limerick, Ireland, Jan 2005.
 - [144] S. Tosi and T. Conway, "Detector Target Response Optimization for Multitrack Digital Data Storage", *IEEE Trans. Magn.*, vol. 42, no. 7, pp. 1926–1928, July 2006.
 - [145] A.H.J. Immink, J. Riani, S. Van Beneden, J.W.M. Bergmans, M. Ciacci, A. Nowbakht Irani, W.M.J. Coene, A.M. van der Lee, and D. Bruls, "Adaptation and Timing Recovery for Two-Dimensional Optical Storage," in *Proceedings of SPIE, ODS 2004*, Apr. 4, 2004.
 - [146] A.K. Ma and M.W. Marcellin, "Timing Recovery For Two-Dimensional Modulation Codes," in *ICC '92*, June 1992, pp. 1361–1365.
 - [147] T. Conway, "A Partial Response Read Channel for Two Dimensional Optical Data Storage", *IEEE Trans. Consumer Electron.*, vol. 50, no. 4, pp. 1107–1112, Nov. 2004.
 - [148] K.D. Fisher, J.M. Cioffi, W.L. Abbott, P.S. Bednarz, and C.M. Melas, "An Adaptive RAM-DFE for Storage Channels", *IEEE Trans. Commun.*, vol. 39, no. 11, pp. 1559–1568, Nov. 1991.
 - [149] J. Riani, J. W. M. Bergmans, and A. H. J. Immink, "Asynchronous LMS Adaptive Equalization: Analysis and Efficient Realization", *Submitted to IEEE Trans. on Sign. Proc.*, 2005.
 - [150] R. Krishnamoorthi, "Two-Dimensional Viterbi-like Algorithms," M.Sc. thesis, Univ. of Illinois at Urbana-Champaign, 1998.
 - [151] W. Weeks, *Full-Surface Data Storage*, Ph.D. thesis, Univ. of Illinois at Urbana-Champaign, 2000.

- [152] A.P. Hekstra, W.M.J. Coene, and C. Baggen, "Iterative Stripewise Trellis-Base Symbol Detection Method and Device", *Patent Application*, , no. EPO02292937.6; IB2003/005208, Nov. 2003.
- [153] J.W.M. Bergmans, S. Mita, and M. Izumita, "Characterization of Digital Recording Channels", *Philps Journal of Research*, vol. 44, no. 1, pp. 57-96, 1989.
- [154] L.C. Barbosa, "A Model for Magnetic Recording Channels with Signal Dependent Noise", *IEEE Trans. Magn.*, vol. 31, pp. 1062-1064, Mar. 1995.
- [155] J. Moon, "Discrete-Time Modeling of Transition-Noise-Dominant Channels and Study of Detection Performance", *IEEE Trans. Magn.*, vol. 27, pp. 4573-4578, Nov. 1991.
- [156] J.-G. Zhu and N.H. Bertram, "Recording and Transition Noise Simulations in Thin Film Media", *IEEE Trans. Magn.*, vol. 24, pp. 2706-2708, Nov. 1988.
- [157] A. Kavčić and A. Patapoutian, "A Signal-Dependent Autoregressive Channel Model", *IEEE Trans. Magn.*, vol. 35, no. 5, pp. 2316-2318, Sept. 1999.
- [158] Z. Jin, K. Zhang, G.H., and H.N. Bertram, "Experimental Study of Off-Track Dependence of Medium Noise Using a Mode Projection Method", *IEEE Trans. Magn.*, vol. 36, pp. 2154-2156, Sept. 2000.
- [159] X. Xing and H.N. Bertram, "Error Rate Analysis of Partial Response Channels in the Presence of Texture Noise", *IEEE Trans. Magn.*, vol. 35, pp. 2070-2079, May 1999.
- [160] K. Cai, G. Mathew, J. Bergmans, and Z. Qin, "A Generalized Braat-Hopkins Model for Optical Recording Channels," in *Consumer Electronics, 2003, ICCE, 2003 IEEE International Conference on*, 17 - 19 June 2003, pp. 324-325.
- [161] R. Hermann, "Volterra Modeling of Digital Saturation Recording Channels", *IEEE Trans. Magn.*, vol. 26, no. 5, pp. 2125-2127, Sept. 1990.
- [162] Y.S. Cho and N.J. Lee, "An Estimation Technique for Nonlinear Distortion in High-Density Magnetic Recording Channels", *IEEE Trans. Magn.*, vol. 34, no. 1, pp. 40-44, Jan. 1998.
- [163] M.Y. Lin, *Synchronous/Asynchronous Adaptive Signal Processing For Recording Channels*, Ph.D. thesis, National University of Singapore, 2002.
- [164] S. Haykin, *Adaptive Filter Theory*, Prentice Hall Inc., 1986.

-
- [165] A. Gersho and T.L. Lim, "Adaptive Cancellation of Intersymbol Interference for Data Transmission", *Bell System Technical Journal*, vol. 60, no. 9, pp. 1997-2021, Nov. 1981.
 - [166] K. Wesolowski, "On the Performance and Convergence of the Adaptive Canceller of Intersymbol Interference in Data Transmission", *IEEE Trans. Commun.*, vol. 33, no. 5, pp. 57-96, May 1985.
 - [167] T. Perkins and Z. A. Keirn, "A Window-Margin-Like Procedure for Evaluating PRML Channel Performance", *IEEE Trans. Magn.*, vol. 31, no. 2, pp. 1109-1114, March 1995.
 - [168] P. Wambacq and H. Mannaert, *Handboek Signaalverwerking*, Uitgeverij Acco, 1998.
 - [169] R.M. Mersereau, "The Processing of Hexagonal Sampled Two-Dimensional Signals", *Proc. IEEE*, vol. 67, no. 6, pp. 930-949, June 1979.
 - [170] A. Kavčić and J.M.F. Moura, "The Viterbi Algorithm and Markov Noise Memory", *IEEE Trans. Inform. Theory*, vol. 46, no. 1, pp. 291-301, Jan. 2000.
 - [171] J. Stander and A. Patapoutian, "Performance of a Signal-Dependent Autoregressive Channel Model", *IEEE Trans. Magn.*, vol. 36, no. 5, pp. 2197-2199, Sept. 2000.
 - [172] P.A.M. Dirac, *The Principles of Quantum Mechanics*, Clarendon, Oxford University Press, 1958.
 - [173] T. Watanabe, K. Saito, and K. Seo, "Study of Error Propagation due to Dust for Thin-Cover Coat Disk Systems," in *Optical Memory and Optical Data Storage Topical Meeting, 2002. International Symposium on*, July7-11 2002, pp. 126-128.
 - [174] P.F. Odgaard and M.V. Wickerhauser, "Time Localisation of Surface Defects on Optical Discs," in *Proceedings of the 2004 IEEE International Conference on Control Applications*, Sept.2-4 2004, pp. 111-116.
 - [175] M. Ciacci, "Design of Several Signal Processing Blocks for the TwoDOS Receiver", *Thesis, Master of Technological Design, University of Technology Eindhoven*, 2004.
 - [176] M. Ciacci, "Equalizer based Tilt Estimation for Two-Dimensional Optical Storage," in *Proceedings of the 25-th Symp. Inf. Th. Benelux.*, Kerkrade, The Netherlands, 2-4 june 2004, pp. 249-256.

- [177] R.D. DeGroat, M. Vis, and W.G. Bliss, "Experimental Characterization of Media Noise Based on Equalized, Synchronized Drive Data", *IEEE Trans. Magn.*, vol. 37, pp. 633-638, Mar. 2001.
- [178] X. Zhang, T.M. Duman, and E.M. Kurtas, "Information Rates of Binary-Input Intersymbol Interference Channels With Signal-Dependent Media Noise", *IEEE Trans. Magn.*, vol. 39, pp. 599-607, Jan. 2003.
- [179] R. Wood, "Detection and Capacity Limits in Magnetic Media Noise", *IEEE Trans. Magn.*, vol. 34, pp. 1848-1850, July 1998.
- [180] J. Fitzpatrick, H.N. Bertram, and X. Che, "The Relationship of Medium Noise to System Error Rate in a PRML Channel", *IEEE Trans. Magn.*, vol. 30, pp. 3990-3995, Nov. 1994.
- [181] T.R. Oenning and J. Moon, "The Effect of Jitter Noise on Binary Input Intersymbol Interference Channel Capacity," in *Communications, 2001, ICC 2001, IEEE International Conference on*, 11-14 June 2001, vol. 8, pp. 2416-2420.
- [182] W. Zeng and J. Moon, "Modified Viterbi Algorithm for a Jitter-Dominant $1 - D^2$ Channel", *IEEE Trans. Magn.*, vol. 28, pp. 2895-2897, Sept. 1992.
- [183] A. Kavčić and M. Srinivasan, "The Minimum Description Length Principle for Modeling Recording Channels", *IEEE J. Select. Areas Commun.*, vol. 19, pp. 719-729, Apr. 2001.
- [184] A. Kavčić and J. M. F. Moura, "Correlation structures for optimizing information criteria," in *Proc. IEEE Inform. Theory Workshop on Detection, Estimation, Classification and Imaging*, (Santa Fe, NM), Feb. 1999.
- [185] C. Fu, Z. Jin, H.N. Bertram, Y. Wu, and D. Guarisco, "Measurements and Analysis of Transition Noise in Perpendicular Media", *IEEE Trans. Magn.*, vol. 39, pp. 2606-2608, Sept. 2003.
- [186] W.C. Williams, B.I. Finkelstein, and T.W. McDaniel, "Spot and Mark-Size Characterization in Magneto-Optic Recording", *IEEE Trans. Magn.*, vol. 24, pp. 2323-2325, Nov. 1988.
- [187] X. Xing and H.N. Bertram, "Analysis of Transition Noise in Thin Film Media", *IEEE Trans. Magn.*, vol. 33, pp. 2959-2961, Sept. 1997.
- [188] W. Zeng, A. Kavčić, and R. Motwani, "Extraction of Timing Error Parameters From Readback Waveforms", *IEEE Trans. Magn.*, vol. 42, pp. 194-199, Feb. 2006.

-
- [189] H.N. Bertram and X. Che, "General Analysis of Noise in Recorded Transitions in Thin Film Recording Media", *IEEE Trans. Magn.*, vol. 29, pp. 201-208, Jan. 1993.
- [190] H. Pozidis, "Decomposition of Noise Sources in Recording Applications Using Symbol-Rate Readback Samples", *IEEE Trans. Magn.*, vol. 40, pp. 2320-2322, July 2004.
- [191] A. Stek, G. W. de Jong, T. P. H. G. Jansen, J. R. M. Bergervoet, and P. H. Woerlee, "Circuit Design and Noise Considerations for Future Blu-ray Disc Optical Storage Technology", *Digest of Technical Papers. ISSCC 15-19 Feb.*, vol. 1, pp. 136-137, 2004.
- [192] A. Taratorin, D. Cheng, and E. Marinero, "Media Noise, Nonlinear Distortions, and Thermal Stability in High Density Recording", *IEEE Trans. Magn.*, vol. 36, no. 1, pp. 80-85, Jan. 2000.
- [193] I. Ozgunes, B.V.K. Vijaya Kumar, and M.H. Kryder, "Effect of Transition Noise on the Signal-to-Noise Ratio of Magneto-optic Read Channels", *IEEE Trans. Magn.*, vol. 32, pp. 3291-3304, July 1996.
- [194] T. Oenning and J. Moon, "Modeling the Lorentzian Magnetic Recording Channel with Transition Noise", *IEEE Trans. Magn.*, vol. 37, pp. 583-591, Jan. 2001.
- [195] L. Isserlis, "On a Formula for the Product-Moment Coefficient of any Order of a Normal Frequency Distribution in any Number of Variables", *Biometrika*, vol. 12, pp. 134-139, 1918.
- [196] O. E. Agazzi and N. Seshradsi, "On the use of Tentative Decisions to Cancel Intersymbol Interference and Nonlinear Distortion (with Application to Magnetic Recording Channels)", *IEEE Trans. Inform. Theory*, vol. 43, no. 2, pp. 394-408, Mar. 1997.
- [197] A.H.J. Immink, W.M.J. Coene, A.M. van der Lee, C. Busch, A.P. Hekstra, J.W.M. Bergmans, J. Riani, S. Van Beneden, and T. Conway, "Signal Processing and Coding for Two-Dimensional Optical Storage," in *Global Telecommunications Conference, 2003. GLOBECOM '03. IEEE*, Dec.1-5, 2003, vol. 7, pp. 3904-3908.
- [198] L. Huang, G. Mathew, and Tow Chong Chong, "Reduced Complexity Viterbi Detection for Two-Dimensional Optical Recording", *IEEE Trans. Consumer Electron.*, vol. 51, no. 1, pp. 123-129, Feb. 2005.

- [199] J.G. Proakis, "Adaptive Nonlinear Filtering Techniques for Data Transmission," in *Proc. IEEE Symp. Adaptive Processes, Decisions, Contr.*, 1970, pp. XV.2.1–XV.2.5.
- [200] M.S. Mueller and J. Salz, "A Unified Theory of Data-Aided Equalization", *Bell System Technical Journal*, vol. 60, no. 9, pp. 2023-2038, Nov. 1981.
- [201] K. Wesolowski and J.G. Proakis, "A Simplified Two-Stage Equalizer with a Reduced Number of Multiplications for Data Transmission over Voiceband Telephone Links", *IEEE J. Select. Areas Commun.*, vol. SAC-2, no. 5, pp. 731-742, Sept. 1984.
- [202] E. Biglieri, A. Gersho, R.D. Gitlin, and T.L. Lim, "Adaptive Cancellation of Nonlinear Intersymbol Interference for Voiceband Data Transmission", *IEEE J. Select. Areas Commun.*, vol. SAC-2, no. 5, 1984.
- [203] S. Mita, M. Izumita, N. Doi, and Y. Eto, "Automatic Equalizer for Digital Magnetic Recording Systems", *IEEE Trans. Magn.*, vol. MAG-23, no. 5, pp. 3672-3674, Nov. 1987.
- [204] J.W.M. Bergmans, S. Mita, and M. Izumita, "A Simulation Study of Adaptive Reception Schemes for High-Density Digital Magnetic Recording", *IEEE Trans. Magn.*, vol. 27, no. 1, pp. 717-723, Jan. 1991.
- [205] R. Wood, "New Detector for (1,k) Codes Equalized to Class II Partial Response", *IEEE Trans. Magn.*, vol. 25, no. 5, pp. 4075-4077, Sept. 1989.
- [206] R. Kohno, H. Imai, and M. Hatori, "Automatic Equalizer Including a Decoder of Error-Correcting Code and Its Development", *Electronics and Communications in Japan, Part 1*, vol. 68, no. 11, pp. 66-77, 1985.
- [207] P.D. Karabinis and F.E. Thau, "An Adaptive Algorithm for Linear/Nonlinear Distortion Cancellation in Digital Communications Systems," in *ICC 1990*, Atlanta, USA, Apr. 16-19, 1990, pp. 1428–1433.
- [208] S. Serfaty, J.L. LoCicero, and G.E. Atkin, "Cancellation of nonlinearities in bandpass QAM systems", *IEEE Trans. Commun.*, vol. 38, no. 10, 1990.
- [209] S. Serfaty, "Performance of combined Trellis Coded Modulation and Nonlinear Cancellation", *IEEE Trans. Commun.*, vol. 41, no. 1, 1993.
- [210] O.E. Agazzi and N. Seshradsi, "When Can Tentative Decisions be Used to Cancel (Linear or NonLinear) Intersymbol Interference? (With Application to Magnetic Recording Channels)," in *ICC 1995*, Seattle, USA, June 18-22, 1995, pp. 647–652.

-
- [211] W.M.J. Coene, D.M. Bruls, A.H.J. Immink, A.M. van der Lee, A.P. Hekstra, J. Riani, S. Van Beneden, M. Ciacci, J.W.M. Bergmans, and M. Furuki, "Two-Dimensional Optical Storage," in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2005*, May 2005.
 - [212] T. Conway, R. Conway, and S. Tosi, "Signal Processing for Multitrack Digital Data Storage", *IEEE Trans. Magn.*, vol. 41, no. 4, pp. 1333-1339, Apr. 2005.
 - [213] J.K. Nelson, A.C. Singer, and U. Madhow, "Multi-Directional Decision Feedback for 2D Equalization," in *ICASSP*, July 10–13, 2004, vol. IV, pp. 921–924.
 - [214] M.A. Neifeld, K.M. Chugg, and B.M. King, "Parallel Data Detection in Page-Oriented Optical Memory", *Optics Letters*, vol. 21, no. 18, pp. 1481-1483, Sept. 1996.
 - [215] J.F. Heanue, K. Gurkan, and L. Hesselink, "Signal Detection for Page-Access Optical Memories with Intersymbol Interference", *Applied Optics*, vol. 35, no. 14, pp. 2431-2438, May 1996.
 - [216] Naveen Singla, J.A. O'Sullivan, R.S. Indeck, and Y. Wu, "Iterative Decoding and Equalization for 2-D Recording Channels", *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2328-2330, Sept. 2002.
 - [217] M. Marrow and J.K. Wolf, "Detection of 2-Dimensional Signals in the Presence of ISI and Noise," in *International Symposium on Information Theory and its Applications 2004*, Oct. 10–13, 2004, vol. 1, pp. 104–108.
 - [218] K.C. Indukumar, "State Reduction in Viterbi Detector Using Preliminary Decisions", *IEEE Trans. Magn.*, vol. 37, no. 4, pp. 1899-1901, July 2001.
 - [219] D.M. Bruls, A.H.J. Immink, A.M. van der Lee, W.M.J. Coene, J. Riani, S. Van Beneden, M. Ciacci, J.W.M. Bergmans, and M. Furuki, "Two-dimensional Optical Storage: High-Speed Read-Out of a 50 GByte Single-Layer Optical Disc with a 2D Format using $\lambda = 405\text{nm}$ and $\text{NA}=0.85$," in *International Symposium on Optical Memory 2004*, Jeju, Korea, Oct. 13, 2004.
 - [220] T.H.M. Akkermans, "Digital Control in Optical Disc Drives," in *Consumer Electronics, 2001. ICCE. International Conference on*, June 19-21 2001, pp. 246–247.
 - [221] M. Marrow and J.K. Wolf, "Iterative Detection of 2-Dimensional ISI Channels," in *IEEE Information Theory Workshop 2003*, March 31 - April 4 2003, vol. 1, pp. 131–134.

-
- [222] Y. Wu, J.A. O'Sullivan, and R.S. Inceck, "Iterative Detection and Decoding for Separable Two-Dimensional Intersymbol Interference", *IEEE Trans. Magn.*, vol. 39, no. 4, pp. 2115-2120, July 2003.
 - [223] A.M. van der Lee, D. Bruls, C. Busch, A. H. J. Immink, W. M. J. Coene, and A. P. Hekstra, "Two-Dimensional Optical Storage", *Jpn. J. Appl. Phys.*, vol. 43, no. 7B, pp. 4912-4914, 2004.
 - [224] J.W.M. Bergmans, "Effect of Loop Delay on Stability of Discrete-Time PLL", *IEEE Trans. Circuits Syst. I*, vol. 42, no. 4, pp. 229-231, Apr. 1995.

Acknowledgment

In the first place I would like to express my heartfelt thanks to my promotor Prof. Jan Bergmans. Without his support, encouragements and the enlightening discussions it would have been impossible to finish this work. His enthusiasm and deep insight into the field of signal processing for data storage inspired me greatly. He reserved a lot of valuable time for explanations and reviews of documents. Furthermore, I also would like to thank him for giving me the opportunity to start with this Ph.D and to work on the TwoDOS project.

I would like to thank to my colleagues in the TwoDOS room at the TU/e for the numerous exchanges of thoughts and ideas about both optical storage as 'real-world' problems. Jamal Riani is thanked for the pleasant cooperation and the exhaustive discussions. With his mathematical skills and theoretical insights, he contributed to many parts of this work and to the TwoDOS project. Furthermore, I would like to thank Massimo Ciacci for his inspiring enthusiasm and the practical contributions to this work. The tracking of parameter variations within the TwoDOS receiver are entirely based on his work and also other characterization results are inspired by the work of Massimo. Ali Nowbahkt is also for the creation of a nice atmosphere in the TwoDOS room and for his work on characterization which is in part incorporated in Chapter 3. These colleagues, together with other people at TU/e, created a nice working environment and are warmly appreciated for that.

The first years of my Ph.D were characterized by the enlightening and inspiring cooperation with numerous people at the Natlab of Philips. First of all, I would like to gratefully thank Andre Immink for his guidance, patience, encouragements and the enriching discussions. His enthusiasm towards research work has inspired me greatly. Further, I would like to thank all the people involved in the TwoDOS project for their cooperation and support, especially dr. W.M.J. Coene, the TwoDOS project manager, for the guidance of this innovative project. Dominique Bruls and Alexander van der Lee, who built the experimental set-up, are thanked for the numerous experimental results on which substantial parts of this thesis are based. Appreciation also goes to the other people involved in the European TwoDOS team with people from Philips ODTG, University of Limerick, Lancaster University and HW Communications.

Finally, very important to me, I would like to thank the people in my personal life: parents, brothers and friends for their love and encouragements; especially, thanks to Lotte for the continuous support.

Curriculum Vitae

Steven Van Beneden was born in Deurne, Belgium, on November 10, 1978. He received his high-school education at Sint-Jan Berchmans college in Malle, Belgium. In 1996, he started the study in electrical engineering at the Katholieke Universiteit Leuven, where he graduated (with distinction) in 2001. For his master thesis with the title “Counteracting of nonlinearities in DSL modems” he did an internship at Alcatel Research in Antwerp in the summer of 2000.

In 2001, he started a Master of technological design of ICT-based systems at University of Technology Eindhoven, The Netherlands under the supervision of Prof. Bergmans. The study included a 1 year internship at Philips Research laboratories (Nat.Lab). For his Master’s thesis, “Design of an efficient evaluation environment for TwoDOS”, he received a nomination for the UFE/OCE award for best thesis. After the Master of technological design, the invitation to continue the work within the TwoDOS project, resulted in the start of the Ph.D work, again under the supervision of Prof. Bergmans.

In December 2006, he started working at Fortis in Brussels, Belgium as a quantitative analyst at the fixed income trading desk, where he develops interest rate models and algorithms for quantitative trading.

