MODELING AND DETECTION FOR HOLOGRAPHIC DATA STORAGE

SEYED IMAN MOSSAVAT

(MSc, Sharif University of Technology)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2007

To my mother...

For her love, patience and support...

Acknowledgements

I am truly indebted to several people for their help during my research and in the preparation of this thesis. Many thanks go to my first supervisor, Dr. George Mathew, for his guidance and support during the first year of my research. I am also indebted to Dr. Nallanathan Arumugam, for supervising me during the second year of my research.

I am most grateful to Dr. Chun, my co-supervisor in Data Storage Institute (DSI). He has been a great help during my research there. Many thanks go to my colleagues and fellow students in DSI. In particular, I learned a lot from the fruitful discussions with Ashwin Kumar and Fabian.

I am deeply indebted to the Agency for Science, Technology and Research (A*STAR) for the award of the international graduate scholarship. In particular, I would like to thank Dr. Mirarefi of University of Illinois at Urbana-Champaign who introduced me and several other Iranian students to the vibrant educational environment of Singapore. He was also a big support during my studies in Singapore. I would like to thank the staff members of the Graduate Studies Office and the Electrical and Computer Engineering department of the National University of Singapore as well as staff members of DSI who helped me in one way or another.

I would like to express my heartfelt gratitude to Professor Bergmans, head of the Signal Processing Systems group in Technical University of Eindhoven, the Netherlands, for his invaluable help regarding the preparation of this thesis. I would like to extend my gratitude to my PhD supervisor, Dr. Bert de Vries, who offered me his kind support when I arrived in the Netherlands and during the several months I was preparing this thesis.

Last but not least, I would like to thank my mother for her endless love and support.

S. Iman Mossavat

July 2007

Contents

Summary

Conventional storage technologies such as hard disk drive, compact disc, digital versatile disk, and blu-ray disc rely on the track-based paradigm i.e. they store information along tracks that are well separated in order to eliminate inter-track interference. This storage paradigm is two-dimensional (2-D); however it uses the second dimension only loosely. Holographic data storage (HDS), on the other hand, breaks the density bottleneck of conventional storage technologies by utilizing the page-oriented paradigm that stores information in the form of 2-D holograms. Vast storage densities are achievable by multiplexing several holograms throughout the volume of the media. In addition, the page-oriented nature of HDS allows for high data rates by retrieving the entire hologram with a single flash of light. Thus, HDS is a promising technology for the increasing demands of information systems. In

this thesis, we study the signal processing aspects related to HDS.

Signal processing techniques are commonly used to meet the stringent requirements on data reliability in storage systems. Typical examples of signal processing algorithms are equalizers, detectors, modulation codes, and error correction codes. From the signal processing perspective, HDS has two key attributes that distinguish it from conventional storage technologies. The first attribute is the page-oriented nature of the HDS which results in higher computational complexities for signal processing algorithms as well as for modeling the HDS channel. Furthermore, there is no natural ordering in a 2-D page; thus it is difficult to generalize a major class of signal processing algorithms that rely on the sequential nature of the data in the track-based storage paradigm. The second attribute that requires attention in the design of signal processing algorithms is the nonlinear nature of the HDS. This feature introduces additional complexity to analysis and modeling of the HDS channel. Furthermore, the signal processing algorithm designer should consider the channel nonlinearity to achieve better performance. In this thesis, we address both attributes in modeling and detection for the HDS.

Design of signal processing algorithms for the HDS heavily relies on accurate channel modeling. In addition, extensive simulations are usually needed to evaluate the performance of such algorithms; thus, the computational complexity of the channel model is important. Because of the channel nonlinearity, linear models are not capable of describing the channel accurately. Various researchers worked

on nonlinear channel models for the HDS. While accurate models for the HDS are available; their computational complexities are prohibitively high. On the other hand, a notably efficient channel model for the HDS exists. This channel model, called the discrete magnitude-squared channel model (DMC), is obtained through exploiting the mathematical structure of the discretization of the channel. In spite of its complexity reduction, this model does not address the optical noise accurately.

In our work, we exploit the band-limited nature of the optical noise to develop accurate models with reduced computational complexities. Furthermore, we point out a flaw in the statistical analysis of the post-detector effects of the optical noise. Our simulations show consistency with the corrected statistical analysis.

In an ideal HDS system, the page that represents the information during data recording phase should be spatially matched with the detector page used in data retrieval phase. In practice, page translation error is inevitable, and it results in misalignment of pixels on these two pages. Pixel misalignment severely deteriorates data reliability in the HDS systems. In our work, we have extended the DMC model for systems with pixel misalignment.

Several researchers investigated equalization and detection techniques for the HDS. The nonlinear nature of the HDS and the absence of natural ordering in the page-oriented storage paradigm, add a great deal of complexity to this task. In

particular, various researchers studied the extension of the one-dimensional Viterbi algorithm for the page-oriented HDS channel. While most solutions are suffering from high complexities, none of them addresses the absence of the natural ordering effectively.

Various linear and nonlinear hard-decision detectors are developed for the HDS channel; whereas little afford has been devoted to developing soft-decision detectors for the HDS channel. The complexity of the existing soft-decision detector is prohibitively high. Soft-decision detectors are of fundamental importance because of their integral role in iterative reception schemes with near-optimal bit-error-rate performance.

In order to address the aforementioned issues we designed a reception scheme that fits the characteristics of the nonlinear HDS channel. Our reception scheme is based on the soft-decision BCJR detector designed by extending an existing reduced-complexity BCJR detector for linear 2-D channels. Exploiting the HDS channel structure, our detector tackles the absence of natural ordering by breaking down the 2-D detection across the page into one-dimensional detection along columns and rows of the page. The complexity of our scheme is much less than the existing soft-decision detector for the HDS channels. We propose a novel partial response signal to limit the complexity further. As an added advantage, our scheme can handle high levels of pixel misalignment.

List of Figures

Δ_s	SLM pixel width
Δ_c	CCD pixel width
α	SLM linear fill factor
β	CCD linear fill factor
ϵ	amplitude contrast ratio
λ	wavelength
ω	normalized aperture width
δ_x	pixel misalignment along axis 'x'
D_N	Nyquist aperture width
f_L	focal length
$\{d_{i,j}\}$	binary input data page
$p(x,y)$	SLM pixel shape
$s(x,y)$	SLM output wavefront
$h_A(x,y)$	aperture impulse response
$h(x,y)$	optical PSF
$n^o(x,y)$	optical noise wavefront
$r(x,y)$	information-bearing wavefront
$z(x,y)$	incident wavefront
$I_{i,j}$	corresponding CCD output

$K_{n^o}(x, y)$	optical noise auto-covariance
$P(x)$	probability of x
$\mathbf{M}_{:,j}$	the j-th column of matrix $\mathbf{M}$
$\mathbf{M}_{i,:}$	the i-th row of matrix $\mathbf{M}$
$n^e_{i,j}$	electronics noise
σ_n^2	electronics noise variance
$\mathbf{H}^{\delta_x,\delta_y}$	discrete channel matrix corresponding to misalignments δ_x and δ_y in 'x' and 'y' directions, respectively
$y_{i,j}$	intermediate signal
$L(y_{i,j})$	LLR values for the signal $y_{i,j}$
L^c	superscript 'c' corresponds to column detector
L^r	superscript 'r' corresponds to row detector
L^E	superscript 'E' corresponds to extrinsic information
$\mathcal{S}$	set of state transition on the binary trellis
$\mathcal{S}^+$	set of state transitions caused by $+1$ on the row detector (binary) trellis
$\mathcal{S}^-$	set of state transitions caused by -1 on the row detector (binary) trellis
$\hat{\mathcal{S}}_Y$	set of possible state transitions on the column detector trellis corresponding to Y

α^c	BCJR forward recursion for column detector
α^r	BCJR forward recursion for row detector
β^c	BCJR backward recursion for column detector
β^r	BCJR backward recursion for row detector
γ^c	BCJR branch transition probability for column detector
γ^r	BCJR branch transition probability for row detector
$z(\hat{s}, s)$	noiseless channel output corresponding to state transition $(\hat{s}, s)$
$s_{i,j}$	equalization target signal
$\widetilde{\gamma}_{i,j}$	target coefficients
$\mathbf{\Gamma}$	matrix representation of target coefficients
$\mathbf{\Lambda}$	vector representation of target coefficients
$\mathbf{c}$	vector representation of equalizer coefficients
$\tilde{s}_{i,j}$	equalizer output
$\mathbf{I}$	equalizer input
$\mathbf{d}$	target input data pattern
ξ	MSE
$2L+1$	width of the IPI support
$2Q+1$	width of the equalizer support

1-D	one-dimensional
2-D	two-dimensional
$4-\mathrm{F_L}$	4-focal-length
APP	a-posteriori probability
AWGN	additive white Gaussian noise
BCJR	Bahl-Cocke-Jelinek-Raviv
BER	bit-error-rate
BSA	Bragg selectivity angle
CCD	charge-coupled-device
CD	compact disk
CT	channel truncation
DCM	discrete channel matrix
DF	decision-feedback
DMC	discrete magnitude-squared channel
DVD	digital versatile disk
ENDC	electronics-noise dominated channel
HDD	hard disk drive
IMSDFE	iterative magnitude-squared DF equalizer
HDS	holographic data storage

IPI	inter-pixel interference
ISI	inter-symbol interference
LDPC	low-density parity-check
LLR	log-likelihood ratio
LMMSE	linear MMSE
LUT	look-up-table
MAP	maximum-a-posteriori
ML	maximum-likelihood
MMSE	minimum mean-square equalization
MSE	mean-square error
NMSE	normalized MSE
PDFE	pseudo DF equalizer
PR	partial-response
PSF	point spread function
QMMSE	quadratic MMSE
SNR	signal-to-noise ratio
SLM	spatial light modulator

Chapter 1

Introduction

1.1 Motivations for Holographic Data Storage

Data storage systems play an integral role in the advances of the information era. Various technologies have been developed to answer diverse needs of various consumers such as entertainment industries, on-line storage service providers, and medical systems. Magnetic storage systems such as hard disk drive (HDD) mostly target for high densities, whereas optical storage such as compact disk (CD) and digital versatile disk (DVD) provide removable storage. Density and data rates of data storage systems grow rapidly in response to increasing demands of information technology. However, as a result of fundamental physical limitations, it is not clear whether the current storage technologies are able to sustain their density growth.

One of the major limitations of current technologies is that they all use a two-dimensional (2-D) recording/retrieval paradigm in which data is stored along well separated tracks. The spacing among tracks limits the achievable density. One solution for breaking the density bottleneck suffered by current storage technologies (i.e. CDs, DVDs, and HDDs) is to break the track-based paradigm. Holographic data storage (HDS) is an example of a non-track-based paradigm. HDS systems store information in the form of 2-D holograms. We will see later that several holograms can be recorded throughout the volume of the recording media allowing for ultra-high densities to be achieved. In addition, the entire hologram is recorded / retrieved with a single flash of light allowing for high data rates to be achieved.

In this work we investigate signal processing aspects of the HDS. Signal processing played a crucial role in achieving reliable storage at high densities. One can view the data storage channel as a noisy communication channel, where retrieving information is prone to errors. In almost all scenarios, errors are more likely to happen when storage density increases. On the other hand, data storage has stringent reliability requirements and the challenge for signal processing is to reduce the storage bit-error-rate (BER) to an acceptable level (usually around 10^{-12}) while achieving high density and high data rates.

We first give a brief description of the HDS and then proceed to survey the existing literature on modeling and detection for the HDS to motivate the research

reported in this thesis. In the end we conclude by a summary of main contributions and the organization of the thesis.

1.2 Introduction to Holographic Data Storage

Holographic data storage (HDS) stores information in the form of 2-D holograms. In order to have a clearer picture we give a basic account of holography. In holography two coherent laser beams are used. One is called the object beam and carries the information (light from the scene or object), and the other is the reference beam. These two beams interfere and the interference pattern which is stored in some photosensitive medium is called the *hologram*. Now if the stored hologram is illuminated with the reference beam used during recording, the object beam is reconstructed (with a loss in signal power). In HDS each hologram corresponds to one page of data. So we call HDS a *page-oriented* data storage technology. By virtue of a phenomenon called Bragg selectivity [1], several holograms can be multiplexed throughout the same storage medium by changing some attribute of the reference beam such as its wavelength or angle, resulting in wavelength multiplexing or angular multiplexing respectively. Other multiplexing techniques for the HDS are phase-code multiplexing, shift multiplexing, or spatial multiplexing [1]. Multiplexing makes HDS a *volumetric* data storage technology.

After describing the basic concepts used in HDS. We proceed to introduce a

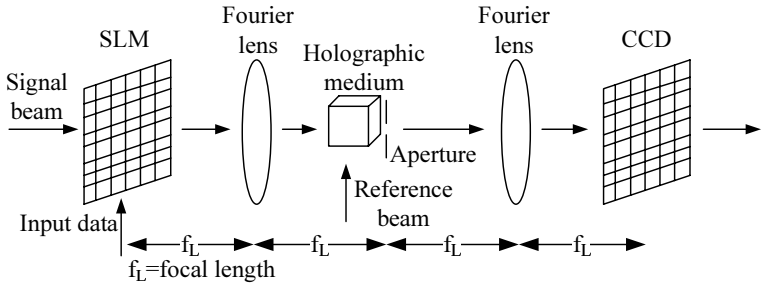


Figure 1.1: Schematic of the holographic data storage (In the 4-focal-length architecture)

typical HDS architecture with *angular multiplexing* called the 4-focal-length ($4-f_L$) architecture [2]. As Figure 1.1 illustrates, in the recording phase data bits are presented by a device called spatial light modulator (SLM) which modulates the amplitude of the object beam. The object beam passes through a lens and the Fourier transform of the SLM image appears on the focal plain (Fourier Plain) of the lens. An aperture, which is placed at the center of the focal plain of the lens, passes the low frequency content of the SLM image and rejects the high frequency portion. This filtered beam reaches the recording medium, where it interferes with the reference beam. The recording medium stores the interference pattern. During data retrieval, the hologram is illuminated with the same reference beam used during the recording phase and the object beam is reconstructed. Each hologram can be randomly accessed by changing the angle of the reference beam. This reconstructed image is inverse Fourier transformed by the second lens. Finally a detector array such as charge-coupled-device (CCD) converts the information

bearing object beam to electronic signal.

Since HDS is a volumetric and page-oriented data storage technology, it is potentially capable of achieving high storage densities and high data rate parallel recording and retrieval of information. Theoretically, densities up to $1/\lambda^3$ are possible for a laser light of wavelength λ [3]. Recently InPhase technologies has demonstrated a HDS system with 500 Gbit/in^2 with a write user rate of 23 MBytes /sec and a read user rate of 13 MBytes /sec [4].

1.3 Holographic Data Storage Channel

Any data storage channel can be viewed as an imperfect communication channel which is susceptible to information loss due to noise, finite bandwidth, and nonlinear distortions. In this section we give a qualitative account of the most salient characteristics of the HDS systems that influence the system fidelity. A quantitative model for HDS systems is presented in Chapter 2.

HDS systems may suffer from a vast array of impairments [5]. However, in order to manage the model complexity, most channel models limit their scope to the dominant channel impairments Viz. crosstalk and noise. Few models consider pixel misalignment (page translation error) in addition to aforementioned impairments. We qualitatively describe these impairments here; A quantitative channel model is given in Chapter 3.

Since HDS is a page-oriented, volumetric storage technology, two types of crosstalk (interference) may arise. The page-oriented nature of holographic data storage leads to intra-page crosstalk, which we refer to as 2-D inter-pixel interference (IPI). On the other hand, the volumetric nature of the HDS which allows for several holograms (pages) to be recorded in one medium may give rise to inter-page interference. Most channel models focus on 2-D IPI, assuming that inter-page interference is negligible [6, 7, 8, 9]. Two-dimensional IPI is the result of filtering the high frequency portion of the SLM image by the aperture. This is the counterpart of the one-dimensional (1-D) inter-symbol interference (ISI) encountered in conventional storage technologies. We will discuss the 2-D IPI in more detail in Section 1.3.1.

Noise sources in the HDS are the optical noise (scatter, laser speckle) and the electronics noise [10]. Optical noise is modeled as a stationary complex-valued circularly symmetric colored Gaussian noise and electronics noise is simply modeled as a real-valued white Gaussian noise. While modeling electronics noise is straightforward, modeling optical noise needs more care. We will look into this issue in more detail later.

The signal power of the replica of the reconstructed object beam in the data retrieval phase is inversely proportional to the square of the number of multiplexed holograms [11]. Hence, increasing the storage density by increasing the number of

holograms leads to low signal-to-noise ratio (SNR). This is an important observation as detector bit-error-rate (BER) relies heavily on the available SNR.

Another major difficulty of the HDS is its quadratic nonlinearity. This specific type of nonlinearity stems from the fact that the SLM modulates the amplitude of the object beam, while the CCD detects the intensity of the reconstructed object beam over its pixels area. While some models assume a linear channel for the HDS [2, 12], more accurate models such as those of Chugg et al. [9] and Keskinoz and Kumar [8] incorporate a quadratic nonlinearity. Chugg et al. [9] accurately model channel nonlinearity, however, their model is computationally demanding as it uses a 4-D kernel to compute the CCD read-back signal.

The discrete magnitude-squared channel (DMC) model of Keskinoz and Kumar [6, 7, 8] is of fundamental importance. In addition to being efficient, their noise-free channel modeling provides us with useful insights on the mathematical structure of the HDS channel. Most importantly, they show that one can view the information-bearing component of the detector read-back signal as the total response of a bank of magnitude-squared sub-channels, where each magnitude-squared sub-channel consists of a 2-D *separable* discrete linear time-invariant channel followed by the magnitude square operation. They call their model discrete magnitude-squared channel (DMC) model. The separability property of the underlying 2-D linear channel means that the 2-D IPI introduced by this linear channel can be viewed as originating from a concatenation of two 1-D linear ISI channels.

Furthermore, by utilizing principal component analysis, Keskinoz and Kumar approximate the model by one dominant sub-channel. This approximation permits a compromise between complexity and accuracy in their model.

Although the DMC model greatly reduces the model complexity and provides us with useful insights, it does not address optical noise accurately. He [13] presents a more accurate treatment of the optical noise while using techniques of [8] to maintain efficiency. Yet, He's [13] model relies on the assumption that the optical noise power is low. Furthermore, despite the fact that the optical noise is not white, He's [13] model does not capture the correlation characteristics of the post-detector effect of the optical noise for the sake of achieving computational efficiency.

A further important challenge in HDS is the pixel misalignment. If we assume an equal number of pixels on SLM and CCD, it is ideal for the pixels on these two devices to be spatially matched, i.e. for each pixel on the CCD to be exactly in front of the corresponding SLM pixel. In practice it is impossible to achieve perfect pixel alignment due to a variety of adverse factors. As [11] reports, the effect of pixel misalignment on BER is substantial. Yet, this has not received enough attention in HDS channel modeling and most models take perfect pixel alignment between SLM and CCD as granted. One exception is the work of Heanue et al. [12] that accommodates pixel misalignment into their *linear* model of the HDS. Menetrier and Burr [11] investigate pixel misalignment more precisely by using a numerical approach based on fast Fourier transforms as in Bernal et al [14]. However, they

give no detailed method for simulating channels with pixel misalignment.

Let us summarize our brief survey on HDS channel models. The HDS is a quadratic, 2-D IPI channel. Existing HDS models capture quadratic nonlinearity but they are either too computationally demanding or need further accuracy improvements regarding the optical noise. None of the nonlinear HDS models captures pixel misalignment despite its importance. Another observation about HDS channels is that using smaller aperture width or multiplexing more holograms result in increased storage density; however they lead to higher IPI and lower SNR respectively.

1.3.1 Detection for Holographic Data Storage

Before presenting our survey on signal processing techniques for holographic data storage, let us define the notion of *reception scheme*. A reception scheme takes as input the CCD read-back signal and makes decisions on the transmitted / stored data bits. Important characteristics of a reception scheme are its complexity and bit-error-rate (BER). Reception schemes may have several components such as equalizers, detectors, and error correction decoders. The simplest reception scheme is a slicer or a threshold detector which decides whether a bit is zero or one by comparing the CCD read-back signal against a threshold. This reception scheme is very simple but usually fails to provide acceptable BER if ISI is present. A more

complicated reception scheme is made by equalization of the CCD signal before the detector to remove or reduce ISI. More sophisticated reception schemes use maximum-likelihood (ML) or maximum-a-posteriori (MAP) detectors to further enhance the BER. Decoders are other possible components of a reception scheme which exploit a known structure embedded earlier in the data bits by an error correction code to detect / correct errors.

Before proceeding to review the existing detection schemes, let us look into the 2-D IPI more closely. Two-dimensional interference is fundamentally different from 1-D interference because of two issues: for an ISI span of L, the number of interfering symbols in 1-D ISI is L, whereas in 2-D IPI the number of interfering pixels is L^2. This potentially leads to higher modeling and detection complexities. However, what makes 2-D interference fundamentally different is the fact that there is no natural ordering of data in two dimensions. This natural ordering is of fundamental importance in designing ML or MAP detectors. MAP detectors provide optimal BER performance, i.e. they yield the best BER performance possible for a given SNR. If there is no prior information about the data bits to be detected, then ML detectors are optimal too. Complexity of ML detection is much lower than that of MAP detection. The Viterbi algorithm is an example of a ML detector and the Bahl-Cocke-Jelinek-Raviv (BCJR) algorithm is an example of a MAP detector. Both detectors are originally developed for 1-D channels and exploit the natural ordering of the data. By comparison, in the 2-D IPI case no

such ordering exists, which hampers the generalization of the widely used Viterbi or BCJR detectors.

In most reception schemes, linear equalizers are probably the most common type of equalizers used to mitigate ISI. A common criterion for designing equalizers is minimization of the mean-square error (MSE). In minimum mean-square error (MMSE) equalization the equalizer is designed such that the energy of the error between the equalizer output and some target signal is minimized. If the target signal is designed such that the interference is eliminated completely, we have *full response* equalization. If the target signal is designed such that a controlled amount of interference is permitted, we have *partial response* equalization. A common practice is to choose the input data bits as the target signal, resulting in an equalizer that produces an output which is as similar as possible (in the MSE sense) to the input data bits. We refer to such targets as *linear full response* targets. Since the common full response MMSE with fixed equalizer coefficients does not rely on natural ordering of the data, it is straightforward to extend such 1-D equalizers to their 2-D counterparts. Chugg et al. [9] and Keskinoz and Kumar [15] investigate the design and performance of linear minimum mean-square error (LMMSE) equalizers with linear targets for the 2-D IPI in HDS systems. When combined with a threshold detector, their results show that LMMSE improves the BER performance that is limited by an error floor at high SNRs.

He and Mathew [16] designed a low-complexity quadratic MMSE (QMMSE)

equalizer. They also investigated the effectiveness of full response equalization that uses nonlinear (quadratic) transformation of data bits to construct the target signal. QMMSE significantly improves the BER performance over LMMSE and the error floor problem is effectively improved. However, the specific target signal that they design does not bring significant BER improvement.

Devising a decision-feedback (DF) loop in reception schemes is a common technique. For HDS systems, two major designs are the pseudodecision-feedback equalizer (PDFE) [17] and the iterative magnitude-squared decision-feedback equalizer (IMSDFE) [7, 8]. The design strategies behind both schemes are similar. First, initial decisions for the data bits are computed. Later these decisions are refined iteratively by computing the interference of neighboring bits on each data bit using the knowledge of the channel nonlinear characteristics. Both PDFE and IMSDFE provide superior BER performance in comparison with LMMSE based reception schemes as they incorporate the knowledge of the channel nonlinearity in their structure. However, both schemes break down if there are too many errors in the initial decisions as a result of error propagations; hence, they require a relatively large SNR.

Several researchers investigated Viterbi-based detection for HDS. Heanue et al. [12] designed a reception scheme for HDS based on a 1-D Viterbi algorithm. The complexity of their detector grows exponentially with $2L^2 + 3L + 1$ for an IPI span of $(2L+1) \times (2L+1)$. For a moderate IPI span of 3×3 i.e $L = 1$, the detector

has $2^6 = 64$ states while an IPI span of 5×5 i.e. $L = 2$, leads to a detector with $2^{15} = 32768$ states! For an IPI span of 3×3, they use DF to reduce the number of states to $2^4 = 16$. Their scheme operates the Viterbi algorithm on a row-by-row basis. This means that in order to detect a certain row, they assume that the upper row is known (or correctly detected). Using DF makes their algorithm susceptible to error propagation. In addition, row-by-row detection is not optimal; hence they have not effectively addressed the problem of 2-D IPI.

Instead of using DF, some researchers opt for partial-response (PR) equalization to reduce the IPI span. The general principle behind all these designs is to equalize the data page in order to eliminate the IPI along one direction while using Viterbi along the other direction. PR equalization allows for a controlled amount of IPI and requires less equalization effort, which in turn results in less noise coloring. Since Viterbi detection is based on the white noise assumption, noise coloring has an adverse impact on the BER performance of the Viterbi detection. Reception schemes based on 1-D PR equalization and Viterbi detection for HDS are investigated by Vadde and Kumar [18]. In their scheme, first they apply a zero-forcing equalizer to eliminate IPI along one dimension. Then the 1-D PR equalization is used along the other dimension to reduce the span of ISI in order to further control the complexity of the following Viterbi detection. Two-dimensional quadratic PR equalization was investigated by He and Mathew [16]. They formulate the optimality criterion for PR quadratic-equalization target which results in

2 dB performance gain over their QMMSE scheme.

All previous algorithms are hard-decision algorithms, i.e. they decide on the the bit values. Soft-decision detectors, on the other hand, estimate the probability that a bit is one (or zero). A notable soft-decision detector for HDS is the one of [19] which works based on the same principles as turbo decoding [20, 21]. More specifically, the likelihood of each data bit is updated based on the likelihood of its neighboring data bits. The likelihood information propagates throughout the 2-D page with iterations. Their algorithm is designed such that it allows for parallel implementation. However, the complexity of the algorithm of [19] tends to be very high: for an IPI span of $L \times L$, the detector complexity is exponential in L^2.

In [22] a reduced-complexity BCJR detector for a specific class of linear 2-D channels was described. As already mentioned, BCJR is an optimal symbol-by-symbol MAP detector that produces soft decisions. For an ISI span of $L \times L$ the detector complexity in [22] is exponential in L. The complexity reduction of [22] applies to linear channels that are separable, i.e. for which the 2-D ISI can be viewed as a concatenation of ISI along the rows and the ISI along the columns. Their scheme effectively solves the problem of natural ordering in 2-D data detection for a specific class of linear channels.

1.4 Motivations and Main Contributions

From the previous survey of channel models for HDS, we conclude that the current efficient models do not treat the optical noise accurately. Since the optical noise is not white and the detector array is nonlinear (quadratic), computing the post-detector effects of the optical noise is not straightforward. We devote a major portion of our research on developing an accurate, yet computationally efficient model for the post-detector effects of the optical noise in the HDS channel.

Another contribution of our research is extending the DMC model originally developed for pixel-aligned HDS channels for the case of pixel misaligned HDS channels. Recall that pixel misalignment severely deteriorates BER performance [11]. However, little effort has been devoted to incorporate pixel misalignment into HDS channel models. On the other hand, such misalignments almost always exist. Hence, there is value in incorporating this impairment into HDS channel models.

We observe that the DMC model (and its extension for pixel misaligned channels) has the separability property. We use this property to effectively address the issue of 2-D IPI by extending the soft-decision BCJR detector of [22] originally developed for linear separable channels. The complexity of our extension is comparable to [22] and far smaller than the current soft-decision detector of [19]. In order to further reduce the complexity, we have designed a new PR target signal that utilizes a quadratic nonlinearity in its structure and derived an analytical expression

for the corresponding optimal detector. However, finding the optimal quadratic PR target remains as a seemingly difficult open problem. Hence, we propose a candidate for the target signal which seems to work with comparable performance to the best target signal found by brute force search. Reception schemes that utilize the quadratic target signal for channels with high misalignment level tend to achieve an acceptable BER performance whereas other reception schemes fail to provide acceptable BER performance.

1.5 Outline of the Thesis

The rest of the thesis is organized as follows. We study the $4-\mathrm{f_L}$ architecture and the BCJR algorithm in more detail in Chapter 2. We dedicate Chapter 3 to the accurate modeling of the HDS channels. In Chapter 4 our soft-decision nonlinear 2-D reception scheme along with the proposed quadratic target signal is described. Chapter 5 concludes the report with some comments on possible directions for further work.

Chapter 2

Preliminaries

In Chapter 1 we presented an overall view of the holographic data storage (HDS). In this chapter, we give a more detailed account of the HDS system. We introduce the 4-focal-length ($4-\mathrm{F_L}$) architecture in Section 2.1. We will derive an accurate model for this architecture in Chapter 3. In Section 2.2 we present the background on 1-D and 2-D MAP detection. The concepts presented in Section 2.2 enable us to describe our quadratic 2-D BCJR detector in Chapter 4.

2.1 Holographic Data Storage Systems

2.1.1 System Architecture

The 4-focal-length ($4-\mathrm{F_L}$) architecture [2] is a commonly used architecture in practical HDS systems. We dedicate this section to describing the system architecture.

The interested reader may refer to [1] for more information on the system components. Furthermore, a detailed explanation of the underlying optical principles is given in [23].

Figure 1.1 illustrates the $4 - \mathrm{F_L}$ architecture. A coherent source is used to provide the object, the reference, and the playback laser beams. The spatial light modulator (SLM) represents the data bits as a 2-D checkerboard pattern of dark and bright pixels which either blocks or permits the object beam. One can view the SLM as a 2-D matrix of miniature shutters that are controlled by the data bits and creates an ON/OFF pattern [1]. Each of the two lenses in the $4 - \mathrm{F_L}$ architecture performs the Fourier transform operation [23]: if an object is placed in the front focal plane of the lens, the Fourier transform of the object's image is formed on the rear focal plane. As Figure 1.1 shows, the SLM is placed on the front focal plane of the first (left-hand side) Fourier lens. The two lenses are placed at the distance equal to twice their focal lengths such that the rear focal plane of the first lens coincides with the front focal plane of the second (right-hand side) Fourier lens. The detector array is placed on the rear focal plane of the second Fourier lens. This configuration allows for SLM to be exactly imaged on the detector array. The storage medium is placed prior to the first Fourier lens focal plane. Consequently, it is the Fourier transform of the SLM image that is stored as a hologram in the storage medium. The advantage of storing the Fourier transforms is the improved resilience to burst errors because each data bit is distributed over

the entire page by the Fourier transform [2]. The most widely used detector array in HDS systems is the charge-coupled-device (CCD) detector which is an array of coupled capacitors [1]. The CCD pixels integrate the intensity of the laser light temporally and spatially over their area and generate a varying voltage which is used as the read-back signal.

During the recording phase, the SLM impresses the data on the object beam. The Fourier transform of the SLM image is formed on the focal plane of the first lens. The object beam then interferes with another beam called the reference beam inside the storage medium which is placed just before the focal plane, and the interference pattern is recorded inside the crystal by some chemical or physical change in the photosensitive medium [1].

During the retrieval phase, the system may access any page by illuminating the storage medium with the reference beam that was used to record that page. The reference beam deflects off the corresponding hologram, thus the object beam is reconstructed. The object beam passes through the aperture which is placed on the center of the focal plane of the first lens thus permitting the lower frequencies (and rejecting the higher frequencies) of the SLM image. The second lens then reconstructs the low-pass filtered replica of the SLM image on its rear focal plane where the CCD is placed.

Holograms are grouped into stacks. In each stack, several holograms are multiplexed within the same volume of the medium. Holograms are randomly addressed by using their corresponding reference beam during retrieval. Various multiplexing techniques are available. One of the most widely used multiplexing techniques is the angle multiplexing which works based on a physical phenomenon called the Bragg effect [1]. In angle multiplexing, multiple holograms are stored by changing the angle between the object beam and a reference beam. To prevent crosstalk among holograms stored at the same location, each reference beam should be separated by the Bragg selectivity angle (BSA) [24, 25]. The BSA of the recording medium is a function of thickness of the recording medium, among other things such as desirable signal-to-noise ratio (SNR). In particular, the BSA decreases as the thickness of the storage medium increases.

The number of holograms that can be multiplexed in the same location directly influences the achievable density; hence, a great deal of research is dedicated to developing efficient multiplexing techniques such as peristrophic multiplexing [26], shift multiplexing [27],wavelength multiplexing [28], and phase-code multiplexing [29].

The role of the aperture is to limit the inter-stack interference and to block the scattered light. The aperture is placed at the center of the rear focal plane of the first Fourier lens in order to minimize the blockage of useful signal (i.e. achieve minimum intra-page crosstalk) [2]. The aperture introduces intra-page crosstalk or

IPI by rejecting the high-frequency content of the image. A small aperture allows for higher storage density at the cost of higher IPI.

2.1.2 Density Limitations

Various factors influence the achievable density. In simple terms, we can increase the density by increasing the number of data bits in each hologram, the number of holograms per stack, or the number of stacks per unit volume. Increasing the number of data bits per hologram requires high precision optics and reduces the system tolerance to misalignments and optical aberrations. Increasing the number of holograms per stack decreases the SNR since the efficiency of the diffraction is inversely proportional to the square of the number of multiplexed holograms [1]. Increasing the number of stacks per unit volume requires using smaller apertures and increases the IPI. As the IPI increases, the required SNR for achieving reliable data storage and retrieval also increases. This limits the number of holograms that can be multiplexed within the same location and reduces the achievable density. Hence, there is a trade-off in choosing the aperture size. Usually HDS systems choose a Nyquist aperture D_N

$$D_N = \lambda f_L / \Delta, \tag{2.1}$$

where Δ denotes the pixel width of the SLM, λ is the wavelength of the laser light, and f_L is the lens's focal length.

2.2 Soft-Decision Detection

The discussion in Section 2.1.2 implies that as density increases the reception scheme should handle higher IPI while less SNR is available. In addition, the nonlinear nature of the HDS calls for more sophisticated reception schemes, thus iterative reception schemes based on joint detection and decoding are interesting options for the HDS because of their near-capacity performance for linear 1-D channels. An integral part of such an iterative reception scheme is a soft-decision MAP detector. In this section, we first study the MAP detector for 1-D channels and then proceed to look into MAP detection for 2-D channels.

2.2.1 MAP Detection for 1-D Channels

The BCJR algorithm provides an efficient method to implement symbol-by-symbol MAP detection by exploiting the natural ordering of the symbols [30]. In this section, we look into the conventional 1-D BCJR detector in order to provide the necessary background for presenting the 2-D BCJR detector.

Figure 2.1 illustrates the channel model. Binary information $u \in \{\pm 1\}$ are transmitted through a band-limited noisy communication channel. The linear time invariant system denoted by h models the effect of the finite band-width i.e. the inter-symbol interference (ISI). Noise is modelled by a white Gaussian process n.

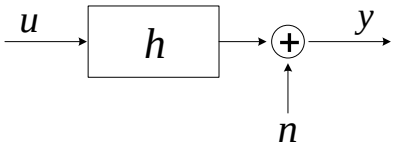


Figure 2.1: The one-dimensional, linear channel with additive white Gaussian noise

The received signal is denoted by y. Mathematically,

$$y_k = \sum_{i=0}^{L-1} h_i u_{k-i} + n_k, \tag{2.2}$$

where y_k is the k-th received signal, L is the channel memory length, u_k is the k-th bit, and n_k is the k-th noise value. We describe the BCJR algorithm following the tutorial paper [31]. The symbol-by-symbol MAP BCJR detector computes the log-likelihood ratio (LLR) of each data bit. Mathematically,

$$L(u_k) = \log\left(\frac{P(u_k = +1|\mathbf{y})}{P(u_k = -1|\mathbf{y})}\right), \tag{2.3}$$

where $\mathbf{y}$ is the entire observed sequence of received signal. Note that we are able to compute the probabilities $P(u_k = +1|\mathbf{y})$ and $P(u_k = -1|\mathbf{y})$ given $L(u_k)$ since

$$P(u_k = +1|\mathbf{y}) + P(u_k = -1|\mathbf{y}) = 1. \tag{2.4}$$

The BCJR detector relies on the trellis diagram to compute $L(u_k)$. The trellis diagram has 2^L states to represent the channel memory. For each possible state, there are two possible branches depending on the input data bit. We consider the

transition of the state from the $(k-1)$-th bit to the k-th bit. Using the Bayes rule, we can write Equation (2.3) as

$$L(u_k) = \log\left(\frac{\sum_{(\hat{s},s)\in\mathcal{S}^+} P(s_{k-1}=\hat{s}, s_k = s, \mathbf{y})}{\sum_{(\hat{s},s)\in\mathcal{S}^-} P(s_{k-1}=\hat{s}, s_k = s, \mathbf{y})}\right). \tag{2.5}$$

where we denote the set of all possible transitions by $\mathcal{S}$ and the state at the k-th bit by s_k. We partition $\mathcal{S}$ into two subsets at each k: $\mathcal{S}^+$ is the set of all the state transitions from s_{k-1} to s_k caused by $u_k = +1$ and $\mathcal{S}^-$ is the set of all the state transitions from s_{k-1} to s_k caused by $u_k = -1$. From now on, we refer to $P(s_{k-1}=\hat{s}, s_k = s, \mathbf{y})$ as $P(\hat{s}, s, \mathbf{y})$ in order to use a shorter notation. In order to compute $P(\hat{s}, s, \mathbf{y})$ for any possible transition $(\hat{s}, s) \in \mathcal{S}$, we apply the Bayes rule for the second time to get

$$P(\hat{s}, s, \mathbf{y}) = \alpha_{k-1}(\hat{s})\gamma_k(\hat{s}, s)\beta_k(s), \tag{2.6}$$

where

$$\alpha_{k-1}(\hat{s}) = P(s_{k-1}=\hat{s}, \mathbf{y}_1^{k-1}), \tag{2.7}$$

$$\gamma_k(\hat{s}, s) = P(s_k = s, y_k | s_{k-1}=\hat{s}), \tag{2.8}$$

$$\beta_k(s) = P(\mathbf{y}_{k+1}^{n} | s_k = s), \tag{2.9}$$

where $\mathbf{y}_a^b = (y_a, y_{a+1}, ..., y_b)$. Using the Bayes rule, we can compute these three terms recursively. In more detail,

$$\alpha_k(s) = \sum_{\hat{s}} \alpha_{k-1}(\hat{s})\gamma_k(\hat{s}, s), \tag{2.10}$$

with the initial conditions

$$\alpha_0(s=0)=1 \text{ , and } \alpha_0(s\neq 0)=0. \tag{2.11}$$

Similarly,

$$\beta_{k-1}(\hat{s})=\sum_{s}\beta_k(s)\gamma_k(\hat{s},s), \tag{2.12}$$

with the initial conditions

$$\beta_{n_{bits}}(s=0)=1 \text{ , and } \beta_{n_{bits}}(s\neq 0)=0, \tag{2.13}$$

where n_{bits} is the number of data bits in the sequence. The initial conditions model our knowledge that the first and the last data bits are zero. It remains to compute $\gamma_k(\hat{s},s)$, for which we have

$$\gamma_k(\hat{s},s)=P(s|\hat{s})P(y_k|\hat{s},s)=P(u_k)P(y_k|\hat{s},s), \tag{2.14}$$

where the value of u_k corresponds to the transition from $\hat{s}$ to s. The term $P(u_k)$ is the a priori probability of u_k. In iterative detection and the decoding schemes, this term allows for passing on information from the decoder to the detector. The term $P(y_k|\hat{s},s)$ is the conditional probability of y_k knowing the state transition $(\hat{s},s)$ occurred. This term depends on the channel characteristics, we will see in Chapter 4 that by using a similar term we are able to incorporate the knowledge of the HDS channel characteristics in our detector.

2.2.2 Maximum Likelihood Detection for 2-D Nonlinear Channels

The page-oriented nature of the HDS makes maximum likelihood (ML) ad MAP detection difficult since a natural ordering of pixels is absent. A conceptually simple page-wise optimal method is to generate a look-up-table (LUT) containing all possible data pages together with their corresponding expected retrieved page as corrupted by the finite band-width, nonlinear HDS channel. In practice, however, this method is infeasible as for page size of $N_{bits} \times N_{bits}$ the required LUT would contain $2^{N_{bits} \times N_{bits}}$ entries.

A computationally feasible method to approximate the likelihood of bits is given by [19] that does not assume the channel to be linear. We briefly describe their method here. Let us denote the data bits and the channel output, respectively, by $d_{i,j}$ and $\hat{I}_{i,j}$. The entire data page is denoted by $\mathbf{D} = [d_{i,j}]_{N_{bits} \times N_{bits}}$. Let the support $\mathbf{S}_{i,j}$ of the channel output $\hat{I}_{i,j}$ be the subset of input data bits such that $P(\hat{I}_{i,j}|\mathbf{S}_{i,j})$ can approximate $P(\hat{I}_{i,j}|\mathbf{D})$ with sufficient accuracy. Assume that the size of the support is $(2L+1) \times (2L+1)$. Also define the neighborhood of $d_{i,j}$ to be $\mathbf{N}_{i,j} = \mathbf{S}_{i,j} - \{d_{i,j}\}$. It is important to note that the $\mathbf{N}_{i,j}$ can take on one of $2^{(2L+1)^2-1}$ values in the space $\Omega = \{0,1\}^{(2L+1)^2-1}$. The following iterative likelihood

propagation rule is proposed by [19]:

$$L_U^{(k)}(d_{i,j}) = \sum_{\mathbf{N}_{i,j}\in\Omega} C_{i,j}(\mathbf{S}_{i,j}) \prod_{d_{l,m}\in\mathbf{N}_{i,j}} L^{(k-1)}(d_{l,m}) \tag{2.15}$$

$$L^{(k)}(d_{i,j}) = (1-\beta)L^{(k-1)}(d_{i,j}) + \beta L_U^{(k)}(d_{i,j}), \tag{2.16}$$

where $L^{(k)}(d_{i,j})$ is the estimate of the likelihood $P(\hat{I}_{i,j}|d_{i,j})$ at k-th iteration, and $C_{i,j}(\mathbf{S}_{i,j})$ is a set of $2^{(2L+1)^2}$ combining coefficients commonly chosen as $P(\hat{I}_{i,j}|\mathbf{S}_{i,j})$ which is completely characterized by the channel. The filtering parameter $\beta \in (0, 1]$ controls the trade-off between the convergence and the accuracy of the iterative algorithm. One may choose to set $L^{(0)}(d_{l,m}) = P(d_{l,m})$. This technique is commonly referred to as "propagation of the extrinsic information" [20] [21].

2.2.3 MAP Detection for 2-D Separable Linear Channels

The iterative likelihood propagation algorithm proposed by [19] does not require the channel to be linear, however, the complexity of this algorithm is exponential in $(2L+1)^2$. By assuming a *Separable* linear channel, Wu et al. [22] extended the BCJR detection technique to efficiently approximate the a-posteriori probability (APP) of the data bits corrupted by the separable 2-D ISI and the additive white Gaussian noise.

Figure 2.2 illustrates the separable 2-D channel. We denote the input data bits by $d_{i,j}$. These bits pass through a *separable* linear 2-D ISI channel characterized by the matrix $\mathbf{H} = [h_{i,j}]$. The ISI span is $L \times L$ which means that L^2 bits interfere

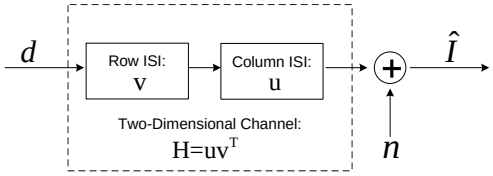


Figure 2.2: The two-dimensional, separable, linear channel with additive white Gaussian noise

for each channel output. Finally, the white Gaussian noise $n_{i,j}$ is added to produce the channel output $\hat{I}_{i,j}$ as

$$\begin{aligned} \hat{I}_{i,j} &= [h_{i,j}] \otimes [d_{i,j}] + n_{i,j} \\ &= \sum_{k_1=0}^{L-1} \sum_{k_2=0}^{L-1} h_{k_1,k_2} d_{i-k_1,j-k_2} + n_{i,j}, \end{aligned} \tag{2.17}$$

where $\otimes$ denotes the discrete 2-D convolution. The fundamental assumption here is that $\mathbf{H}$ is separable, i.e.

$$\mathbf{H} = \mathbf{u}\mathbf{v}^T,$$

or

$$h_{i,j} = u_i v_j, \tag{2.18}$$

where $\mathbf{u}$ and $\mathbf{v}$ are $L \times 1$ vectors and $\mathbf{v}^T$ is the transpose of $\mathbf{v}$. The k-th element of $\mathbf{u}$ and $\mathbf{v}$ are denoted by u_k and v_k respectively. If we substitute Equation (2.18)

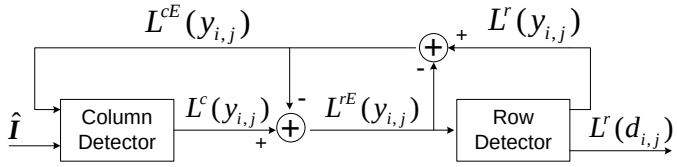


Figure 2.3: The two-dimensional BCJR detector for separable linear channels

into Equation (2.17) we get

$$\begin{aligned} \hat{I}_{i,j} &= \sum_{k_1=0}^{L-1} u_{k_1} \sum_{k_2=0}^{L-1} v_{k_2} d_{i-k_1,j-k_2} \\ &= \sum_{k_1=0}^{L-1} u_{k_1} y_{i-k_1,j}, \end{aligned} \tag{2.19}$$

where

$$y_{i,j} = \sum_{k=0}^{L-1} v_k d_{i,j-k}. \tag{2.20}$$

Hence, Equations (2.19) and Equation (2.20) suggest that we may consider the 2-D ISI channel as the concatenation of two one-dimensional channels representing the row and the column ISI respectively. First, data bits $d_{i,j} \in \{\pm 1\}$ pass through the row ISI channel, characterized by $\mathbf{v}$, and an intermediate output $y_{i,j}$ is produced. In the next stage, the symbols $y_{i,j}$ pass through the column ISI channel, characterized by $\mathbf{u}$.

Let us emphasis two issues here: First, the row ISI channel output $y_{i,j}$ is not binary any more; in fact, it can take 2^L values. We denote the set of these values by $\mathcal{Y}$. Second, we are only able to observe $\hat{I}_{i,j}$, since $y_{i,j}$ is an intermediate signal.

First, let us define the matrices $\mathbf{Y} = [y_{i,j}]$ and $\hat{\mathbf{I}} = [\hat{I}_{i,j}]$. The iterative detector

that approximates $P(d_{i,j}|\hat{\mathbf{I}})$ is illustrated in Figure 2.3. We describe the components in the following. For demonstration purposes, with no loss of generality, we assume that the ISI span is 2×2 i.e. $L = 2$ and $\mathcal{Y} = \{-1.5, -0.5, 0.5, 1.5\}$ similar to [22].

Column Detector

Since $y_{i,j}$ is not binary and it can take four values, the column detector trellis has four states with four branches departing from and arriving at each state. The column detector takes a column of $\hat{\mathbf{I}}$ and calculates the log-likelihood ratio (LLR) of each element of the corresponding column in matrix $\mathbf{Y}$. The MAP algorithm for binary trellises can be extended to a MAP algorithm for non-binary trellises [32]. A brief description of the non-binary MAP is given in the following. The LLR for $y_{i,j}$ is defined as

$$\begin{aligned} L^c(y_{i,j} = Y|\hat{\mathbf{I}}_{:j}) &= \log\left(\frac{P(y_{i,j} = Y|\hat{\mathbf{I}}_{:j})}{P(y_{i,j} = -1.5|\hat{\mathbf{I}}_{:j})}\right) \\ &= \log\frac{\sum_{(\hat{s},s)\in\hat{\mathcal{S}}_Y} P(\hat{s}, s, \hat{\mathbf{I}}_{:j})}{\sum_{(\hat{s},s)\in\hat{\mathcal{S}}_{-1.5}} P(\hat{s}, s, \hat{\mathbf{I}}_{:j})}, \end{aligned} \tag{2.21}$$

where $\hat{\mathbf{I}}_{:j}$ represents the j-th column of the observation matrix, $\hat{s}$ and s respectively represent the starting state and the ending state of a branch of the column ISI trellis. The set $\hat{\mathcal{S}}_Y$ contains the transitions $(\hat{s}, s)$ corresponding to Y. The superscript 'c' represents the information produced or used by the column detector, and similarly the symbol 'r' plays the same role for row detector. Obviously, $L^c(y_{i,j} = -1.5|\hat{\mathbf{I}}_{:j})$ always equals zero. The choice of the base -1.5 is arbitrary.

Similar to Equation (2.6), the joint probability $P(\hat{s}, s, \hat{\mathbf{I}}_{:j})$ in Equation (2.21) consists of three terms

$$P(\hat{s}, s, \hat{\mathbf{I}}_{:j}) = \alpha_{k-1}^{c}(\hat{s})\gamma^{c}(\hat{s}, s)\beta_{k}^{c}(s). \tag{2.22}$$

The forward recursion is given by

$$\alpha_{k}^{c}(s) = \sum_{\hat{s}} \alpha_{k-1}^{c}(\hat{s})\gamma_{k}^{c}(\hat{s}, s). \tag{2.23}$$

Similarly, the backward recursion is given by

$$\beta_{k-1}^{c}(\hat{s}) = \sum_{s} \beta_{k}^{c}(s)\gamma_{k}^{c}(\hat{s}, s). \tag{2.24}$$

The branch transition probability for the column ISI trellis $\gamma_{k}^{c}(\hat{s}, s)$ is calculated as

$$\gamma_{k}^{c}(\hat{s}, s) = P(\hat{I}_{i,j}|\hat{s}, s) \exp\left(L^{cE}(y_{i,j} = Y)\right), \tag{2.25}$$

where the extrinsic information $L^{cE}(y_{i,j} = Y)$ is provided by the row detector and the superscript "E" is used to represent the extrinsic information. Assuming $n_{i,j}$ to be a white zero-mean Gaussian process with variance σ^2, we get

$$P(\hat{I}_{i,j}|z(\hat{s}, s)) = A \exp \frac{(\hat{I}_{i,j} - z(\hat{s}, s))^2}{2\sigma^2}, \tag{2.26}$$

where A is some normalization factor that does not affect the computations and $z(\hat{s}, s)$ denotes the noiseless column ISI channel output corresponding to the state transition from $\hat{s}$ to s. For the first iteration, there is no information available from the row detector; hence $L^{cE}(y_{i,j} = Y)$ is initialized to zero. This follows from

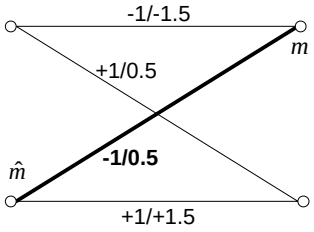


Figure 2.4: Binary row ISI trellis

the assumption that the four input symbols to the column detector are equiprobable. After computing the column LLR defined in Equation (2.21), the extrinsic information

$$L^{rE}(y_{i,j} = Y) = L^{c}(y_{i,j} = Y|\hat{\mathbf{I}}_{:j}) - L^{cE}(y_{i,j} = Y), \tag{2.27}$$

is passed to the constituent row detector as Figure 2.3 illustrates. Let us define the matrix $\mathbf{L^{rE}} = [L^{rE}(y_{i,j})]$.

Row Detector

The inputs to row ISI are $d_{i,j} \in \{\pm 1\}$. The row detector calculates the LLR of the row ISI input symbols

$$\begin{aligned} L^{r}(d_{i,j}|\mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:}) &= \log\left(\frac{P\left(d_{i,j}=1|\mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:}\right)}{P\left(d_{i,j}=-1|\mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:}\right)}\right) \\ &= \log\frac{\sum_{(\hat{m},m)\in\mathcal{S}^{+}} P(\hat{m},m,\mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:})}{\sum_{(\hat{m},m)\in\mathcal{S}^{-}} P(\hat{m},m,\mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:})}, \end{aligned} \tag{2.28}$$

where $\mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:}$ represents the extrinsic information for the i-th row given by Equation (2.27), and $\hat{m}$ and m respectively represent the starting state and the ending

state of a branch of the row ISI trellis. Furthermore, $\mathcal{S}^+$ and $\mathcal{S}^-$ are defined on the row detector trellis similar to the one-dimensional BCJR. For the row detector, we directly apply the BCJR algorithm described in Section 2.2 to the binary row ISI trellis illustrated in Figure 2.4. In Equation (2.28), the value of $P(\hat{m}, m, \mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:})$ can be calculated using the BCJR algorithm.

Since the row ISI trellis illustrated in Figure 2.4 is binary, we are able to apply the BCJR algorithm described in Section 2.2 directly. In order to calculate the value of $P(\hat{m}, m, \mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:})$ in Equation (2.28) using the one-dimensional BCJR algorithm we need to modify the computation of the branch transition probability $\gamma^r_k(\hat{m}, m)$ since there is no direct observation of the row ISI output. To do so, the branch transition probability is computed based on the extrinsic information obtained from the column detector. To explain more, consider the labeled trellis branch $(\hat{m}, m)$ in Figure 2.4

$$\gamma^r_k(\hat{m}, m) = \frac{\exp\left(L^{rE}(y_{i,j} = -0.5)\right)}{\sum_{Y\in\{-1.5,-0.5,0.5,1.5\}} \exp\left(L^{rE}(y_{i,j} = Y)\right)}, \tag{2.29}$$

where $L^{rE}(y_{i,j} = Y$ is given by Equation (2.27). Furthermore, the LLR of the row ISI output symbols are

$$\begin{aligned} L^r(y_{i,j} = Y|\mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:}) &= \log\left(\frac{P\left(y_{i,j} = Y|\mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:}\right)}{P\left(y_{i,j} = -1.5|\mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:}\right)}\right) \\ &= \log\frac{\sum_{(\hat{m},m)\in\tilde{\mathcal{S}}_Y} P(\hat{m}, m, \mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:})}{\sum_{(\hat{m},m)\in\tilde{\mathcal{S}}_{-1.5}} P(\hat{m}, m, \mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:})}, \end{aligned} \tag{2.30}$$

where $\tilde{\mathcal{S}}_Y$ denote the set of transitions $(\hat{m}, m)$ on the row detector trellis corresponding to Y. The detection is performed iteratively between the column detector and

the row detector. The information passed from the row detector to the column detector is the difference of Equation (2.27) and Equation (2.30):

$$L^{cE}(y_{i,j} = Y) = L^{r}(y_{i,j} = Y|\mathbf{L}^{\mathbf{rE}}_{\mathbf{i},:}) - L^{rE}(y_{i,j} = Y). \tag{2.31}$$

Once a predefined number of iterations are performed, Equation (2.28) gives the LLR of the channel input symbols.

2.3 Conclusions

Our review of the $4 - \mathrm{F_L}$ architecture in this chapter is the basis of the accurate model we propose in Chapter 3. Although various hard-decision reception schemes for HDS are designed, the problem of soft-decision detection for the HDS did not receive much attention. One reason could be lack of a natural ordering of pixels in a 2-D array which prevents the generalization of the low-complexity 1-D MAP detector, i.e. BCJR, for 2-D channels. In this chapter, we described the 1-D BCJR algorithm and showed how the lack of natural ordering is tackled for 2-D separable linear channels. Separability property allows us to reduce the high complexity problem of the 2-D MAP detection into a much simpler problem of 1-D MAP detection.

Chapter 3

Accurate Modeling of Holographic Data Storage

Design of equalizers, detectors and codes for the HDS usually relies on extensive simulations in order to measure their performance. The accuracy of such performance measurements is influenced by the accuracy of the channel model that is used for the simulations. Hence, an accurate yet computationally efficient channel model is required. In this paper we study the channel model for the HDS with 4-focal-length (4-f_L) architecture [2].

Since the HDS is a 2-D nonlinear channel, the analysis and efficient generation of the read-back signal is not easy. Typical models capture the ISI, the channel nonlinearity, the optical noise and the detector electronics noise. As a result of the

HDS nonlinearity, the effect of the Gaussian optical noise on the detector read-back signal is no longer Gaussian. Furthermore, the nonlinearity of the channel lead to presence of various *components* in the detector read-back signal. As we will see in Section 3.1.2, the optical noise has a band-limited nature thus results these components to be correlated along the pixels.

Chugg et al. [9] develop an accurate yet computationally demanding model for the HDS. On the other hand, Keskinoz and Kumar [6, 7, 8] develop a computationally efficient model for the HDS called discrete magnitude-squared channel model (DMC). We believe that the discrete magnitude-squared channel(DMC) model is of fundamental importance since it provides us with insights about the structure of noiseless channel output that comes in handy for design signal processing algorithms such as the one presented in Chapter 4. However, as we will see later in Section 3.4 the optical noise modeling in DMC is not accurate. In addition, the DMC model is originally developed for pixel aligned channels. An interesting question would be whether we are able to extend this structure for pixel mis-aligned channels as well.

He [13] presents a more accurate treatment of the optical noise while using techniques of [8] to achieve efficiency. Yet, He's [13] model relies on the assumption that the optical noise power is low. Furthermore, as we will see later in Section 3.4 He's [13] model is not accurate as well. In summary, the accurate model of [9] is not efficient, and the efficient models of [13] and [8] are not accurate man does not

consider pixel misalignment.

In this chapter we extend the DMC model for the channels with pixel misalignment. Furthermore, we investigate the CCD read-back signal in the presence of the optical noise and study its statistics. Furthermore, we develop efficient simulation methods for CCD read-back signal.

This chapter is organized as follows: In Section 3.1 we introduce the HDS channel model, following the same notations and assumptions as Keskinoz and Kumar [8] with the difference that we incorporate the optical noise using the approach of Chugg et al. [9]. In Section 3.2 we introduce our efficient numerical method for simulating the HDS channel. In addition, we present the analytical results on the statistics of detector array read-back values. Section 3.3 presents the simulation results. Section 3.4 presents a detailed comparison between our model and other HDS models. We present our conclusions in Section 3.5.

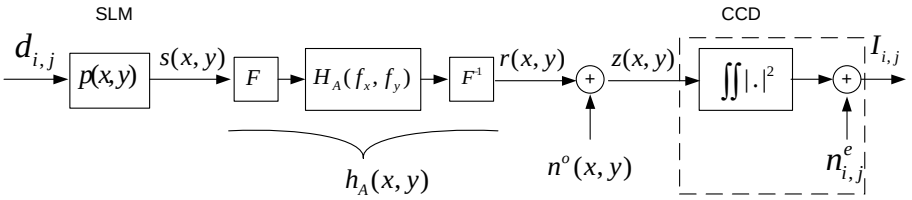


Figure 3.1: Physical Model of the Holographic Data Storage.

3.1 Channel Modeling

We present the mathematical derivation of the channel model for holographic data storage using similar notations and assumptions as Keskinoz et al. [8]. The model incorporates ISI, electronics and optical noise. Although we follow [8] closely, we borrow the optical noise analysis from Chugg et al. [9]. The physical parameters that our model takes into account are width of the frequency plane aperture, the SLM finite contrast ratio, and non unity fill factors for SLM and CCD. Pixel sizes of the SLM and the CCD are identical. We assume a defect free and linear recording medium.

Figure 3.1 illustrates the HDS channel. Let us denote the input binary data by $d_{i,j}$ and the SLM finite contrast ratio by ϵ. The SLM represents one and zero binary values by two amplitude levels of 1 and $1/\epsilon$ respectively. The SLM's pixel shape is described by $p(x, y)$ (usually rectangular). The SLM output, $s(x, y)$, is Fourier transformed. As the aperture windows the Fourier transform of the signal beam, it acts as a low-pass filter with a cutoff frequency determined by the aperture

width. The second lens system (denoted F^{-1}) inverse Fourier transforms the wave-front emerging from the medium during retrieval and produces information bearing wave-front, $r(x, y)$. The incident wave-front, $z(x, y)$ is generated by addition of a complex zero-mean colored Gaussian noise denoted by $n^o(x, y)$ that represents the optical noise. The CCD integrates the magnitude squared of the amplitude wave-front $z(x, y)$ over its pixels. The effect of the electronics noise is captured by the white Gaussian process denoted by $n^e_{i,j}$ and the CCD read-back signal is denoted by $I_{i,j}$.

3.1.1 Linear Subsystem

Between the SLM output and the CCD input the system can be thought to be linear if we ignore nonlinearities of the medium. Therefore, linear system theory can be applied up to the point of the incidence of light on the CCD. In this section we model this linear subsystem in the absence of the optical noise. We will treat the effect of that noise in the next subsection.

Because of the linear, shift-invariant assumption, the aperture and the Fourier transform lenses can be characterized by the apertures impulse response $h_A(x, y)$. For a square aperture of width D centered at the origin in the frequency plane, the aperture's impulse response, $h_A(x, y)$, is given as

$$h_A(x, y) = \acute{h}_A(x)\acute{h}_A(y), \tag{3.1}$$

where

$$\acute{h}_A(x) = \frac{D}{\lambda f_L}\text{sinc}(\frac{xD}{\lambda f_L}), \tag{3.2}$$

λ is the wavelength of the light, and f_L is the lens's focal length. In order to omit λ, f_L, and D from our simulations later, we use the normalized aperture width ω, defined as

$$\omega = D/D_N, \tag{3.3}$$

where $D_N = \lambda f_L/\Delta_s$ is known as the Nyquist aperture width [33, 8] where Δ_s is the pixel-to-pixel spacing (assumed to be the same for both the SLM and the CCD). Consequently, we re-write Equation (3.2) as $\acute{h}_A(x) = (\omega/\Delta_s)\text{sinc}(\omega x/\Delta_s)$. The Fourier transform $\acute{H}_A(f_x)$ of $\acute{h}_A(x)$ is an ideal low-pass filter with cutoff frequency $\omega/2\Delta_s$; i.e.,

$$\acute{H}_A(f_x) = \begin{cases} 1 & |f_x| \leq \frac{\omega}{2\Delta_s}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.4}$$

The SLM's pixel shape function, $p(x, y)$, can be expressed as

$$p(x, y) = \Pi(\frac{x}{\alpha\Delta_s})\Pi(\frac{y}{\alpha\Delta_s}) \tag{3.5}$$

where α is the SLM's linear fill-factor (i.e., square root of the areal fill-factor for a square pixel) and $\Pi(x)$ represents the unit rectangle function. Although we have identical pixel spacing for CCD and SLM, we keep using the subscript in order to be

able to distinguish between the SLM and the CCD pixel spacing. Mathematically, the output of the SLM, $s(x, y)$, is described by

$$s(x,y) = \sum_{k} \sum_{l} d_{k,l} p(x - k\Delta_s, y - l\Delta_s), \tag{3.6}$$

where k and l represent the pixel locations along the x and y directions, respectively, and $k = 0$ and $l = 0$ represent the center pixel in the page. Each page is assumed to contain $N \times N$ pixels. The wave-front of the light leaving the SLM $s(x, y)$ passes through this linear, shift-invariant system whose impulse response is $h_A(x, y)$. Hence, the information bearing wave-front, denoted by $r(x, y)$, is

$$\begin{aligned} r(x,y) &= s(x,y) \star h_A(x,y) \\ &= \left[\sum_{k,l} d_{k,l} p(x - k\Delta_s, y - l\Delta_s) \right] \star h_A(x,y) \\ &= \sum_{k,l} d_{k,l} h(x - k\Delta_s, y - l\Delta_s), \end{aligned} \tag{3.7}$$

where $\star$ represents 2-D (2-D) convolution of continuous signals. Furthermore, the pixel response before the CCD, $h(x, y)$, is defined as

$$h(x,y) = p(x,y) \star h_A(x,y). \tag{3.8}$$

We call $h(x, y)$ the optical point spread function (PSF). Note that $h(x, y)$ is separable and symmetric since $p(x, y)$ and $h_A(x, y)$ are both separable and symmetric.

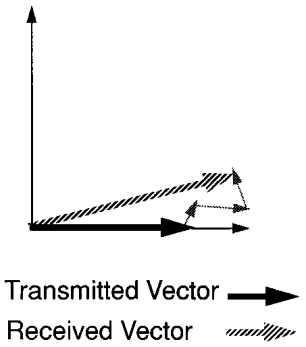


Figure 3.2: Optical noise due to independent scatterers in holographic data storage.

Mathematically,

$$h(x,y) = \acute{h}(x)\acute{h}(y) \tag{3.9}$$

$$\acute{h}(x) = \acute{h}_A(x) \star \Pi(\frac{x}{\alpha\Delta_s}). \tag{3.10}$$

Since $\acute{h}_A(x)$ defined in Equation (3.2) is low-pass and band-limited, Equation (3.10) implies that $\acute{h}(x)$ is low-pass and band-limited too; i.e. the $\acute{H}(f_x)$ is a low-pass filter with cutoff frequency $\omega/2\Delta_s$. We also assume that $\acute{h}(x)$ is zero outside the interval $|x| \leq (L+0.5)\Delta_s$; hence the ISI span is $(2L+1)\Delta_s$.

3.1.2 Optical Noise

Similar to [34], we assume that the optical noise is predominantly due to scatter of the reference beam from a collection of independent scatterers. As shown in Figure 3.2, assume that the information bearing wave-front, $r(x,y)$, is represented

by a vector of length A and zero phase. The noise can therefore be represented by a sum of random phasors. The vector representing the signal at the detector will have an amplitude and a phase that are, in general, different from the amplitude and the phase of the transmitted vector.

We model this noise through the addition of $n^o(x,y)$, to get the incident light on the CCD, $z(x,y)$

$$\begin{aligned} z(x,y) &= r(x,y) + n^o(x,y) && (3.11)\\ &= \sum_{k,l} d_{k,l} h(x - k\Delta_s, y - l\Delta_s) + n^o(x,y), && (3.12) \end{aligned}$$

where with no loss of generality we assume a real information bearing wave-front, $r(x,y)$, with zero phase. In addition, $n^o(x,y)$ is a stationary complex circular Gaussian process with mean of zero and auto-covariance given by $K_n(x,y) = \mathbf{E}\{n^o(x,y)n^{o*}(0,0)\}$. The statistics of the optical noise are assumed to be determined by the optical PSF, $h(x,y)$. In particular, we model $n^o(x,y)$ as spatially white noise (intensity E_0) filtered by the system optical PSF, and as a result,

$$K_{n^o}(x,y) = E_0 h(x,y) \star h(-x,-y). \tag{3.13}$$

Consequently,

$$S_{n^o}(f_x,f_y) = E_0|H(f_x,f_y))|^2 = E_0|\acute{H}(f_x)\acute{H}(f_y)|^2, \tag{3.14}$$

where $S_{n^o}(f_x,f_y)$ is the power spectrum of $n^o(x,y)$. This means that $n^o(x,y)$ is a

low-pass and band-limited process with cutoff frequency of $\omega/2\Delta_s$.

3.1.3 Detector Array Modeling

The CCD array integrates intensity of the incident light, $z(x, y)$, spatially and temporally to produce the read-back signal $I_{i,j}$. For simplicity, we assume that the intensity variations on the CCD pixels are sufficiently smooth in time so that only the spatial integration matters. As a result of spatial integration in the CCD, read-back signal $I_{i,j}$ is given as

$$\begin{aligned} I_{i,j} &= \int_{(i-\beta/2)\Delta_c}^{(i+\beta/2)\Delta_c} \int_{(j-\beta/2)\Delta_c}^{(j+\beta/2)\Delta_c} |z(x,y)|^2 dxdy + n^e_{i,j} \\ &= \int_{-\beta\Delta_c/2}^{+\beta\Delta_c/2} \int_{-\beta\Delta_c/2}^{+\beta\Delta_c/2} |z(x+i\Delta_c, y+j\Delta_c)|^2 dxdy + n^e_{i,j}, \end{aligned} \tag{3.15}$$

where β is the CCD linear fill-factor, and Δ_c is the CCD pixel width which is assumed to be equal to SLM pixel width

$$\Delta_s = \Delta_c = \Delta. \tag{3.16}$$

Furthermore, $n^e(i, j)$ is the electronics noise of the CCD, which is Gaussian and white with variance σ_e^2. In order to model the misalignment of the CCD and the SLM, we assume that the CCD integrates a shifted signal $z(x + \delta_x, y + \delta_y)$; hence

$$I_{i,j} = \int_{-\beta\Delta/2}^{+\beta\Delta/2}\int_{-\beta\Delta/2}^{+\beta\Delta/2} |z(x+\delta_x+i\Delta, y+\delta_y+j\Delta)|^2 dxdy + n^e_{i,j} \quad (3.17)$$

$$= \int_{-\beta\Delta/2}^{+\beta\Delta/2}\int_{-\beta\Delta/2}^{+\beta\Delta/2} |\sum_{k,l} d_{k,l} h(x+\delta_x+(i-k)\Delta, y+\delta_y+(j-l)\Delta) + n^o(x+\delta_x+i\Delta, y+\delta_y+j\Delta)|^2 dxdy + n^e_{i,j}. \quad (3.18)$$

We restrict our study to global misalignments (i.e. page translation error); so δ_x and δ_y represent page-wide misalignments in the x and y directions respectively. Equation (3.18) completely characterizes the physical model. In the next step, we develop efficient algorithms to compute $I_{i,j}$.

3.2 Efficient Simulation of Detector Read-Back Signal

The complexity of computing Equation (3.18) is high. In order to generate $I_{i,j}$ efficiently, we decompose it into four major components. Using the band-limited property and the separability property of the optical PSF, $h(x, y)$, we devise efficient methods to compute/approximate each component . Similar to [13, 9], we expand the squaring operation expression in Equation (3.18) to get

$$I_{i,j} = a_{i,j} + b_{i,j} + c_{i,j} + n^e_{i,j}, \quad (3.19)$$

where

$$a_{i,j} = \int_{-\beta\Delta/2}^{+\beta\Delta/2}\int_{-\beta\Delta/2}^{+\beta\Delta/2} |r(x+\delta_x+i\Delta, y+\delta_y+j\Delta)|^2 dxdy, \quad (3.20)$$

$$b_{i,j} = \int_{-\beta\Delta/2}^{+\beta\Delta/2}\int_{-\beta\Delta/2}^{+\beta\Delta/2} |n^o(x+\delta_x+i\Delta, y+\delta_y+j\Delta)|^2 dxdy, \quad (3.21)$$

$$c_{i,j} = \int_{-\beta\Delta/2}^{+\beta\Delta/2}\int_{-\beta\Delta/2}^{+\beta\Delta/2} 2\Re\{r(x+\delta_x+i\Delta, y+\delta_y+j\Delta) \times n^{o^*}(x+\delta_x+i\Delta, y+\delta_y+j\Delta)\} dxdy, \quad (3.22)$$

where $a_{i,j}$ is the information-bearing component of $I_{i,j}$, $b_{i,j}$ is the post-detector effect of the optical noise intensity, $c_{i,j}$ is the post-detector effect of the cross product of the information-bearing wave-front and the optical noise wave-front, and $\Re\{x\}$ denotes the real part of the complex variable x. Furthermore, let us write the complex optical noise as $n^o(x,y) = n_r^o(x,y) + \mathrm{j}n_i^o(x,y)$, where $n_r^o(x,y)$ is the real part and $n_i^o(x,y)$ is the imaginary part of the $n^o(x,y)$ respectively. Now we can write $c_{i,j}$ as

$$c_{i,j} = \int_{-\beta\Delta/2}^{+\beta\Delta/2}\int_{-\beta\Delta/2}^{+\beta\Delta/2} 2r(x+\delta_x+i\Delta, y+\delta_y+j\Delta) \times n_r^o(x+\delta_x+i\Delta, y+\delta_y+j\Delta) dxdy. \quad (3.23)$$

He [13] also considered this decomposition of the CCD read-back signal and develop methods for efficient computation/approximation of these terms *separately*. However, in order to be more precise we take the dependency of the CCD read-back signal into account. To explain more, we discuss a major constraint on the

three components $a_{i,j}$, $b_{i,j}$, and $c_{i,j}$ in Section 3.2.1. Afterwards, in Section 3.2.2 we derive the model for $a_{i,j}$ following the methodology used in [8]. We show that in the presence of pixel misalignments (i.e. $\delta_x \neq 0$ or $\delta_y \neq 0$), the structure of the HDS channel remains intact. Furthermore, we illustrate the separability property of the underlying discrete linear time-invariant filter that is useful for developing reduced complexity 2-D BCJR detectors [35]. In Chapter 4 we explain how to exploit this property as the foundation for design of a 2-D quadratic BCJR detector. In Section 3.2.3 we derive discrete time models for $b_{i,j}$ and $c_{i,j}$. In order to generate valid triples of $\{a_{i,j}, b_{i,j}, c_{i,j}\}$, we generate samples of the optical noise. This increases the complexity of our model, but guaranties statistical consistency among these components. We use the separability of the optical PSF to reduce the complexity of our computation.

3.2.1 Relation Among $a_{i,j}$, $b_{i,j}$, and $c_{i,j}$

Now we proceed to investigate the relationship between $x_{i,j}$, $b_{i,j}$ and $c_{i,j}$. The most important constraint to be considered is stated in the following theorem.

Theorem 3.1. For $a_{i,j}$, $b_{i,j}$, and $c_{i,j}$ given by Equation (3.20), Equation (3.21), and Equation (3.23) respectively, the following inequalities hold

$$0 \leq a_{i,j} + b_{i,j} + c_{i,j} \tag{3.24}$$

$$|c_{i,j}| \leq a_{i,j} + b_{i,j} \tag{3.25}$$

Proof. It is easy to prove Equation (3.24) since

$$a_{i,j} + b_{i,j} + c_{i,j} = \int_{(i-\beta/2)\Delta}^{(i+\beta/2)\Delta} \int_{(j-\beta/2)\Delta}^{(j+\beta/2)\Delta} |z(x,y)|^2 dxdy \geq 0 \tag{3.26}$$

In order to prove Equation (3.25), we recall that

$$\begin{aligned} |r(x,y) + n^o(x,y)|^2 &= |r(x,y)|^2 + |n^o(x,y)|^2 + 2\Re\{r(x,y)n^{o^*}(x,y)\} &\geq 0 \\ |r(x,y) - n^o(x,y)|^2 &= |r(x,y)|^2 + |n^o(x,y)|^2 - 2\Re\{r(x,y)n^{o^*}(x,y)\} &\geq 0 \end{aligned}$$

Hence

$$|2\Re\{r(x,y)n^{o^*}(x,y)\}| \leq |r(x,y)|^2 + |n^o(x,y)|^2 \tag{3.27}$$

Integrating the Inequality Equation (3.27) over the pixel area proves Equation (3.25).

□

In summary, we observe that the CCD read-back signal components are heavily constrained. This observation implies lack of precision in Computing methods that generate the CCD read-back signal components separately.

3.2.2 Efficient Simulation of $a_{i,j}$

We start by presenting an efficient method to compute the information-bearing component of the CCD read-back signal Equation (3.20) using same techniques as [8].

Plugging Equation (3.7) into Equation (3.20) we get

$$\begin{aligned} a_{i,j} &= \sum_{k,l}\sum_{m,n} d_{k,l}d_{m,n}\int_{-\Delta/2}^{+\Delta/2}\int_{-\Delta/2}^{+\Delta/2} \\ &\quad \times h(x+\delta_x+(i-k)\Delta, y+\delta_y+(j-l)\Delta) \\ &\quad \times h(x+\delta_x+(i-m)\Delta, y+\delta_y+(j-n)\Delta)dxdy. \end{aligned} \tag{3.28}$$

Since $h(x,y)$ is separable in terms of x and y, we can simplify Equation (3.28) further to

$$a_{i,j} = \sum_{k,l}\sum_{m,n} d_{k,l}d_{m,n}G^{\delta_x}_{i-k,i-m}G^{\delta_y}_{j-l,j-n}, \tag{3.29}$$

where $G^{\delta}_{k,m}$ is defined as

$$G^{\delta}_{k,m} = \int_{-\Delta/2}^{+\Delta/2} \acute{h}(\tau+\delta+k\Delta)\acute{h}(\tau+\delta+m\Delta)d\tau. \tag{3.30}$$

Note that $\mathbf{G}^{\delta} = [G^{\delta}_{k,m}]$ depends on the amount of misalignment, δ. These equations imply that for each amount of pixel misalignment along either axis there is a corresponding matrix $\mathbf{G}^{\delta}$. We proceed to simplify Equation (3.29) further. Using eigenvalue decomposition techniques, we may express entries in $\mathbf{G}^{\delta}$ as

$$G^{\delta}_{k,m} = \sum_{r=1}^{R} \lambda^{\delta}_r (\mathbf{v}^{\delta}_r)_k (\mathbf{v}^{\delta}_r)_m \tag{3.31}$$

where R, the rank of the DCM, is at most (2L+1), λ^{δ}_r is the r-th eigenvalue associated with eigenvector $\mathbf{v}^{\delta}_r$, and $(\mathbf{v}^{\delta}_r)_k$ represents the k-th component of eigenvector

$\mathbf{v}_r^{\delta}$. Inserting Equation (3.31) into Equation (3.29) we get

$$\begin{aligned} a_{i,j} &= \sum_{k,l}\sum_{m,n} d_{k,l}d_{m,n}\sum_{r=1}^{R}\sum_{s=1}^{R}\lambda_r^{\delta_x}\lambda_s^{\delta_y}(\mathbf{v}_r^{\delta_x})_{i-k}(\mathbf{v}_r^{\delta_x})_{i-m}(\mathbf{v}_s^{\delta_y})_{j-l}(\mathbf{v}_s^{\delta_y})_{j-n} \\ &= \sum_{r=1}^{R}\sum_{s=1}^{R}(d_{i,j}\otimes[\sqrt{\lambda_r^{\delta_x}\lambda_s^{\delta_y}}(\mathbf{v}_r^{\delta_x})_i(\mathbf{v}_s^{\delta_y})_j])^2, \end{aligned} \tag{3.32}$$

where $\otimes$ denotes discrete 2-D convolution. We can see that in the presence of pixel misalignments we still have the DCM structure as in [8]. In more detail, Equation (3.32) shows that one can view the $a_{i,j}$ as the total response of a bank of magnitude-squared sub-channels, where each magnitude-squared sub-channel consists of a discrete linear time-invariant channel followed by the magnitude square operation.

We approximate Equation (3.31) as

$$G_{k,m}^{\delta} \approx \lambda_{\max}^{\delta}(\mathbf{v}_{\max}^{\delta})_k(\mathbf{v}_{\max}^{\delta})_m. \tag{3.33}$$

where $\lambda_{\max}^{\delta}$ is the largest eigenvalue and $\mathbf{v}_{\lambda_{\max}}^{\delta}$ is its corresponding eigenvector. From now on, we denote these quantities by λ^{δ} and $\mathbf{v}^{\delta}$ respectively. Inserting Equation (3.33) into Equation (3.29) we get

$$a_{i,j} \approx \sum_{k,l}\sum_{m,n} d_{k,l}d_{m,n}H_{k,l}^{\delta_x,\delta_y}H_{m,n}^{\delta_x,\delta_y} \tag{3.34}$$

$$= \left(H_{i,j}^{\delta_x,\delta_y}\otimes[d_{i,j}]\right)^2, \tag{3.35}$$

where

$$H_{i,j}^{\delta_x,\delta_y} = \sqrt{\lambda^{\delta_x}\lambda^{\delta_y}}\left(v^{\delta_x}\right)_i\left(v^{\delta_y}\right)_j. \tag{3.36}$$

It is worthwhile to note that by storing N different vectors of $\mathbf{q}^{\delta_i}$ we can construct channel models for N^2 combinations of misalignments along x and y directions. In order to show that Equation (3.35) is an accurate approximation of Equation (3.29), we compute the $a_{i,j}$ values based on Equation (3.29) and Equation (3.35) respectively. Following [8], we use the normalized mean squared error (NMSE) to quantify the approximation accuracy. The NMSE is defined as the ratio (mean-squared error between the intensity sequences coming from the HDS as computed by Equation (3.29) and the simplified DCM output as computed by Equation (3.35))/(mean power of the HDSs intensity sequence computed by Equation (3.29)). The normalized mean squared error (NMSE) results that are presented in Section 3.3.1 show that NMSE is at most 6.31%. A very interesting property of $\mathbf{H}^{\delta_x,\delta_y} = [H_{i,j}^{\delta_x,\delta_y}]$ is its separability. Mathematically,

$$\mathbf{H}^{\delta_x,\delta_y} = \sqrt{\lambda^{\delta_x}\lambda^{\delta_y}}\mathbf{v}^{\delta_x}(\mathbf{v}^{\delta_y})^T. \tag{3.37}$$

As figure 3.3 illustrates, the 2-D linear IPI channel may be considered as the concatenation of a linear row ISI channel and a linear column ISI channel. Information bits are denoted by $d_{i,j}$. The row ISI channel is determined by $\mathbf{v}^{\delta_x}$ and represents the interference among bits $d_{i,j}$ that are placed along neighboring rows on the same column. The output of the row ISI is denoted by $y_{i,j}$. The $y_{i,j}$ that are placed along neighboring columns on the same row interfere according to $(\mathbf{v}^{\delta_y})^T$ that represents the column ISI channel. The information-bearing component $a_{i,j}$

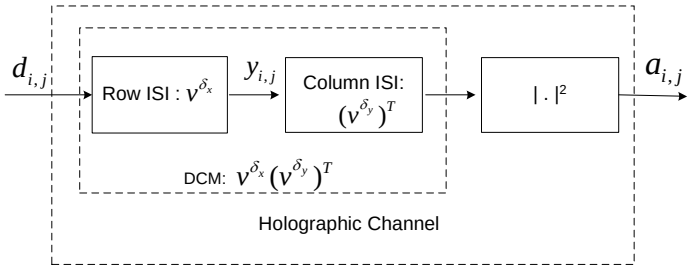


Figure 3.3: Separable DCM of information-bearing component of the CCD read-back signal in a HDS system with pixel misalignment.

is the magnitude-square of the column ISI output. In [35] we present a reduced complexity BCJR detector for the electronics noise dominated HDS channels with pixel misalignment based on the separability property discussed here.

3.2.3 Efficient Simulation of $b_{i,j}$ and $c_{i,j}$

As we observe in Section 3.2.1 there is dependency among the CCD read-back signal components. Hence in our approach we generate samples of the optical noise $n^{\varsigma}_{k,l} = n^o(k\varsigma, l\varsigma)$, where ς is the sampling step. As the optical noise auto-covariance is $K_{n^o}(x,y) = E_0 h(x,y) \star h(-x,-y)$, the auto-covariance of the process $n^{\varsigma}_{k,l}$ is

$$\mathbf{E}\{n^{\varsigma}_{k,l} n^{\varsigma^*}_{0,0}\} = \mathbf{E}\{n^o(k\varsigma, l\varsigma) n^{o^*}(0,0)\} = K_{n^o}(k\varsigma, l\varsigma). \tag{3.38}$$

Hence we generate the samples by filtering a complex zero-mean white Gaussian

process with the filter

$$h^{\varsigma}_{k,l} = \varsigma\sqrt{E_0}h(k\varsigma, l\varsigma). \tag{3.39}$$

In order to simulate the post-detector effect of scatter noise, $b_{i,j}$, based on Equation (3.21), we need to perform 2-D numerical integrations. However, using a sufficiently fine integration step, ς, lead to a very large matrix representing $n^{\varsigma}_{i,j}$ that requires huge memory and takes a lot of time to apply the coloring filter $h^{\varsigma}_{i,j}$. In more detail,a data page of $N \times N$ pixels and integration step of $\varsigma = \Delta/M$ lead to $n^{\varsigma}_{i,j}$ page of size

$$MN \times MN. \tag{3.40}$$

Furthermore, if the IPI span is $(2L+1) \times (2L+1)$ then size of h^{ς} is

$$(2L+1)M \times (2L+1)M. \tag{3.41}$$

Consequently, using a fine integration step lead to prohibitive complexity. We show an efficient method for performing 2-D integration of band-limited signals using a large sampling step.

Theorem 3.2. Assume that $E(x, y)$ is the response of the following two-dimensional integrator system

$$E(x,y) \;=\; \int_{x-\beta\Delta/2}^{x+\beta\Delta/2} \int_{y-\beta\Delta/2}^{y+\beta\Delta/2} S(\tau_1, \tau_2) d\tau_2 d\tau_1, \tag{3.42}$$

where $S(x, y)$ which is a band-limited signal with Nyquist sampling rate of T_s. The samples $e_{i,j} = E(iT_s, jT_s)$ are

$$e_{i,j} = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} S_{k_1,k_2} w_{i-k_1} w_{j-k_2}, \tag{3.43}$$

where $S_{k_1,k_2} = S(iT_s, jT_s)$ and

$$w_i = \int_{-\beta\Delta/2}^{+\beta\Delta/2} \mathrm{sinc}(\frac{\tau_1 + iT_s}{T_s}) d\tau_1. \tag{3.44}$$

Proof. Let us first look at the reconstruction of $S(x, y)$ based on its samples $S_{i,j}$ as

$$S(x, y) = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} S_{k_1,k_2} \ \mathrm{sinc}(\frac{x - k_1T_s}{T_s}, \frac{y - k_2T_s}{T_s}). \tag{3.45}$$

$$\begin{aligned} E(iT_s, jT_s) = & \int_{iT_s-\beta\Delta/2}^{iT_s+\beta\Delta/2} \int_{jT_s-\beta\Delta/2}^{jT_s+\beta\Delta/2} \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} S_{k_1,k_2} \\ & \times \mathrm{sinc}(\frac{\tau_1 - k_1T_s}{T_s}, \frac{\tau_2 - k_2T_s}{T_s}) d\tau_2 d\tau_1. \end{aligned} \tag{3.46}$$

Changing the order of summation and integration and changing the integration variables we get

$$\begin{aligned} E(iT_s, jT_s) &= \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} S_{k_1,k_2} \\ &\quad \times \int_{-\beta\Delta/2}^{+\beta\Delta/2} \int_{-\beta\Delta/2}^{+\beta\Delta/2} \mathrm{sinc}(\frac{\tau_1 + (i - k_1)T_s}{T_s}, \frac{\tau_2 + (j - k_2)T_s}{T_s}) d\tau_2 d\tau_1 \\ &= \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} S_{k_1,k_2} w_{i-k_1} w_{j-k_2}, \end{aligned} \tag{3.47}$$

where

$$w_i = \int_{-\beta\Delta/2}^{+\beta\Delta/2} \mathrm{sinc}(\frac{\tau_1 + iT_s}{T_s}) d\tau_1. \tag{3.48}$$

where we exploit the separability of $\text{sinc}(x, y) = \text{sinc}(x)\text{sinc}(y)$ to drive Equation (3.48). □

Equation (3.47) shows that 2-D integration is reduced to 2-D convolution as the result of finite band-width. Replacing the separable two-dimensional convolution with the concatenation of two one-dimensional convolutions in Equation (3.47) brings more complexity reduction. In more detail, for a filter w_i size $N_f \times N_f$, the complexity is reduced from $O(N_f^2)$ for two-dimensional convolution to $O(N_f)$ for one-dimensional convolution. In addition, despite the fact that the function $\text{sinc}(1/T_s)$ decays slowly, the filter coefficients, w_i decays much faster as i increases because the positive and negatives lobes of the sinc function cancel each other in the integration Equation (3.48). Figure 3.4 shows filter $\acute{w}_i$ for CCD fill-factor $\beta = 1$ and $T_s = \Delta/2$.

In addition to generic complexity reduction gained by Theorem 3.2, it also brings substantial complexity reduction in the context of computing $b_{i,j}$ and $c_{i,j}$ as we explain in the following.

Derivation of the $b_{i,j}$

In this section we study the post-detector effect of the optical scatter noise; namely the process $b_{i,j}$ which is defined in Equation (3.21).

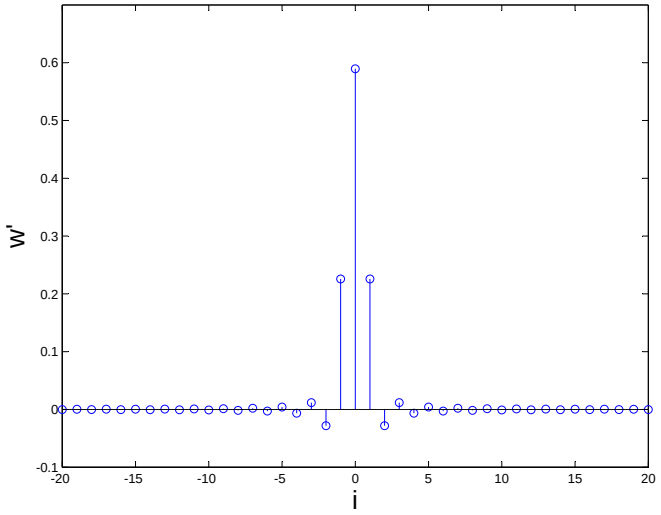


Figure 3.4: Integrator filter $\acute{w}_i$ for $\beta = 1$ and $\varsigma = T_s = \Delta/2$.

We can view Equation (3.21) as sampling of the time-invariant system

$$B(x,y) \quad = \quad \int_{x-\beta\Delta/2}^{x+\beta\Delta/2} \int_{y-\beta\Delta/2}^{y+\beta\Delta/2} |n^o(\tau_1,\tau_2)|^2 d\tau_2 d\tau_1 \tag{3.49}$$

with rate Δ; i.e. $b_{i,j} = B(i\Delta, j\Delta)$. Let us denote the intensity of the optical scatter noise by $I^{on}(x,y) = |n^o(x,y)|^2$. Since the cutoff frequency of $I^{on}(x,y)$ is ω/Δ with Nyquist aperture width, $\omega = 1$, we have $T_s = \Delta/2$ which is much larger than a typical ς required for precise numerical integration. This results in a much smaller $n^{\varsigma}_{i,j}$ page, and lower computation complexity for generating the optical noise sample using the filter $h^{\varsigma}_{i,j}$.

Efficient Simulation of the $c_{i,j}$

Let us express the Post-detector effect of the cross product field as [13]

$$c_{i,j} = 2 \sum_{k,l=-L}^{L} d_{i-k,j-l} J_{k,l,i,j}, \tag{3.50}$$

where

$$\begin{aligned} J_{k,l,i,j} &= \int_{-\beta\Delta/2}^{+\beta\Delta/2} \int_{-\beta\Delta/2}^{+\beta\Delta/2} h(k\Delta + x + \delta_x, l\Delta + y + \delta_y) \\ &\quad \times n_r^o(x + \delta_x + i\Delta, y + \delta_y + j\Delta) dx dy. \end{aligned} \tag{3.51}$$

The discrete random process $J_{k,l,i,j}$ is zero-mean and Gaussian. Efficient computation of the $c_{i,j}$ boils down to efficient computation of the $J_{k,l,i,j}$. Vector $\mathbf{J}_{i,j}$ the coefficients $J_{k,l,i,j}$ for the pixel (i,j). Since vector $\mathbf{J}_{i,j}$ contains zero-mean jointly Gaussian random variables, it is fully described by its first and second order statistics. Hence, we start by investigating its auto-covariance

$$K_{\mathbf{J}_{i,j}} = \mathbf{E}\{\mathbf{J}_{i,j}\mathbf{J}_{i,j}^T\}. \tag{3.52}$$

By spectral decomposition of $K_{\mathbf{J}_{i,j}}$ numerically, we observed that this matrix has only one significant eigenvalue and the rest of eigenvalues are negligible. So $K_{\mathbf{J}_{i,j}}$ can be very well approximated in the (minimum mean-squared sense [36]) by

$$K_{\mathbf{J}_{i,j}} \approx \lambda_{max} \mathbf{p}\mathbf{p}^T, \tag{3.53}$$

where λ_{max} is the maximum eigenvalue of $K_{\mathbf{J}_{i,j}}$ and $\mathbf{p}$ is its corresponding eigenvector. The normalized mean-squared error of the approximation is 5.5×10^{-30}.

Which is actually beyond the precision of the software we used in our simulations (MATLAB). This shows that we may represent $\mathbf{J}_{i,j}$ as

$$\mathbf{J}_{i,j} = \mathbf{q}\hat{n}_{i,j}, \tag{3.54}$$

where $\mathbf{q} = \eta\mathbf{p}$ and $\hat{n}_{i,j}$ is a zero-mean Gaussian process with variance λ_{max}/η^2. In other words, for a given pixel (i, j), the coefficients $J_{k,l,i,j}$ are all factors of one random variable $\hat{n}_{i,j}$. This insight is hard to get from Equation (3.51).

The difficulty we face here is that $\hat{n}_{i,j}$ has no physical meaning and hence there seem to be no apparent connection between $\hat{n}_{i,j}$ and the optical scatter noise. Such a connection is important as long as we want to generate $b_{i,j}$ and $c_{i,j}$ consistently. In addition, the relationship between vector $\mathbf{p}$ and the function $h(x, y)$ is not clear. Note that computing $K_{\mathbf{J}_{i,j}}$ and its spectral decomposition is time consuming. This relationship is important since it potentially enables us to generate $\mathbf{p}$ more efficiently for different system parameters. In order to derive a physically meaningful representation for $\mathbf{J}_{i,j}$ we approximate $J_{k,l,i,j}$ in Equation (3.51) by substituting a constant $q_{k,l}^{\delta_x,\delta_y}$ instead of $h(k\Delta + x + \delta_x, l\Delta + y + \delta_y)$. We omit the superscript and simply write $q_{k,l}$. Using this approximation Equation (3.51) we get the approximation for $J_{k,l,i,j}$ as

$$\tilde{J}_{k,l,i,j} = q_{k,l}\tilde{n}_{i,j} \tag{3.55}$$

where

$$\tilde{n}_{i,j} = \iint_{-\beta\Delta/2}^{+\beta\Delta/2} n_r^o(x + \delta_x + i\Delta, y + \delta_y + j\Delta)dxdy. \tag{3.56}$$

Since $h(x, y)$ does not fluctuate rapidly over the pixel area, this is a reasonable approximation. We would like to determine $q_{k,l}$ such that the mean-squared error

$$MSE = \mathbf{E}\{(J_{k,l,i,j} - \tilde{J}_{k,l,i,j})^2\} \tag{3.57}$$

is minimized. Inserting Equation (3.51) and Equation (3.55) into Equation (3.57) and set the expectation of the derivative of MSE with respect to $q_{k,l}$ to zero results in

$$q_{k,l} = \frac{A_{k,l}}{\sigma^2_{\tilde{n}}}, \tag{3.58}$$

where

$$\begin{aligned} A_{k,l} &= \frac{1}{2}\iiiint_{-\beta\Delta/2}^{\beta\Delta/2} h(k\Delta + x + \delta_x, l\Delta + y + \delta_y) \\ &\quad \times K_{n^o}(x - \acute{x}, y - \acute{y})dxdyd\acute{x}d\acute{y}, \end{aligned} \tag{3.59}$$

and $\sigma^2_{\tilde{n}}$ is, the variance of $\tilde{n}_{i,j}$, equal to

$$\begin{aligned} \sigma^2_{\tilde{n}} &= \mathbf{E}\{\tilde{n}^2_{i,j}\} \\ &= \iiiint_{-\beta\Delta/2}^{\beta\Delta/2} \mathbf{E}\{n^o_r(x + \delta_x + i\Delta, y + \delta_y + j\Delta) \\ &\quad \times n^o_r(\acute{x} + \delta_x + i\Delta, \acute{y} + \delta_y + j\Delta)\}dxdyd\acute{x}d\acute{y} \\ &= \iiiint_{-\beta\Delta/2}^{\beta\Delta/2} K_{n^o_r}(x - \acute{x}, y - \acute{y})dxdyd\acute{x}d\acute{y} \\ &= (\beta\Delta)^2 \iint_{-\beta\Delta}^{\beta\Delta} (1 - \frac{|x|}{\beta\Delta})(1 - \frac{|y|}{\beta\Delta})K_{n^o_r}(x, y)dxdy \\ &= \frac{(\beta\Delta)^2}{2} \iint_{-\beta\Delta}^{\beta\Delta} (1 - \frac{|x|}{\beta\Delta})(1 - \frac{|y|}{\beta\Delta})K_{n^o}(x, y)dxdy, \end{aligned} \tag{3.60}$$

where the last equality is due to the fact that the optical scatter noise $n^o(x, y)$ is a complex symmetric Gaussian process; hence its power is equally divided between its real and imaginary parts and hence $K_{n_r^o}(x, y) = 0.5 * K_{n^o}(x, y)$. Let $\mathbf{q}^{\mathrm{MMSE}}$ denote the vectorized representation of $q_{k,l}$. The NMSE

$$\frac{\mathbf{E}\{(\mathbf{p} - \mathbf{q}^{\mathrm{MMSE}})^2\}}{\mathbf{E}\{\mathbf{q}^{\mathrm{MMSE}^2}\}} = 0.0019 \tag{3.61}$$

signifies that the optimal $q_{k,l}$ that minimizes the MSE, also result in accurate second order statistics. We expect this to happen in fact.

Considering the finite band-width of the optical noise, we are allowed to exploit Theorem 3.2 to efficiently calculate $\tilde{n}_{i,j}$ based on Equation (3.56); mathematically,

$$\tilde{n}_{i,j} = \sum_{k_1=-\infty}^{\infty} \sum_{k_2=-\infty}^{\infty} n^{\delta}_{k_1,k_2} w_{i-k_1} w_{j-k_2}, \tag{3.62}$$

where integrator filter coefficients, w_i, are similar to Equation (3.48).

The Nyquist sampling distance for the optical scatter noise is $T_s = \Delta/\omega$ which is half of the Nyquist sampling distance for intensity of the optical scatter noise used in calculating Equation (3.49). However, for this increased sampling distance the corresponding integrator filter coefficients vanish much slower. The reason is that for sampling distance $T_s = \Delta/\omega$ each filter coefficient in Equation (3.48) is the area under one lobe of the 'sinc' function. Since we have already computed samples of the optical scatter noise with sampling rate of $\Delta/(2\omega)$, it is more efficient to use the smaller sampling distance $T_s = \Delta/2\omega$ since the filter coefficients vanish much faster.

Statistics of the $b_{i,j}$

Throughout our research, we use the analytical results of Chugg et al. [9] on the first and second order statistics of the process $b_{i,j}$ to validate our approach. Interestingly, our numerical results appear to be very different to what their theory projects. We find out that a flaw in the derivation of the second order statistics of $b_{i,j}$ in [9]. we present their results as well as our correction. Later, in Section 3.3.2 we will observe that our simulation results match the corrected analytical results.

Let us first look into the statistics of the integrand, $|n^o(x,y)|^2$. The mean of $|n^o(x,y)|^2$ is

$$\mathbf{E}\{|n^o(x,y)|^2\} = \mathbf{E}\{n^o(x,y)n^{o^*}(x,y)\} = K_{n^o}(0,0), \tag{3.63}$$

where $K_{n^o}(x,y)$ is the covariance of the optical scatter noise process given by Equation (3.13). In order to compute the auto-covariance of $|n^o(x,y)|^2$, $K_{|n^o|^2}(x,y) = \mathbf{E}\{|n^o(x,y)|^2|n^o(0,0)|^2\} - \mathbf{E}^2\{|n^o(x,y)|^2\}$, Chugg et al. resort to Reed's theorem for complex Gaussian random variables [37] and get

$$K_{|n^o|^2}(x,y) = |K_{n^o}(x,y)|^2 . \tag{3.64}$$

Based on Equation (3.64) we conclude that,

$$S_{|n^o|^2}(f_x,f_y) = S_{n^o}(f_x,f_y) \star S_{n^o}(-f_x,-f_y). \tag{3.65}$$

Equation (3.65) and the band-limited property of $n^o(x,y)$ shows that $|n^o(x,y)|^2$ is also a low-pass and band-limited process with cutoff frequency of ω/Δ.

After discussing the statistics of $|n^o(x,y)|^2$, we present the results of Chugg et al. [9] on the statistics of the process $b_{i,j}$:

$$\begin{aligned}
\mathbf{E}\{b_{i,j}\} &= \beta^2\Delta^2 K_{n^o}(0,0) && (3.66)\\
K_b(i,j) &= \mathbf{E}\{e_{i,j}e_{0,0}\} - \mathbf{E}^2\{e_{0,0}\} \\
&= \frac{1}{(2\beta\Delta)^2}\iint_{-\beta\Delta}^{\beta\Delta}(1-\frac{|x|}{\beta\Delta})(1-\frac{|y|}{\beta\Delta}) \\
&\quad \times |K_{n^o}(i\Delta+x, j\Delta+y)|^2 dxdy. && (3.67)
\end{aligned}$$

The auto-covariance function Equation (3.67) is incorrect. In their derivation they use the equality

$$\begin{aligned}
\iiiint_{-\beta\Delta/2}^{\beta\Delta/2} f(x-\acute{x}, y-\acute{y})dxdyd\acute{x}d\acute{y} &= \frac{1}{4(\beta\Delta)^2}\iint_{-\beta\Delta}^{\beta\Delta}(1-\frac{|x|}{\beta\Delta})(1-\frac{|y|}{\beta\Delta}) \\
&\quad \times f(x,y)dxdy, && (3.68)
\end{aligned}$$

where they actually considered $\beta = 1$. The correct equality is

$$\begin{aligned}
\iiiint_{-\beta\Delta/2}^{\beta\Delta/2} f(x-\acute{x}, y-\acute{y})dxdyd\acute{x}d\acute{y} &= (\beta\Delta)^2\iint_{-\beta\Delta}^{\beta\Delta}(1-\frac{|x|}{\beta\Delta})(1-\frac{|y|}{\beta\Delta}) \\
&\quad \times f(x,y)dxdy. && (3.69)
\end{aligned}$$

For a simple check one may substitute the function $f(x,y) = 1$ to see that in Equation (3.68) the right hand side does not match the left hand side;whereas in Equation (3.69) they both equal $(\beta\Delta)^4$. We present a detailed proof for Equation (3.69) next that you may safely escape.

Proof. Let us define $u = x - \acute{x}$ and $v = y - \acute{y}$. Hence,

$$\iiiint_{-\beta\Delta/2}^{\beta\Delta/2} f(x-\acute{x}, y-\acute{y})dxdyd\acute{x}d\acute{y} \;=\; \iiiint_{\Xi} f(u,v)dxdydudv, \quad (3.70)$$

where Ξ is the integration region. Since it is a four-dimensional space, it is impossible to visualize it. From the four variables, we can take u and v as free variables and $x = u + \acute{x}$, $y = v + \acute{y}$. Hence, the right hand side of Equation (3.70) is equal to

$$\begin{aligned}\iiiint_{\Xi} f(u,v)dxdydudv \;&=\; \iint_{-\beta\Delta/2}^{\beta\Delta/2} f(u,v) \\ &\quad \times \left(\iint_{\acute{\Xi}_{u,v}} dxdy \right) dudv, \end{aligned} \quad (3.71)$$

where $\acute{\Xi}_{u,v}$ is an integration region in the two-dimensional space which depends u and v. In addition,

$$\iint_{\acute{\Xi}_{u,v}} dxdy = \text{Area of the region } \acute{\Xi}_{u,v}, \quad (3.72)$$

since the integrand is unity. For a fixed u and v, the equalities $x = u + \acute{x}$ and $y = v + \acute{y}$ and the fact that all variables should be in the range $[-\beta\Delta/2, \beta\Delta/2]$, we conclude that

$$\acute{\Xi}_{u,v} = \begin{cases} \max(-\beta\Delta/2, u - \beta\Delta/2) \le x \le \min(\beta\Delta/2, u + \beta\Delta/2), \\ \max(-\beta\Delta/2, v - \beta\Delta/2) \le y \le \min(\beta\Delta/2, u + \beta\Delta/2). \end{cases} \quad (3.73)$$

This area is the intersection of two squares. Both has a width of $\beta\Delta$, one is centered at the origin in the (x, y) plane and one is centered at the point (u, v) in the (x, y)

plane. Hence,

$$\begin{aligned} \text{Area of the region } \acute{\Xi}_{u,v} &= (\beta\Delta)^2 - (\beta\Delta)|u| - (\beta\Delta)|v| + (\beta\Delta)^2|uv| \\ &= (\beta\Delta)^2(1 - \frac{|u|}{\beta\Delta})(1 - \frac{|v|}{\beta\Delta}). \end{aligned} \tag{3.74}$$

If we substitute Equation (3.74) into Equation (3.72) we get

$$\iint_{\acute{\Xi}_{u,v}} dxdy = (\beta\Delta)^2(1 - \frac{|u|}{\beta\Delta})(1 - \frac{|v|}{\beta\Delta}). \tag{3.75}$$

If we substitute Equation (3.75) into Equation (3.71), the theorem is proved. □

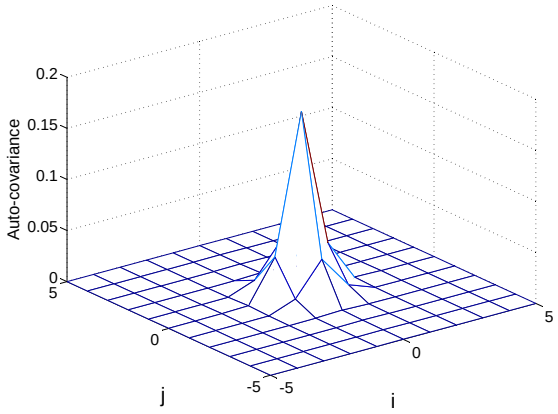


Figure 3.5: Auto-covariance of $K_b(i,j)$ for $E_0 = 1$, $\omega = 1$, $\alpha = \beta = 1$, $\Delta = 1$.

Using 3.69 in their derivation, we get the correct second order statistics as

$$\begin{aligned} K_b(i,j) &= (\beta\Delta)^2 \iint_{-\beta\Delta}^{\beta\Delta} (1 - \frac{|x|}{\beta\Delta})(1 - \frac{|y|}{\beta\Delta}) \\ &\quad \times |K_{n^o}(i\Delta + x, j\Delta + y)|^2 dxdy. \end{aligned} \tag{3.76}$$

Figure 3.5 gives a 2-D view of $K_b(i,j)$ as a function of i and j.

Next we present a reduced complexity method to compute Equation (3.76). Since $h(x,y) = \acute{h}(x)\acute{h}(y)$ is separable we can write $K_{n^o}(x,y) = E_0 h(x,y) \star h(-x,-y)$ as

$$K_{n^o}(x,y) = K^1_{n^o}(x) K^1_{n^o}(y), \tag{3.77}$$

where

$$K^1_{n^o}(x) = \sqrt{E_0}\acute{h}(x) \star \acute{h}(-x). \tag{3.78}$$

In order to simplify the 2-D numerical integration in Equation (3.76) we use the

separability of $K_{n^o}(x,y)$ to write $K_b(i,j)$ as

$$K_b(i,j) \;=\; K_e^1(i)K_e^1(j), \tag{3.79}$$

where

$$K_e^1(i) = \beta\Delta \int_{-\beta\Delta}^{\beta\Delta} (1 - \frac{|x|}{\beta\Delta})|K_{n^o}^1(i\Delta + x)|^2 dx. \tag{3.80}$$

Following a tedious mathematical derivation He [13] proved the inequality

$$K_b(0,0) \leq \beta^4\Delta^4|K_{n^o}(0,0)|^2. \tag{3.81}$$

We can simply prove the inequality in the following way noting that $|K_{n^o}(x,y)| \leq |K_{n^o}(0,0)|$ and upper bounding the integrand in 3.76.

3.3 Numerical Results

First, we investigate the accuracy of the simplified model of the information-bearding component $a_{i,j}$ in Section 3.3.1. Next, we proceed to present the simulation results for $b_{i,j}$ in Section 3.3.2. We validate our computation method against the analytical results of $b_{i,j}$. In the end, we investigate the complexity reduction provided by our methods.

In a our simulations we the HDS parameters as following:

$\Delta = 1$	Pixel spacing for the SLM/CCD
$\omega = 1$	Normalized aperture width
$\alpha = 1$, $\beta = 1$	The fill-factor for the SLM/CCD pixels
$2L + 1 = 21$	The IPI span
$E_0 = 1$	Refer to auto-covariance of $K_{n^o}(x, y)$in (3.13)
$\epsilon = 100$	The SLM contrast ratio
$N = 1024$	The number of pixels along each dimension of the data page

(3.82)

3.3.1 Accuracy of $a_{i,j}$ Simplification

In order to assess the accuracy of the one branch DCM model we adopt the approach of [8] which quantifies accuracy based on the normalized mean-squared error (NMSE). The NMSE is defined as the ratio (mean-squared error between the intensity sequences coming from the HDS as computed by (3.29) and the simplified DCM output as computed by (3.35))/(mean power of the HDSs intensity sequence computed by (3.29)).

Several different factors such as SLM contrast ratio, normalized aperture width, SLM and CCD fill factors, and finally pixel misalignment affect the accuracy of the one branch DCM. For a channel with no pixel misalignments, the NMSE results are studied by [8]. We are particularly interested in the effect of pixel misalignment on the NMSE. To investigate further NMSE plots as a function of δ_x/Δ,

the normalized pixel misalignment along axis x, are presented. The value of δ_x varies in the range of 0 - 0.5Δ, where $\delta = 0.5\Delta$ is the most severe case of pixel misalignment along an axis. To study the effect of each factor, the rest of factors are fixed to typical values and three plots are sketched for $\delta_y = 0.1\Delta$, $\delta_y = 0.25\Delta$, and $\delta_y = 0.5\Delta$.

Figure 3.6 illustrates the effect of the pixel misalignment on the NMSE for various normalized aperture widths (ω). The SLM/CCD fill factors are 100% and SLM contrast ratio is set to 100. We observe that the pixel misalignment highly impact the NMSE. First of all, in Figure 3.6(a) we observe that for a limited pixel misalignment the NMSE is maximum at Nyquist aperture ($\omega = 1$) (The NMSE results in [8] show this as well); nevertheless for higher values of the pixel misalignment, larger apertures lead to higher NMSE values. Nyquist and sub-Nyquist apertures ($\omega \leq 1$) tend to maintain a constant error level as the pixels lose their alignment. However, for super-Nyquist apertures ($\omega > 1$) the NMSE steeply increases as the pixel misalignment increases. The maximum NMSE of 3.63% occurs at $\omega = 1.2$, $\delta_x = 0.5\Delta$, and $\delta_y = 0.5\Delta$ for the given set of parameters.

Figure 3.7 illustrates the effect of the pixel misalignment on the NMSE for various CCD fill factors (β). The SLM fill factor is 100% and the SLM contrast ratio is set to 100. We use Nyquist aperture ($\omega = 1$). We observe that the behavior of the NMSE is consistent for pixel aligned and pixel mis-aligned channels. In other words, increase of the CCD fill factor leads to substantial increase of the NMSE.

However, for the given set of system parameters, the NMSE does not significantly change as pixel misalignment increases. The maximum NMSE is 3.63% for the given set of parameters. The maximum NMSE of 1.24% occurs at $\beta = 1$, $\delta_x = 0.5\Delta$, and $\delta_y = 0.5\Delta$ for the given set of parameters.

Figure 3.8 illustrates the effect of the pixel misalignment on the NMSE for various SLM fill factors (α). The CCD fill factor is 100% and SLM contrast ratio is set to 100. We use Nyquist aperture ($\omega = 1$). We observe that the NMSE slightly decreases as the SLM fill factor increases. The maximum NMSE of 1.24% occurs at $\beta = 1$, $\delta_x = 0.5\Delta$, and $\delta_y = 0.5\Delta$ for the given set of parameters.

Figure 3.9 illustrates the effect of the pixel misalignment on the NMSE for various SLM contrast ratios (ϵ). The SLM/CCD fill factors are 100% and Nyquist aperture is used. We observe that the increase of the SLM contrast ratio leads to substantial increase of the NMSE. However, for the given set of system parameters, the NMSE does not significantly change as pixel misalignment increases. The maximum NMSE of 1.26% occurs at $\epsilon = 100$, $\delta_x = 0.5\Delta$, and $\delta_y = 0.5\Delta$ for the given set of parameters.

We summarize the above results: 1) For the Nyquist aperture the NMSE does not change much as pixel misalignment varies. The maximum NMSE is around 1.8% and occurs when $alpha = 0.8$. 2) The NMSE increases steeply as pixel misalignment increases for $\omega \geq 1$. 3) The NMSE increases as normalized aperture

width, CCD fill factor or SLM contrast ratio increase. The only exception is the pixel aligned channel, where NMSE decreases when ω increases from 1 to 1.2. 5) The NMSE increases as SLM fill factor decrease. 4) For reasonable configurations, $\omega = 1.2$, $\alpha = 0.8$, $\beta = 1$, $\epsilon = 100$ lead to the maximum NMSE 0f 6.31% at $\delta_x = \delta_y = 0.5\Delta$. We conclude that for Nyquist or sub-Nyquist apertures, the one branch DCM is accurate to around 2.0% NMSE. For $\omega > 1$, the accuracy depends on the pixel misalignment and the SLM fill factor and special care should be taken with small SLM fill factors and high pixel misalignments.

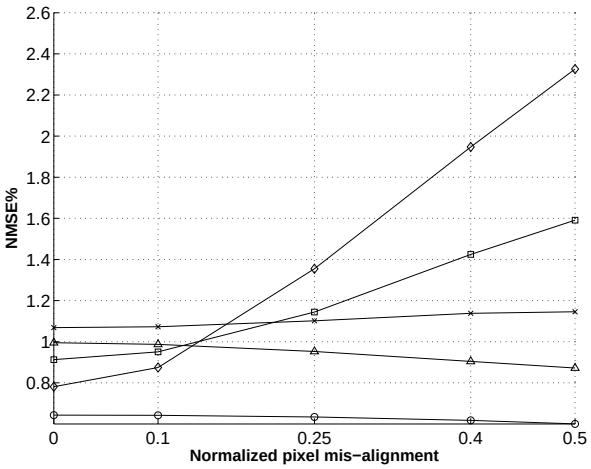
3.3.2 Validation of $b_{i,j}$

The system parameters are $\omega = 1$, $\Delta = 1$, $E_0 = 1$, $\alpha = \beta = 1$. The analytical value for mean of $b_{i,j}$ is

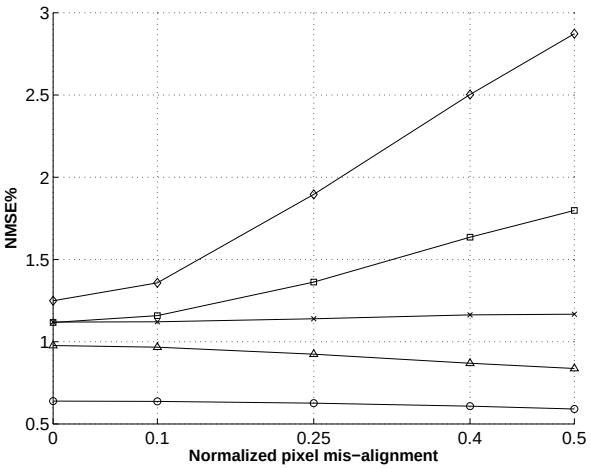
$$\mathbf{E}\{b_{i,j}\} = \beta^2\Delta^2 K_{n^o}(0,0) = 0.5925. \tag{3.83}$$

Our simulations show $\mathbf{E}\{b_{i,j}\} = 0.5905$ which is consistent with the theory. Matrices in 3.84 and 3.85 correspond to analytical results and statistical averaging over simulated $b_{i,j}$ respectively.

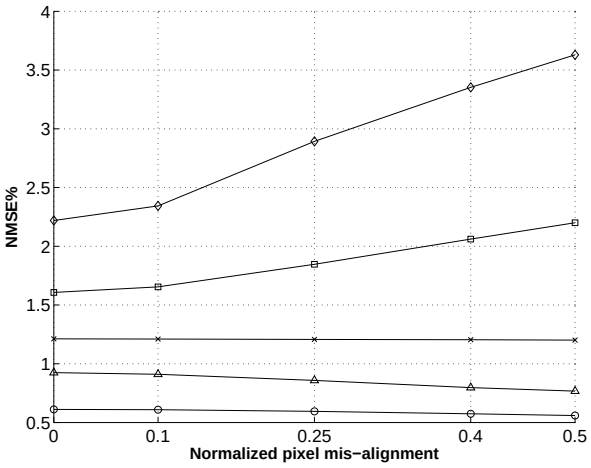
$$K_{i,j}^{Analytical} = \begin{pmatrix} 0.0098 & 0.0418 & 0.0098 \\ 0.0418 & 0.1772 & 0.0418 \\ 0.0098 & 0.0418 & 0.0098 \end{pmatrix} \tag{3.84}$$

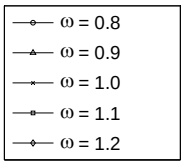


(a) $\delta_y = 0.1\Delta$



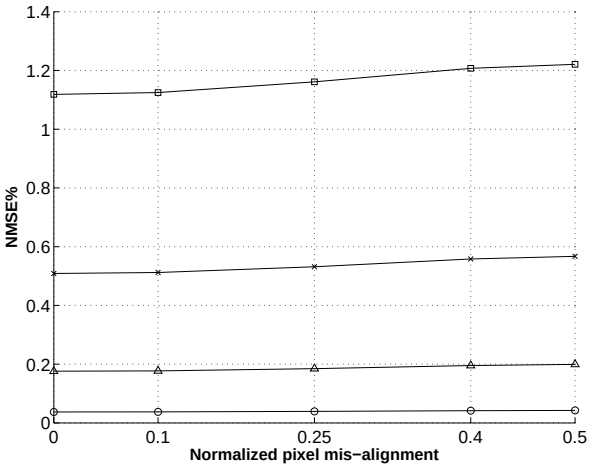
(b) $\delta_y = 0.25\Delta$



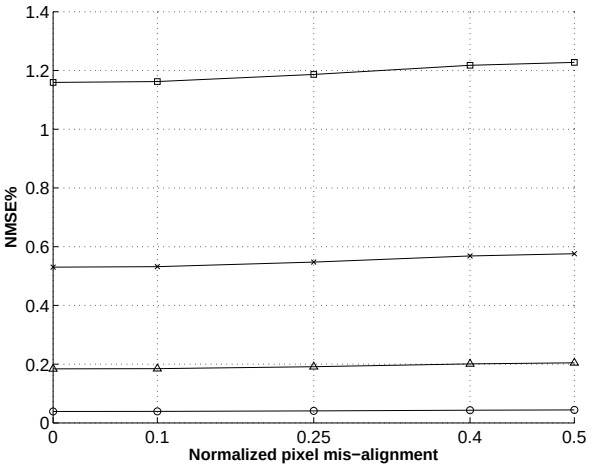


(c) $\delta_y = 0.5\Delta$

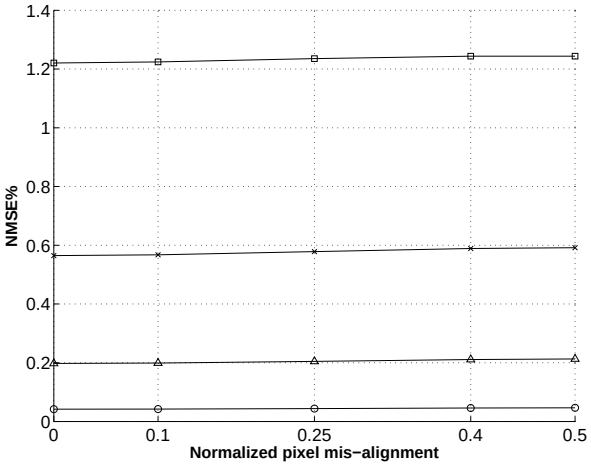
Figure 3.6: NMSE as a function of the normalized pixel misalignment δ_x/Δ for various normalized aperture widths, $\beta = 1$, $\alpha = 1$, $\epsilon = 100$, and δ_y ,(a) 0.1Δ, (b) 0.25Δ, (c) 0.5Δ.

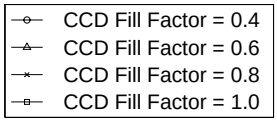


(a) $\delta_y = 0.1\Delta$



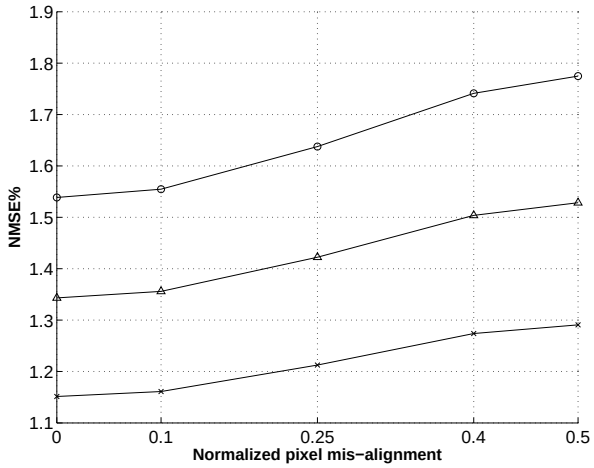
(b) $\delta_y = 0.25\Delta$



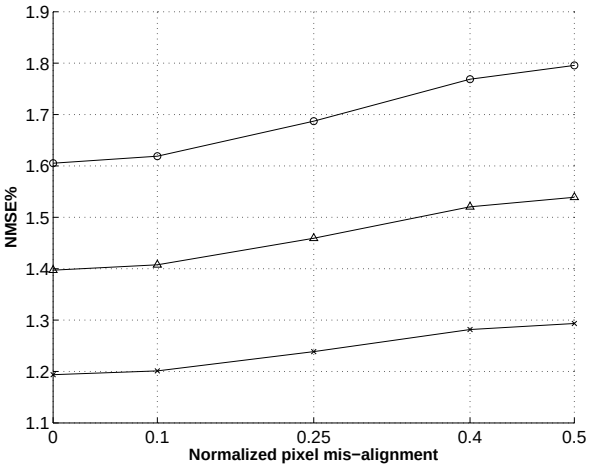


(c) $\delta_y = 0.5\Delta$

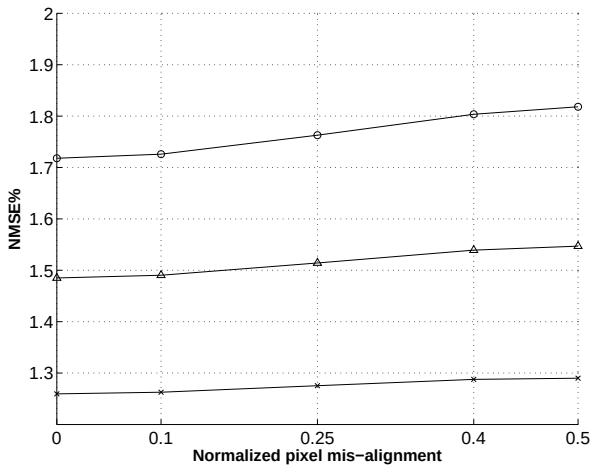
Figure 3.7: NMSE as a function of the normalized pixel misalignment, δ_x/Δ for various CCD fill factors, $\omega = 1$, $\alpha = 1$, $\epsilon = 100$, and δ_y ,(a) 0.1Δ, (b) 0.25Δ, (c) 0.5Δ.

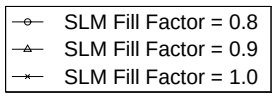


(a) $\delta_y = 0.1\Delta$



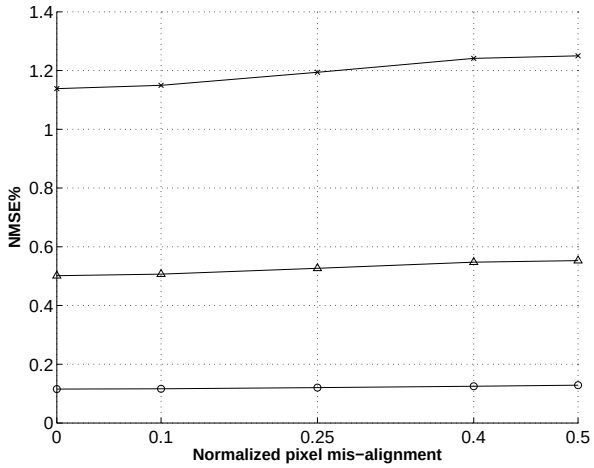
(b) $\delta_y = 0.25\Delta$



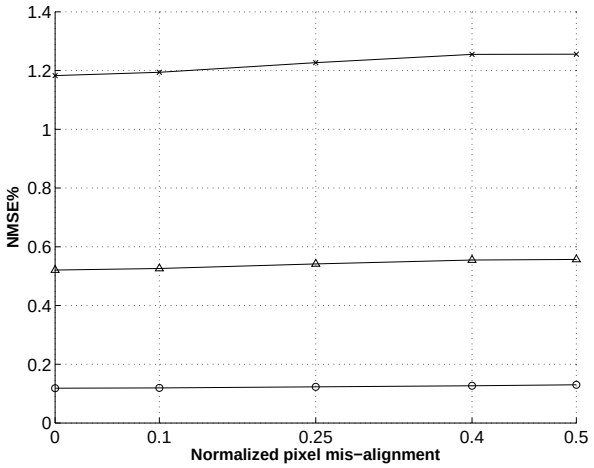


(c) $\delta_y = 0.5\Delta$

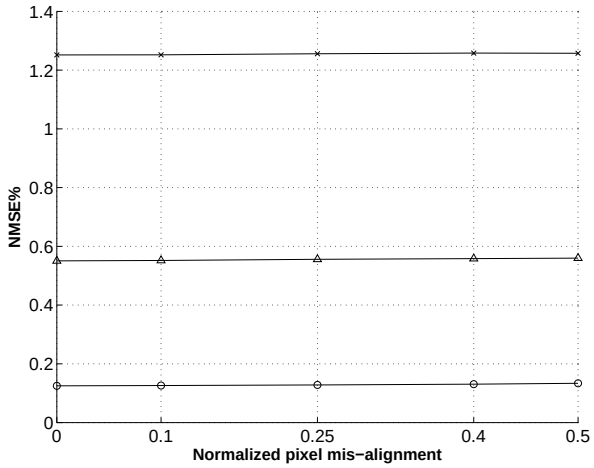
Figure 3.8: NMSE as a function of the normalized pixel misalignment, δ_x/Δ for various SLM fill factors, $\omega = 1$, $\beta = 1$, $\epsilon = 100$, and δ_y ,(a) 0.1Δ, (b) 0.25Δ, (c) 0.5Δ.

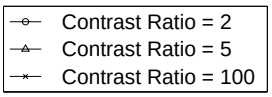


(a) $\delta_y = 0.1\Delta$



(b) $\delta_y = 0.25\Delta$





(c) $\delta_y = 0.5\Delta$

Figure 3.9: NMSE as a function of the normalized pixel misalignment, δ_x/Δ for various contrast ratios, $\omega = 1$, $\alpha = 1$, $\beta = 1$, and δ_y ,(a) 0.1Δ, (b) 0.25Δ, (c) 0.5Δ.

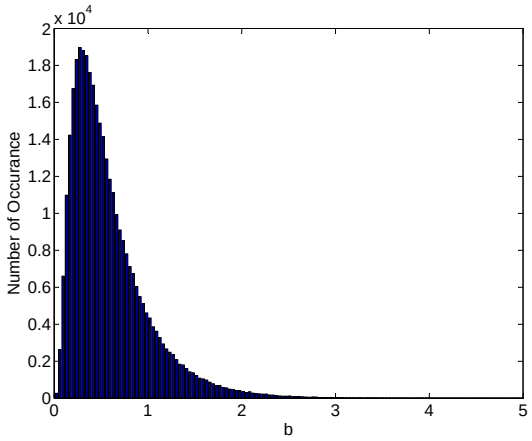


Figure 3.10: Histogram of $b_{i,j}$ for $E_0 = 1$, $\omega = 1$, $\alpha = \beta = 1$, $\Delta = 1$.

$$K_{i,j}^{Averaging} = \begin{pmatrix} 0.0095 & 0.0412 & 0.0095 \\ 0.0418 & 0.1767 & 0.0418 \\ 0.0095 & 0.0412 & 0.0095 \end{pmatrix} \tag{3.85}$$

Figure 3.10 illustrates the histogram of $b_{i,j}$. The complexity of computing the information-bearing component of the CCD read-back signal, $a_{i,j}$ is same as that of Keskinoz and Kumar [8]. Using the separability property of the underlying linear channel in the DCM model enables us to characterize the channel for N^2 pixel misalignment lays by storing only N vectors. Furthermore, we can use the separability property to reduce the complexity of computing the two-dimensional convolution in (3.35) from $O((2L+1)^2)$ to $O(2L+1)$.

Recalling the discussion in 3.2.3, the major bottleneck regarding the computing of $b_{i,j}$ and $c_{i,j}$ is related to generating the samples of the optical noise. Based on

the simulation parameters given in 3.82, the data page size is 1024×1024. If we intend to compute (3.21) and (3.23) by numerical integration; we need at least a sampling rate of $T_s = \Delta/5$ (i.e. $M = 5$ points per pixel) for acceptable accuracy. Consequently, based on (3.40) and (3.41) the size of the optical noise page would be 5120×5120 and the filter size would be 105×105 respectively. However, utilizing our method in Section 3.2.3 needs a sampling rate of $T_s = \Delta/2$ (i.e. $M = 2$ points per pixel) which results in a much smaller noise page size of 2048×2048 and coloring filter size of 42×42.

3.4 Model Comparison

We end this article by comparing our model with that of Keskinoz and Kumar [8] and He [13].

Keskinoz and Kumar [8] model for the HDS with electronics and optical noise is given as

$$\begin{aligned} I_{i,j} &= |d_{i,j} \otimes h_{i,j} + n^o_{i,j}|^2 + n^e_{i,j}, \\ &= |d_{i,j} \otimes h_{i,j}|^2 + |n^o_{i,j}|^2 + 2\Re(d_{i,j} \otimes h_{i,j}) n^{o^*}_{i,j} + n^e_{i,j}, \end{aligned} \tag{3.86}$$

where $I_{i,j}$ denotes the CCD read-back signal, $d_{i,j}$ denotes the data bits, $h_{i,j}$ is the DCM (for pixel aligned HDS), $n^o_{i,j}$ denotes the optical noise and $n^e_{i,j}$ denotes the CCD electronics noise. They assume that $n^o_{i,j}$ is white and Gaussian. However,

since the optical noise is colored as Equation (3.13) suggests, Equation (3.86) is not accurate. Furthermore, the term $|n^o_{i,j}|^2$ is exponentially distributed whereas the distribution of its counterpart, $b_{i,j}$, which is shown in Figure 3.10, does not appear to be exponential. Based on these arguments, we believe that Equation 3.86 is accurate as long as the optical noise is negligible (electronics noise dominated channel). The main difference between our model and He [13]'s model is that He [13] ignores the dependency among the three components of the read-back signal and investigates methods to generate them separately. He [13] derived simple expressions for the first and second order statistics of each component for a typical configuration of parameters as

$$\mathbf{E}\{a_{i,j}\} \approx 0.5\Delta^2, \tag{3.87}$$

$$\text{var}\,(a_{i,j}) \approx 0.1\Delta^4, \tag{3.88}$$

$$\mathbf{E}\{b_{i,j}\} = \beta^2\Delta^2 K_{n^o}(0,0), \tag{3.89}$$

$$\text{var}\,(b_{i,j}) \leq \beta^4\Delta^4 \left(K_{n^o}(0,0)\right)^2, \tag{3.90}$$

$$\mathbf{E}\{c_{i,j}\} = 0 \tag{3.91}$$

$$\text{var}\,(c_{i,j}) = 0.5\beta^4\Delta^4 K_{n^o}(0,0). \tag{3.92}$$

Based on these equations He [13] argues that when the power of the optical noise is low, the variance of $b_{i,j}$ is much smaller compared to those of $a_{i,j}$ and $c_{i,j}$. In other words, the relative variations in $b_{i,j}$ about its mean are much less compared to those of $a_{i,j}$ and $c_{i,j}$. Hence, He [13] approximates $b_{i,j}$ as a constant noise floor located

at its mean value in his channel model. In order to compute $c_{i,j}$, He [13] develops an efficient approximation method; however, as we have already stated $c_{i,j}$ is a colored process whereas in his method it is considered white. Furthermore, there is a dependency between $b_{i,j}$ and $c_{i,j}$ which is not considered in [13]. Comparing to [13] and [8], our model is more complex. However, its derivation does not involve major approximations; hence it is more accurate and follow the statistics of the actual read-back signal closely.

If we ignore the connection between $c_{i,j}$ and $b_{i,j}$, we may generate $c_{i,j}$ independently by a fast routine which does not require to compute $\tilde{n}_{i,j}$ in Equation (3.56). In order to find auto-covariance of the $\hat{n}_{i,j}$ over i and j we need to compute the auto-covariance $c_{i,j}$ and then compute the analytical relationship between $K_{c_{i,j}}$ and $K_{\hat{n}_{i,j}}$. Unlike $\tilde{n}_{i,j}$ which is the average of the optical scatter noise, $\hat{n}_{i,j}$ has no physical meaning , and we can not incorporate the optical noise information; so it does not result in $c_{i,j}$ which is statistically consistent with $b_{i,j}$. If power of the optical scatter noise is low, we can treat $b_{i,j}$ as a constant and this second method is useful. However, for high levels of noise where we want $b_{i,j}$ and $c_{i,j}$ to be consistent, we incorporate the optical noise samples to generate $b_{i,j}$ and $c_{i,j}$.

3.5 Conclusion

An accurate model for the HDS is of fundamental importance since it allows for accurate performance evaluation of signal processing algorithms such as equalizers, detectors etc. The discrete magnitude-squared channel model of Keskinoz and Kumar [8] is a one of they most important models for the HDS channels. This importance stems from the fact that The DCM model provides us with a useful insight into the underlying structure of the inter-pixel interference(IPI) in the HDS. As we will see in the Chapter 4, this model inspires us to design a robust detector for HDS. However, this model only works for pixel aligned channels. We have extend the DCM model of Keskinoz and Kumar for channels with pixel misalignment.

Analysis and simulation of the detector read-back signal in the presence of the optical noise is difficult because of its quadratic nonlinearity. As another contribution of this work, we present a new method for efficient and accurate simulation of the detector read-back signal in the presence of the optical noise.

Chapter 4

Soft-Decision Nonlinear Two-Dimensional Reception Scheme for Holographic Data Storage

Our discussion in Section 2.1.2 shows that as the storage density increases, the inter-pixel interference (IPI) increases as a result of using smaller apertures [2]. Furthermore, SNR is inversely proportional to the square of the number of holograms multiplexed in each stack [1, 11]. Consequently, low SNRs are inevitable as density increases. Accordingly, developing soft-decision detectors for HDS channels is of fundamental importance, since such detectors are readily integrable with low-density parity-check (LDPC) codes into iterative reception schemes that can

achieve near-optimal BER performance at low SNRs. Most existing detection techniques for HDS [9], [8], [38] produce hard decisions and hence do not fit this bill. A notable exception is [19]. Unfortunately, complexity of the detector of [19] tends to be very high.

An important soft-decision detector is BCJR [30] which is an optimal symbol-by-symbol MAP detector. Unfortunately, the absence of natural ordering in the 2-D page of data renders the generalization of BCJR algorithm difficult as it relies on the sequential nature of the data. In Section 2.2.3 we explained how [22] exploit the separability property to tackle the lack of natural ordering in linear 2-D channels. In a linear separable channel, the 2-D ISI can be viewed as a concatenation of the ISI along the rows and the ISI along the columns. The work reported in this chapter is based on our observation in Section 3.2.2: The HDS channel, while nonlinear, has a similar separability property. This key observation permits us to extend the 2-D BCJR detector of [22] to deal with the quadratic nonlinearity of the HDS channel at no additional complexity.

For an ISI span of $(2L+1) \times (2L+1)$, the detector complexity in [19] is exponential in $(2L+1)^2$ while complexity of the detector in [22] is exponential in $2L+1$ (Refer to Sections 2.2.2 and 2.2.3 respectively). Although, the resulting complexity is much lower than that of [19], it can still be very high at high densities (i.e. for large ISI spans). To limit complexity further, we resort to partial-response (PR) techniques that limit the ISI span prior to detection. To this end, we introduce

a new target signal that involves the same nonlinear mechanism as in the HDS channel.

A further important challenge in HDS is pixel misalignment. In practice it is impossible to achieve perfect pixel alignment due to a variety of adverse factors. The effect of pixel misalignment is substantial. As [11] showed, pixel misalignment significantly deteriorates the detection performance and even sometimes brings the achievable density to zero. Reference [11] presented a nonlinear algorithm to mitigate the effects of pixel misalignment. However, this algorithm fails to work with an acceptable performance for high levels of pixel misalignment. Unlike the approach of [11], our technique works well even for severe misalignments. Furthermore, [11] uses decision feedback; consequently, error propagation may arise at low SNRs. Conversely, combinations of BCJR detectors and LDPC codes are known for their excellent performance at low SNRs. We also test linear PR targets along with the 2-D BCJR detector of [22] (Refer to Section 2.2.3), and observe a poor BER performance for severe misalignment. This clearly illustrates the necessity of accommodating nonlinearity in the 2-D BCJR for HDS channels.

This chapter is organized as follows: In Section 4.1 we review the separability property of the HDS system which is of fundamental importance in our detector design. In Section 4.2 we discuss our reception technique including the 2-D quadratic BCJR detector and the quadratic partial response (PR) signal. In Section 4.4 we discuss the numerical results and finally we present our conclusions in Section 4.5.

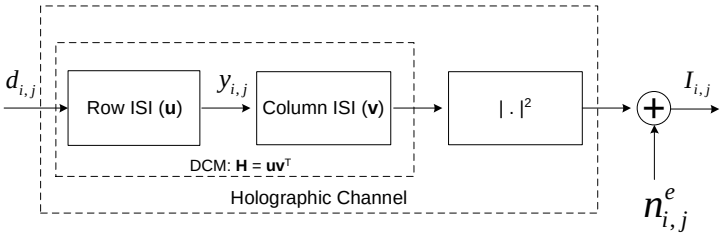


Figure 4.1: Electronics-Noise Dominated HDS Channel Model.

4.1 Channel Model

We use the discrete magnitude-squared channel model derived in Section 3.2.2 to simulate CCD read-back values in HDS. We consider detector electronics noise only, which is zero-mean, additive, white, and Gaussian. In other words, we assume an electronics-noise dominated channel (ENDC) [8] . The simplified channel model with no optical noise is illustrated in Figure 4.1. We denote the input data bits by $d_{i,j}$. These bits pass through a linear 2-D ISI channel characterized by a discrete channel matrix (DCM) $\mathbf{H}^{\delta_x,\delta_y}$ (We use the shorthand $\mathbf{H}$ in Figure 4.1). The ISI span is $(2L+1)\times(2L+1)$ which means that $(2L+1)^2$ pixels interfere for each read-back value of the CCD. As shown in the Equation (3.37), $\mathbf{H}^{\delta_x,\delta_y}$ is separable. Separability of the DCM $\mathbf{H}^{\delta_x,\delta_y}$ allows us to consider 2-D ISI as the concatenation of two channels representing row and column ISI respectively similar to Section 2.2.3. The only difference is that now the magnitude of the column ISI channel is squared to generate the noiseless channel output. White Gaussian noise $n^e_{i,j}$ is added afterward to produce the CCD read-back value $I_{i,j}$. Now the CCD read-back signal $I_{i,j}$ in

$$\begin{pmatrix} 0.0005 & -0.0027 & -0.0169 & -0.0027 & 0.0005 \\ -0.0027 & 0.0156 & 0.0972 & 0.0156 & -0.0027 \\ -0.0169 & 0.0972 & \mathbf{0.6045} & 0.0972 & -0.0169 \\ -0.0027 & 0.0156 & 0.0972 & 0.0156 & -0.0027 \\ 0.0005 & -0.0027 & -0.0169 & -0.0027 & 0.0005 \end{pmatrix}$$

$$\begin{pmatrix} 0.0027 & -0.0043 & 0.0288 & 0.0288 & -0.0043 \\ -0.0043 & 0.0069 & -0.0456 & -0.0456 & 0.0069 \\ 0.0288 & -0.0456 & \mathbf{0.3029} & 0.3030 & -0.0456 \\ 0.0288 & -0.0456 & 0.3030 & 0.3032 & -0.0456 \\ -0.0043 & 0.0069 & -0.0456 & -0.0456 & 0.0069 \end{pmatrix}$$

Figure 4.2: Discrete channel matrix for the pixel-aligned channel $\mathbf{H}^{\mathbf{0.0,0.0}}$ (Up) and the pixel-misaligned channel $\mathbf{H}^{\mathbf{0.5,0.5}}$ (Down).

the absence of the optical noise is :

$$I_{i,j} = \left| [h_{i,j}^{\delta_x,\delta_y}] \otimes [d_{i,j}] \right|^2 + n_{i,j}^e \tag{4.1}$$

where $\otimes$ denotes 2D convolution.

From this point on, we will refer to a HDS by its corresponding DCM. We consider two HDS channels. The first channel is a perfectly pixel-aligned HDS channel, referred to as *pixel-aligned HDS*. Second channel is a HDS channel with half pixel misalignment in both x and y directions, simply referred to as *pixel-misaligned HDS*. Figure 4.2 illustrates the DCM of these channels. The entry at the center of the DCM is presented in bold face. All the other entries correspond to interfering pixels. Note that $\mathbf{H}^{0.5,0.5}$ represents the most severe case of pixel misalignment where we can see that 4 pixels contribute almost equally to the read-back value.

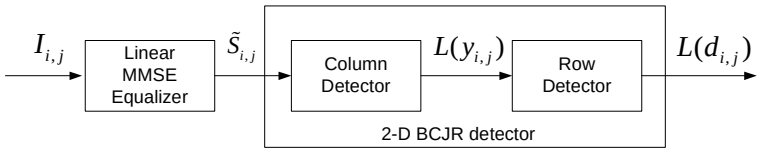


Figure 4.3: Reception for nonlinear separable channel.

4.2 Reception Technique

As shown in Figure 4.3, our reception technique is comprised of a linear minimum-mean-squared-error (MMSE) equalizer and the 2-D BCJR detector. Linear MMSE equalization for HDS channels is investigated in detail by [9].

First, we equalize the channel output, $I_{i,j}$. The equalizer output $\tilde{s}_{i,j}$ is an estimate of the target signal $s_{i,j}$ to be described later in Section 4.2.2. For the time being, it is enough to know that the HDS channel and the target signal follow the same model and the only difference between them is the size of their DCM. The 2-D BCJR detector in Figure 4.3 is slightly different than the one described in Section 2.2.3.

Let us first explain the BCJR detector.

4.2.1 The Quadratic Reduced-Complexity 2-D BCJR Detector

The BCJR detectors we use here are based on the 2-D BCJR detectors in [22] explained in Section 2.2.3. Detectors in [22] were developed for separable linear

additive white gaussian noise channels illustrated in Figure 2.2.

As Figure 4.1 suggests, the DCM is separable. However, the ENDC HDS channel is not linear because of the squaring operation. Thus, we have to modify the detectors to work with the nonlinear HDS channel. First we use a BCJR detector similar to the column detector in Section 2.2.3 to produce the log-likelihood ratio (LLR) values for the intermediate signal $y_{i,j}$. We only have to modify the branch values of the column detector trellis based on the nonlinearity of the HDS channel. In other words, we only need to substitute the values $z(\hat{s}, s)$ by their squares. The remaining column detector equations and the entire row detector equations are kept unchanged.

In Figure 4.3, we denote the output of the column detector by $L(y_{i,j})$. The column detector then passes $L(y_{i,j})$ to another binary BCJR detector to compute the LLR of data bits, denoted by $L(d_{i,j})$. We refer to this detector as the row detector and its structure is exactly the same as the row detector in [22]. We decide on the bit values based on the sign of $L(d_{i,j})$.

The original row and column BCJR detectors exchange extrinsic information $L^{cE}(y_{i,j} = Y)$ and $L^{rE}(y_{i,j} = Y)$ and exhaust a pre-defined number of iterations to compute the final LLR. This operation is computationally demanding, and it is sensible to do so only if the BCJR detectors are exchanging information with an error correction code. Such an scheme is described in [22]. Otherwise, exchanging

extrinsic information does not significantly improve the detection BER since the same information in used in each iteration. Our simulations certify this fact; thus, we use only one iteration and the BCJR detectors in Figure 4.3 do not exchange information.

4.2.2 New Magnitude-Squared Partial Response Signal

In order to limit the complexity of the 2D-BCJR we choose a partial response signal that has an $S \times S$ support for $S < 2L + 1$. We present the following 2-D signal

$$s_{i,j} = |\widetilde{\gamma}_{i,j} \otimes [d_{i,j}]|^2 \tag{4.2}$$

as the equalization target. The structure of Equation 4.2 is identical to that of Equation 4.1, where $\widetilde{\gamma}_{i,j}$ are target coefficients that control the shape of the target signal. We present the target coefficients $\widetilde{\gamma}_{i,j}$ in matrix form, and we constrain the matrix to be separable, i.e.

$$\mathbf{\Gamma} = [\widetilde{\gamma}_{i,j}] = \mathbf{x}\mathbf{y}^T. \tag{4.3}$$

Vectors $\mathbf{x}$ and $\mathbf{y}$ are $S \times 1$. For the rest of this paper, we refer to such a target signal by its underlying matrix $\mathbf{\Gamma} = [\widetilde{\gamma}_{i,j}]$.

Nonlinearity is incorporated so that the signal can to be very close to the channel output. Hence, less equalization effort is needed and better noise-whitening

is achieved. This will improve the performance of the BCJR detector. Furthermore, since $\mathbf{\Gamma}$ is separable, we can still use the simplified BCJR detector of Figure 4.3.

We equalize the channel to a target signal with a support size of 2×2. Given this support size, the column detector traverses a non-binary trellis with four states and the row detector traverses a binary trellis with two states.

4.3 Equalizer and Target Optimization

The error signal between the equalizer output and the corresponding target signal is

$$e_{i,j} = s_{i,j} - \tilde{s}_{i,j}. \tag{4.4}$$

In order to derive the expression for mean-squared-error (MSE) and the optimal equalizer coefficients, we use a vector format to represent the variables. As [9] suggests, we represent equalizer, target, and their inputs and outputs by vectors instead of matrices. We can represent a matrix by a vector using any arbitrary convention. Assume that the equalizer support size is $(2Q+1) \times (2Q+1)$. Vector $\mathbf{c}_{(2Q+1)^2 \times 1}$ represents the equalizer coefficients and $\mathbf{I}_{(2Q+1)^2 \times 1}$ is the equalizer input. Hence, the equalizer output (2D-BCJR input) $\tilde{s}_{i,j}$ is

$$\tilde{s}_{i,j} = \mathbf{c}^T \mathbf{I}. \tag{4.5}$$

We use the vector $\mathbf{\Lambda}_{S^2\times 1}$ to represent the target coefficients in $\mathbf{\Gamma}$, and the vector $\mathbf{d}_{S^2\times 1}$ to represent the target input data bits, we have:

$$s_{i,j} = (\mathbf{\Lambda}^T\mathbf{d})^2 = \mathbf{\Lambda}^T\mathbf{d}\mathbf{d}^T\mathbf{\Lambda}. \tag{4.6}$$

The MSE is

$$\begin{aligned} \xi_{\mathbf{\Lambda}} &= E\left[(\mathbf{\Lambda}^T\mathbf{d}\mathbf{d}^T\mathbf{\Lambda} - \mathbf{c}^T\mathbf{I})^2\right] \\ &= E\left[(\mathbf{\Lambda}^T\mathbf{d}\mathbf{d}^T\mathbf{\Lambda})^2\right] + \mathbf{c}^T\mathbf{R}\mathbf{c} - 2\mathbf{c}^T\mathbf{P}_{\mathbf{\Lambda}}\mathbf{\Lambda} \end{aligned} \tag{4.7}$$

where $\mathbf{R} = E\left[\mathbf{I}\mathbf{I}^T\right]$ and $\mathbf{P}_{\mathbf{\Lambda}} = E\left[\mathbf{I}\mathbf{\Lambda}^T\mathbf{d}\mathbf{d}^T\right]$. Although a nonlinear target is used, $\xi_{\mathbf{\Lambda}}$ is still convex in terms of $\mathbf{c}$ for a given $\mathbf{\Lambda}$. So we take the gradient with respect to $\mathbf{c}$ and obtain

$$\nabla_c\xi_{\mathbf{\Lambda}} = 2\mathbf{R}\mathbf{c} - 2\mathbf{P}_{\mathbf{\Lambda}}\mathbf{\Lambda}. \tag{4.8}$$

Setting this gradient to zero and solving for $\mathbf{c}$ we get

$$\mathbf{c} = \mathbf{R}^{-1}\mathbf{P}_{\mathbf{\Lambda}}\mathbf{\Lambda}. \tag{4.9}$$

Note that $\mathbf{P}_{\mathbf{\Lambda}}$ depends on $\mathbf{\Lambda}$. This dependency is not desirable if we wish to compute equalizer coefficients for various targets. However, there is a simple way to overcome this problem. Assume that vector $\mathbf{\Lambda}$ is expressed as a linear combination of some basis $\{\mathbf{v}_i\}$ that contains S^2 linearly independent vectors,

$$\mathbf{\Lambda} = \sum_{i=1}^{S^2} a_i\mathbf{v}_i \tag{4.10}$$

where a_i are scalars. Then

$$\mathbf{P}_{\mathbf{\Lambda}} = \sum_{i=1}^{S^2} a_i \mathbf{P}_{\mathbf{v}_i} \tag{4.11}$$

where $\mathbf{P}_{\mathbf{v}_i} = E\left[\mathbf{I}\mathbf{v}_i^T \mathbf{d}\mathbf{d}^T\right]$. Hence, if we compute $\mathbf{P}_{\mathbf{v}_i}$ for the entire basis, we can efficiently compute the $\mathbf{P}_{\mathbf{\Lambda}}$ for any vector $\mathbf{\Lambda}$ of length $S^2 \times 1$.

For example, consider $\mathbf{H}^{0.5,0.5}$ which is the DCM of the pixel-misaligned HDS in Figure 4.2. This matrix has four entries that are significantly larger than other entries. Consequently, the channel output is mostly dominated by 4 bits corresponding to these entries. So the 2×2 target coefficient matrix

$$\mathbf{\Gamma}_{\mathrm{CT}} = \begin{pmatrix} 0.3029 & 0.3030 \\ 0.3030 & 0.3032 \end{pmatrix} \tag{4.12}$$

which is simply obtained by truncating $\mathbf{H}^{0.5,0.5}$ is intuitively a promising candidate. We refer to target signals with such a coefficient matrix as *Channel Truncation* (CT) target signals.

As we have not yet developed an analytical way to find the optimal target, we perform a brute force search for a coefficients that yield the best BER performance. We search the space of 2×2 separable matrices; such matrices have the general form

$$\begin{pmatrix} a^2 & ab \\ ab & b^2 \end{pmatrix} \tag{4.13}$$

where a and b are scalars. In order to limit the search complexity, we constrain a to be 1 and b to be smaller than 1. We increase b from zero to one and estimate the

corresponding BER by simulation. This will largely reduce the search complexity, but it may lead to loss of optimality as well. In spite of posing these constraints, our results in the next section still illustrate that magnitude-squared targets achieve superior performance. A separate b is chosen for each SNR. Here, SNR is defined as

$$\text{SNR} = 10 \log \left(\frac{1}{\sigma_n^2} \right) \tag{4.14}$$

where σ_n^2 is the electronics noise variance. It is worthwhile to note that for electronics-noise dominated channels, $\sqrt{\sigma_n}$ is proportional to number of recorded pages as [11] stated.

4.4 Numerical Results

In our simulations, we use unity fill factors for SLM and CCD, normalized pixel width, Nyquist aperture width, and SLM contrast ratio of 100. A MMSE equalizer of kernel size 5×5 is used for all equalizations. We present the BER performance in Figures 4.4 and 4.5. We have also plotted the BER performance of BCJR with a linear 2-D PR target and the BER performance of a full response equalizer with threshold detection.

For convenience we refer to the *best* target found as *optimal* target. We should bear in mind that because of constraining the search space, our results are not optimal, still they show the significant gains of using magnitude-squared target

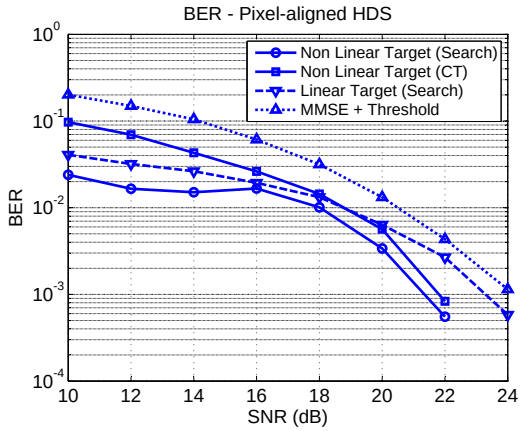


Figure 4.4: BER performance of BCJR detection with linear and nonlinear PR targets, and MMSE-threshold detection for pixel-aligned HDS.

signals. Also note that the discrete channel matrices of the two HDS channels we study are diagonally symmetrical as Figure 4.2 suggests.

We can see that the optimal nonlinear target gives the best performance among different reception techniques/targets. For the pixel-aligned channel, the CT target is far away from optimality at low SNR. However, the performance gap between the optimal nonlinear target and the CT target reduces at high SNR for pixel-aligned channel. For the pixel-aligned channel the CT target outperforms the optimal linear target at high SNR. For the pixel-misaligned channel the CT target always offers superior BER performance. In fact, for the pixel-misaligned channel, threshold detection and linear target PR fail due to the high amount of ISI. However, for nonlinear targets the BER decays slowly and reaches a floor beyond SNR of 36 dB.

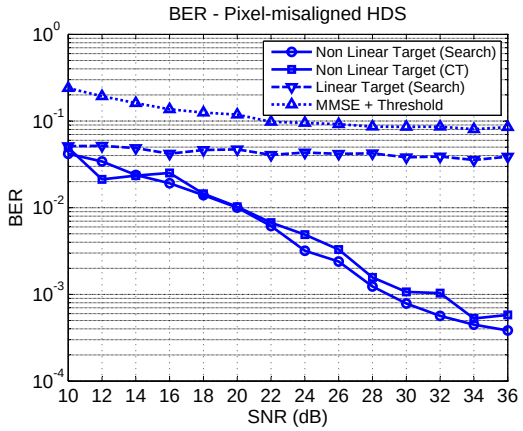


Figure 4.5: BER performance of BCJR detection with linear and nonlinear PR targets, and MMSE-threshold detection for pixel-misaligned HDS.

4.5 Conclusion

We extended the low-complexity, 2-D BCJR detector of [22] to the nonlinear HDS channels. We exploited the separability property of the holographic data storage channel for this purpose. With simple adjustments, our 2-D BCJR detector is able to handle channel nonlinearity at no additional complexity. We present a new partial response target signal that mimics the nonlinear behavior of the channel. This new partial response enables us to detect at low complexity even in the face of severe pixel misalignment. By comparison, linear targets fail when severe misalignment exists.

Chapter 5

Conclusions and Further Work

5.1 Conclusions

In this work, we studied modeling and detection for the holographic data storage (HDS). In more detail, we derived reduced-complexity computational methods to simulate the HDS channel accurately. Furthermore, we extended the discrete magnitude-squared channel (DMC) model to incorporate pixel misalignment. In the detection part, we focused on developing soft-decision maximum-a-posteriori detectors for the two-dimensional nonlinear HDS channel. We designed a novel nonlinear partial-response (PR) target signal which enables us to alleviate the adverse effects of pixel misalignment effectively.

We may partition this thesis into three parts. In part 1, which consists of

Chapters 1 and 2, we presented a brief survey of the literature on the related topics. This survey served as the ground for motivating our research. In addition, we presented the required preliminaries such as a detailed description of the $4 - f_L$ architecture, the BCJR detector for one-dimensional linear channels with additive white Gaussian noise (AWGN), and the BCJR detector for two-dimensional *separable* linear channels with AWGN. In part 2, which consists of Chapter 3, we presented an accurate channel model for HDS systems with pixel misalignment along with an efficient simulation approach for the optical noise. In part 3, which consists of Chapter 4, we presented a reduced-complexity two-dimensional BCJR detector modified for the quadratic nonlinearity of the HDS channel. We presented a novel PR target signal which enabled us to combat the adverse effects of pixel misalignment. The contributions of this thesis are discussed further in the following.

While it is essential for a channel model to be accurate and efficient, the existing channel models for the HDS do not achieve both requirements simultaneously. A notable channel model for the HDS is the DMC model. The simple structure of this model brings efficiency as well as handy insights which are useful for developing signal processing algorithms for the HDS channel. Despite its merits, this model does not accurately address the optical noise. The quadratic nonlinearity as well as the page-oriented nature of the channel renders the analysis and simulation of the detector read-back signal in the presence of the optical noise difficult. We

presented a new method for efficient and accurate simulation of the detector read-back signal in the presence of the optical noise. Furthermore, we corrected a flaw in the statistical analysis of the effects of the optical noise on the detector read-back signal. Our simulation results are consistent with the corrected statistical analysis.

The DMC model does not incorporate pixel misalignment. Since pixel misalignment has a substantial effect on the bit-error-rate performance and it is inevitable in practice, we extended the discrete magnitude-squared channel model for pixel misaligned channels. We observe that the two-dimensional inter-symbol interference in the extended model is separable i.e. one can view the two-dimensional HDS channel as concatenation of two one-dimensional channels.

The page-oriented nature of the HDS results in additional complexity for the detectors. In more detail, the number of interfering pixels increases significantly and the natural ordering of the data often used by Viterbi or BCJR detectors is lost. The quadratic nonlinearity of the channel also makes the detection problem more challenging. In the detection part, we showed how to exploit the separability property of the HDS channels to tackle the absence of natural ordering in a two-dimensional data page. Using this property, we extended an existing reduced-complexity two-dimensional BCJR detector for separable linear channels to the nonlinear HDS channel. Furthermore, we presented a new PR target signal with quadratic nonlinearity similar to the channel. This quadratic PR target enables

us to detect with an acceptable bit-error-rate even in the face of severe pixel misalignment.

5.2 Directions for Further Work

The contributions of this thesis need further work in several directions. We discuss the issues that require further attention in the following.

In the modeling part, we need further investigation on the effect of the optical noise on the bit-error-rate performance. This investigation allows us to understand how much accuracy we need for modeling the optical noise in the HDS; This understating, in turn, leads to a better trade-off between accuracy and efficiency of the HDS model.

In the detection part, we need to develop an analytical approach to find the optimum quadratic PR target for the electronics-noise dominated channels. The channel truncation target inspired by intuition shows near-optimum performance. Furthermore, we need to develop novel reception schemes customized for channels with correlated optical noise.

Bibliography

[1] H. Coufal, G. Sincerbox, and D. Psaltis, *Holographic Data Storage*. Springer-Verlag New York, Inc. Secaucus, NJ, USA, 2000.

[2] V. Vadde and B. Kumar, "Channel modeling and estimation for intrapage equalization in pixel-matched volume holographic data storage," *Applied Optics*, vol. 38, pp. 4374–4386, 1999.

[3] P. van Heerden, "Theory of optical information storage in solids," *Applied Optics*, vol. 2, no. 4, pp. 393–400, 1963.

[4] A. H. B. S. Ken Anderson, Edeline Fotheringham and K. Curtis, "High speed holographic data storage at 500 Gbit/in.2," *InPhase Technologies White Papers*, 2006.

[5] L. Ramamoorthy, S. Nabavi, and B. Kumar, "Physical channel model for holographic data storage systems," *Lasers and Electro-Optics Society, 2004. LEOS 2004. The 17th Annual Meeting of the IEEE*, vol. 2, 2004.

[6] M. Keskinoz and B. Kumar, "Efficient modeling of volume holographic storage channels (VHSC)[4090-38]," *Proceedings-SPIE The International Society For Optical Engineering*, pp. 205–210, 2000.

[7] M. Keskinoz and B. Kumar, "Efficient modeling and iterative magnitude-squared decision feedback equalization (DFE) for volume holographic storage channels," *IEEE International Conference On Communications*, vol. 9, pp. 2696–2700, 2001.

[8] M. Keskinoz and B. Kumar, "Discrete Magnitude-Squared Channel Modeling, Equalization, and Detection for Volume Holographic Storage Channels," *Applied Optics*, vol. 43, no. 6, pp. 1368–1378, 2004.

[9] K. Chugg, X. Chen, and M. Neifeld, "Two-dimensional equalization in coherent and incoherent page-oriented optical memory," *Journal of the Optical Society of America A*, vol. 16, no. 3, pp. 549–562, 1999.

[10] C. Gu, F. Dai, and J. Hong, "Statistics of both optical and electrical noise in digital volumeholographic data storage," *Electronics Letters*, vol. 32, no. 15, pp. 1400–1402, 1996.

[11] L. Menetrier and G. Burr, "Density implications of shift compensation post-processing in holographic storage systems," *Applied Optics*, vol. 42, no. 5, pp. 845–860, 2003.

[12] J. 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, 1996.

[13] H. An, "DETECTION FOR HOLOGRAPHIC RECORDING SYSTEMS," Master's thesis, National University of Singapore, 2005.

[14] M. Bernal, G. Burr, H. Coufal, and M. Quintanilla, "Balancing interpixel cross talk and detector noise to optimize areal density in holographic storage systems," *Applied Optics*, vol. 37, no. 23, pp. 5377–5385, 1998.

[15] M. Keskinoz and B. Kumar, "Application of linear minimum mean-squared-error equalization for volume holographic data storage," *Applied Optics*, vol. 38, no. 20, pp. 4387–4393, 1999.

[16] A. He and G. Mathew, "Nonlinear equalization for holographic data storage systems," *Applied Optics*, vol. 45, no. 12, pp. 2731–2741, 2006.

[17] B. King and M. Neifeld, "Parallel detection algorithm for page-oriented optical memories," *Applied Optics*, vol. 37, no. 26, pp. 6275–6298, 1998.

[18] V. Vadde and B. Vijaya Kumar, "Performance comparison of equalization and low-pass coding for volume holographic data storage," *Optical Data Storage 2000*, pp. 197–204.

[19] X. Chen, K. Chugg, and M. Neifeld, "Near-Optimal Parallel Distributed Data Detection for Page-Oriented Optical Memories," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 4, no. 5, pp. 866–879, 1998.

[20] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes.," *Communications, 1993. ICC 93. Geneva. Technical Program, Conference Record, IEEE International Conference on*, vol. 2, 1993.

[21] S. Benedetto, D. Divsalar, G. Montorsi, F. Pollara, and D. di Elettronica, "Serial concatenation of interleaved codes: performance analysis, design, and iterative decoding," *IEEE Transactions on Information Theory*, vol. 44, no. 3, pp. 909–926, 1998.

[22] Y. Wu, J. O'Sullivan, N. Singla, and R. Indeck, "Iterative Detection and Decoding for Separable Two-Dimensional Intersymbol Interference," *IEEE Transactions on Magnetics*, vol. 39, no. 4, p. 2115, 2003.

[23] J. Goodman, *Introduction to Fourier Optics*. Roberts & Co, 2004.

[24] E. Ramberg, "Holographic information storage," *RCA Review*, vol. 33, no. 1, pp. 5–53, 1972.

[25] C. Gu, J. Hong, I. McMichael, R. Saxena, and F. Mok, "Cross-talk-limited storage capacity of volume holographic memory," *Journal of the Optical Society of America A*, vol. 9, no. 11, pp. 1978–1983, 1992.

[26] K. Curtis, A. Pu, D. Psaltis, *et al.*, "Method for holographic storage using peristrophic multiplexing," *Optical Letters*, vol. 19, no. 13, pp. 993–994, 1994.

[27] D. Psaltis, M. Levene, A. Pu, and G. Barbastathis, "Holographic storage using shift multiplexing," Sept. 23 1997. US Patent 5,671,073.

[28] G. Rakuljic, V. Leyva, and A. Yariv, "Optical data storage by using orthogonal wavelength-multiplexed volume holograms," *Optical Letters*, vol. 17, no. 20, pp. 1471–1473, 1992.

[29] J. Heanue, M. Bashaw, and L. Hesselink, "Encrypted holographic data storage based on orthogonal phase code multiplexing," Aug. 17 1999. US Patent 5,940,514.

[30] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate (Corresp.)," *IEEE Transactions on Information Theory*, vol. 20, no. 2, pp. 284–287, 1974.

[31] W. Ryan, "A turbo code tutorial," *on h ttp://www. ece. arizona. edu/ ryan*, 1998.

[32] J. Berkmann, "On turbo decoding of nonbinary codes," *IEEE Communications Letters*, vol. 2, no. 4, pp. 94–96, 1998.

[33] M. Bernal, G. Burr, H. Coufal, and M. Quintanilla, "Balancing interpixel cross talk and detector noise to optimize areal density in holographic storage systems," *Applied Optics*, vol. 37, no. 23, pp. 5377–5385, 1998.

[34] J. Heanue, M. Bashaw, and L. Hesselink, "Channel codes for digital holographic data storage," *Journal of the Optical Society of America A*, vol. 12, pp. 2432–2439, 1995.

[35] S. I. Mossavat and J. Bergmans, "A Reduced Complexity Two-dimensional BCJR Detector for Holographic Data Storage Systems with Pixel Misalignment," in *Proceedings of the SPS-DARTS 2007, The third annual IEEE BENELUX/DSP Valley Signal Processing Symposium*, (Antwerp, Belgium), March 2007.

[36] A. Jain, *Fundamentals of digital image processing.* Prentice-Hall, Inc. Upper Saddle River, NJ, USA, 1989.

[37] I. Reed, "On a moment theorem for complex Gaussian processes," *IEEE Transactions on Information Theory*, vol. 8, no. 3, pp. 194–195, 1962.

[38] A. He and G. Mathew, "Nonlinear equalization for holographic data storage systems," *Applied Optics*, vol. 45, no. 12, pp. 2731–2741, 2006.