

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

NOTE TO USERS

This reproduction is the best copy available.

UMI[®]

Technical Advances in Volume Holographic Memories

by

Brian King

Copyright © Brian King 2000

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

In Partial Fulfillment of the Requirements
For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2 0 0 0

UMI Number: 3010199

Copyright 2000 by
King, Brian Michael

All rights reserved.

UMI[®]

UMI Microform 3010199

Copyright 2001 by Bell & Howell Information and Learning Company.

All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ©
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Brian King entitled Technical Advances in Volume Holographic Memories

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Mark A. Neifeld
Mark A. Neifeld

4/10/00
Date

Michael W. Marcellin
Michael W. Marcellin

4/10/00
Date

William E. Ryan
William E. Ryan

4/10/00
Date

Raymond K. Kostuk
Raymond K. Kostuk

11-10-00
Date

B. Roy Frieden
B. Roy Frieden

11/28/00
Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

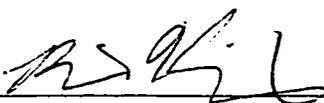
Mark A. Neifeld
Dissertation Director Mark A. Neifeld

4/10/00
Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED:  _____

ACKNOWLEDGMENTS

Traveling the road from aspirations to completion of a PhD has proven to be a difficult journey fraught with many unforeseen challenges. Needless to say, it has been a period of fantastic learning, long hours, and interesting discussions.

My eternal gratitude to Professor Neifeld and Wu-chun Chou. Without them I would never have completed this degree; nor would it have meant very much without their keen insight. In addition, the unwavering love and support from my family has constantly amazed me; they have provided the key final component to unlock the PhD equation.

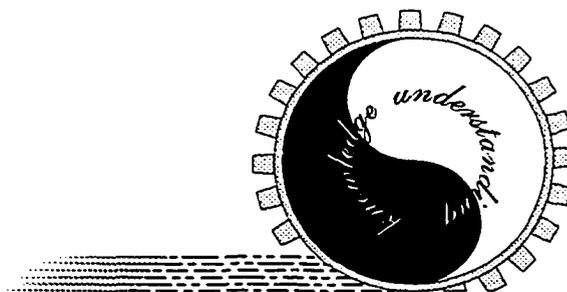
Many thanks to all my friends who have made my time in Tucson truly wonderful. Special thanks to stagecoach tilter Casey for his advice and insight on many matters both technical and social, foreign and domestic, surly and sumptuous. Thank you to Billy and Cindy for giving me a home away from home. Their kindness to such a disoriented confused anti-social person is greatly appreciated.

Sexy Erich and his traveling circus act always guaranteed a great show. Vociferous debates and exercise helped purify my spirit, improve my quality of life, and I strongly suspect, improve the quality of my research career here.

Willy will go down in the annals of history for his sparkling personality and gratuitous use of the word "spanked". His love of puzzles no doubt rivals my own, and will likely wind him up with a short stay in a Mexican jail after a bar bet gone horribly horribly wrong.

Last, but by no means least, sock monkey and Evil monkey have been of extraordinary help throughout my graduate career; though in hindsight, I suspect I may just have been a pawn in their greater plan of world domination.

Brian King
Tucson, Arizona
December 2000



知識累積如聚沙成塔

思想更替似彩雲變幻

發揚前人成果

願此文得為來者之礎石

TABLE OF CONTENTS

LIST OF FIGURES	9
LIST OF TABLES	13
ABSTRACT	14
1. INTRODUCTION	16
2. OPTICAL DETECTION IN THE PRESENCE OF BLUR . .	21
2.1. Introduction	21
2.2. System Modeling	24
2.3. Decision Schemes	33
2.3.1. Simple Threshold	34
2.3.2. 1:2 Modulation Code	36
2.3.3. Wiener Filter	36
2.3.4. PDFE	38
2.4. Simulation Results	41
2.4.1. Gaussian Noise	43
2.4.2. Systematic Error Dominated System	45
2.4.3. Imperfect Channel Knowledge	53
2.5. Coherent Channel Model	57
2.5.1. Fixed Pattern Noise Dominated Coherent System	64
2.5.2. Imperfect Coherent Channel Knowledge	67
2.6. Implementation	68
2.7. Summary	76
3. APODIZATION	79
3.1. Introduction	79
3.2. Bragg diffraction	80
3.2.1. Ideal volume phase grating	82
3.2.2. Angular Selectivity	85
3.3. Bragg Selectivity Function	86
3.3.1. Material Effects and the BSF	87
3.3.2. Modulation depth	90
3.3.3. Bragg Profile Function	92

TABLE OF CONTENTS – *Continued*

3.4. Apodization Theory	94
3.4.1. Experiment	97
3.4.2. Cross Talk noise sources	101
3.5. Noise-to-Signal Ratio	102
3.5.1. Apodizer Shapes	106
3.5.2. NSR results	107
3.6. Apodized M/#	113
3.7. Summary	117
4. BINARY ENUMERATION	119
4.1. Introduction	119
4.2. Recording Dynamics in Holographic Storage	120
4.2.1. Recording time constant	122
4.2.2. Erasure time constant	123
4.2.3. M/#	123
4.3. Noise-limited detection	125
4.4. Sparse data pages	126
4.5. Information-theoretic capacity	128
4.6. Enumeration Code	129
4.6.1. Encoding algorithm	132
4.6.2. Decoding algorithm	133
4.7. Experimental Results	135
4.8. Discussion	140
4.9. Summary	141
5. GRAYSCALE ENUMERATION	144
5.1. Introduction	144
5.2. Background	145
5.3. Permutation Codes	147
5.4. Enumeration of Permutation Codes	148
5.5. Shortened Non-binary Permutation Codes	149
5.6. Encoding Algorithm	154
5.7. Decoding Algorithm	155
5.8. Extended Detection Algorithm	156
5.8.1. Algorithm Description	159

TABLE OF CONTENTS – *Continued*

5.9. Proof of ML Decoding of Shortened Permutation Codes	162
5.10. Application to VHM	166
5.10.1. System Model	167
5.10.2. Capacity Gain	168
5.11. Summary	172
6. GRAYSCALE ENUMERATION EXPERIMENTS	174
6.1. Introduction	174
6.2. Testing Procedure	176
6.3. Grayscale M/#	181
6.4. Experimental Results to Estimated Capacity	185
6.4.1. Local Threshold Detection	188
6.5. Results	189
6.6. Error Propagation	192
6.6.1. Implementation Details	199
6.6.2. Simulation Results	200
6.6.3. Reverse coding applied to the IBM experiment	201
6.7. Summary	207
7. CONCLUSIONS	209
REFERENCES	215

LIST OF FIGURES

2.1. Parallel Optical Binary Detection System. (a) Block Diagram (b) System Model	25
2.2. Incoherent Pixel Profiles at Detection Plane	29
2.3. Superposition of Incoherent Pixel Profiles at Detection Plane . .	31
2.4. 0 and 1 Level Histograms for Incoherent (a) Sinc ² and (b) Gaussian Channels with no noise	32
2.5. BER of Detection Schemes on Sinc ² Resolved Channel	35
2.6. BER of Detection Schemes on Gaussian Resolved Channel	35
2.7. BER of Detection Schemes on Sinc ² Beyond Resolution Channel	40
2.8. BER of Detection Schemes on Gaussian Beyond Resolution Channel	41
2.9. System Capacity in the Presence of Magnification Error	48
2.10. System Capacity in the Presence of Rotation Error	49
2.11. System Capacity in the Presence of Shift Error	50
2.12. Tolerance of Decision Schemes to Knowledge of System Blur Width	55
2.13. Tolerance of Decision Schemes to Knowledge of System Magnification	55
2.14. Tolerance of Decision Schemes to Knowledge of System Rotation	56
2.15. Tolerance of Decision Schemes to Knowledge of Shift	56
2.16. Coherent Sinc Rayleigh Resolved Pixel Profile at Detection Plane	58
2.17. 0 and 1 Level Intensity Histogram for Coherent Sinc Channel . .	60
2.18. BER of Detection Schemes on Coherent Sinc Channel	63
2.19. Coherent Sinc System Capacity in the Presence of Magnification Error	65
2.20. Coherent Sinc System Capacity in the Presence of Rotation Error	65
2.21. Coherent Sinc System Capacity in the Presence of Shift Error . .	67
2.22. PDFE Implementation Block Diagram	70
2.23. PDFE VLSI Transistor Scaling Law	73
2.24. PDFE VLSI Clock Rate Scaling Law	74
2.25. Proof of Concept Fabricated PDFE Chip Plot	75
3.1. Crystal geometry and coordinate system for holographic recording in the 90 degree configuration. An example shows the superposition of scattered wavelets off of the phase grating.	84
3.2. Bragg angular selectivity for a 1 cm long crystal with $\lambda = 0.5 \mu\text{m}$ and a recording reference beam angle of 0 degrees.	86

LIST OF FIGURES – *Continued*

3.3. Apodizer profiles for (a) no apodization (b) reconstruction beam apodization (c) recording beam apodization.	95
3.4. Effective BSF Fourier transform kernel for (a) no apodization (b) reconstruction beam apodization (c) recording beam apodization.	96
3.5. Normalized amplitude diffraction efficiency versus normalized multiplexing detuning for (a) no apodization (b) reconstruction beam apodization (c) recording beam apodization.	96
3.6. Rectangle apodization. Experimental (a) reference beam profile, (b) resulting modulation depth, and (c) Bragg selectivity function in a 5mm sample of KNSBN. The recording beam ratio was 10:1.	98
3.7. Triangle apodization. Experimental (a) reference beam profile, (b) resulting modulation depth, and (c) Bragg selectivity function in a 5mm sample of KNSBN. The recording beam ratio was 10:1.	99
3.8. Gaussian apodization. Experimental (a) reference beam profile, (b) resulting modulation depth, and (c) Bragg selectivity function in a 1cm sample of Fe:LiNbO ₃ . The recording beam ratio was 10:1.	100
3.9. NSR in upper right quadrant of $m = M_r$ page.	106
3.10. Apodizer transmittance and spectrum of (a) rectangle, (b) hamming, (c) raised cosine.	108
3.11. Apodizer transmittance and spectrum of (a) tukey, (b) trapezoid, (c) triangle.	109
3.12. NSR (random page phase) for 1 cm crystal with $\alpha = 0.55 \text{ cm}^{-1}$	110
3.13. NSR (coherent page phase) for 1 cm crystal with $\alpha = 0.55 \text{ cm}^{-1}$	111
3.14. Experimental jitter histogram from repeated measurements on pointing accuracy and repeatability of New Focus picomotor. The standard deviation is $7.6 \mu\text{rad}$, which corresponds to 3.3% of a Bragg null spacing for a 0.5 cm crystal. The quadratic curve and the right axis refer to the Bragg selectivity function centered at the first null.	112
3.15. Jitter dominated IPI noise; $L = 0.45 \text{ cm}$, $\alpha = 0.55 \text{ cm}^{-1}$, 3.3% angular jitter.	113
3.16. $M/\#$ versus absorption in a $1 \times 1 \times 1 \text{ cm}^3$ Fe:LiNbO ₃ crystal for no apodization, reconstruction beam apodization, and recording beam apodization.	118

LIST OF FIGURES – *Continued*

4.1. Capacity of a VHM with “on” pixel <i>a priori</i> probability π_1 . Capacity is relative to an equal priors ($\pi_1 = 1/2$) memory.	127
4.2. Information-theoretic (IT) capacity of a VHM with M pages. The solid curve shows the memory capacity with an “on” pixel <i>a priori</i> probability π_1 that maximizes the capacity and the dashed curve represents the memory capacity when π_1 is forced to be 0.5. The circle symbol curve along with the right vertical axis shows the maximizing value of π_1 as a function of M	130
4.3. Block diagram of volume holographic memory experiment.	136
4.4. Retrieved holographic data pages for (a) $\pi_1 = 1/4$ and (b) $\pi_1 = 1/2$	137
4.5. Bit error rate for different page readout exposures for the sparse and dense memories. The error bars on the data points indicate the 95% confidence interval.	139
5.1. Block diagram describing the channel of interest. k bits of user data are encoded into a n symbol codeword in which each symbol can take one of L possible values. The codeword is stored and later retrieved as a real-valued vector. The detection scheme estimates the transmitted codeword which is then decoded to the estimated user data.	151
5.2. User bit-error-rate versus additive white Gaussian channel signal-to-noise-ratio for various detection/enumerative permutation codes.	153
5.3. 3-level sort-based decision example. The “1” symbols are assigned to the locations with the 5 smallest received values. The “2” symbols are assigned to the 3 locations with the next largest received values. The “3” symbol is assigned to the location with the largest received value.	158
5.4. VHM storage capacity, T , for the case of L -ary pixel values. π_1 is the <i>a priori</i> probability of a pixel being “on”. π_1^* represents the choice of π_1 that maximizes the memory capacity.	170
5.5. Memory capacity of various codes at a user BER of 10^{-12}	171
6.1. Block diagram of the IBM DEMON platform	175
6.2. (a) example SLM image and (b) reconstructed hologram for a 3-ary valued pixel case.	180
6.3. Histogram of 3-level hologram.	181
6.4. Experimental data and curve fit for XA32 data set	182

LIST OF FIGURES – *Continued*

6.5. Experimental estimated VHM capacity	191
6.6. Histogram of the number of bit errors incurred when decoding random codewords from code A46 with weight 2 error patterns. . .	195
6.7. Block diagram of reverse encoding scheme. (a) Type I: parity pixels violate sparse constraint. (b) Type II: parity pixels meet sparse constraint due to encoding with enumeration code E2. . . .	198
6.8. Unprotected BER performance of the A4 set of codes	200
6.9. Reverse coding on the A46 ($n = 90$, $L = 4$) code using an inner RS(255, K , t) code.	202
6.10. Comparison of theoretical and experimental input/output error- rates for data sets XA2, XA3, XA4, and XB4.	204
6.11. Extrapolated reverse code performance for XA4 sets. The solid curve represents the type I code and the dashed curve demarks the type II code.	205
6.12. Extrapolated reverse code performance for XA3 sets. The solid curve represents the type I code and the dashed curve demarks the type II code.	206

LIST OF TABLES

2.1. Channel Capacities of Decision Schemes for Gaussian Noise Dominated Case	46
2.2. Magnification 90% Tolerance Width	51
2.3. Rotation 90% Tolerance Width	52
2.4. Shift 90% Tolerance Width	52
2.5. 90% Tolerance Values in Prior Knowledge Study	57
2.6. Coherent Sinc Channel Capacities for Gaussian Noise Dominated Case	64
2.7. Coherent Magnification 90% Tolerance Width	66
2.8. Coherent Rotation 90% Tolerance Width	66
2.9. Coherent Shift 90% Tolerance Width	66
2.10. Coherent 90% Tolerance Values in Prior Knowledge Study	68
2.11. PDFE Design Options	72
3.1. Apodizer functions	107
5.1. VHM capacity gain associated with graylevel signaling compared to binary $\pi_1 = 0.5$ signaling.	172
6.1. IBM experimental set/parameter descriptions	177
6.2. IBM experimental set/parameter descriptions	178
6.3. Experimental capacity parameters	190
6.4. Change in capacity and readout rate relative to case of binary pixels and no modulation coding (local threshold detection).	193
6.5. Inner code parameters	197

ABSTRACT

Volume holographic memories (VHMs) are a candidate technology for next-generation high-density and high data-rate digital storage. Capacities greater than 1 Terabit are promised, available at read-out rates exceeding 1 Gigabit per second.

The capacity target will be achieved through two mechanisms. First, retrieval in a VHM reconstructs a holographic page (a two-dimensional image) captured on a CCD (charge-coupled device) camera. Each page represents on the order of one million bits of data, by encoding the data as bright and dark pixels in the 1024 x 1024 stored/retrieved image. Second, due to the thickness of the recording medium, a large number of such pages can be recorded in the same volume of material. In this dissertation we address some of the difficult technical issues that either currently limit the VHM system design, or are expected to become a limiting factor in the future.

The first such concern involves how to process the simultaneous optical arrival of one million pixels. In high-density storage, there will be significant cross-talk between pixels which limits the storage capacity. We develop a novel highly-parallel focal-plane processor, which can significantly improve the system capacity by performing reliable detection in the presence of optical blur and alignment errors introduced by the imaging system. A fabricated proof-of-concept VLSI design is described.

Another fundamental noise source is caused by the cross-talk between holographic pages. Reconstruction of the desired data page reconstructs every page

in the memory, albeit at a very low relative diffraction efficiency. As the number of multiplexed pages increases, the cross-talk from the other pages can constitute a significant optical field noise source. Apodization seeks to either suppress this noise source or control it such that system tolerances can be relaxed.

Bright data pixels are stored by altering the material properties of the crystal. However, dark pixels require no adjustment to the crystal; they are implicitly stored. This asymmetric storage cost drives a capacity improvement by biasing the data pages to contain more dark pixels and fewer bright pixels. An increased number of pages can be stored at the same reconstruction fidelity. We propose a novel modulation code to encode and decode these sparse data pages. Experimental results are presented showing the improvement in capacity.

If the data page is composed of non-binary or grayscale pixels, then a further capacity enhancement is possible. The previous binary modulation code is extended into an arbitrary grayscale modulation code and a low-complexity maximum-likelihood decoder is developed as well as a mathematical proof of correctness. Extensive experimental results verify that the proposed method is practical and offers a substantial capacity improvement.

CHAPTER 1

INTRODUCTION

Holography is often accredited to the pioneering work of Dennis Gabor¹ in 1948 when he discovered a way to record the phase information of an optical field in an irradiance measurement. Once it became common knowledge that the notion of a coherent source interfering with an optical field could record the both the amplitude and phase information, the new emerging science of holography: did not go far! Only once the laser was invented did holography become an extensive subject of research. With the work of Leith and Upatnieks²⁻⁴ in 1962 – 1964 conventional holography was finally established as a practical and practiced scientific discipline. The 14 year delay was mostly due to the lag in the quality of the components necessary for holography.

In 1963 van Heerden^{5,6} published two articles back to back which proposed a basic understanding of volume holographic optical storage and made suggestions on exactly how to achieve large volumetric storage densities. It is from these articles that the commonly bandied theoretical volumetric density limit of V/λ^3 bits/cm³ is discussed. In 1968 Gabor⁷ developed much of the relevant theory on special properties of “thick” holograms. Furthermore, 1969 saw the publication of Kogelnik’s coupled-wave theory⁸ which precisely predicted the diffraction properties of thick holograms.

By 1969 holography was a practical technique; the mathematical and theoretical tools for volume holographic storage were in place. So why did it take so long for volume holographic storage to actually be realized? Again, like the laser

spurring holography forward, the supporting O/E and E/O high space-bandwidth product devices such as the spatial light modulator (SLM) and charge-coupled device (CCD) camera have finally made it practical to imprint and retrieve the promised huge amounts of data. Holographic data storage pulls together a number of newly emerging technologies and integrates them into a powerful system. But as a consequence of the reliance on a number of still maturing technologies, the overall system design can be difficult due to the demanding operating requirements placed on nearly all the components in the system.

Nevertheless, with the necessary components in various stages of readiness, research in volume holographic memories (VHMs) is now being propelled forward by the tantalizing promise of enormous storage capacities and large aggregate data rates, available with low latency. With many recent advancements⁹⁻¹⁹ comes a need for improved digital coding and signaling schemes tailored to the unique page-access operating environment. Digital data to be stored is typically encoded as a 2D array of binary pixels and is presented to the recording medium through a spatial light modulator (SLM). The key to large data rates is the 2D parallel nature of the data. Storage and retrieval of megapixel (1024×1024) data pages has been demonstrated.⁹ A retrieval page rate of 1 kHz produces an aggregate data rate of 1 Gbit/sec. This high-rate highly-parallel data places difficult processing requirements on many of the components in the system. The hope of Tbit capacities and Gbit/sec retrieval rates, however, encourages forward progress. This dissertation is by no means an exception! We consider various present and predicted future issues, and offer our research as a stepping stone; one step further along the path.

The highly parallel nature of the data places difficult processing requirements on the memory interface causing potential bottlenecks within conventional serial solutions. A parallel memory interface is needed to support the large aggregate data rate. To make matters more difficult, a high-density VHM will suffer from large amounts of pixel-to-pixel cross talk due to the blur induced by the imaging optics. In chapter 2, we propose a focal plane processor design for a smart pixel array that will in parallel, detect and de-blur the data. One obvious advantage of this approach is the elimination of the need to move the megabit page off the CCD chip with each pixel being digitized at many bits of resolution. With reliable detection and data estimation (in the presence of optical system alignment errors) performed well, the pixel can be exported as a single bit. A full discussion of the detection abilities of the proposed parallelizable design are presented in the chapter, including a figure showing the proof-of-concept fabricated chip design.

Another avenue to larger storage capacities is to multiplex more holograms in the same location. Diffraction efficiency can establish a limit on the number of multiplexed holograms, but in the near future with improved materials, we expect that the reconstruction cross talk between pages will prove to limit the system retrieval for very large number of holograms. Apodization is a technique that seeks to quell the cross talk noise, by reshaping the cross talk function to reduce the noise contributions accumulating from the number of pages in the memory. Chapter 3 discuss this subject in detail.

Data pages, until recently, have contained on average a comparable number of bright and dark pixels. Bright pixels are stored by forming a phase grating in the holographic material. However, dark pixels require no adjustment to the crystal;

they are implicitly stored. This asymmetric storage cost urges us to consider storing less “on” pixels per page. Doing so allows more pages to be stored in the memory when diffraction efficiency limits the size of the memory. A careful balance between sparser pages (about 25% “on” pixels) and the reduced information per page leads to the development of the enumerative permutation code described in chapter 4. Approximately a 15% gain in capacity can be achieved by adjusting the average number of “on” and “off” pixels per page. Experimental results confirm this line of research.

Chapter 5 extends this idea of sparsity to grayscale pixels. The increased signaling levels per pixel increase the information content per page while the sparsity reduces the per page information. Thus the two terms tend to somewhat balance each other providing up to about a 50% theoretical gain in capacity. We extend the previous binary enumeration code into an arbitrary grayscale modulation code. A low-complexity encoding and decoding scheme are developed. In addition, the detection algorithm is proven to achieve maximum-likelihood performance.

In chapter 6 we discuss results of implementing the proposed codes in a real holographic system testing device, IBM’s DEMON platform. We see about a 10% improvement in extrapolated capacity using 3-level sparse pixels as compared to no coding using equal numbers of “off” and “on” binary pixels. The improvement is reduced due to the effect of error propagation on the long-length modulation codewords. We introduce the concept of reverse coding to address this concern. Application of reverse codes to the experimental data estimates a 35% capacity improvement.

We begin by considering the blur introduced between pixels due to the finite spatial bandwidth allowed by the optical system. High density storage typically requires a small Fourier aperture for the object beam, resulting in significant blurring on the CCD detection plane.

CHAPTER 2

OPTICAL DETECTION IN THE PRESENCE OF BLUR

2.1 Introduction

There are a lot of possible noise sources that may corrupt the retrieved holographic data page. In this chapter we consider the case when a specific subset dominate the storage and retrieval processes. In order to place more pixels in one page the pixels must either be reduced in size or the effective page extent must be increased. Alternatively, the Fourier aperture of the object beam may be decreased (thereby reducing the used photorefractive interaction volume). Of course, a combination of these methods may be applied, but in any case, the net result is that in the detection plane the pixels begin to receive a significant contribution of energy from their neighbors. We refer to this cross talk as inter-pixel interference or IPI. As high-density storage improves, the IPI will likely be a limiting factor in the system design.²⁰ The optics relaying the information-bearing pixel array through the holographic medium to the detection plane are subject to a finite optical bandwidth and hence provide the optical mechanism for the optical blur or cross talk. To make matters worse, there may be fixed alignment errors in the optical system, inducing a fixed-pattern noise on the retrieved page. Examples of this include a slight rotational misalignment between the SLM and CCD planes, a sub-pixel transverse shift between the two planes, or a small fixed magnification error. We also consider that the CCD detection array is operating

in a regime where the thermal electronic noise is the dominate noise among the possible noise sources corrupting the retrieved electrical signal.

Posing the detection problem in its most general form, one can search the communications community literature for proposed solutions. They have addressed a serial version of this detection problem under the topic of maximum likelihood sequence detection (MLSD). In related research, the optical storage community has recently presented high performance MLSD approaches that are appropriate for page-oriented memories based on the Viterbi²¹ algorithm.²²⁻²⁵ Some of these algorithms offer excellent performance but are computationally expensive. We seek reduced complexity techniques suitable for a VLSI smart pixel implementation.

In this chapter we discuss such a low-complexity solution that recovers a significant amount of the performance loss due to the noise sources and incurs a low complexity cost. One of the key insights to achieving the improvement in data detection is to note that the 2D pixel array represents a massively parallel stream of digital data where the cross talk induced on a single pixel/channel is almost completely a function of the local neighborhood to that pixel. We hypothesize that a small processing element around a pixel which communicates with only the neighboring processing units may have the ability to significantly improve the detection estimate of the page data. These small units will thus operate in a massively parallel fashion with only local connectivity requirements. Of course, because of the small detection plane pixel pitch, we have a very small amount of VLSI area in which to do all the necessary computations; hence, the calculation unit must be of very low-complexity.

We will consider both incoherent and coherent optical memories corrupted by thermal detector noise as well as fixed pattern noise. We examine the effect of two different SLM fill factors and channel point spread functions (PSFs) on the system metrics of bit error rate (BER) and storage capacity. To address the detection problem we propose a page-parallel processing algorithm based on the communication technique of decision feedback equalization (DFE).^{26,27} The advantages of our proposed pseudo-decision feedback equalizer^{28,29} (PDFE) scheme include its ease of implementation and completely parallel design. Due to the local nature of PDFE processing, this algorithm can offer good performance in the presence of an optical channel with a space-variant point spread function in addition to small amounts of systematic fixed pattern errors. We compare and contrast the performance of the PDFE algorithm against competitive detection methods with respect to an ideal upper bound on system capacity.³⁰

In Section 2.2 we introduce a simple memory channel model. We trace the digital data originating on the SLM through an incoherent optical channel to detection at the CCD plane. Next we examine a number of common decision schemes and our proposed PDFE method, all of which allow us to consider BER parameterized by the noise level, SLM fill factor, and PSF. From the scaling law relating BER and number of pages in a holographic memory³¹ we develop the Gaussian noise-limited capacities associated with these various decision methods. Section 2.4 presents the simulated performance of the decision techniques when we vary the noise level, SLM pixel pitch, PSF, and introduce systematic errors in the optical system. We also consider the performance degradation due to imperfect channel knowledge. In Section 2.5 we consider coherent optical channels and

rederive the augmented system model. Following the format for the incoherent case we analyze the performance of the PDFE and the other algorithms under the same noise and systematic error conditions. In the last section we introduce estimates of chip area and clock rate for a parallel VLSI implementation of the PDFE algorithm. Analog versus digital approaches are discussed and the design parameters for a fabricated MOSIS 2-micron chip PDFE implementation are presented.

2.2 System Modeling

We begin by modeling our optical memory system as an incoherent optical channel corrupted by Gaussian detector noise. The coherent case is more complicated and will be considered in Section 2.5. The total system behavior can be characterized by the deterministic intensity PSF and the statistics of the random noise. Fig. 2.1a shows the basic system architecture. During recording, a binary amplitude SLM transmits the two-dimensional page of binary data into the optical recording system governed by the incoherent PSF $h_{chan}(x, y)$. The SLM pixel shape, $p(x, y)$, can be incorporated with the channel PSF to yield an overall incoherent system PSF defined as

$$h(x, y) \triangleq p(x, y) \otimes h_{chan}(x, y), \quad (2.1)$$

where \otimes represents the two-dimensional convolution operator. In order to easily distinguish the system and channel PSFs we will refer to the system PSF as the pixel spread function (PxSF).

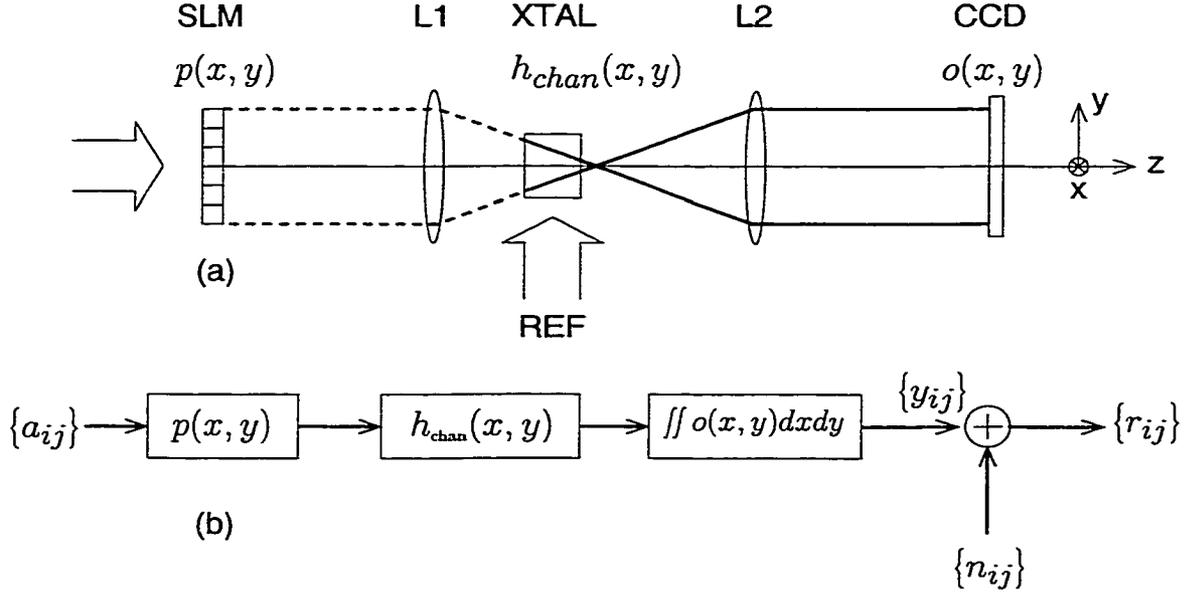


Figure 2.1: Parallel Optical Binary Detection System. (a) Block Diagram (b) System Model

A 2D detector array (*e.g.*, CCD) integrates in space and time the retrieved data page photons at the output plane of the system. Fig. 2.1b shows the mathematical model for the overall system, where a_{ij} denotes the data bit associated with row i and column j , $p(x, y)$ is the SLM pixel profile, $o(x, y)$ is the continuous spatial intensity at the output plane of the optical system, y_{ij} is the integrated intensity from the CCD at location (i, j) , n_{ij} is the associated detector noise for that pixel, and r_{ij} is the noisy CCD output. We model the noise as zero mean additive white Gaussian with variance σ_n^2 . We consider two candidate functions for the incoherent PSF:

$$\text{sinc}^2(x, y) = \left(\frac{\sin(\pi x/\sigma_b)}{\pi x/\sigma_b} \frac{\sin(\pi y/\sigma_b)}{\pi y/\sigma_b} \right)^2, \text{ and}$$

$$\text{Gaussian}(x, y; \sigma_b^2) = \frac{1}{2\pi\sigma_b^2} \exp \left\{ -\frac{x^2 + y^2}{2\sigma_b^2} \right\}.$$

The central limit theorem implies that an optical system that is subject to a large number of small independent aberrations will be characterized by a Gaussian PSF. When the aberrations have been corrected the system will be diffraction limited and described by a sinc^2 PSF. Note that σ_b is used to represent the channel width parameter of the PSF: σ_b represents null spacing for the sinc^2 PSF, and it represents standard deviation for the Gaussian PSF. This should not be confused with σ_n^2 representing the variance of the additive Gaussian noise.

From the PxSF we can relate the binary input data to the real-valued received data through a single convolution sum as shown in eq. 2.2, where the effective channel taps, $\{h_{kl}\}$, are defined by eq. 2.3.

$$r_{ij} = \sum_{kl=-\infty}^{\infty} a_{ij-kl} h_{kl} + n_{ij} \quad (2.2)$$

$$h_{kl} = \int \int_{x_l^-, y_k^-}^{x_l^+, y_k^+} h(x, y) dx dy, \quad (2.3)$$

where $x_l^\pm = lD \pm \Gamma/2$ and $y_k^\pm = kD \pm \Gamma/2$, $D = \text{SLM pixel pitch}$, $\Gamma = \text{CCD pixel size}$, and double subscripted sums and indices are understood to be a notational convenience for representing a sum or index over two variables.

In order to make distance ratios intuitively simple we set the pixel pitch of the SLM to unity ($D = 1$) and express all distances in units of the SLM pitch. Similarly we express all power quantities in units of the system input power. Because the number of pixels per page will remain unchanged throughout this chapter, a constant system input power is also equivalent to a constant input power per pixel. Furthermore, we assume the optical channel is non-absorptive and that the SLM delivers the same power regardless of its fill factor. Simulations

indicate that under these conditions the optimum CCD fill factor will always be unity so Γ is set to 1.

We model the SLM pixel profile $p(x, y)$ as a square spaced at the SLM pitch and we can therefore consider the PxSF as a 2D $\text{rect}()$ convolved with either a $\text{sinc}^2()$ or a $\text{Gaussian}()$ term. Eqs. 2.4 and 2.5 show the PxSF (pixel spread function) for the Gaussian and sinc^2 incoherent cases of interest, respectively.

$$h(x, y) = \frac{1}{\Delta^2} \text{rect}(x/\Delta, y/\Delta) \otimes \text{Gaussian}(x, y; \sigma_b^2) \quad (2.4)$$

$$h(x, y) = \frac{1}{\Delta^2} \text{rect}(x/\Delta, y/\Delta) \otimes \frac{1}{\sigma_b^2} \text{sinc}^2(x/\sigma_b, y/\sigma_b) \quad (2.5)$$

The width of the PxSF is of obvious consequence in determining the difficulty of resolving the data page on the CCD. We consider each of these channels operating at two blur widths: the Sparrow resolution width and 25% wider. The Sparrow resolution criterion specifies that the incoherent intensity superposition of two neighboring pixels yields an intensity profile that has a point of inflection at the spatial position halfway between the two pixels. Given the channel PSF the overall system PSF (PxSF) will be determined by the fill factor of the SLM. We consider linear SLM fill factors between $\Delta/D = 0$ and $\Delta/D = 0.9$. So of all the possible combinations of channel PSF (Gaussian or sinc^2), channel blur width (resolved or 25% beyond resolution), and SLM fill factor (between 0 and 0.9), only two cases result in a Sparrow resolved PxSF: Gaussian resolved $\Delta/D = 0$ and sinc^2 resolved $\Delta/D = 0$. The remaining cases result in output plane data that is approximately 25% ($\Delta/D = 0$) beyond the Sparrow resolution criterion and further from the resolution criterion as the SLM fill factor increases. Fig. 2.2

shows a one-dimensional slice through the center of the PxSF for the (a) Gaussian resolved, (b) sinc^2 resolved, (c) Gaussian beyond resolution, and (d) sinc^2 beyond resolution channels. The abscissa is in units of the SLM pitch. The solid line corresponds to $\Delta/D = 0$ and the dashed line corresponds to $\Delta/D = 0.9$. Note that for the sinc^2 channels with $\Delta/D = 0.9$ that the PxSF is no longer zero at integer multiples of the sinc width so that even with impulse sampling at the neighboring pixels cross talk can never be eliminated. Fig. 2.3 shows the superposition of two neighboring pixel profiles for both PSFs and resolutions with $\Delta/D = 0$. Note that the Sparrow resolved blur size is decided by the channel PSF only, so the effect of the finite SLM pixel size is not incorporated in the channel resolution blur width, only the system blur width.

Alternatively, we can consider the resolution criterion in terms of the Nyquist condition for a SLM and CCD pixel spacing of D . We define the bandwidth of the optical transfer function as the 90% energy bandwidth. For both the Gaussian and sinc^2 channel, we can compute the ratio of the 90% energy channel bandwidth to the Nyquist bandwidth. The sinc^2 resolved PSF has a bandwidth that is 113.5% of the Nyquist bandwidth. Similarly, the bandwidths for the sinc^2 not resolved, Gaussian resolved, and Gaussian not resolved PSFs are 91.2%, 81.8%, and 65.7%, respectively.

Eq. 2.2 suggests a discrete matrix representation of the overall memory channel and establishes a 1:1 mapping between the input binary data page, $\{a_{ij}\}$, and the discrete set of real-valued received values, $\{r_{ij}\}$. For example, using an SLM with square pixels and a linear fill factor of 0.9 propagating through a Sparrow

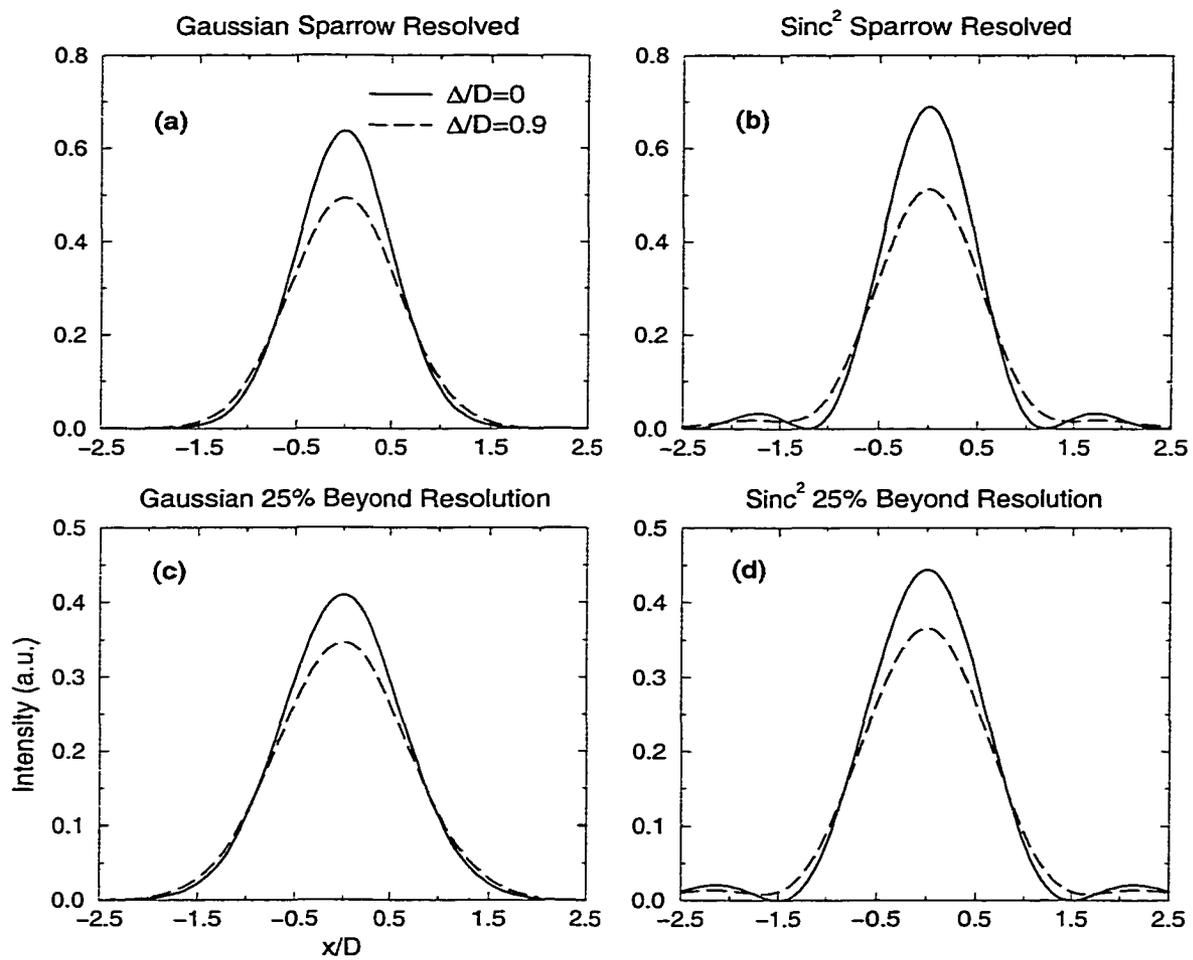


Figure 2.2: Incoherent Pixel Profiles at Detection Plane

resolved sinc^2 channel gives us the following cross talk matrix which we refer to as the discrete channel model (DCM).

$$\{h_{kl}\} = \begin{array}{|c|c|c|c|c|} \hline 0.00033 & 0.0027 & 0.011 & 0.0027 & 0.00033 \\ \hline 0.0027 & 0.022 & 0.091 & 0.022 & 0.0027 \\ \hline 0.011 & 0.091 & 0.39 & 0.091 & 0.011 \\ \hline 0.0027 & 0.022 & 0.091 & 0.022 & 0.0027 \\ \hline 0.00033 & 0.0027 & 0.011 & 0.0027 & 0.00033 \\ \hline \end{array}$$

The DCM indicates how much of an input pixel's energy is blurred into the local neighborhood. Notice that for this channel only 39% of an on-center pixel is transmitted to the corresponding CCD pixel and the remaining 61% is effectively dispersed over the nearest 24 pixels. The use of a 5×5 model is justified by the localized nature of the cross talk allowing truncation of the theoretically infinite extent of the cross talk to a more practical limit.

Fig. 2.3 demonstrates a convenient way to visualize blurring. The overlapping tails of the intensity superposition are responsible for the IPI incurred on neighboring pixels. A histogram of the received target pixel intensity values when a one or zero is transmitted provides a valuable alternative view of the channel blurring operation. Fig. 2.4a shows the histogram of 30 pages of size 128×128 (491,520 pixels' worth of data) for the sinc^2 resolved channel with $\Delta/D = 0$ and fig. 2.4b shows the histogram for the Gaussian resolved channel with $\Delta/D = 0$. From these figures we see that for the Gaussian resolved case there exists no choice for a simple threshold that will allow us to always correctly decide if a one or a zero was transmitted; while, for the sinc^2 resolved channel a simple threshold does indeed exist. When we incorporate the effect of detector thermal noise, the histograms are further broadened reducing the performance of a simple threshold detection scheme. The communication-theoretic approach can be seen as jointly

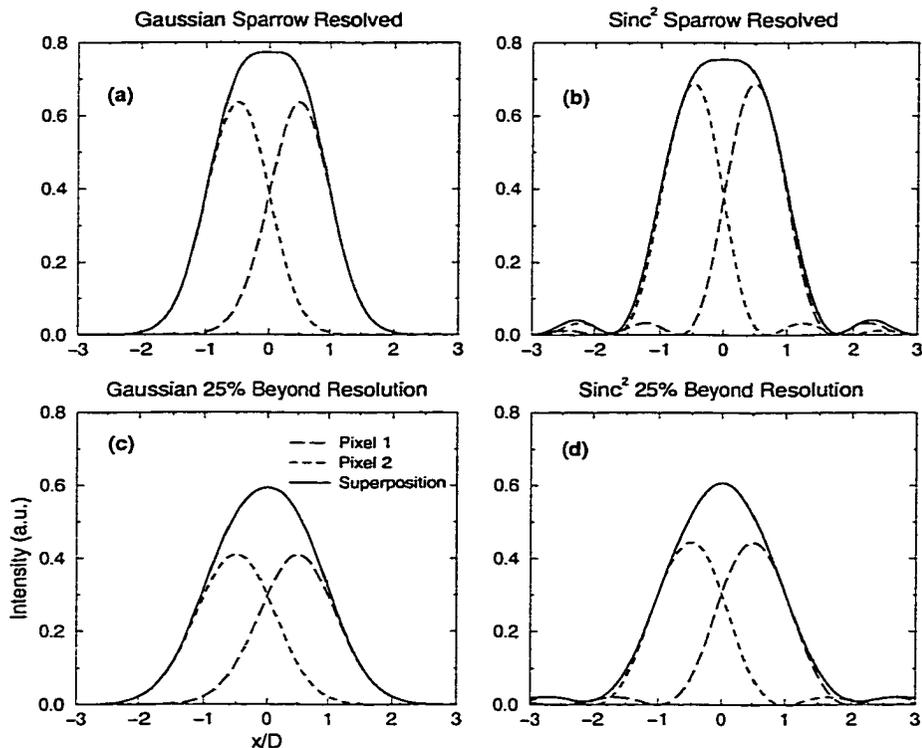


Figure 2.3: Superposition of Incoherent Pixel Profiles at Detection Plane

estimating the binary page using the noise statistics and the received intensity values. The goal of such an approach will be to provide a post-detection bit error rate (BER) of 10^{-4} or less. As we will discuss later, 10^{-4} was chosen on the basis of requirements for an error correcting code (ECC) following the detection stage.

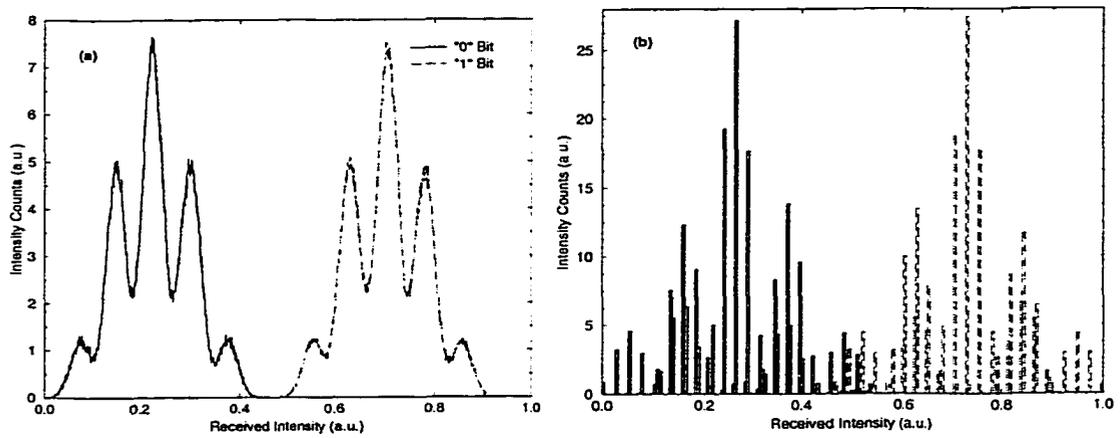


Figure 2.4: 0 and 1 Level Histograms for Incoherent (a) Sinc² and (b) Gaussian Channels with no noise

2.3 Decision Schemes

We define a decision scheme as an algorithm that yields binary estimates of the data page using only the received intensity values. The approaches considered include a simple threshold, 1:2 modulation code, Wiener filter, and PDFE. The basic simulation procedure for any of these algorithms consists of:

1. Selecting a SLM pixel pitch (Δ/D), channel PSF, channel blur width (σ_b), and noise level (σ_n)
2. Generating a random 128×128 page of data
3. Computing the received blurred intensity page corrupted by noise at the detection plane
4. Applying the decision algorithm
5. Counting the number of errors in the estimated page
6. Repeating steps 2-5 until a statistically significant number of errors and trials have been counted

The above process will be repeated for the various combinations of parameters listed in step 1. In all simulations the CCD fill factor is unity and it is pixel matched to the SLM. In order to have a significant measure of the probability of error we tested as many pages as necessary to count at least 100 errors. So for a BER of 10^{-5} this implies testing at least ten million pixels (611 pages). Also, we will need to curve fit the data to efficiently find important numbers such as

the noise level where the decision scheme performs at a BER of 10^{-4} . All curve fitting applied in this chapter is done through cubic spline interpolation.

2.3.1 Simple Threshold

The Simple threshold decision scheme compares the received pixel intensity value to a predetermined threshold value. This threshold is defined by the intersection of the one and zero level probability density functions (PDFs). If the PDFs do not overlap then perfect decisions are possible. If they do overlap then the intersection of the two PDFs selects the threshold that achieves the lowest possible BER for a thresholding scheme. It should be noted that the threshold is selected on the basis of the PDFs associated with a specific pixel. For a space-variant channel this makes the threshold a function of the pixel location, but in any event it is always the optimal threshold. Figs. 2.5 and 2.6 present plots of the post detection BER as a function of the noise level for various decision schemes. The performance of the threshold scheme is denoted by the diamond symbol. The dashed line type with open symbols denotes the SLM fill factor of $\Delta/D = 0.9$ and the solid line type with filled symbols denotes $\Delta/D = 0$. Note the characteristic “waterfall” curve shape for the resolved sinc² channel and how the threshold fails to achieve the target BER for the resolved Gaussian channel. This failure occurs even for a noise-free channel and urges us to consider more powerful schemes to process the received data.

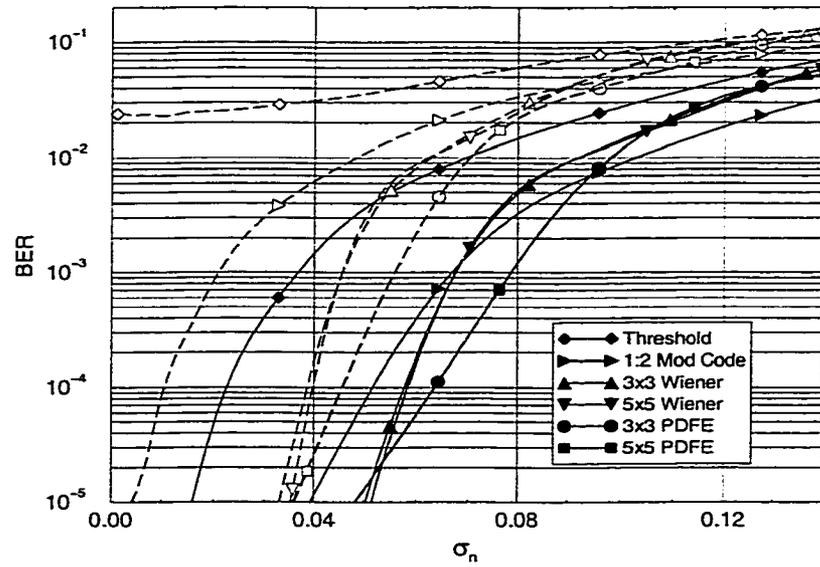


Figure 2.5: BER of Detection Schemes on Sinc^2 Resolved Channel

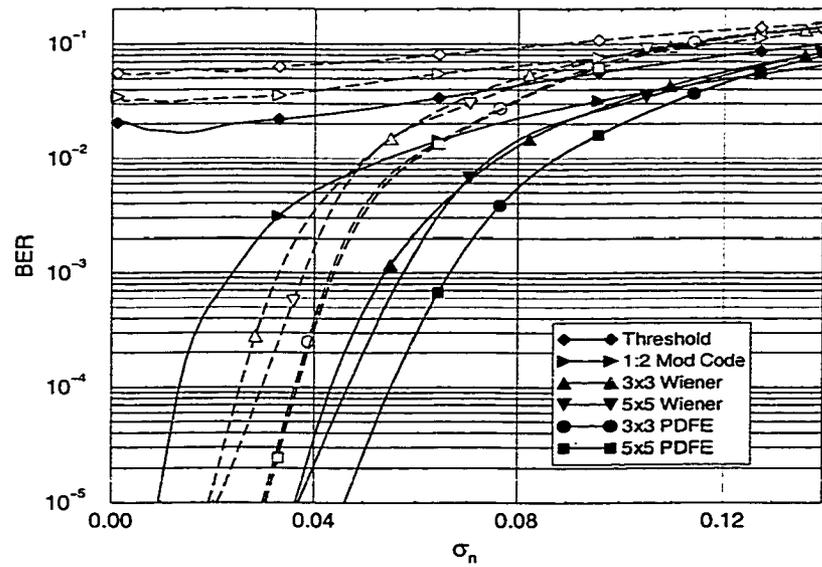


Figure 2.6: BER of Detection Schemes on Gaussian Resolved Channel

2.3.2 1:2 Modulation Code

A $m : n$ modulation code^{32,33} encodes m input bits into 2^m unique states using n bits, where $m < n$. One advantage of such a code is that it is more tolerant to IPI. For the specific case of a 1:2 modulation code a single bit is encoded by arbitrarily choosing the bit pattern 10 to represent a “one” bit and 01 to represent a “zero” bit. Decoding is accomplished via differential detection resulting in improved performance in the presence of unknown amplitude scaling. A further advantage of this technique is the trivial implementational complexity. The major limitation of this modulation code is its low code rate of 1/2. The performance of the 1:2 modulation code is presented in figs. 2.5 and 2.6 as the right-pointing triangle symbol with the dashed and solid lines once again representing two different SLM fill factors. We can see the modulation code outperforms the simple threshold for all noise levels.

2.3.3 Wiener Filter

A channel equalization filter can be applied to the received data to improve decision performance.²⁷ Equalization attempts to undo the blurring effect of the system PSF and to allow simple thresholding of the filter output. By applying various desirable constraints we can derive a number of different versions of the channel equalizer. The simplest version is the channel inverse which has a spectrum that is the inverse of the Fourier Transform of the PSF. Under noise-free conditions and a narrow signal bandwidth this would completely eliminate the effect of the channel. The well-known problems with the inverse filter include

spectral nulls, noise amplification, and the infinite impulse response (IIR) necessary to implement the filter.³⁴ If we choose to minimize the equalizer output MSE using a filter with a small number of taps (finite support filter) followed by a simple threshold we arrive at the FIR Wiener filter²⁷ (WF). It is the optimal linear filter (in the MMSE sense) when the received data is corrupted by additive white Gaussian noise (AWGN). It differs from the inverse filter by decreasing the filter gain appropriately at frequencies with low signal content. This avoids overamplifying the noise present at those frequencies and leads to the optimal linear restoration filter. Because the support of the filter is an arbitrary parameter, the Wiener filter allows us to easily trade off computation and connectivity complexity for BER performance. As we will discuss later, the locally connected Wiener filter also has a final important advantage of being easily adaptable to a space-variant channel.

The computation of the Wiener filter involves the solution of the Wiener-Hopf equations which can be easily derived from the condition that the minimum mean square error (MSE) filter (*i.e.*, Wiener filter), $\{w_{ij}\}$, must satisfy the orthogonality principle:

$$E [(a_{ij} - \hat{a}_{ij}) r_{ij-kl}] \equiv 0 \quad \forall k, l, \quad (2.6)$$

where $E[\cdot]$ is the expectation operator, a_{ij} is pixel (i, j) , \hat{a}_{ij} is the estimated pixel value at the Wiener filter output, and $\{r_{ij}\}$ are the received noisy intensity values. Expressing \hat{a}_{ij} as a convolution of $\{r_{ij}\}$ and $\{w_{ij}\}$ and rearranging the resulting expression yields the Wiener-Hopf equations:

$$\sum_m \sum_n w_{mn} E [r_{ij-kl} r_{ij-mn}] = E [a_{ij} r_{ij-kl}] \quad \forall k, l. \quad (2.7)$$

Since the Wiener-Hopf equations must be satisfied for all values of k and l , eq. 2.7 represents a set of equations that can be represented in the matrix form:

$$K\vec{w} = \vec{c}. \quad (2.8)$$

K will be block Toeplitz when the PSF is space-invariant, providing a simple solution to the well-conditioned linear inverse problem posed by eq. 2.8. The Wiener filter is, of course, constructed based on the specific channel PSF. Note that in eq. 2.7 the channel response, $\{h_{ij}\}$, is hidden inside the received noisy values, $\{r_{ij}\}$.

By thresholding the Wiener filter output we arrive at the data page estimate where the predetermined threshold value is chosen to minimize the BER. Referring once again to figs. 2.5 and 2.6 and noting the Wiener filter curves for 3×3 and 5×5 channel support (symbols triangle and down-pointing triangle, respectively) we can see the improvement in performance over the simple threshold scheme. To further reduce the BER we use the Wiener filter estimate as the starting point for the pseudo-decision feedback equalizer (PDFE) described next.

2.3.4 PDFE

As with the Wiener filter, successful pseudo-decision feedback equalization (PDFE) hinges on the knowledge of the channel response. In particular, given that we have correctly estimated a bit on the received page we can then blur that bit using the channel model and subtract its interfering effect from the remainder of the received data. This is very similar but not identical to the digital communication technique of DFE, hence the name PDFE. This method

ideally cancels all IPI present in the data page. The technique breaks down if there are too many initial errors as these errors will propagate throughout the entire page; however, with an initial BER on the order of 10^{-2} error propagation is not a concern. The main advantage of this approach is that we can make bit decisions based only on the received value for that bit and its neighbors. This can be contrasted with optimal estimation techniques (*e.g.*, MLSD) which require the use of all received values while deciding on each bit. These optimal approaches present a large computational complexity burden on the receiver design that can be avoided via bit-by-bit decision.

The PDFE decisions are performed for each pixel in parallel and involve the local interference estimates, \hat{e}_{ij} , which are computed as the sum of the neighboring pixel estimates weighted by their interference contribution determined by the discrete channel model. We express the interference estimate for pixel (i, j) as a function of the most recent estimates for its neighbors:

$$\hat{e}_{ij} = \sum_{k \neq i} \sum_{l \neq j} \hat{a}_{ij-kl} h_{kl}, \quad (2.9)$$

where \hat{a}_{ij} is the current estimate of pixel (i, j) . Each pixel then forms an estimate of the received data conditioned on a one or zero being transmitted as:

$$\hat{r}_{ij}|0 = \hat{e}_{ij} + (\hat{a}_{ij}|a_{ij} = 0)h_{00} \quad (2.10)$$

$$\hat{r}_{ij}|1 = \hat{e}_{ij} + (\hat{a}_{ij}|a_{ij} = 1)h_{00}. \quad (2.11)$$

For the binary case we consider here \hat{a}_{ij} is either 0 or 1, but the method could easily be extended to handle the case of finite contrast where \hat{a}_{ij} is no longer binary. The conditioned received data estimates are then compared to the

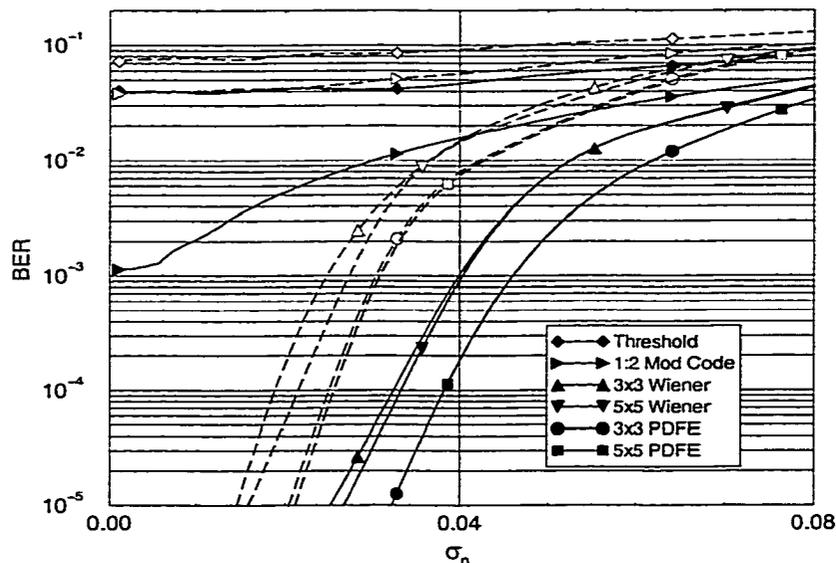


Figure 2.7: BER of Detection Schemes on Sinc^2 Beyond Resolution Channel

actual received data and each pixel chooses the state (*i.e.*, 1 or 0) corresponding to the smaller difference. Thus, each pixel chooses the state that maximizes its conditional probability. We perform these computations in parallel for each pixel. Such a parallel implementation requires a multiply and accumulate architecture with local connectivity whose size equals the channel PSF support, which is the same requirement necessary for implementation of the Wiener filter. Eqs. 2.9-2.11 define an iterative process that can be repeated to further refine the estimate of the data page. Figs. 2.5 and 2.6 summarize the performance of all the detection methods on the two resolved channels. Figs. 2.7 and 2.8 present the same data for the beyond resolution channels. Note that all results involving the PDFE algorithm use two iterations. Next we examine simulation results of the various algorithms acting on the eight optical systems of interest.

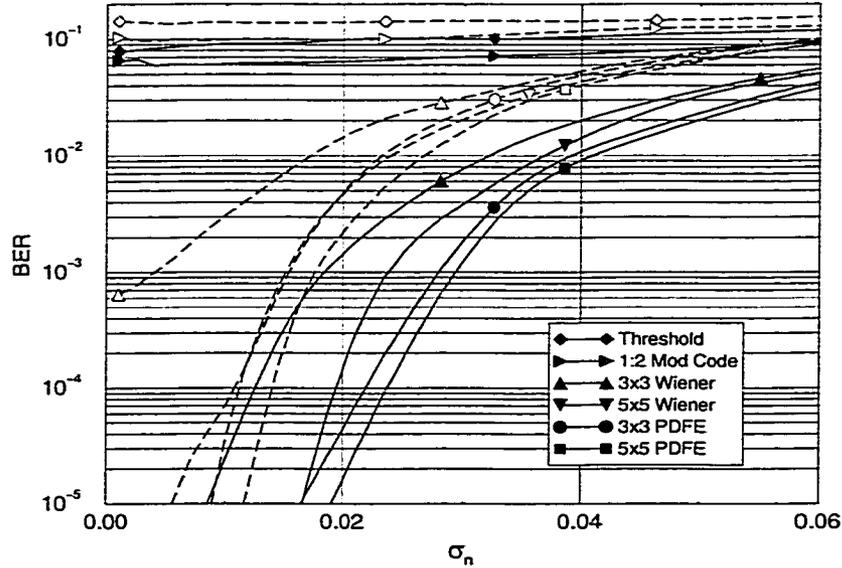


Figure 2.8: BER of Detection Schemes on Gaussian Beyond Resolution Channel

2.4 Simulation Results

We consider the overall memory system interface to include a channel decoder operating on the output from our decision system. We select a Reed-Solomon (15,11,2) 2-error correcting code as a channel decoder with a code rate of $11/15 = 73\%$ allowing a post-detection bit error rate of 10^{-4} and yielding an output bit error rate of 10^{-12} . For any decision scheme there is a specific maximum tolerable noise level, σ_n^* , associated with operating at a BER of 10^{-4} . If we consider, for example, a holographic optical memory, then we can apply the well-known scaling law^{31,35} relating the signal-to-noise ratio (SNR) to the number of stored pages in the memory, M :

$$\frac{P_s}{\sigma_n} = \frac{\Gamma}{M^2}, \quad (2.12)$$

where P_s is the input bit signal power, set in simulation to be unity, and Γ is a system dependent scaling constant. Now we can relate the maximum tolerable

noise level to the maximum number of pages that can be stored in the memory while maintaining a BER $\leq 10^{-4}$ through the equation

$$M^* = \sqrt{\frac{\Gamma}{\sigma_n^*}}. \quad (2.13)$$

Our final step is to reexpress M in terms of the system capacity, C , which is defined as

$$C = MN^2R, \quad (2.14)$$

where N^2 is the number of pixels per stored data page and R is the code rate of the ECC and, when appropriate, the modulation code. Typically R is just the code rate of the Reed-Solomon ECC. Making a substitution we arrive at an equation relating σ_n^* to the overall system capacity,

$$C = N^2R\sqrt{\frac{\Gamma}{\sigma_n^*}}. \quad (2.15)$$

In order to compare the capacities of the various algorithms operating on the different channels we require an absolute capacity baseline. This baseline was derived in ref. 30 and represents an information-theoretic bound for a perfect optical system corrupted only with Gaussian noise. In ref. 30 a simplified formula for the holographic system scaling constant, Γ , is used:

$$\Gamma = \frac{P}{4k_B T N^3 r^2}, \quad (2.16)$$

where P is the total optical readout power, k_B is Boltzmann's constant, T is the operating temperature, N^2 is the number of pixels per page, and r is the readout page rate. Furthermore it is assumed that the page is read out row-wise parallel (*i.e.*, N pixels per clock). Using a total power of 1 Watt, $T = 300$ K, with $N = 128$ and $r = 1$ kHz yields $\Gamma = 28.8 \times 10^6$.

The above equations allow us to reexpress the performance of any of the detection schemes in terms of the relative capacity they achieve with respect to C_{\max} , an information-theoretic upper bound on the capacity of a Gaussian noise limited channel with a perfect optical system. The relative capacity is computed as $C_{\text{ref}} = C/C_{\max}$.

2.4.1 Gaussian Noise

We now examine the BER performance of the various decision systems under the assumption that the additive thermal Gaussian noise level is the limiting noise in the system. Figs. 2.5 through 2.8 show the BER for the PDFE algorithm, Wiener filter alone, simple thresholding, and the 1:2 modulation code. The line type denotes the SLM fill factor with a solid line and filled symbols representing $\Delta/D = 0$ and the dashed line with open symbols representing $\Delta/D = 0.9$. The performance of the Wiener filter and PDFE algorithm using 3×3 and 5×5 support are presented to show the advantage gained for the corresponding increase in connectivity. Higher support sizes are not presented as they never offered any substantial increase in performance. The noise level is represented through the standard deviation of the Gaussian noise, σ_n .

Important trends and results to note include the following. The simple threshold fails to achieve the target 10^{-4} BER in every case except the sinc² resolved $\Delta/D = 0$ case. Similarly, the 1:2 modulation code only reaches the BER goal in the sinc² resolved (for $\Delta/D = 0$ and $\Delta/D = 0.9$) and Gaussian resolved ($\Delta/D = 0$) cases. The Wiener filter meets the BER goal in all eight cases with the exception of the 3×3 Wiener filter on the Gaussian beyond resolution $\Delta/D = 0.9$

case, which is the most difficult of the channels. The PDFE approach works in every case. It is also notable that for the same support size the PDFE algorithm always performs better than its Wiener filter counterpart. This implies that the local iteration of the PDFE approach is providing a non-linear improvement since the Wiener filter alone offers the lowest BER for any linear processing decision scheme. Also, comparing the performance gain achieved through increasing the support from 3×3 to 5×5 we see that it is insignificant for all cases except the sinc^2 resolved case. This can be explained by noting that the blur in the sinc^2 resolved case is small enough that the nine taps in the 3×3 Wiener filter and channel model offer enough degrees of freedom to invert the effect of the channel blur. Adding more taps fails to significantly improve performance since the new taps are nearly zero.

From these BER versus noise level curves, we can find the noise level corresponding to our target 10^{-4} BER goal and map this value of σ_n^* into a capacity using eq. 2.15. We can then express this capacity relative to the information-theoretic capacity bound as $C_{\text{ref}} = C/C_{\text{max}}$. Each curve can therefore be characterized by a single number representing the relative capacity of the associated detection scheme. Table 2.1 shows the relative capacities for the Gaussian noise limited case at six different SLM fill factors. Note that the PDFE algorithm always provides a higher relative capacity than the Wiener filter alone for the same support size. With the exception of the Gaussian not resolved ($\Delta/D = 0.9$, $\Delta/D = 0.8$) cases, the 3×3 PDFE always outperforms the 5×5 Wiener filter. The 1:2 modulation code compares poorly in terms of capacity with the Wiener

filter and PDFE techniques even in the cases where the modulation code has good BER performance. This is due to its low code rate of $1/2$.

In order to reduce the amount of data from here on we will only consider two SLM fill factors, $\Delta/D = 0$ and $\Delta/D = 0.9$. In order to relate this choice to real devices, consider two examples of commercially available SLMs: Displaytech's 256×256 Liquid Crystal SLM with an areal fill factor of 87% and Boulder Nonlinear Systems' 256×256 FLC device with a 60% fill factor. Also, the Holographic Data Storage Symposium (HDSS) specifies the use of a 1024×1024 SLM in their final system with a $10.3 \mu\text{m}$ pixel and $15.6 \mu\text{m}$ pitch yielding a 66% fill factor. Using zero and 81% fill factors provides us with a reasonable bound on the performance with regard to a variety of device choices.

2.4.2 Systematic Error Dominated System

In addition to additive post detector noise we also expect the system to suffer from fixed pattern noise. We have considered three systematic error sources: magnification error, rotation error, and CCD shift error. Magnification error occurs when the system magnification does not perfectly match one SLM pixel to one (or an integer number of) CCD pixel(s). This can be caused by a small shift of either the SLM or CCD along the system's optical axis or mismatch among lenses in a $4F$ system. Rotation error occurs when the SLM and/or CCD are rotated about the optical axis from their intended orientation. Shift error occurs when CCD pixels are laterally misaligned with those of the SLM.

Channel	Fill Factor (%)	Relative Channel Capacity (%)					
		TH	1:2 MC	WF 3 × 3	PDFE 3 × 3	WF 5 × 5	PDFE 5 × 5
Sinc ² Resolved	0	24.4	18.1	38.3	40.5	38.5	40.5
	4	22.9	17.8	38.3	40.0	38.4	40.0
	16	18.0	16.7	37.4	39.6	37.4	39.6
	36	0	14.6	35.6	38.4	35.6	38.4
	64	0	11.4	32.9	36.1	32.9	36.1
	81	0	8.3	31.2	34.1	31.7	34.2
Sinc ² Not Resolved	0	0	0	28.6	31.3	29.0	31.3
	4	0	0	28.2	31.3	28.8	31.3
	16	0	0	27.4	30.4	27.7	30.5
	36	0	0	26.0	28.3	26.6	28.5
	64	0	0	23.4	26.6	24.8	26.7
	81	0	0	22.1	25.2	23.4	25.5
Gaussian Resolved	0	0	9.6	33.4	37.7	34.5	37.7
	4	0	8.4	33.2	37.7	34.1	37.7
	16	0	5.6	32.1	36.3	33.1	36.7
	36	0	0	29.7	34.1	31.7	34.3
	64	0	0	27.2	31.9	29.4	32.1
	81	0	0	25.3	30.1	27.6	30.4
Gaussian Not Resolved	0	0	0	18.0	23.7	22.8	24.6
	4	0	0	17.1	23.2	22.1	24.4
	16	0	0	15.9	22.3	21.4	23.2
	36	0	0	13.3	20.6	19.7	22.1
	64	0	0	6.9	18.1	18.2	19.9
	81	0	0	0	16.1	17.3	19.6

Table 2.1: Channel Capacities of Decision Schemes for Gaussian Noise Dominated Case

Note that magnification and rotation errors can be considered to produce a space-varying optical channel. We can incorporate the effect of such a space-variant PSF by adjusting the detection schemes to use the appropriate local PSF at each pixel. To speed up our simulations we modeled blocks of 16×16 pixels to have the same PSF, computed from the actual space-variant PSF at the center of the block. Therefore for the 128×128 simulation page size we modeled the space-variant channel with 64 different PSFs.

Figs. 2.9-2.11 show the performance of the PDFE algorithm compared with the other techniques, presented in terms of the relative channel capacity. Once again solid lines with filled symbols represent $\Delta/D = 0$, and dashed lines with open symbols represent $\Delta/D = 0.9$. We see that for all three fixed pattern noises the difference in capacity between the PDFE algorithm based on 3×3 support and 5×5 support is minimal for small errors; however, the larger support does make the system more tolerant to large errors.

Similarly, we see trends as in table 2.1, with the 1:2 modulation code offering the lowest capacity due to its low code rate of $1/2$; the notable difference being that in a number of cases the Wiener filter outperforms the PDFE of the same support size for very large fixed pattern errors. This can be explained by looking at the difference in the blurred energy at a target pixel versus the energy at nearby pixels. When the channel blur becomes very wide the desired pixel receives about the same amount of energy as its neighbors. This poses a difficulty for the PDFE algorithm as it tries to exploit the differences in the received energies to determine which pixel states are more likely. The Wiener filter does not suffer from this problem since it incorporates a linear restoration filter before using a threshold

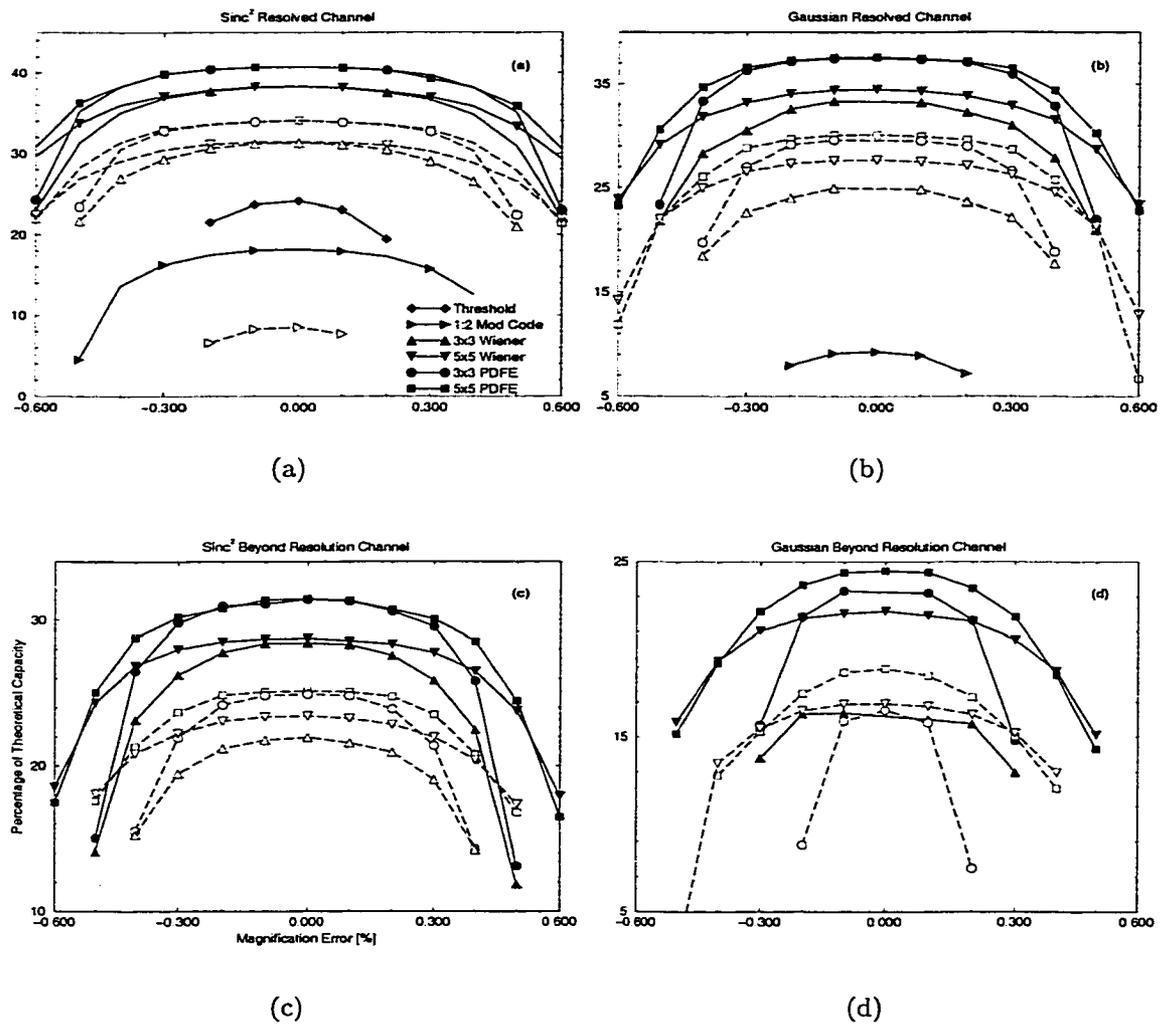


Figure 2.9: System Capacity in the Presence of Magnification Error

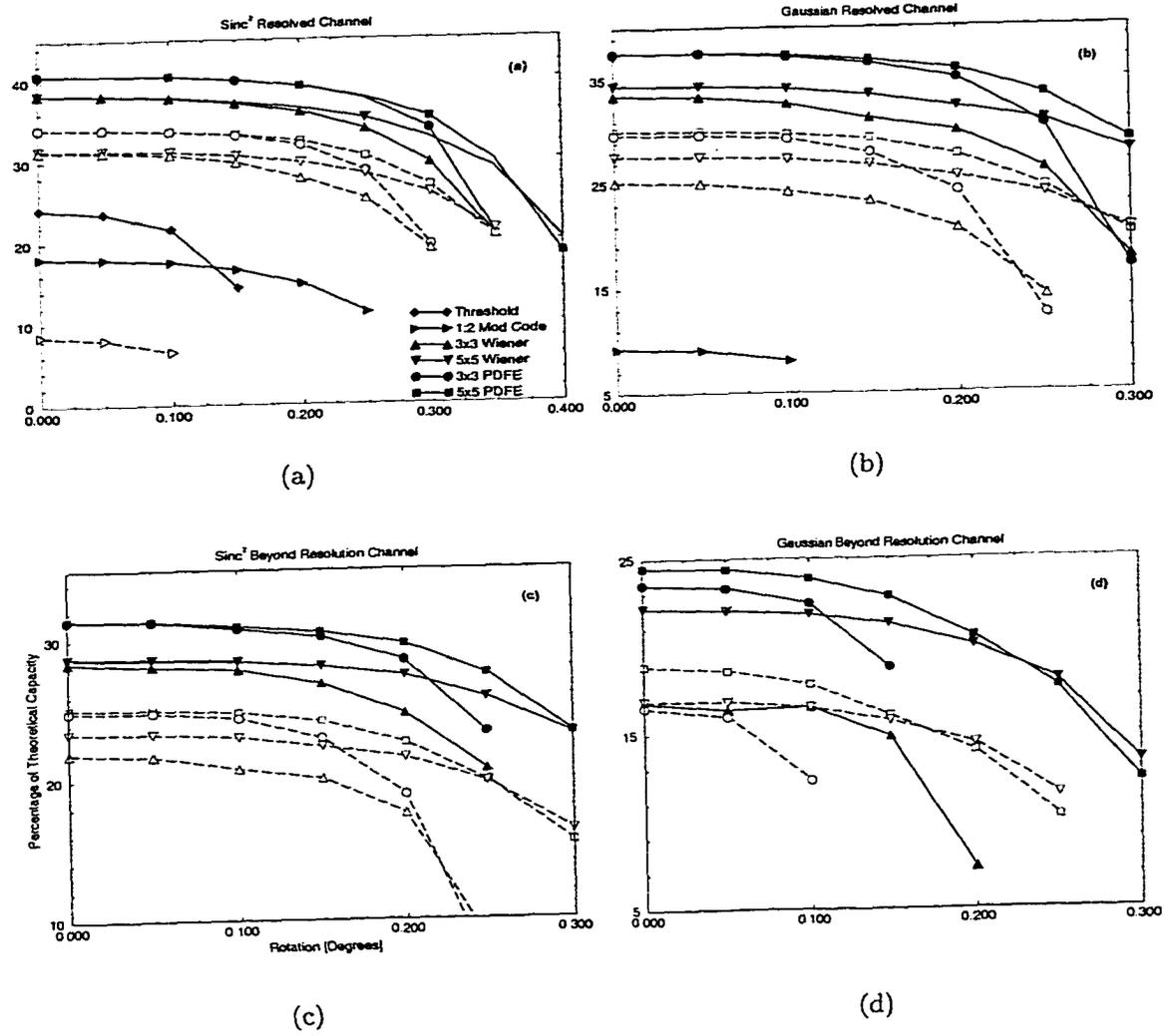


Figure 2.10: System Capacity in the Presence of Rotation Error

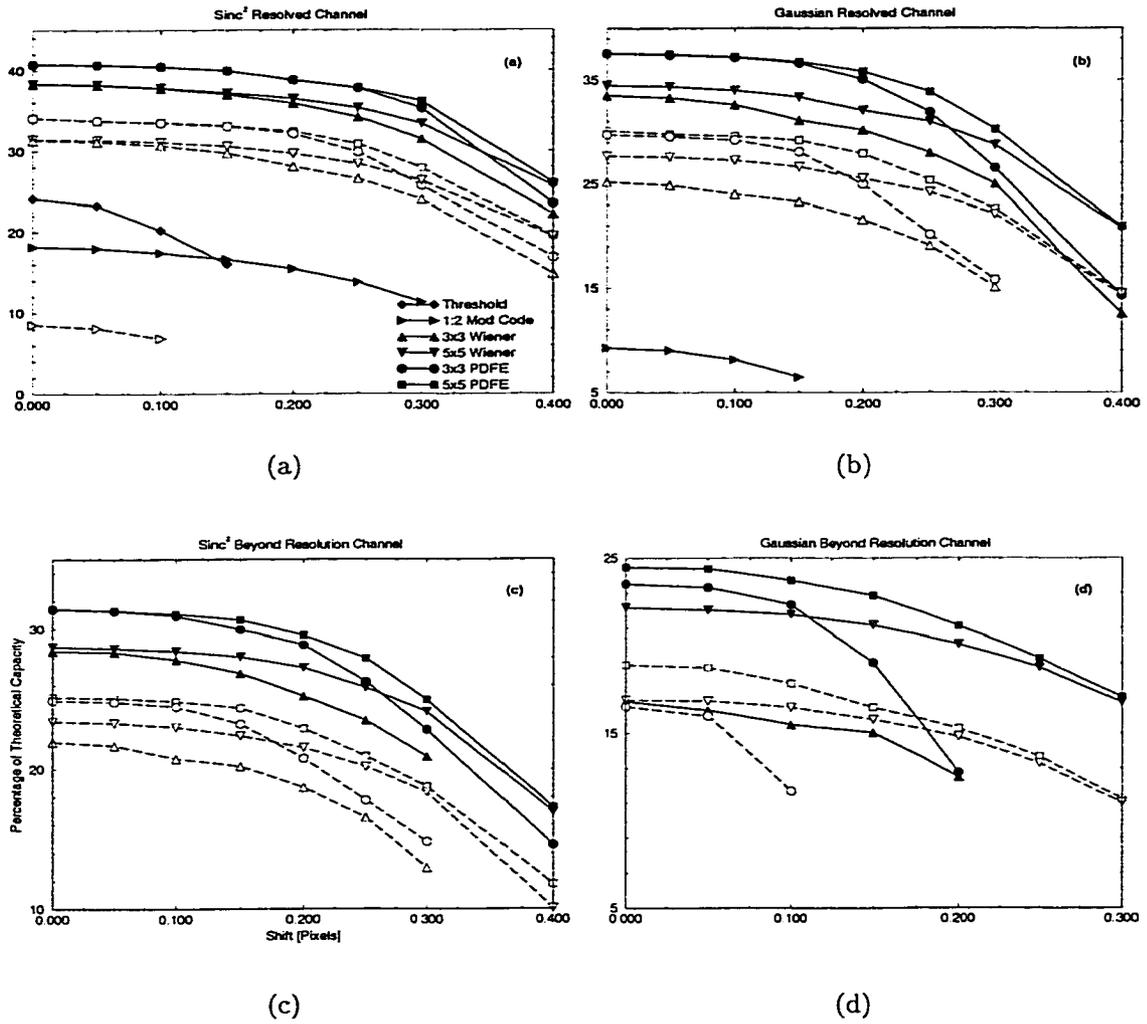


Figure 2.11: System Capacity in the Presence of Shift Error

decision. The filter attempts to redistribute the energy of the pixel back into the center. A similar front-end filter could be applied before every iteration of the PDFE algorithm to alleviate this problem; however, the cost is an increase in the storage requirements and algorithmic complexity.

We can characterize the tolerance of these algorithms to fixed pattern noise by defining an ad-hoc metric, the 90% tolerance width. This is defined with the aid of figs. 2.9-2.11 as the range of the fixed pattern noise parameter over which the capacity remains above 90% of its noise-free value. We denote this width as $M_{90\%}$ for magnification error, $S_{90\%}$ for shift error, and $R_{90\%}$ for rotation error. Mapping figs. 2.9 through 2.11 into tolerance widths yields tables 2.2 through 2.4.

Channel	Magnification 90% Width (% Magnification Error)					
	TH	1:2 MC	WF 3 × 3	PDFE 3 × 3	WF 5 × 5	PDFE 5 × 5
Sinc ² Resolved						
FF = 0	0.17	0.28	0.42	0.47	0.46	0.48
FF = 0.9	0	0.13	0.35	0.39	0.44	0.43
Sinc ² Not Resolved						
FF = 0	0	0	0.33	0.36	0.44	0.41
FF = 0.9	0	0	0.28	0.28	0.38	0.35
Gaussian Resolved						
FF = 0	0	0.16	0.33	0.39	0.43	0.42
FF = 0.9	0	0	0.28	0.30	0.39	0.36
Gaussian Not Resolved						
FF = 0	0	0	0.27	0.22	0.36	0.30
FF = 0.9	0	0	0	0.12	0.33	0.22

Table 2.2: Magnification 90% Tolerance Width

Channel	Rotation 90% Width (Degrees)					
	TH	1:2 MC	WF 3 × 3	PDFE 3 × 3	WF 5 × 5	PDFE 5 × 5
Sinc ² Resolved						
FF = 0	0.10	0.16	0.24	0.27	0.27	0.28
FF = 0.9	0	0.07	0.19	0.23	0.25	0.25
Sinc ² Not Resolved						
FF = 0	0	0	0.18	0.20	0.25	0.24
FF = 0.9	0	0	0.17	0.16	0.22	0.20
Gaussian Resolved						
FF = 0	0	0.09	0.19	0.22	0.25	0.24
FF = 0.9	0	0	0.17	0.18	0.22	0.21
Gaussian Not Resolved						
FF = 0	0	0	0.15	0.12	0.20	0.17
FF = 0.9	0	0	0	0.07	0.18	0.12

Table 2.3: Rotation 90% Tolerance Width

Channel	Shift 90% Width (Pixels)					
	TH	1:2 MC	WF 3 × 3	PDFE 3 × 3	WF 5 × 5	PDFE 5 × 5
Sinc ² Resolved						
FF = 0	0.08	0.16	0.24	0.28	0.27	0.29
FF = 0.9	0	0.07	0.20	0.24	0.25	0.25
Sinc ² Not Resolved						
FF = 0	0	0	0.19	0.22	0.25	0.24
FF = 0.9	0	0	0.17	0.17	0.22	0.21
Gaussian Resolved						
FF = 0	0	0.09	0.20	0.22	0.25	0.25
FF = 0.9	0	0	0.17	0.18	0.23	0.22
Gaussian Not Resolved						
FF = 0	0	0	0.15	0.12	0.21	0.18
FF = 0.9	0	0	0	0.07	0.18	0.13

Table 2.4: Shift 90% Tolerance Width

Because of the front-end processing issue discussed above, we see that the Wiener filter can be slightly more tolerant to the fixed pattern errors for the large blur channels. In the worst and best cases the 5×5 PDFE achieves a 0.22% to 0.48% tolerance to magnification error, 0.12 to 0.28 degree tolerance to rotation error, and a 0.13 to 0.29 pixel tolerance to shift error.

2.4.3 Imperfect Channel Knowledge

It is important to understand how the PDFE performance degrades when our *a priori* knowledge of the optical system becomes imperfect. In order to examine this phenomena we keep the optical system fixed at a set of operating parameters which include channel shape (Gaussian or sinc²), channel blur width (σ_b ; null spacing for sinc² channel and standard deviation for Gaussian channel), magnification error, rotation error, and CCD transverse shift error. The tolerancing study builds the detection algorithms using correct knowledge of all the operating point parameters except the one under test, which is presented to the algorithm either with an additive error (for tolerancing shift and rotation errors) or with a multiplicative error (for tolerancing channel width and magnification error). The multiplicative tolerancing works by expressing the percent change in the error parameter from the operating point. For example, a 0.2% change in a magnification error operating point of 0.1% would correspond to an actual magnification error of 0.1002%.

We toleranced four different parameters: channel PSF width, magnification error, rotation error, and shift error. For each choice of a parameter the algorithm performance was computed as a function of noise level in order to discover

the capacity of the decision algorithm operating under the desired tolerancing conditions. The tolerancing study was conducted at an operating point (OP) of a Gaussian resolved channel with a magnification error of 0.1%, a shift error of 0.05 pixels, and a rotation error of 0.05 degrees.

Figs. 2.12-2.15 show the tolerance of the algorithms designed at the OP indicated above as a function of channel knowledge error. With regard to channel blur width tolerance (fig. 2.12), we see the 3×3 PDFE can operate at 90% or better of its peak performance with a change from -30.61% to 47.7% in the tolerated blur width. The 3×3 Wiener filter can operate over the slightly smaller range from -21.41% to 43.77%. For the 5×5 support case the trend continues with the PDFE operating over the range -30.44% to 37.85% and the Wiener filter on the reduced range of -20.38% to 24.23%. Another trend to notice is that the 3×3 PDFE and Wiener filter are more tolerant than their 5×5 equivalents for all tolerated parameters. The 90% tolerance width for the magnification (fig. 2.13), rotation (fig. 2.14), and shift (fig. 2.15) errors shows very little change between the 3×3 and 5×5 support and the PDFE consistently outperforms the Wiener filter of the same support size. Results suggest that exact channel knowledge is not required to see a benefit from the PDFE algorithm. Table 2.5 summarizes these results.

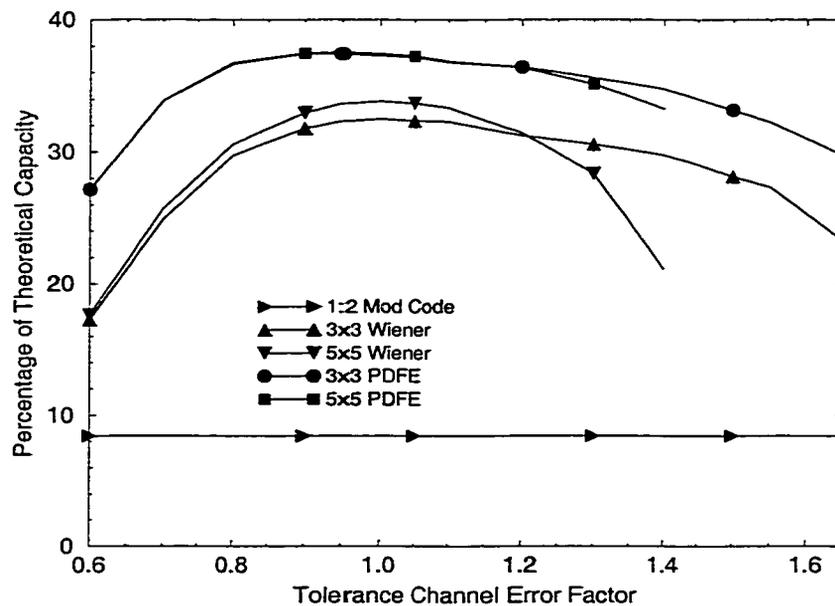


Figure 2.12: Tolerance of Decision Schemes to Knowledge of System Blur Width

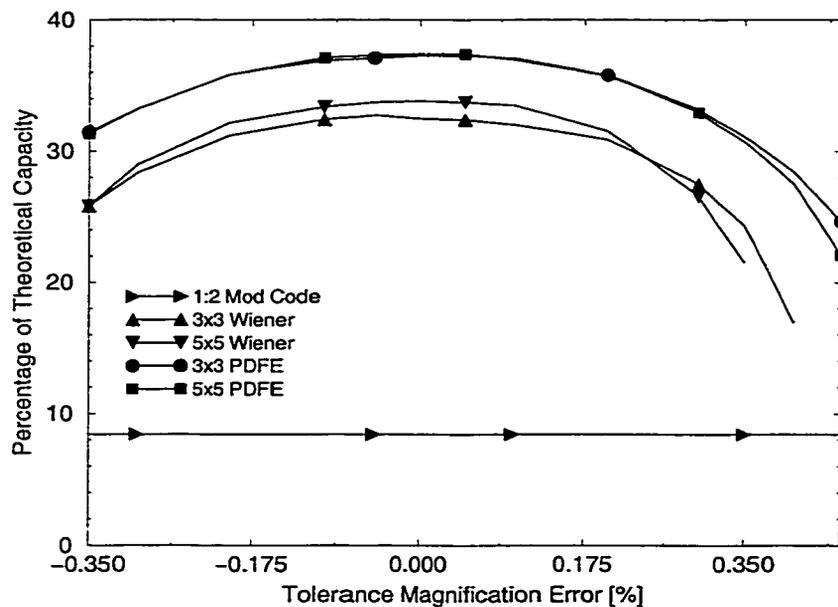


Figure 2.13: Tolerance of Decision Schemes to Knowledge of System Magnification

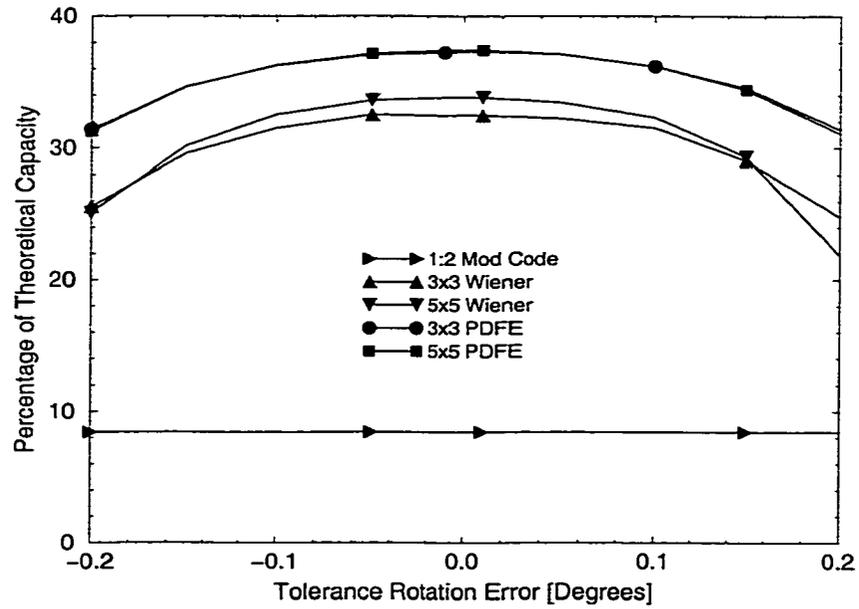


Figure 2.14: Tolerance of Decision Schemes to Knowledge of System Rotation

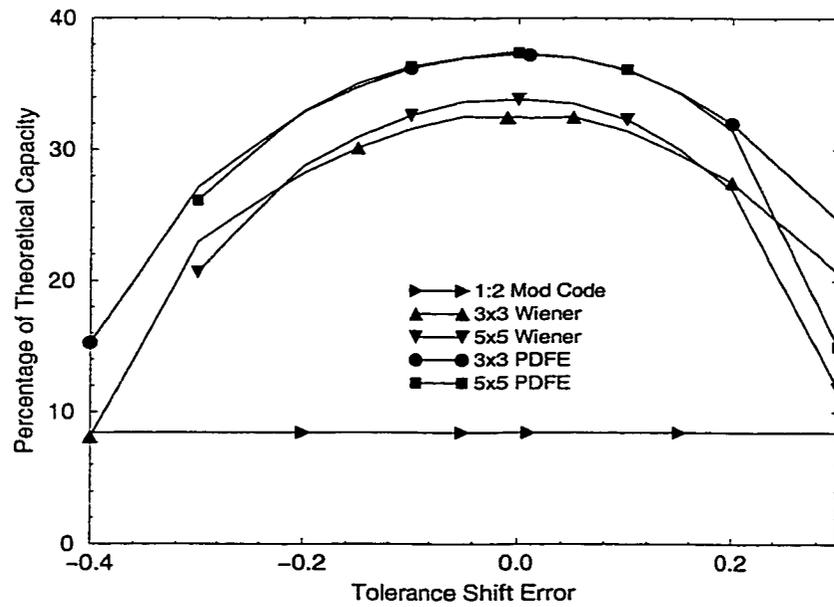


Figure 2.15: Tolerance of Decision Schemes to Knowledge of Shift

Decision Scheme	Channel Blur Width (% Change)		Magnification (% Magnification Error)	
	min	max	min	max
3 × 3 Wiener Filter	-21.41	43.77	-0.28	0.26
3 × 3 PDFE	-30.61	47.70	-0.29	0.29
5 × 5 Wiener Filter	-20.38	24.23	-0.27	0.23
5 × 5 PDFE	-30.44	37.85	-0.29	0.28
	Rotation (Degrees)		Shift (Pixels)	
	min	max	min	max
3 × 3 Wiener Filter	-0.16	0.15	-0.17	0.16
3 × 3 PDFE	-0.17	0.17	-0.18	0.17
5 × 5 Wiener Filter	-0.15	0.14	-0.16	0.14
5 × 5 PDFE	-0.17	0.16	-0.19	0.16

Table 2.5: 90% Tolerance Values in Prior Knowledge Study

2.5 Coherent Channel Model

All the results and discussion presented up to this point have focused on an incoherent channel. Extending these techniques to a coherent optical system requires an extension to the underlying algorithms. The perspective simply moves from intensity to field. The intensity PSF now becomes a field PSF and the intensity pixel profile becomes a field profile. The linear superposition and convolution relationships now hold in electric field. The detected signal however, remains in intensity and so all direct phase information in the received field is lost. Fig. 2.16 shows the pixel profile for a sinc, Rayleigh resolved channel under the two SLM fill factors of interest. We switch to the Rayleigh resolution criterion from the

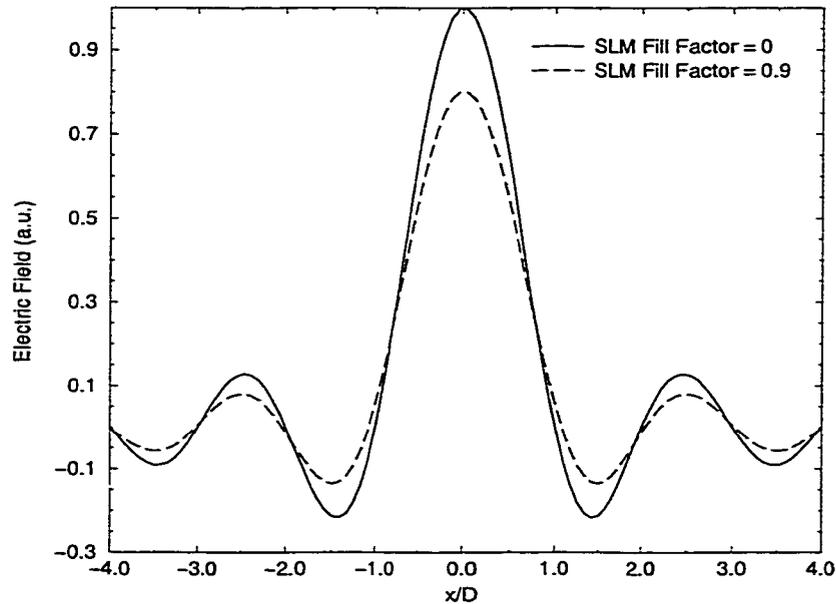


Figure 2.16: Coherent Sinc Rayleigh Resolved Pixel Profile at Detection Plane

Sparrow criterion used for the incoherent channels since the Rayleigh criterion is more appropriate for a coherent optical system.

We can write the received field incident on the CCD in terms of the coherent channel PSF as

$$E_o^c(x, y) = \sum_{ij} a_{ij} p^c(x - jD, y - iD) \otimes h_{chan}^c(x, y) \quad (2.17)$$

$$= \sum_{ij} a_{ij} h^c(x - jD, y - iD) \quad (2.18)$$

$$= \sum_{ij} a_{ij} E_{ij}^c, \quad (2.19)$$

where the superscript c denotes the coherent (field) version of the referenced function, D is the pixel pitch, and the double subscripted sum is a notational convenience for representing a sum over each index. Similarly, E_{ij}^c is a shorthand for the field contribution associated with pixel (i, j) defined in eqs. 2.18 and 2.19.

The incident spatial intensity pattern is therefore

$$I_o^c(x, y) = E_o^c(x, y)E_o^{c*}(x, y) \quad (2.20)$$

$$= \sum_{ij} \sum_{kl} a_{ij}a_{kl}^*E_{ij}^cE_{kl}^{c*} \quad (2.21)$$

$$= I_o(x, y) + \sum_{ij} \sum_{kl \neq ij} a_{ij}a_{kl}^*E_{ij}^cE_{kl}^{c*}. \quad (2.22)$$

We see the intensity from the coherent channel is simply the intensity from the incoherent version of the channel plus the cross-term contributions from all other pixels. The noise-free output of the CCD is the integrated intensity over the pixel elements which we express for pixel (i, j) as

$$y_{ij}^c = \int \int I_o(x, y) dx dy + \sum_{ij} \sum_{kl \neq ij} a_{ij}a_{kl}^* \int \int E_{ij}^c E_{kl}^{c*} dx dy \quad (2.23)$$

$$= y_{ij} + \sum_{ij} \sum_{kl \neq ij} a_{ij}a_{kl}^* x^c(ij; kl), \quad (2.24)$$

where we have defined the pixel integral over the field cross-terms between the pixels (i, j) and (k, l) as $x^c(ij; kl)$. All channel knowledge is contained in $x^c(ij; kl)$. Also note due to the symmetry of the cross-terms $x^c(ij; kl) = x^{c*}(kl; ij)$.

An incoherent channel has the convenient feature that the histogram of the IPI differed for the zero and one levels only by an offset. The coherent channel loses this symmetry as seen in fig. 2.17. This represents the histogram for a coherent sinc channel with a main lobe width equal to twice the CCD pitch. This width corresponds to the Rayleigh resolution criterion.

Because the received intensity can no longer be written as a simple linear superposition of the contributions from each pixel we need to extend the iterative PDFE stage distance computation to correctly model the received intensity as a function of the known channel response. This can be done by exchanging the

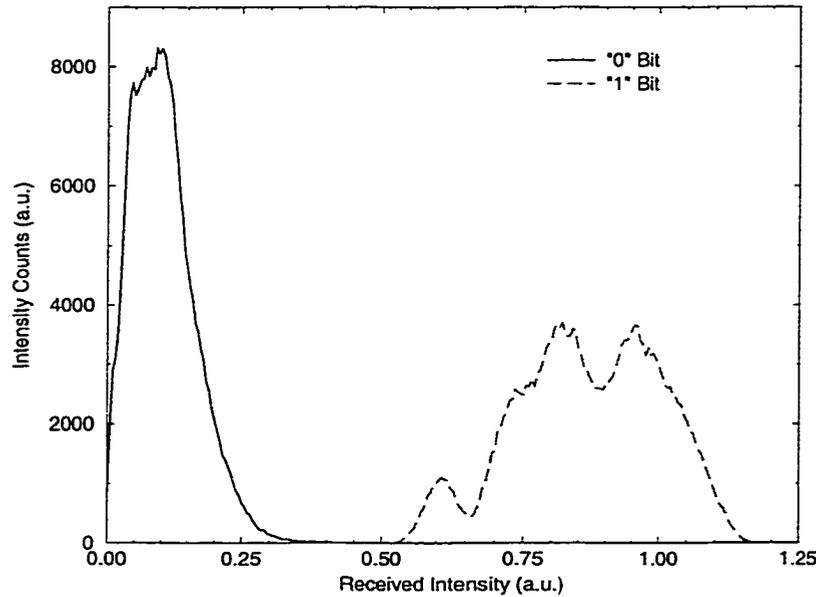


Figure 2.17: 0 and 1 Level Intensity Histogram for Coherent Sinc Channel

linear channel filter for a quadratic channel filter as in eq. 2.24. The quadratic approach, referred to as the QPDFE algorithm, has the associated disadvantage of increased implementation complexity and connectivity.

Before we examine the results for the coherent channel we need to consider what major differences exist between the incoherent and coherent processing. Due to the coherent superposition and the resulting quadratic channel model, the Wiener filter is no longer the minimum MSE linear equalizer when operating on the CCD intensity data. We considered a number of alternative approaches to arrive at a suitable replacement constrained to be a linear filter with compact local support. Among these were the adaline³⁶ (adaptive linear element), forward and inverse estimation filters, and Wiener filters computed from various versions of the intensity PSF. If strong assumptions are made about the phase of the coherent system then operating on the square root of the received intensity data

can be beneficial. In the square root domain the linear superposition holds and we expect the linear schemes proposed for the incoherent system to work with reasonable success.

Although some of the square root domain algorithms did perform fairly well, the intensity domain approaches performed better. There are two reasons for this difference in performance. First, additive thermal noise in intensity loses its simple additive signal independent nature in the square root domain. Second, in the square root domain the arithmetic sign of the true field quantity is lost. This leads to a potentially large difference between the square root of the received intensity value and the model. For these approaches to work better we need to consider techniques that estimate the phase of the received field and can retain the true sign of the field quantity.

The optimal minimum MSE linear filter for the coherent case can be derived as a simple extension of the method used to compute the incoherent Wiener filter. However, the Wiener-Hopf equations become sufficiently complicated to make the numerical computation non-trivial. Because the Wiener filter constructed based on the average intensity PSF performs with a negligible loss compared to the exact minimum MSE filter and because the computation of the average intensity Wiener filter is trivial, we will use the average intensity version as the Wiener filter algorithm for the coherent case and as the initial stage for the QPDFE algorithm. The average intensity PSF is derived by noting for any specific data pattern there is a different measured intensity PSF. Or alternatively, the taps in the intensity PSF are a function of the transmitted data pattern. By computing

the expected (over the data bits) intensity PSF taps we arrive at the average intensity PSF.

Mathematically, the derivation for the average intensity PSF begins with the quadratic form of the received (noise-free) intensity:

$$y_{ij}^c = \sum_{ij} a_{ij} \left\{ \sum_{kl} a_{kl}^* x^c(ij; kl) \right\} \quad (2.25)$$

$$= \sum_{ij} a_{ij} \left\{ a_{ij}^* x^c(ij; ij) + \sum_{kl \neq ij} a_{kl}^* x^c(ij; kl) \right\}. \quad (2.26)$$

We want to express the received intensity as a linear combination of the data bits and the taps of a filter. We can define the quantity in the curly braces of eq. 2.26 as taps of a filter, \tilde{w}_{ij}^c , but the difficulty now lies with the dependence of \tilde{w}_{ij}^c on the complete data pattern. The elements of the average intensity PSF, $\{w_{ij}^c\}$, are found by applying the expectation operation to the taps $\{\tilde{w}_{ij}^c\}$ and remembering $a_{ij} \in \{0, 1\}$ for the binary pulse amplitude modulation case we consider here:

$$w_{ij}^c = \langle \tilde{w}_{ij}^c \rangle \quad (2.27)$$

$$= a_{ij}^* x^c(ij; ij) + \left\langle \sum_{kl \neq ij} a_{kl}^* x^c(ij; kl) \right\rangle \quad (2.28)$$

$$= x^c(ij; ij) + \frac{1}{2} \sum_{kl \neq ij} x^c(ij; kl), \quad (2.29)$$

where the brackets represent expectation over all data patterns (local neighborhoods). Since the channel and SLM pixel profile define the values of $x^c(ij; kl)$ we can immediately compute the taps of the average intensity PSF.

Fig. 2.18 shows the BER performance of the various algorithms on the coherent sinc channel for both $\Delta/D = 0$ (solid lines, filled symbols) and $\Delta/D = 0.9$

(dashed lines, open symbols). The relative performance of the threshold, Wiener filter, and QPDFE (the Quadratic complexity version) is as expected. Note that although the 1:2 modulation code has the best BER performance over any of the approaches it nevertheless has the worst capacity due to its low code rate of 1/2. Basically, it achieves the better BER by reducing the density of one bits which introduces less cross talk but as the simulations show this is an inefficient method for recovering capacity. The noise level achieving the 10^{-4} BER goal, σ_n^* , is once again used to allow us to compute the associated capacities which are presented in table 2.6.

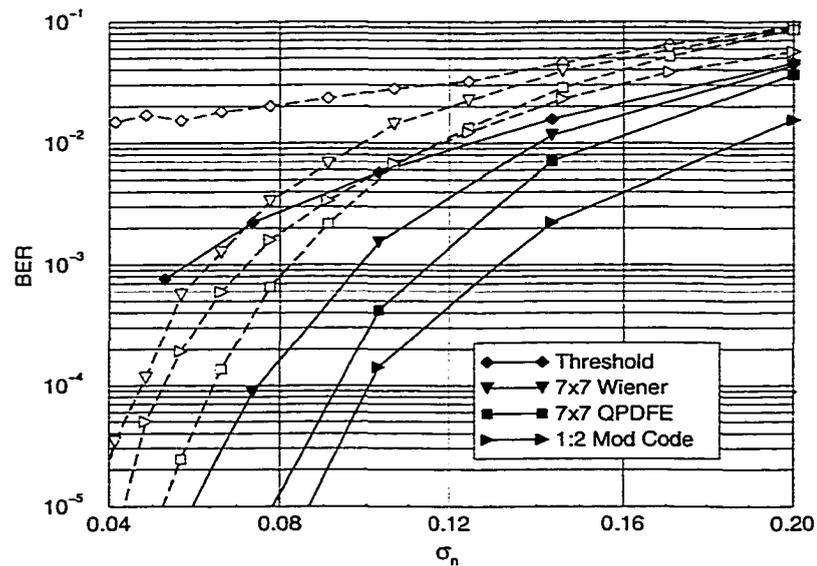


Figure 2.18: BER of Detection Schemes on Coherent Sinc Channel

It is also important to note that the size of the filter for QPDFE and the Wiener filter has increased to 7×7 from the incoherent channel sizes of 3×3 and 5×5 . This is because the energy distribution in the coherent channel falls

Channel	Relative Channel Capacity (%)			
	TH	1:2 MC	WF 7x7	QPDFE 7x7
Coherent Sinc Resolved				
FF = 0	25.38	25.3	43.85	48.71
FF = 81%	6.63	18.25	34.97	40.71

Table 2.6: Coherent Sinc Channel Capacities for Gaussian Noise Dominated Case off slower than its incoherent counterpart for the same value of null width. The larger kernel is necessary to account for more of the channel's energy.

2.5.1 Fixed Pattern Noise Dominated Coherent System

Figs. 2.19-2.21 show the capacity performance of the PDFE, Wiener filter, and 1:2 modulation code in the presence of magnification, rotation, and shift alignment errors. As expected the performance degrades quickly as the error sizes become non-negligible. As a relative comparison we see that all the algorithms degrade in a similar manner implying that they all are equally tolerant to systematic misalignments. For transverse shift errors we see that the performance in the $\Delta/D = 0$ case degrades quicker than the $\Delta/D = 0.9$ case. This difference in capacity decay rate is due to the increased pixel blur in the $\Delta/D = 0.9$ case which dominates the decay for the smaller misalignment errors. For the larger shift errors the fixed pattern error dominates the decay rate. The difference in decay rate is more obvious when we examine the 90% tolerance widths for the three alignment errors as shown in tables 2.7-2.9.

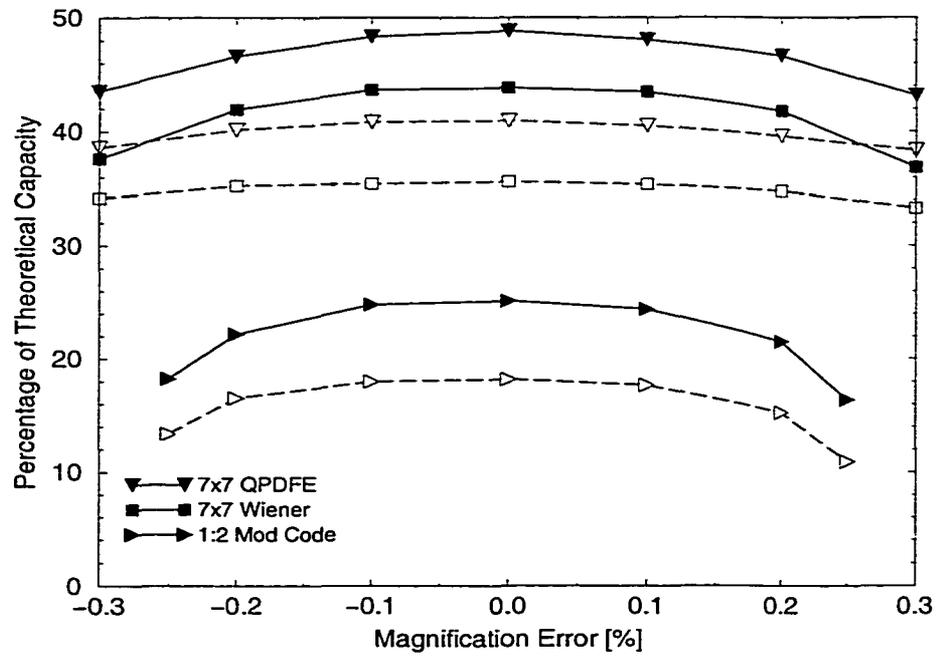


Figure 2.19: Coherent Sinc System Capacity in the Presence of Magnification Error

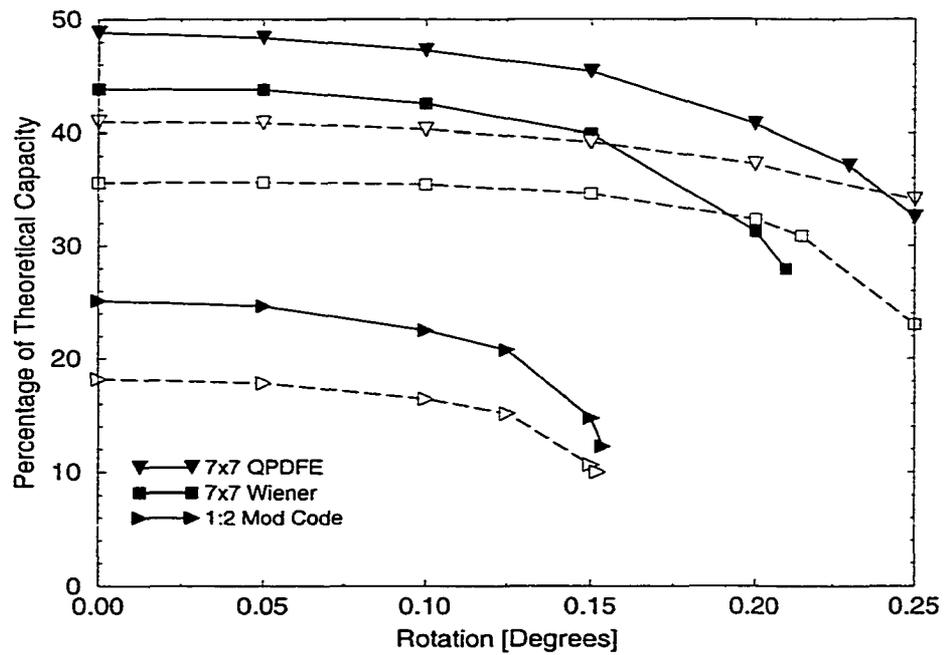


Figure 2.20: Coherent Sinc System Capacity in the Presence of Rotation Error

	Magnification 90% Width (Percentage Magnification Error)		
Channel	1:2 MC	WF 7x7	QPDFE 7x7
Sinc (k=1)			
FF = 0	0.188	0.258	0.288
FF = 81%	0.191	0.348	0.378

Table 2.7: Coherent Magnification 90% Tolerance Width

	Rotation 90% Width (Degrees)		
Channel	1:2 MC	WF 7x7	QPDFE 7x7
Sinc (k=1)			
FF = 0	0.0982	0.155	0.167
FF = 81%	0.101	0.204	0.207

Table 2.8: Coherent Rotation 90% Tolerance Width

	Shift 90% Width (Pixels)		
Channel	1:2 MC	WF 7x7	QPDFE 7x7
Sinc (k=1)			
FF = 0	0.171	0.209	0.181
FF = 81%	0.215	0.246	0.202

Table 2.9: Coherent Shift 90% Tolerance Width

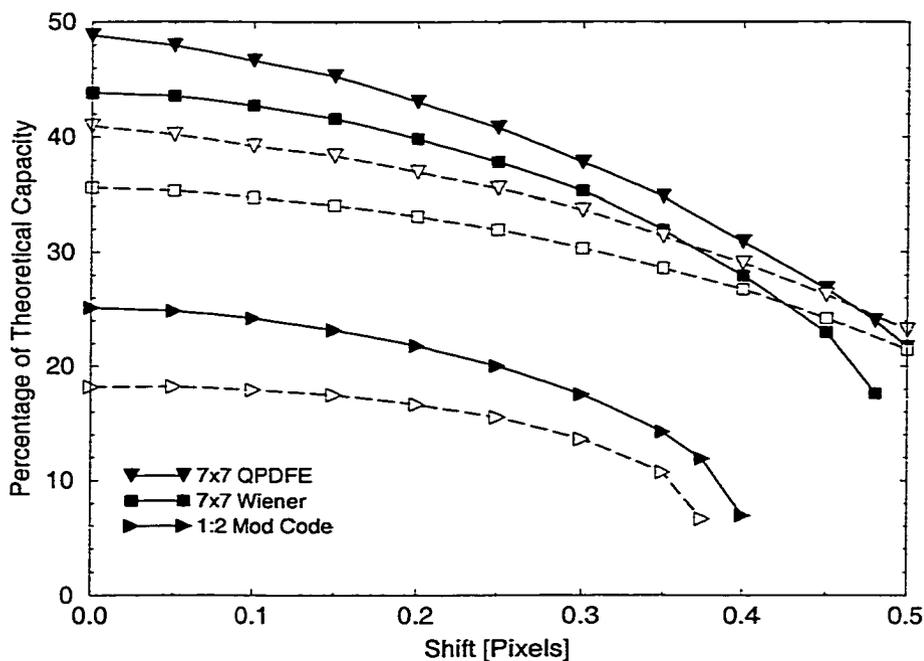


Figure 2.21: Coherent Sinc System Capacity in the Presence of Shift Error

The QPDFE has a larger 90% width than the Wiener filter and 1:2 modulation code for magnification and rotation errors, but for shift we see that the Wiener filter has a slightly larger width. The explanation for this behavior is the same as for the incoherent case. We have not exploited a front-end filter for the QPDFE to concentrate the pixel energy back into the center after every iteration.

2.5.2 Imperfect Coherent Channel Knowledge

In the same fashion as for the incoherent channel we want to examine how the capacity degrades when the *a priori* knowledge of the optical system becomes imperfect. The base optical system ran on a coherent sinc channel ($k = 1$) with a magnification error of 0.1%, a shift error of 0.05 pixels, and a rotation error of 0.05 degrees.

The QPDFE and Wiener filter algorithms were constructed with the correct parameters except for the single parameter under test. The channel blur width, magnification error, rotation error, and shift error were each perturbed around the base operating value and the performance of the QPDFE, Wiener filter and 1:2 modulation code was measured. Table 2.10 summarizes the resultant 90% tolerance widths. We see that the QPDFE algorithm outperforms the Wiener filter and the 1:2 modulation code. The QPDFE can tolerate a channel width error up to 6%, magnification error up to 0.13%, rotation error up to 0.08 degrees, and shift error up to 0.08 pixels.

Decision Scheme	Channel Blur Width (% Change)		Magnification (% Magnification Error)	
	min	max	min	max
7 × 7 Wiener Filter	-5.7	10	-0.149	0.11
7 × 7 QPDFE	-7.2	5.6	-0.143	0.12
	Rotation (Degrees)		Shift (Pixels)	
	min	max	min	max
7 × 7 Wiener Filter	-0.086	0.061	-0.095	0.061
7 × 7 QPDFE	-0.089	0.072	-0.095	0.066

Table 2.10: Coherent 90% Tolerance Values in Prior Knowledge Study

2.6 Implementation

The high aggregate data rates available through POMs drive the need for parallelism in the page decision stage; however, the amount of parallelism can be

limited by large area and heat dissipation requirements. It is important therefore to balance space and time multiplexing in the equalizer design.

In this section we analyze a VLSI implementation of the PDFE decision scheme and show that there exist viable hardware solutions. We choose a design and present the associated layout and fabrication parameters. The goal is to confirm theoretically and experimentally that the PDFE algorithm can be fabricated with currently available technology. We consider two types of designs: analog and digital.

Fig. 2.22 shows the general block diagram of a PDFE design. The digital design uses one or more m -bit analog-to-digital (A/D) converter(s) to digitize the analog output from a N pixel CCD row. The digital values are latched and passed to the digital Wiener filter. The filter output is then thresholded and the estimated data page is latched into a $k \times N$ bit memory. Iterative updates to the estimated data page are performed by comparing the N Wiener filter outputs with the result of the estimated data row processed by the forward channel filter. The filter units have $k \times k$ taps and compute with m bits of output accuracy. The analog design eliminates A/D conversion and computes using analog multipliers and summation units. If the analog design is to be appropriate the fidelity of the analog signal must be of acceptable quality. SPICE simulations show that the precision of our analog multiplier design will be approximately equal to 4 digital bits when operating in a MOSIS 2-micron symmetric CMOS n-well analog process (Orbit SCNA20). A more involved design can achieve even more precision if desired.

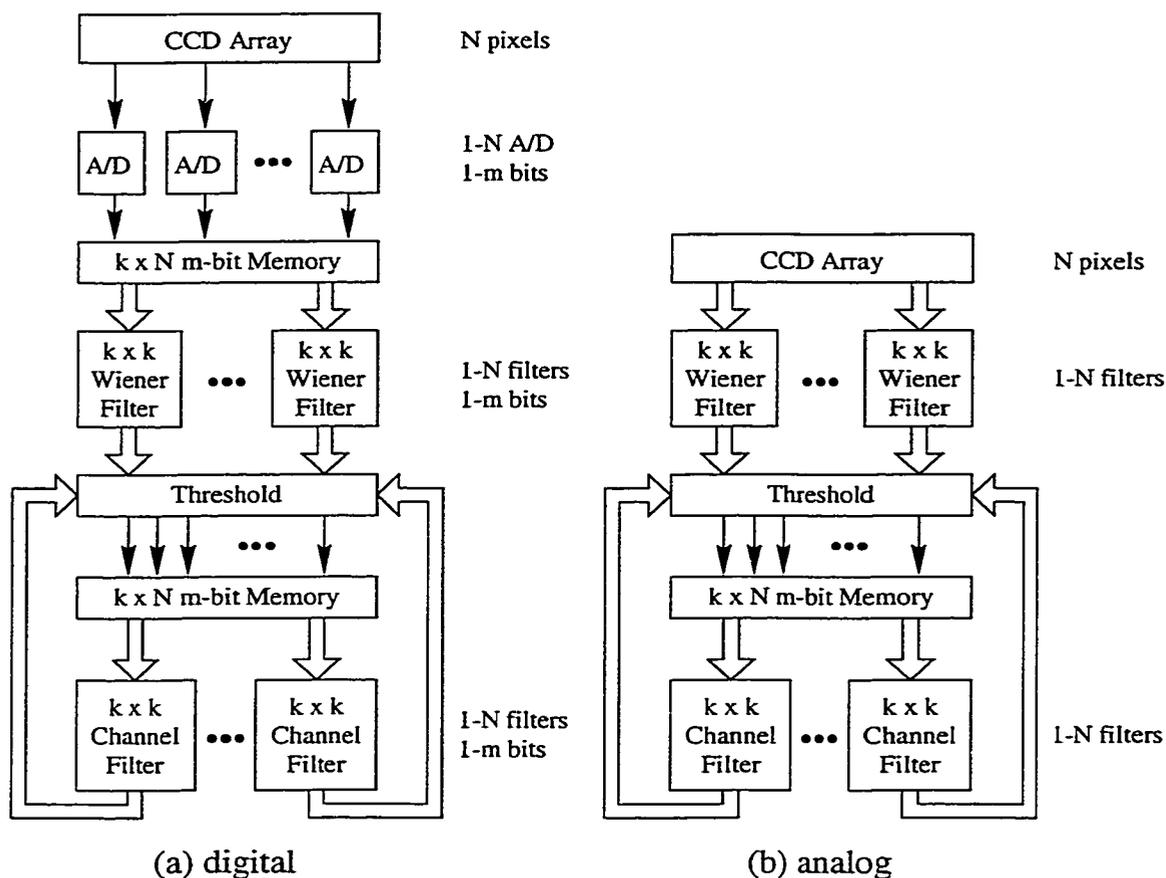


Figure 2.22: PDFE Implementation Block Diagram

Within both types of design we can adjust the amount of parallelism to trade-off the number of transistors with clock rate. This is accomplished by repeating stages (increasing parallelism) or reusing the same stage at a higher clock rate (decreasing parallelism). There are two design limits: row-wise parallel and fully serial; fully parallel processing becomes unmanageable due to the enormous transistor count. As an example of the possible parallel versus serial tradeoffs consider the digital design. Anywhere from 1 to N A/D convertors can be used. If 1 A/D converter is used it will have to run at N times the desired clock rate, while N A/D convertors can be run at exactly the desired row clock rate (*i.e.*, $N \times$ the

page rate). This tradeoff allows us to linearly adjust transistor count for a reduced clock rate. Similarly, the number of forward and inverse channel filters can be varied in the same fashion. Another tradeoff available within the digital design is the replacement of the m -bit wide A/D or filter stage units with a 1-bit wide component that is reused m times sequentially to produce the desired m -bit result.

We consider a design in terms of two metrics: the chip clock rate, and the number of transistors. These two metrics represent a number of actual design constraints. The clock rate imposes a limit on the switching speed of the transistors and also on the power dissipation. The number of transistors is proportional to chip area. Together the clock rate and number of transistors are proportional to the total power consumed. The area cost of the wiring is assumed to be small compared to the area cost of the transistors. This assumption is justified for row-wise parallelism (or less), small channel support and when many levels of metals are available for routing.

The scaling laws we derive for the PDFE implementation revolve around the design of a $N \times N$ CCD read out row-wise parallel at a page rate of 1 kHz (*i.e.*, if $N = 1024$ then all 1024 rows are readout in parallel with a 1 MHz clock). The channel model is of size $k \times k$ and the number of PDFE iterations is n_i .

We consider the scaling laws associated with both analog and digital designs. The designs tradeoff the number of processing units and the use of 1-bit or m -bit digital units. The N -row parallel digital design uses N A/D convertors and N forward and inverse channel filters. Likewise the N -row, m -bit parallel, N -unit, digital design uses N m -bit A/D convertors and N m -bit forward and inverse

channel filters. The N -row, m -bit serial, N -unit digital design uses N 1-bit A/D convertors run at m times the row clock rate as well as N 1-bit channel filters also run at m times the row clock rate. Similarly the 1-unit digital designs use only one A/D converter, one forward channel filter and one inverse channel filter run at N times the clock rate of the N -unit design. Table 2.11 summarizes the design options considered.

Design name	# of filter units	# of bits per unit	# of times unit clocked per page
Analog N -Row	N		N
Analog $N \times N$	N^2		1
Digital m -bit Serial 1 Unit	1	1	N^2
Digital m -bit Serial N Units	N	1	N
Digital m -bit Parallel 1 Unit	1	m	N^2
Digital m -bit Parallel N Units	N	m	N

Table 2.11: PDFE Design Options

In figs. 2.23 and 2.24 we present the transistor count and clock rate scaling laws, respectively, for the design: $k = 3$, $n_i = 3$, $m = 4$, and a page clock rate of 1 kHz. As a bound on parallelism a $N \times N$ fully parallel analog design is represented by the downward triangle symbol. Also shown on the figures are the maximum transistor count and clock rate for a MOSIS 2-micron orbit SCNA20 process and a competitive commercial fabrication process. The figures show that the N parallel analog design is the best choice in terms of both transistor count and clock rate yielding a possible 1024×1024 PDFE design with a 3×3 channel and 3 iterations using approximately 300,000 transistors and running at a clock rate of 1 MHz (achieving the page rate goal of 1 kHz). It should also be noted

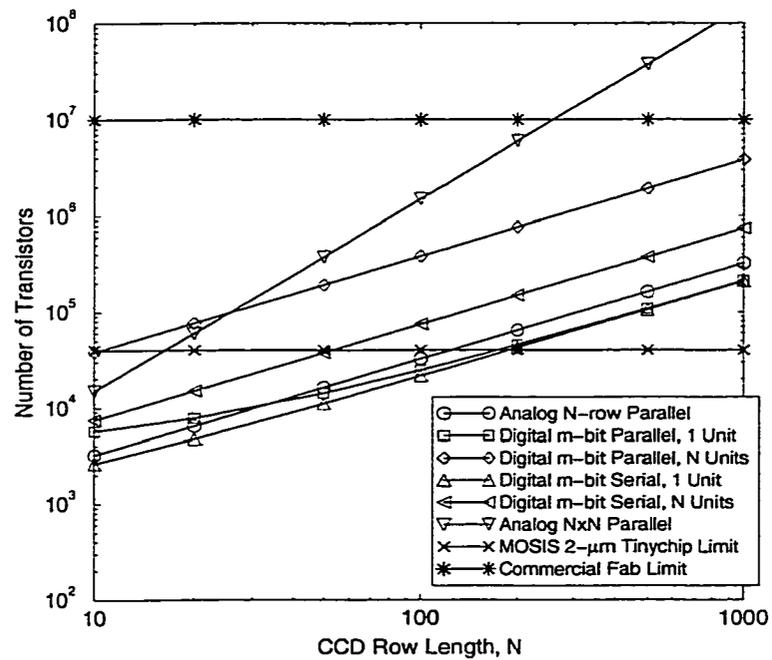


Figure 2.23: PDFE VLSI Transistor Scaling Law

that the chip area required for a transistor will be different for the analog and digital designs.

In the above discussion we have bound the VLSI resources necessary to construct the PDFE device. Other important details such as how to load the Wiener filter taps are suitably ignored and left for a more detailed future consideration. In the proof-of-concept PDFE design described next the Wiener filter taps are loaded electronically one tap at a time through the pins.

Wu-chun Chou, a fellow group member, engineered the mixed-signal VLSI design for the PDFE chip. The full details of the design are described in his dissertation.²⁰ In fig. 2.25, we present the VLSI layout of the MOSIS tinychip. The N -parallel analog design constructs a 9×9 CCD array implementing the PDFE algorithm with a 3×3 channel model and 3 iterations requiring 2250×2220

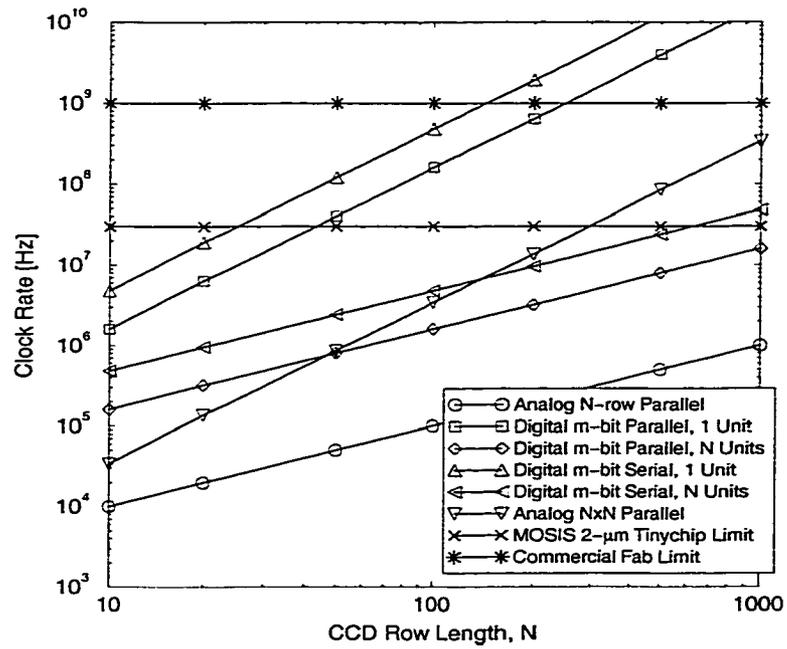


Figure 2.24: PDFE VLSI Clock Rate Scaling Law

μm^2 of area. The pixel active area is $20 \times 26 \mu\text{m}^2$ with a x,y pitch of $60 \times 190 \mu\text{m}$. This design represents the lower left corner of both metric figures.

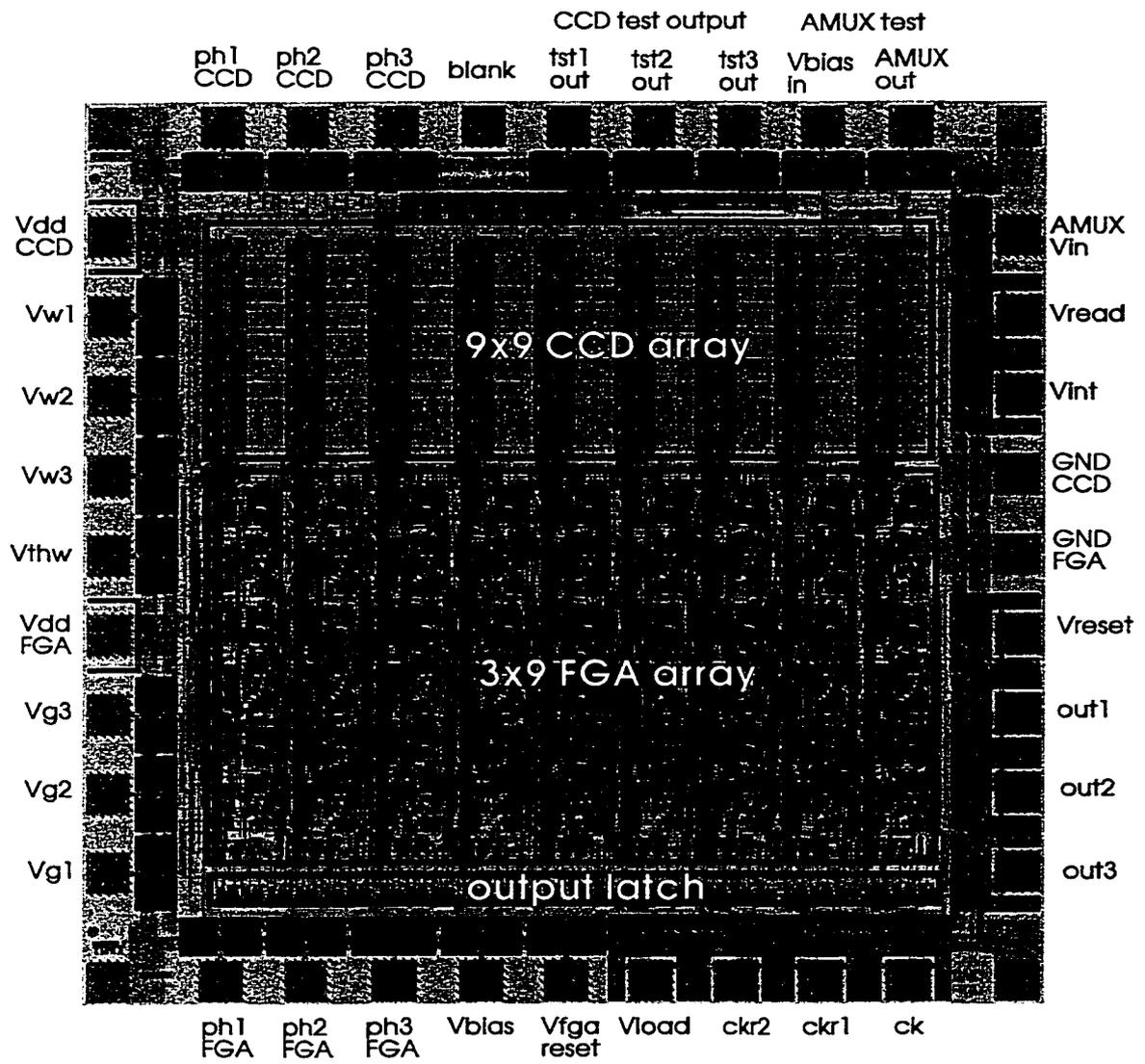


Figure 2.25: Proof of Concept Fabricated PDFE Chip Plot

2.7 Summary

2D page detection under conditions of significant blur is almost always the relevant case for a 2D optical data communication or storage channel. If the blur is insignificant, then the system can achieve a higher density, capacity and/or aggregate data rate by decreasing the pixel dimensions until IPI becomes a significant noise source. It is this case for which the PDFE detection algorithm is intended. Knowledge of the channel PSF is available either through models and/or experimental measurements. Using the channel knowledge we can improve the system performance by mitigating the effect of the cross talk which allows a simple fully parallel bit-by-bit detection scheme.

We first modeled the optical system as incoherent and examined the BER performance of the simple threshold, the 1:2 modulation code, the Wiener filter, and the PDFE algorithm operating on four different channels with two input SLM fill factors. By invoking a scaling law that relates system capacity to SNR, we presented the storage capacities associated with the various algorithms. For the incoherent channel the Wiener filter is the linear filter that achieves the lowest BER (and hence highest capacity since its code rate is 1). We found that the 3×3 Wiener filter offers a 107% increase in capacity as compared with simple threshold detection, when tested on the sinc^2 Sparrow resolved channel. We also found that the non-linear PDFE algorithm significantly improves on the Wiener filter estimate of the data page using the same 3×3 or 5×5 local connectivity. For the sinc^2 resolved channel, PDFE achieves an additional 9.7% capacity gain over the Wiener filter and on the most difficult IPI channel considered (Gaussian 25%

beyond resolution channel), the 3×3 PDFE yields a 162% capacity improvement over the 3×3 Wiener filter for a SLM fill factor of 64%. We also examined the degradation of performance as systematic errors such as rotation, shift, and magnification were introduced and discovered that PDFE can tolerate up to a 0.48% magnification error, 0.28 degree rotation error, and 0.29 pixel shift error (tolerance being determined as the point where the performance drops to 90% of its value for no systematic errors). Tolerancing against prior channel knowledge was conducted at an operating point with 0.1% magnification error, 0.05 pixel shift, and 0.05 degrees of rotation. Notable results included a tolerance of up to 48% error in channel blur width, 0.29% magnification error, 0.17 degrees rotation error, and a pixel shift of 0.17. Tolerance in this case is defined as the point where the performance drops to 90% of its value at the tolerance operating point.

We also considered a coherent optical system and developed a modified PDFE algorithm for use in a system where simple pixel intensity superposition no longer applies. We determined that the previously linear (in number of channel taps) complexity of the algorithm needed to be increased to quadratic (the number of filter taps increases as the square of the number of channel taps) in order to achieve an acceptable performance level. The performance of the modified (QPDFE) algorithm was demonstrated on the coherent channel and significant capacity improvements were found. For example, we recovered an additional 17.3% capacity by using the 7×7 QPDFE over the 7×7 Wiener filter. We found similar tolerances to systematic errors including a 0.38% tolerance to magnification error, 0.21 degree tolerance to rotation, and 0.2 pixel tolerance to shift.

A preliminary analysis of the VLSI resources necessary to realize the PDFE design showed that with current VLSI technology we could construct a 1024×1024 CCD read out row by row. The design would correct for a 3×3 pixel optical channel using three iterations of the PDFE algorithm and operate at 1000 frames per seconds (1 Mbit per second). A proof-of-concept fabrication in a MOSIS $2\text{-}\mu\text{m}$ analog process of a 3×3 PDFE design with 81 pixels was also described.

Notes This work has been published in *Applied Optics* as described in references [28,29]. Wu-chun Chou deserves much praise for his hard work in designing and fabricating the proof-of-concept VLSI implementation! I am very grateful for his help.

CHAPTER 3

APODIZATION

3.1 Introduction

The key to achieving an enormous storage capacity in a volume holographic memory is the ability to multiplex a large number of holograms within the same physical volume. The choice of a multiplexing scheme dictates much of the system architecture necessary to allow extraction of the desired hologram from the collection. There are a large number of proposed multiplexing architectures,^{19,37-41} each of which has its own unique advantages and disadvantages. We will primarily consider angular multiplexing but much of what is presented here can be easily applied to a variety of other multiplexing approaches.

A hologram is created by interfering the data-bearing object beam with a reference beam at a specific choice of the multiplexing parameter, say ψ_0 . Later, when we wish to reconstruct the desired object beam, we use the reference beam corresponding to ψ_0 resulting in a high-fidelity reconstruction of the object beam. As we perturb the multiplexing parameter, in general, the irradiance and quality of the reconstructed image decay. We wish to investigate in considerable detail precisely how the diffraction efficiency changes as we adjust the multiplexing parameter away from ψ_0 .

When there are multiple holograms stored in the same memory, illumination with the desired reference beam reconstructs the desired page but it also produces contributions from all the other holograms albeit at very low diffraction

efficiencies relative to the desired page. Although with proper selection of the multiplexing spacing the hologram-to-hologram cross talk can be significantly reduced, it remains as a noise source that scales with the number of participating holograms. In the case of very high density storage, we expect that the cross talk level will limit acceptable retrieval of the desired data page. It is in this cross talk limited case that we wish to consider in this chapter.

We begin by reviewing Bragg diffraction for thick media and proceed to develop expressions for the diffraction efficiency and hologram-to-hologram cross talk in terms of the underlying system parameters specifically for angular multiplexing. The noise-to-signal (NSR) metric is introduced in sec. 3.5 to characterize the level of cross talk present. Reference beam apodization during recording and/or reconstruction is described as a technique to improve the NSR and improve both the multiplexing density of the memory and the total number of multiplexed holograms. In the last part of this chapter the relationship between apodization and $M/\#$ is derived.

3.2 Bragg diffraction

The first step in understanding volume diffraction of the reference beam is to consider how a single location in the material responds to the incident optical wave. The first-order Born approximation^{7,42} allows us to characterize the response at each point as a spherical scattering center with a scattering coefficient proportional to the local change in the material permittivity and amplitude of the illuminating wave. The field at a point in the output plane can then be

represented as the coherent superposition of the scattered wavelets from all the contributing points in the volume.

Note that the Born approximation is valid only for weak holograms. If the diffraction efficiency is significantly large, then the more rigorous coupled wave theory⁸ must be applied. The small diffraction efficiency assumption can be thought of in two ways. First, that scattering from one location does not scatter again at some other location and two, the reference beam is not significantly attenuated due to the coupling of energy into the diffracted beam, *i.e.*, the energy lost in the reference beam due to the scattering is negligible.

By considering the reciprocal wavevector space (angular spectrum), we can represent the superposition of scattering sources in terms of a transfer function $A(\vec{k}_i, \vec{k}_d)$ that provides the amplitude of the diffracted field with wavevector \vec{k}_d due to the component of the reference beam with incident wavevector \vec{k}_i . $A(\vec{k}_i, \vec{k}_d)$ establishes a convenient mapping between input wavevectors and output diffracted wavevectors. The derivation can be found in refs. [42 and 43]. In the derivation it is assumed that the spatially varying material permittivity can be expressed as:

$$\epsilon(\vec{r}) = \epsilon_0 + \bar{\epsilon}(\vec{r})e^{i\vec{K}_g \cdot \vec{r}} \quad (3.1)$$

where two conditions must be satisfied. First, the amplitude of the modulated permittivity must be much smaller than the bulk average permittivity: $|\bar{\epsilon}(\vec{r})| \ll \epsilon_0$. The second requirement is that the spatial bandwidth of the permittivity modulation be much less than the grating frequency. Both conditions will be in practice satisfied for photorefractive holographic storage.

Accepting the previous assumptions, the wavevector transfer function can be written as:^{42,43}

$$A(\vec{k}_i, \vec{k}_d) = \int_V \frac{\tilde{\epsilon}(\vec{r}) e^{i\vec{K}_g \cdot \vec{r}} E_i(\vec{r}) e^{i(\vec{k}_i - \vec{k}_d) \cdot \vec{r}}}{2i\sqrt{k^2 - k_{d,x}^2 - k_{d,y}^2}} d\vec{r} \quad (3.2)$$

where $E_i(\vec{r})$ is the amplitude of the reference field at the position \vec{r} inside the scattering volume, V .

Note that if the reference beam only consists of one wavevector, \vec{k}_i , that $A(\vec{k}_i, \vec{k}_d)$ defines the angular spectrum (in terms of the variable \vec{k}_d) of the diffracted wave. We can then write the diffracted field as the inverse 2D Fourier transform of the wave's angular spectrum:

$$E_{out}(\vec{r}) = \int \int A(\vec{k}_i, \vec{k}_d) e^{i\vec{k}_d \cdot \vec{r}} dk_{d,x} dk_{d,y} \quad (3.3)$$

3.2.1 Ideal volume phase grating

Understanding diffraction from the ideal thick phase grating will serve as a starting point for the later more complicated developments. Consider the simple case where the permittivity modulation consists of an undamped single grating: $\tilde{\epsilon}(\vec{r}) = 1$ and the grating vector, \vec{K}_g , is defined through the interference of an object monochromatic wave with wavevector \vec{k}_o and a recording reference wave with wavevector \vec{k}_r :

$$\vec{K}_g = \vec{k}_o - \vec{k}_r \quad (3.4)$$

If a crystal of dimensions $W \times H \times L$ (refer to fig. 3.1) with negligible absorption (an ideal assumption) is illuminated with a reconstruction reference beam of wavevector \vec{k}_i , we wish to compute the amplitude of the diffracted field with

wavevector \vec{k}_d . By eq. 3.2, we know the amplitude of the diffracted field in the direction defined by \vec{k}_d is $A(\vec{k}_i, \vec{k}_d)$ which can be simplified using the assumed ideal phase grating:

$$A(\vec{k}_i, \vec{k}_d) \propto \int_V \bar{\epsilon}(\vec{r}) e^{i\vec{K}_g \cdot \vec{r}} E_i(\vec{r}) e^{i(\vec{k}_i - \vec{k}_d) \cdot \vec{r}} d\vec{r} \quad (3.5)$$

$$= \int_V e^{i\vec{K}_g \cdot \vec{r}} e^{i(\vec{k}_i - \vec{k}_d) \cdot \vec{r}} d\vec{r} \quad (3.6)$$

$$= \int_V e^{i(\vec{k}_o - \vec{k}_r) \cdot \vec{r}} e^{i(\vec{k}_i - \vec{k}_d) \cdot \vec{r}} d\vec{r} \quad (3.7)$$

$$= \int_V e^{i((\vec{k}_o - \vec{k}_d) + (\vec{k}_i - \vec{k}_r)) \cdot \vec{r}} d\vec{r} \quad (3.8)$$

$$= \int_V e^{-i\vec{\xi} \cdot \vec{r}} d\vec{r} \quad (3.9)$$

where the Bragg or momentum mismatch vector, $\vec{\xi}$, is defined as the vector difference between the grating, incident, and diffracted wavevectors:

$$\vec{\xi} \triangleq (\vec{k}_d - \vec{k}_o) + (\vec{k}_r - \vec{k}_i) \quad (3.10)$$

To satisfy the Bragg matching condition, the vector triplet $(\vec{K}_g, \vec{k}_d, \vec{k}_i)$ should form a closed triangle such that: $\vec{k}_d = \vec{K}_g + \vec{k}_i$. When this occurs $\vec{\xi}$ will be identically $\vec{0}$. The parentheses in eq. 3.10 are intended to suggest grouping the difference vectors between the object and diffracted wavevector and also the incident and recording reference wavevector. When the difference vectors are identically zero, there is strong diffraction; hence the ideal reconstruction of the original object beam. As a difference accrues between the wavevectors, the diffracted amplitude decreases. To understand conditions placed on the mismatch vector, we continue the simplification of eq. 3.9 by making the typical assumption that the transverse dimensions (width and height) of the volume holographic medium are much greater than the spatial bandwidth of the object beam. In this case

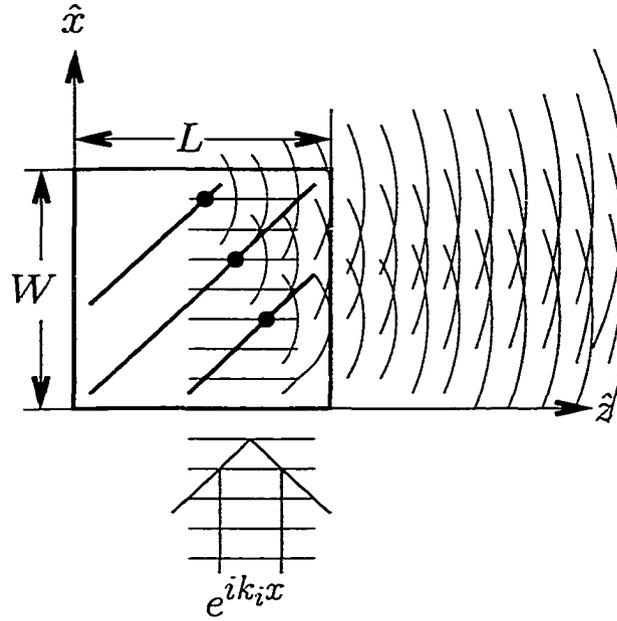


Figure 3.1: Crystal geometry and coordinate system for holographic recording in the 90 degree configuration. An example shows the superposition of scattered wavelets off of the phase grating.

the integrals over the transverse dimensions collapse into delta functions on the components of $\vec{\xi}$ in the \hat{x} and \hat{y} directions. The \hat{z} direction remains producing:

$$A(\vec{k}_i, \vec{k}_d) \propto \int_V e^{-i\vec{\xi} \cdot \vec{r}} d\vec{r} \quad (3.11)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{z=0}^L e^{-i\vec{\xi} \cdot \vec{r}} dx dy dz \quad (3.12)$$

$$= \delta(\xi_x) \delta(\xi_y) \int_0^L e^{-i\xi_z z} dz \quad (3.13)$$

$$= \delta(\xi_x) \delta(\xi_y) \text{sinc}(\xi_z L/2) \quad (3.14)$$

where $\text{sinc}(\varphi)$ is defined as $\text{sinc}(\varphi) \triangleq \frac{\sin \varphi}{\varphi}$.

3.2.2 Angular Selectivity

As an example, let the recording reference beam propagate at an angle θ with respect to the \hat{x} axis. The reconstruction beam is at a slightly different angle: $\theta + \Delta\theta$. The Bragg mismatch in the \hat{z} direction is thus:

$$\xi_z = k \sin(\theta + \Delta\theta) - k \sin \theta \quad (3.15)$$

$$= k \Delta\theta \cos \theta \quad (3.16)$$

where the last step assumes $\Delta\theta \ll \theta$ so that a first order Taylor series expansion of $\sin(\theta + \Delta\theta)$ around θ is valid. The amplitude of the diffracted field in the direction of the object wavevector for a reference beam detuned by a small angle, $\Delta\theta$ is:

$$A(\vec{k}_i, \vec{k}_d) = \text{sinc}(\xi_z L/2) \quad (3.17)$$

$$= \text{sinc}\left(\frac{k \Delta\theta L}{2} \cos \theta\right) \quad (3.18)$$

$$= \text{sinc}\left(\frac{\pi \Delta\theta}{\Delta\Theta}\right) \quad (3.19)$$

where $\Delta\Theta$ is the null spacing of the sinc function. Note that in terms of the wavelength, $\Delta\Theta$ is:

$$\Delta\Theta = \frac{\lambda}{L \cos \theta} \quad (3.20)$$

If $\lambda = 0.5 \mu\text{m}$, $L = 1 \text{ cm}$, $\theta = 0$ then $\Delta\Theta = 5 \times 10^{-5}$ radians or 2.86×10^{-3} degrees (measured externally to the crystal). Fig. 3.2 plots the angular selectivity for this example.

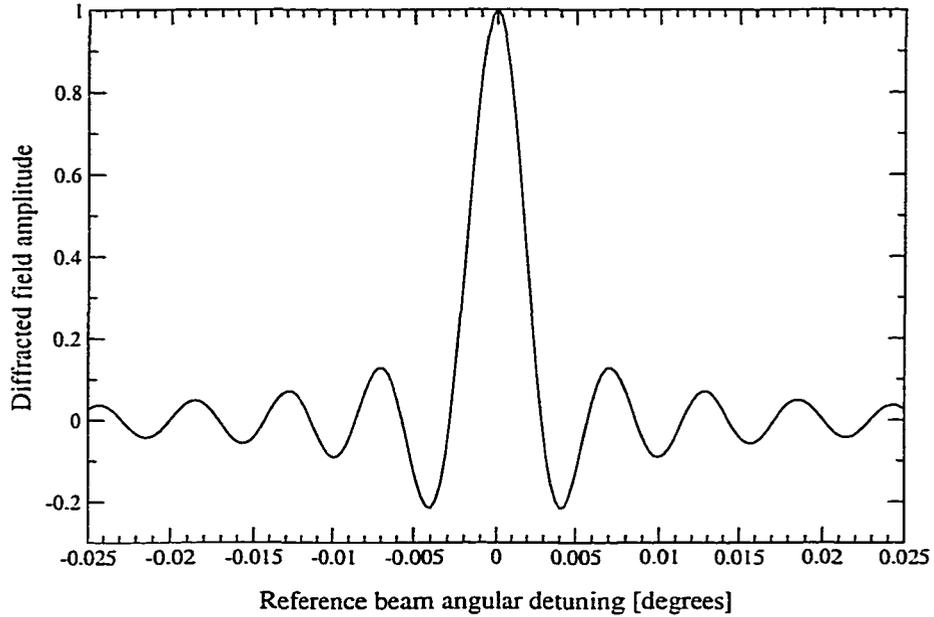


Figure 3.2: Bragg angular selectivity for a 1 cm long crystal with $\lambda = 0.5 \mu\text{m}$ and a recording reference beam angle of 0 degrees.

3.3 Bragg Selectivity Function

The general extension of the angular selectivity is the Bragg selectivity function (BSF). It yields the diffraction efficiency as the reference beam is detuned away from the value used during recording. The detuning may occur due to changes in wavelength (λ -mux) or rotations of the medium (peristrophic-mux), or a variety of others possibilities. Regardless, in every case it is the change in the Bragg momentum vector, ξ , that matters.

In the case of angular multiplexing, we saw that the transverse components of the mismatch were constrained to be zero. In this instance, the BSF is reduced to a one-dimensional function of ξ_z , the momentum mismatch nominally in the direction of the object wavevector. The diffraction efficiency is thus the

BSF evaluated at ξ_z . The strict relationship between ξ_z and the reference beam angle depends on the specific interacting object-reference grating vector, but the dependence is slight. Writing the relation as in eq. 3.16, is accurate to third order for small deviations from θ .

3.3.1 Material Effects and the BSF

The ideal case assumed the spatial modulation envelope of the permittivity to be constant. In practice, this does not occur due to the presence of material absorption among other things.

In this section we wish to incorporate the attenuated power profiles of the object and reference beams during recording and also on reconstruction. We also wish to assume an apodized form for the reference plane wave as described next.

During reconstruction we choose $R(z)$ to be the apodized field profile of the reference beam. It is defined such that $\max_z |R(z)| = 1$, thus the peak magnitude of the wave is R_0 . During recording, the reference beam is defined as:

$$E_r(\vec{r}) = W_0 W(z) e^{i\vec{k}_r \cdot \vec{r}} \quad (3.21)$$

where a small reference beam angle is assumed in order to ignore the geometric shadowing effect in $W(z)$. This is valid to third order deviations in the reference beam angle away zero degrees.

The reference beam interferes with each plane wave component in the object beam. We model each component in electric field as:

$$E_o(\vec{r}) = S_0 e^{i\vec{k}_o \cdot \vec{r}} \quad (3.22)$$

As is common in the literature, we assume all inferred angles, wavelengths, and wavevectors are specified inside the material. We tacitly assume that the space surrounding the hologram is index matched to the crystal. If external angles are desired, then Snell's law must be applied.

On reconstruction, we use a new reference beam (the same as the recording beam for Bragg-matched conditions) of:

$$E_i(x = 0, y, z) = R_0 R(z) e^{i\vec{k}_i \cdot \vec{r}} \quad (3.23)$$

The reference beam diffracts through the volume hologram for $x \in [0, W]$ and $z \in [0, L]$. As it propagates nominally in the \hat{x} direction, it will be attenuated due to the crystal absorption. The reference beam at a point \vec{r} in the medium is expressed as:

$$E_i(\vec{r}) = R_0 R(z) e^{-\frac{\alpha}{2} x} e^{i\vec{k}_i \cdot \vec{r}} \quad (3.24)$$

where α is the irradiance bulk absorption coefficient. The same relationship holds for the recording reference beam, the object wave, and the diffracted wave. Note we will assume only phase holograms so there is no spatial modulation of the absorption coefficient.

The permittivity modulation of the hologram results indirectly from the space charge field established during the holographic exposure. The fast fringe patterns in irradiance are mimicked by phase-shifted spatial electric fields established by electrons bound in doped deep impurity traps. The linear electro-optic effect then produces the permittivity modulation from the space charge field.⁴⁴ For a small amplitude modulation of the permittivity, the material index of refraction approximately duplicates the spatial distribution. For the 90 degree angular

multiplexing configuration in LiNbO₃, the c-axis is oriented at 45 degrees to the x - z plane to maximize the projection of the grating vector on the electro-optic coefficient, r_{13} . In this case, the permittivity modulation is:

$$\bar{\epsilon}(\vec{r})e^{i\vec{K}_g \cdot \vec{r}} = -\epsilon_0 n_o^4 r_{13} E_1(\vec{r}) \quad (3.25)$$

where n_o is the ordinary index of refraction (around 2.286 for $\lambda = 514$ nm) and $E_1(\vec{r})$ is the space charge field generated during the recording process.

Combining eqs. 3.2 and 3.25 we compute the amplitude of the diffracted plane wave component traveling with the correct object wavevector, \vec{k}_o . The deviation of the reconstruction reference beam from the recording beam is captured in the momentum mismatch term $\vec{\xi}$:

$$E_d(\vec{\xi}) = \int_V R_0 R(z) e^{-\frac{\alpha}{2}(L-z+x)} e^{i(\vec{k}_i - \vec{k}_o) \cdot \vec{r}} \bar{\epsilon}(\vec{r}) e^{i\vec{K}_g \cdot \vec{r}} d\vec{r} \quad (3.26)$$

$$= -\epsilon_0 n_o^4 r_{13} \int_V R_0 R(z) e^{-\frac{\alpha}{2}(L-z+x)} e^{i(\vec{k}_i - \vec{k}_o) \cdot \vec{r}} E_1(\vec{r}) d\vec{r} \quad (3.27)$$

$$= -\epsilon_0 n_o^4 r_{13} \int_V R_0 R(z) e^{-\frac{\alpha}{2}(L-z+x)} e^{i(\vec{k}_r - \vec{k}_o) \cdot \vec{r}} E_1(\vec{r}) e^{-i\vec{\xi} \cdot \vec{r}} d\vec{r} \quad (3.28)$$

We see that the diffracted amplitude can be represented as a three-dimensional Fourier transform with \vec{r} and $\vec{\xi}$ forming a reciprocal vector pair. Ignoring some of the normalization constants we can write the relationship as:

$$E_d(\vec{\xi}) \propto \text{FT} \left[R_0 R(z) e^{-\frac{\alpha}{2}(L-z+x)} e^{i(\vec{k}_r - \vec{k}_o) \cdot \vec{r}} E_1(\vec{r}) \right] \quad (3.29)$$

Remembering for the case of angular multiplexing that the transverse mismatch components ξ_x and ξ_y must be identically zero, we can simplify the relationship into a one-dimensional Fourier transform pair:

$$E_d(\xi_z) \propto \int_0^H \int_0^W e^{-\frac{\alpha}{2}x} e^{i(\vec{k}_r - \vec{k}_o) \cdot \hat{x}} E_1(\vec{r} \cdot \hat{x}) dx dy \cdot \text{FT} \left[R_0 R(z) e^{-\frac{\alpha}{2}(L-z)} e^{i(\vec{k}_r - \vec{k}_o) \cdot \hat{z}} E_1(\vec{r} \cdot \hat{z}) \right] \quad (3.30)$$

We define the Bragg selectivity function, $\eta(\xi_z)$, and the Bragg profile function (BPF), $\Gamma(z)$ from eq. 3.30 as:

$$\eta(\xi_z) = \int_0^H \int_0^W e^{-\frac{\alpha}{2}x} e^{i(\vec{k}_r - \vec{k}_o) \cdot \hat{x}} E_1(\vec{r} \cdot \hat{x}) dx dy \cdot \text{FT} \left[R_0 R(z) e^{-\frac{\alpha}{2}(L-z)} e^{i(\vec{k}_r - \vec{k}_o) \cdot \hat{z}} E_1(\vec{r} \cdot \hat{z}) \right] \quad (3.31)$$

$$= \text{FT} [\Gamma(z)] \quad (3.32)$$

The BSF and BPF will be simplified once we have a more specific form for the space charge field, $E_1(\vec{r})$. The first step in finding $E_1(\vec{r})$ is to compute the modulation depth established during recording.

3.3.2 Modulation depth

The space charge field will be proportional to the interference grating formed between the apodized reference beam (defined in eq. 3.21) and one of the plane wave components present in the object beam (eq. 3.22).

Interfering two plane waves with wavevectors \vec{k}_o and \vec{k}_r produces an interference pattern with a grating vector of:

$$\vec{K}_g = \vec{k}_o - \vec{k}_r \quad (3.33)$$

The local irradiance at a location \vec{r} in the crystal is:

$$I(\vec{r}) \propto |E_o(\vec{r})e^{-\frac{\alpha}{2}z} + E_r(\vec{r})e^{-\frac{\alpha}{2}x}|^2 \quad (3.34)$$

$$= I_0(\vec{r}) + E_o(\vec{r})E_r^*(\vec{r})e^{-\frac{\alpha}{2}(z+x)} + E_r(\vec{r})E_o^*(\vec{r})e^{-\frac{\alpha}{2}(z+x)} \quad (3.35)$$

$$= I_0(\vec{r}) \left(1 + \frac{S_0 W_0 W(z)e^{-\frac{\alpha}{2}(z+x)}}{I_0(\vec{r})} e^{i\vec{K}_g \cdot \vec{r}} + \frac{S_0 W_0 W^*(z)e^{-\frac{\alpha}{2}(z+x)}}{I_0(\vec{r})} e^{-i\vec{K}_g \cdot \vec{r}} \right) \quad (3.36)$$

$$= I_0(\vec{r}) \left(1 + m(\vec{r})e^{i\vec{K}_g \cdot \vec{r}} + m^*(\vec{r})e^{-i\vec{K}_g \cdot \vec{r}} \right) \quad (3.37)$$

$$m(\vec{r}) = \frac{S_0 W_0 W(z)e^{-\frac{\alpha}{2}(z+x)}}{I_0(\vec{r})} \quad (3.38)$$

where $I_0(\vec{r})$ is defined as:

$$I_0(\vec{r}) = |E_o(\vec{r})|^2 e^{-\alpha z} + |E_r(\vec{r})|^2 e^{-\alpha x} \quad (3.39)$$

$$= S_0^2 e^{-\alpha z} + W_0^2 |W(z)|^2 e^{-\alpha x} \quad (3.40)$$

For short exposure times the space charge field amplitude will grow linearly with the modulation depth $m(\vec{r})$:

$$E_1(\vec{r}; t) \propto m(\vec{r})e^{i\vec{K}_g \cdot \vec{r}t} \quad (3.41)$$

Notice from eq. 3.37 that there are possibly two interference gratings formed. One that is Bragg-matched to reconstruction with the recording beam ($m(\vec{r})$) and a second matched to reconstruction with the phase conjugate of the recording wave ($m^*(\vec{r})$). In a thick hologram, the conjugate wave will not diffract when reconstructing with a reference beam propagating in the approximate same direction as the original reference beam, allowing us to ignore the $m^*(\vec{r})$ grating term.

3.3.3 Bragg Profile Function

From the modulation depth given in eqs. 3.36 and 3.37, and the grating amplitude (eq. 3.41), we can simplify the Bragg profile function:

$$\Gamma(z) \propto \int_0^H \int_0^W R_0 R(z) e^{-\frac{\alpha}{2}(L-z+x)} e^{i(\bar{k}_r - \bar{k}_o) \cdot \bar{r}} E_1(\bar{r}) dx dy \quad (3.42)$$

$$= \int_0^H \int_0^W R_0 R(z) e^{-\frac{\alpha}{2}(L-z+x)} e^{-i\bar{K}_g \cdot \bar{r}} m(\bar{r}) e^{i\bar{K}_g \cdot \bar{r}} dx dy \quad (3.43)$$

$$= H \int_0^W R_0 R(z) e^{-\frac{\alpha}{2}(L-z+x)} m(x, z) dx \quad (3.44)$$

$$= H \int_0^W \frac{S_0 R_0 W_0 R(z) W(z) e^{-\alpha x} e^{-\frac{\alpha}{2}L}}{S_0^2 e^{-\alpha z} + W_0^2 |W(z)|^2 e^{-\alpha x}} dx \quad (3.45)$$

$$= H \left(\frac{S_0}{W_0} \right) \int_0^W \frac{R_0 R(z) W(z) e^{-\alpha x} e^{-\frac{\alpha}{2}L}}{(S_0/W_0)^2 e^{-\alpha z} + |W(z)|^2 e^{-\alpha x}} dx \quad (3.46)$$

$$= H \left(\frac{S_0}{W_0} \right) R_0 R(z) W(z) e^{-\frac{\alpha}{2}L} \int_0^W \frac{e^{-\alpha x} dx}{(S_0/W_0)^2 e^{-\alpha z} + |W(z)|^2 e^{-\alpha x}} \quad (3.47)$$

$$= H \left(\frac{S_0}{W_0} \right) \frac{R_0 R(z) e^{-\frac{\alpha}{2}L}}{|W(z)|} \left[W + \frac{1}{\alpha} \ln \frac{|W(z)|^2 + (S_0/W_0)^2 e^{-\alpha z}}{|W(z)|^2 + (S_0/W_0)^2 e^{-\alpha z} e^{\alpha W}} \right] \quad (3.48)$$

$$= H \left(\frac{S_0}{W_0} \right) \frac{R_0 R(z) e^{-\frac{\alpha}{2}L}}{|W(z)|} \left[W + \frac{1}{\alpha} \ln \frac{1 + A(z) e^{-\alpha z}}{1 + A(z) e^{-\alpha(z-W)}} \right] \quad (3.49)$$

$$A(z) \triangleq (S_0/W_0)^2 / |W(z)|^2 \quad (3.50)$$

The integral in eq. 3.47 can be solved in closed form as long as $W(z) \neq 0$:

$$\int_0^W \frac{e^{-\alpha x}}{C e^{-\alpha x} + D} dx = \frac{1}{\alpha C} [\alpha W + \ln(C + D) - \ln(C + D e^{\alpha W})] \quad (3.51)$$

It is clear that when $W(z) = 0$ the integral simplifies to:

$$\int_0^W \frac{e^{-\alpha x} dx}{(S_0/W_0)^2 e^{-\alpha z} + |W(z)|^2 e^{-\alpha x}} = \frac{1 - e^{-\alpha W}}{(S_0/W_0)^2 \alpha e^{-\alpha z}} \quad (3.52)$$

$A(z)$ is the local apodized beam ratio. The BPF and BSF defined in eqs. 3.31 and 3.32 can now be expressed in their full form:

$$\eta(\xi_z) \propto \int_0^L \left(\frac{S_0}{W_0} \right) \frac{R_0 R(z)}{|W(z)|} \left[W + \frac{1}{\alpha} \ln \frac{1 + A(z)e^{-\alpha z}}{1 + A(z)e^{-\alpha(z-W)}} \right] e^{-i\xi_z z} dz \quad (3.53)$$

$$= \text{FT}[\Gamma(z)] , \quad (3.54)$$

where $\Gamma(z)$ is defined to be zero outside the interval $[0, L]$ and also for any z such that $W(z) = 0$.

We refer to $\Gamma(z)$ as the Bragg profile function (BPF) because it is the Fourier reciprocal function to the BSF. Because the BPF is limited to a spatial extent of L , we see an immediate consequence is that the BSF can not provide a narrower main lobe than $\xi_0 = 2\pi/L$. This can be made clearer by examining the Fourier transform pair:

$$\Gamma(z) = 1 \quad (3.55)$$

$$\eta(\xi_z) = \text{sinc}(\xi_z L/2) \quad (3.56)$$

ξ_0 represents the smallest null spacing possible for the given length L . It thus establishes a practical minimum separation between successive holograms. Comparing this result to sec. 3.2.1 we see that this case occurs when the overall system acts like an ideal thick phase hologram.

To provide some real numbers, consider angular multiplexing in a $1 \times 1 \times 1$ cm³ crystal. The ideal BSF for the $\theta = 0$ page has periodic nulls at integer multiples of $\Delta k = \sin^{-1}(\lambda/L)$. This represents a nearly constant angular spacing of $\lambda/L \approx 5 \times 10^{-5}$ radians or 2.86×10^{-3} degrees. There can be up to 14,000 multiplexed holograms in a ± 20 degree span. Of course, such a large angular

range places additional non-trivial design requirements on the optical system implementation which we tacitly assume have been adequately satisfied (in this chapter at least).

3.4 Apodization Theory

The periodic nulls of the BSF are not actually nulls in reality due to the presence of absorption. Absorption can not be eliminated as it is one of the key operating mechanisms responsible for the photorefractive effect. The first few absorption-corrupted nulls can often produce an unacceptably large diffraction efficiency leading to significant interpage cross talk. In practice it is common to reduce the cross talk by storing pages not on the first nulls of the BSF but on the second, third, or fourth. This solution markedly reduces the multiplexing density.

Another technique that has been proposed to reduce the cross talk without sacrificing the density is apodization.^{13,17,45-47} The reference beam is shaped during recording and reconstruction in order to provide a suitably improved BSF. Apodization can be applied in three ways:

1. during recording only: $R(z) = 1$
2. during reconstruction only: $W(z) = 1$
3. during both recording and reconstruction

Figure 3.3 shows the apodization profiles for (a) no apodization, (b) reconstruction only, and (c) recording only, where the apodizers are chosen to flatten

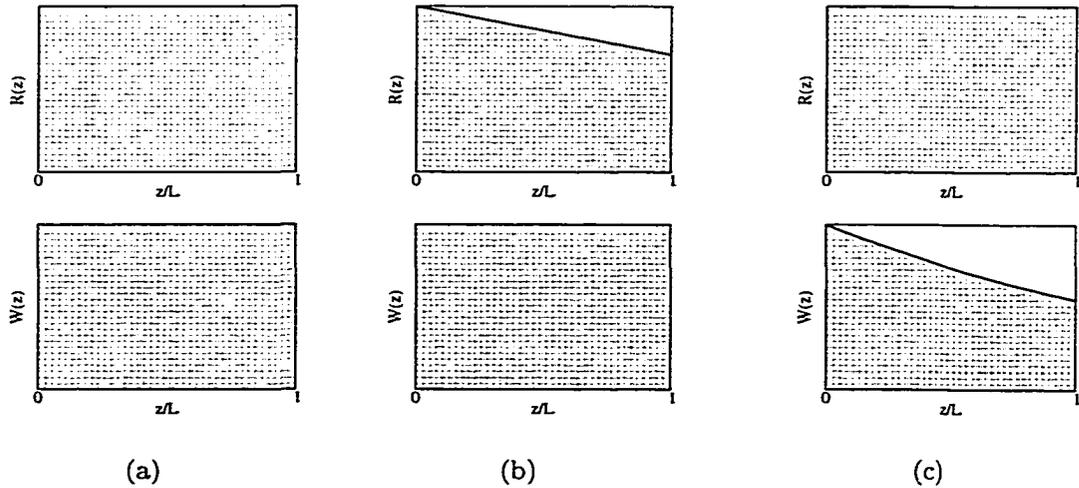


Figure 3.3: Apodizer profiles for (a) no apodization (b) reconstruction beam apodization (c) recording beam apodization.

out their associated $\Gamma(z)$. Figure 3.4 shows $\Gamma(z)$ for the three apodizers. For these figures the assumed beam ratio is 1, $L = W = H = 1$ cm, and $\alpha = 0.73$ cm⁻¹. Notice that the reconstruction apodizer merely corrects for the nearly constant slope present in the no apodization case. The recording apodizer accounts for the exponential absorption of the signal beam as it propagates in the $+z$ direction.

The BSF for the three apodizers are shown in figure 3.5. The non-flat profile of the no apodization $\Gamma(z)$ results in a relative amplitude diffraction efficiency of 5.817% and 2.894% at the first and second nulls, respectively. Reconstruction apodization corrects most of this by flattening out $\Gamma(z)$. Note that recording apodization can precisely flatten $\Gamma(z)$ and recover the perfect Bragg nulls.

During reconstruction, all the pages in the memory contribute a small diffracted component superimposed with the Bragg-matched page. This coherent superposition constitutes a data-dependent coherent field noise source that we refer to

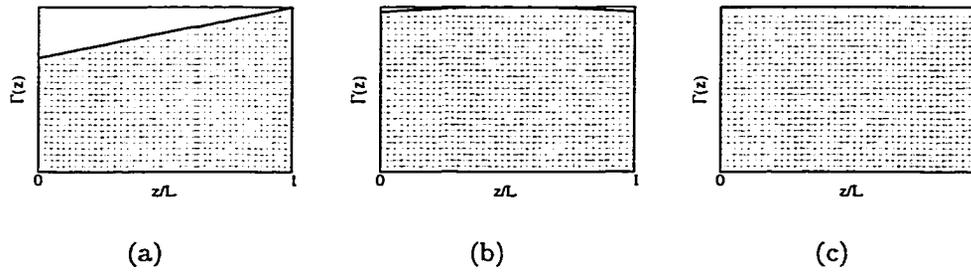


Figure 3.4: Effective BSF Fourier transform kernel for (a) no apodization (b) reconstruction beam apodization (c) recording beam apodization.

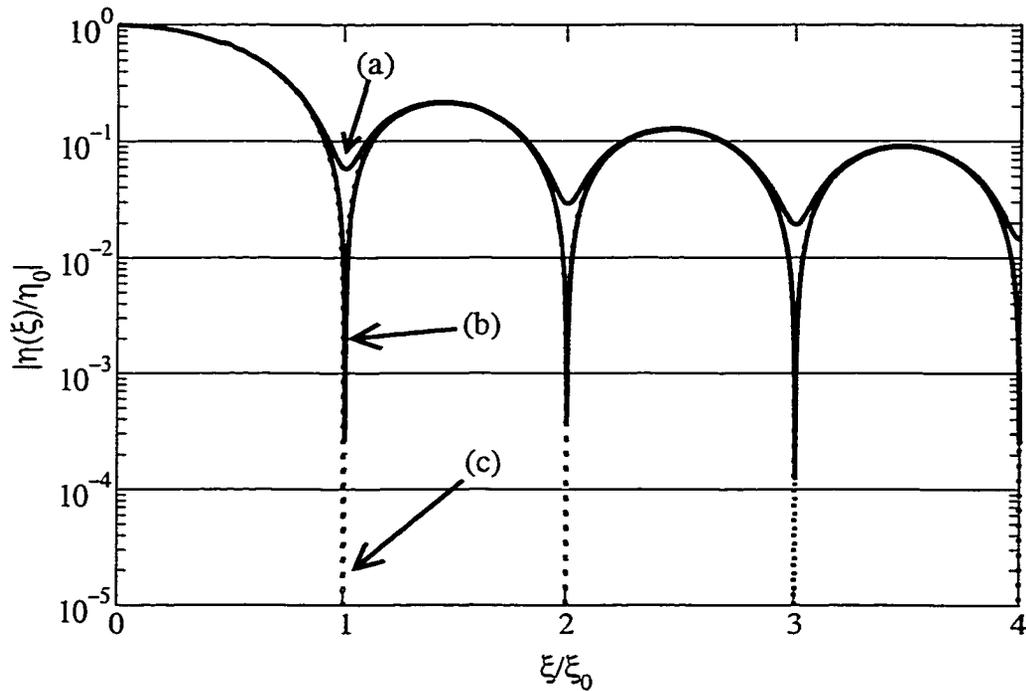


Figure 3.5: Normalized amplitude diffraction efficiency versus normalized multiplexing detuning for (a) no apodization (b) reconstruction beam apodization (c) recording beam apodization.

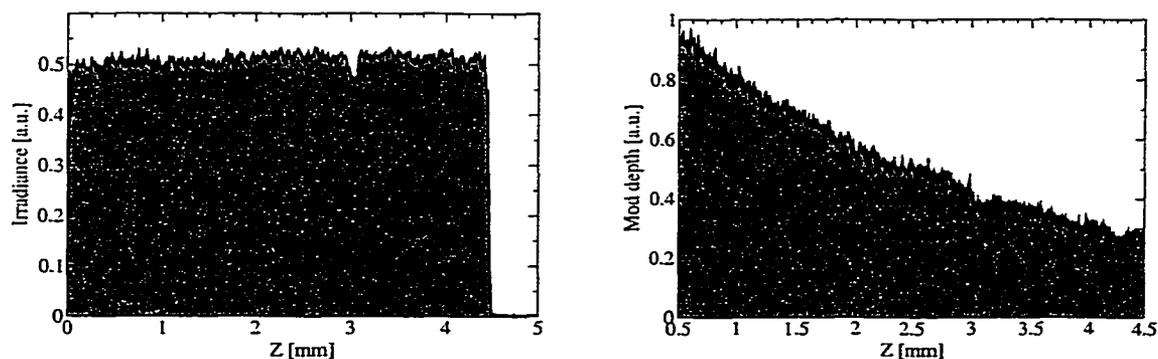
as interpage cross talk. When no apodization is performed the nearby few multiplexed pages dominate the cross talk noise and force a loss in multiplexing density in order to allow reliable page retrieval. The above proposed recording or reconstruction apodizers strongly reduce the cross talk noise, but there is an associated loss in diffraction efficiency due to apodization; apodization lowers the system $M/\#$ as is discussed later in sec. 3.6.

3.4.1 Experiment

To verify that we could modify the Bragg selectivity function by changing the reference beam profile, we conducted some experiments. The optical setup was constructed in a 90 degree angular-multiplexing configuration. A Newport rotation stage (0.001 deg accuracy) was used for coarse angular selection as well as for angular scans at an approximate constant velocity. A New Focus picomirror mount (0.1 mdeg accuracy) was used for fine angular control.

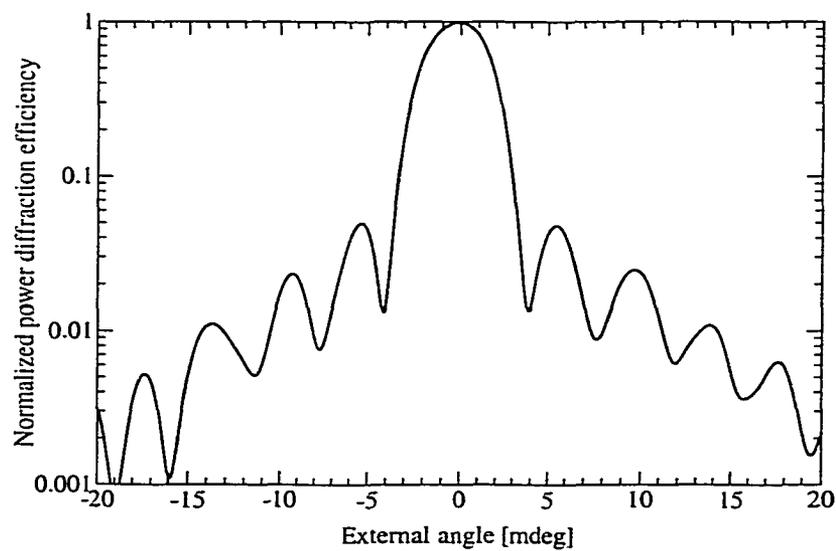
The photorefractive crystal was placed 3 mm behind the Fourier plane of the object arm, resulting in Fresnel holograms. For this experiment, a 5 mm \times 5 mm \times 5mm KNSBN crystal was used. The crystal was chosen for its fast recording time and low absorption ($\alpha = 0.55 \text{ cm}^{-1}$).

Figs. 3.6 and 3.7 show the reference beam profile, hologram modulation depth, and Bragg selectivity function for rectangular and triangular apodization. The non-uniform (decaying exponential) modulation depth in the rectangular apodization clearly shows the effect of the crystal absorption. The “filling-in” of the Bragg nulls as a result is likewise apparent in fig. 3.6c. For the case of triangle apodization, there are no solid conclusions. The periodic Bragg nulls can not clearly be



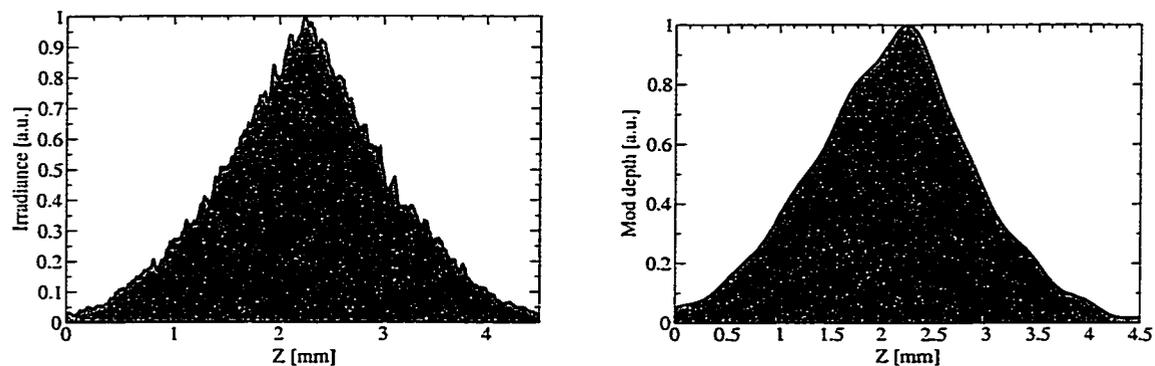
(a) Reference beam profile

(b) Modulation depth



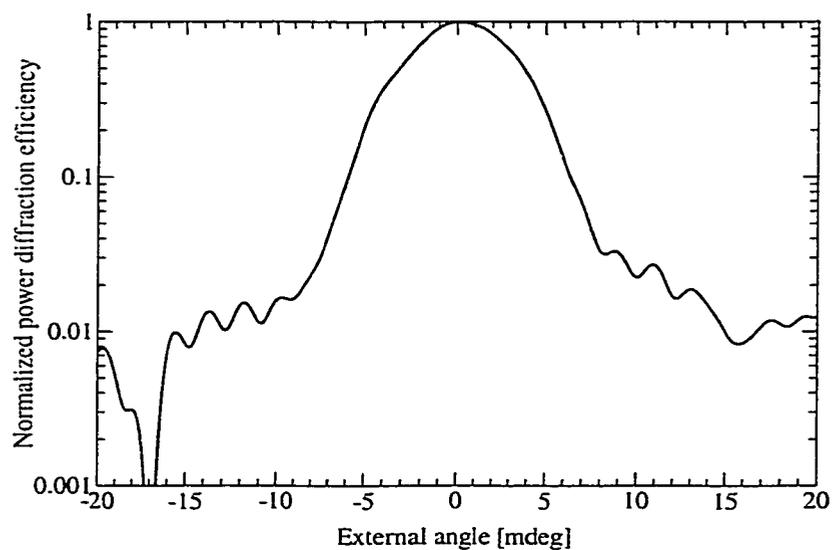
(c) Bragg selectivity function

Figure 3.6: Rectangle apodization. Experimental (a) reference beam profile, (b) resulting modulation depth, and (c) Bragg selectivity function in a 5mm sample of KNSBN. The recording beam ratio was 10:1.



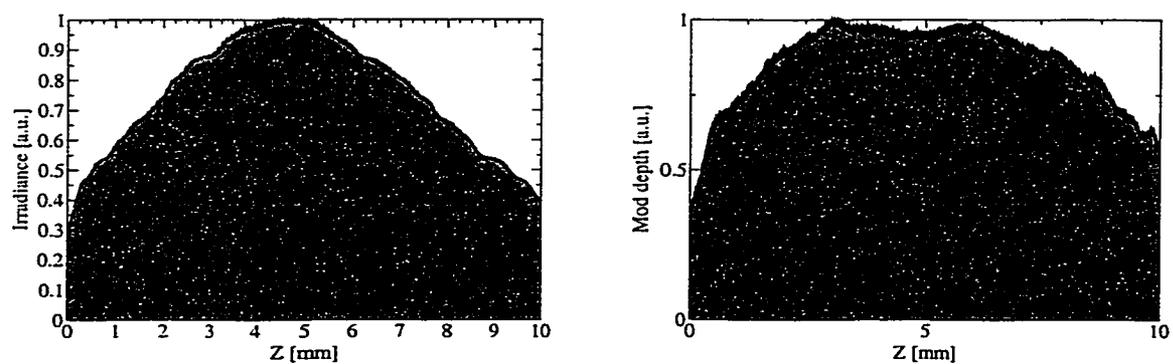
(a) Reference beam profile

(b) Modulation depth



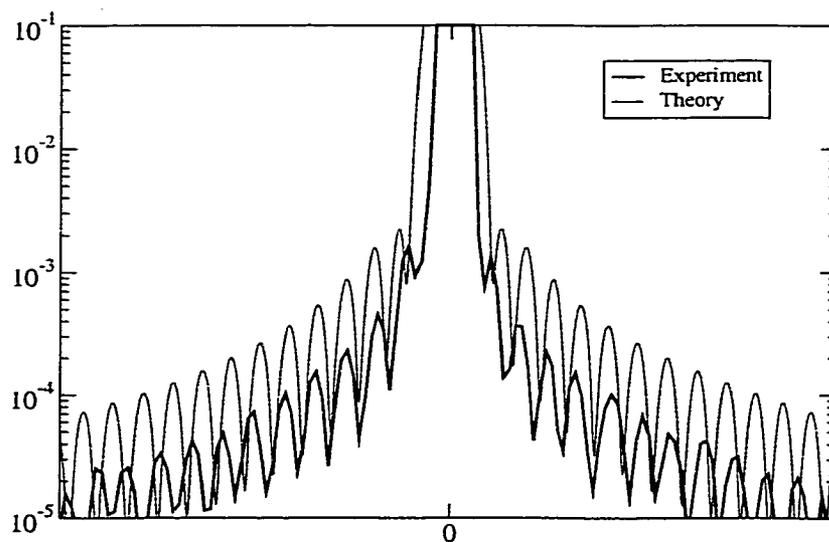
(c) Bragg selectivity function

Figure 3.7: Triangle apodization. Experimental (a) reference beam profile, (b) resulting modulation depth, and (c) Bragg selectivity function in a 5mm sample of KNSBN. The recording beam ratio was 10:1.



(a) Reference beam profile

(b) Modulation depth



(c) Bragg selectivity function

Figure 3.8: Gaussian apodization. Experimental (a) reference beam profile, (b) resulting modulation depth, and (c) Bragg selectivity function in a 1cm sample of Fe:LiNbO₃. The recording beam ratio was 10:1.

distinguished. Though, we do see the profile and modulation depth behave as we expect.

Fig. 3.8 shows the results of using no intentional apodization (implicitly a Gaussian apodizer). The available high beam power allows very good resolution on the sidelobes of the BSF. On top of the experimental curve is plotted the theoretical prediction for the 1 cm Fe:LiNbO₃ crystal with absorption coefficient 3.8 cm⁻¹.

3.4.2 Cross Talk noise sources

So far we have discussed apodization with respect to the effects of the crystal absorption during the recording and readout processes. Even when the BSF is the ideal sinc function, there will still be an intrinsic amount of cross talk that grows with the number of pages in the memory. It is a directly a result of the non-zero angular span of object wavevectors in the \hat{x} direction, as well as a small change in the null spacing with increasing off-axis reference beam angles. For small object spatial bandwidths and reference angular range, this cross talk can be negligibly small. But when the number of multiplexed holograms is very large, then the additive field noise will offer a significant contribution.

In addition to intrinsic cross talk noise, there are primarily two other mechanisms by which cross talk can limit the system performance. First, if absorption is not corrected, the neighboring holograms will cross talk significantly. As discussed earlier, besides apodization, this effect can be reduced by spacing the holograms on a larger angular spacing, but this may not be a good system level

choice due to the extra reference beam angular bandwidth that is sacrificed. The larger angles also increase the effect of the intrinsic cross talk.

A second cross talk source is a consequence of non-ideal angular pointing accuracy in the reference beam. Errors in the reference beam angle will decrease the diffraction efficiency of the desired page and increase the cross talk from the other holograms. The cross talk is very sensitive to angular beam jitter, and therefore very small pointing errors will quickly magnify the cross talk levels.

We propose apodization as a technique that:

1. compensates for the non-uniform grating profile (and hence restore the sinc nulls)
2. increases tolerance to angular jitter by decreasing the slope of the BSF around the nulls

We measure the effectiveness of apodization by introducing the noise-to-signal ratio (NSR) which characterizes the cross talk noise by the first and second moments on the detection plane irradiance signal. We discuss a variety of apodization shapes to reduce the NSR in the presence of intrinsic cross talk noise, absorption, jitter, and both a coherent and incoherent page-to-page phase model. Afterwards in sec. 3.6, we derive the relationship between apodization and the holographic system metric $M/\#$.

3.5 Noise-to-Signal Ratio

The diffracted field present at the detection plane can be written as a superposition of every stored hologram weighted by their diffraction efficiency. For

retrieval of page j we write the field amplitude as:

$$E = \sum_{m=-M_1}^{M_1} \eta(\xi_z(m; j)) E_O^{(m)} e^{i\phi_m} \quad (3.57)$$

where the shorthand notation of $\xi(m; j)$ refers to the component of the Bragg mismatch vector in the \hat{z} direction between pages m and j . The large transverse dimensions (with respect to the spatial bandwidth of the object beam) require the other components of $\vec{\xi}$ to be identically zero. This introduces a small shift in the diffracted images that we can ignore; however it also imparts a small deviation to ξ_z as well, which is accounted for in our simulation model.

The last term in eq. 3.57 may or may not be present depending on the model. The term $e^{i\phi_m}$ represents a random phase present upon retrieval of a page. When we incorporate this term, we will refer to the cross talk as incoherent (as the average effect of the term is to average out the cross talk). The phase is assumed to be uniform over the full range of 2π . When the term is not present, the cross talk is said to be coherent.

For Fourier holograms, we can write the object beam as a collection of plane waves with direction cosines corresponding to the associated pixel location:

$$E_O^{(m)} = a_{mlp} e^{i\vec{k}_{lp} \cdot \vec{r}} \quad (3.58)$$

$$\vec{k}_{lp} = k(\cos \alpha_p \cos \beta_l \hat{z} - \sin \alpha_p \cos \beta_l \hat{y} - \sin \beta_l \hat{x}) \quad (3.59)$$

where α_p and β_l are the direction angles between the pixel at location (l, p) on the SLM and the center of the Fourier transform lens a distance f away from the SLM. a_{mlp} is the SLM pixel value for page m , pixel location (l, p) . In this chapter we will only consider a binary amplitude SLM so a_{mlp} is either 0 or 1.

The Bragg mismatch vector associated with reconstructing the grating corresponding to page m and pixel (l, p) ($\vec{k}_r = \vec{k}^{(m)}, \vec{k}_o = \vec{k}_{lp}$), using the reference beam for page j ($\vec{k}_i = \vec{k}^{(j)}$) is:

$$\vec{\xi} = \vec{k}_o - \vec{k}_d + \vec{k}_i - \vec{k}_r \quad (3.60)$$

$$= \vec{k}_{lp} - \vec{k}_d + \vec{k}^{(j)} - \vec{k}^{(m)} \quad (3.61)$$

where the requirements for non-zero diffraction imply that $\xi_x = 0$ and $\xi_y = 0$. In order to achieve this, a slight rotation of \vec{k}_d is necessary which imparts a small change to ξ_z . This slight dependence on the \hat{x} component of the object wavevector, \vec{k}_{lp} , is responsible for different columns on the same page having slightly different mismatches and hence slightly different cross talk values. We can ignore the y dependence of the mismatch as it is in the degenerate direction (with an exception for fractal multiplexing). As a result of this degeneracy, all the pixels in the same column on the same object page will have similar cross talk properties. The amplitude of each individual cross talk contribution is just the Bragg selectivity function evaluated at ξ_z .

The important thing to note is that the cross talk properties vary across columns of pixels in the detection plane as well as from page to page. Previous studies^{13,46,48} have analyzed the cross talk over all pages and all columns and found that the worst cross talk occurs at approximately $m = M_r \triangleq \pm 0.9M_1$ for the column of pixels with the largest wavevector component in the \hat{x} direction. We will refer to all cross talk measurements with respect to this reference location of page M_r . The diffracted field, E , at a pixel in the detection plane for the worst

case cross talk position can be written as:

$$E \propto \sum_{m=-M_1}^{M_1} a_{mlp} \eta(\xi_z) e^{i\phi_m} \quad (3.62)$$

$$\xi_z = (\vec{k}_{lp} - \vec{k}_d + \vec{k}^{(M_r)} - \vec{k}^{(m)}) \cdot \hat{z} \quad (3.63)$$

where (l, p) corresponds to the corner pixel on the SLM.

Because of the requirement that $\xi_x = 0$ and $\xi_y = 0$, there will only be one pixel grating from each page that contributes to the reconstructed field for the page and pixel under consideration. We can simplify the notation, by implicitly assuming all the other necessary indices and vector computations, only leaving the page index:

$$E \propto \sum_m a_m \eta_m e^{i\phi_m} \quad (3.64)$$

By separating out page M_r from the sum, we can write the output field as the desired data embedded in an additive field noise:

$$E \propto a_{M_r} + \sum_{m \neq M_r} a_m \eta_m e^{i\phi_m} \quad (3.65)$$

$$= a_{M_r} + E_n \quad (3.66)$$

Because the data bits are random, the field noise is likewise random. The first and second moments will strongly define how easy or difficult it is to distinguish the case of $a_{M_r} = 1$ from $a_{M_r} = 0$. We characterize this by defining the noise-to-signal ratio as:

$$\text{NSR} \triangleq \frac{\sqrt{\sigma_1^2 + \sigma_0^2}}{\mu_1 - \mu_0} \quad (3.67)$$

where μ_j and σ_j^2 are the first and second moments of the detected pixel irradiance. j refers to the stored 0 or 1 pixel value.

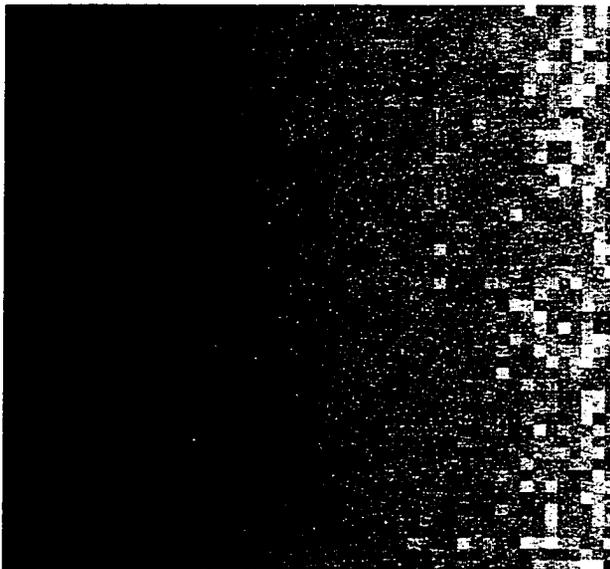


Figure 3.9: NSR in upper right quadrant of $m = M_r$ page.

The pixel irradiance is proportional to the square of the electric field:

$$I \propto |E|^2 \quad (3.68)$$

$$= a_{M_r} + |E_n|^2 + 2a_{M_r} \text{Re}[E_n] \quad (3.69)$$

Fig. 3.9 shows a simulation of the NSR with respect to every pixel in the upper right quadrant of the $m = M_r$ page. Notice that columns have the same NSR value as we expect due to the degeneracy in the \hat{y} direction.

3.5.1 Apodizer Shapes

We considered a variety of apodized reference beam shapes in order to characterize the effect of many parameters relevant to apodization, such as sidelobe suppression, main lobe width, and sidelobe decay rate.

In this study we only consider amplitude apodization functions. Phase apodization is a possible alternative technique which we did not investigate. Table 3.1 lists the considered apodizer shapes which are described by their normalized amplitude transmittance function, $t(z)$, defined over the interval $[-\frac{1}{2}, +\frac{1}{2}]$. Note that some of the apodizers are parameterized by $\beta \in [0, 1]$.

Figs. 3.10-3.11 plot the transmittance function and spectrum for the apodizers considered. The spectrum is shown as the log-magnitude in dB. Notice that the peak value of the spectrum represents the lost power when using an absorptive apodizers to construct the apodized electric field profile.

Apodizer Type	$t(z)$
Rectangle	1
Trapezoid	$\begin{cases} 1 - 2 z /\beta & z \geq \frac{1}{2}(1 - \beta) \\ 1 & \text{else} \end{cases}$
Triangle	$1 - 2 z $
Hamming	$0.5435 + 0.4565 \cos 2\pi z$
Gaussian	$e^{-\frac{1}{2}(2\beta z)^2}$
Raised Cosine	$\cos^\beta \pi z$
Tukey	$\begin{cases} 1 + \cos \left[\frac{\pi(2 z -\beta)}{1-\beta} \right] & z \geq \beta/2 \\ 1 & \text{else} \end{cases}$

Table 3.1: Apodizer functions

3.5.2 NSR results

The noise statistics on the irradiance random variable will determine when cross talk noise limits the storage of additional pages. Consider the NSR when the Bragg selectivity function is determined by the absorption corrupted modulation depth. In fig. 3.12 we plot the NSR as a function of the number of pages in

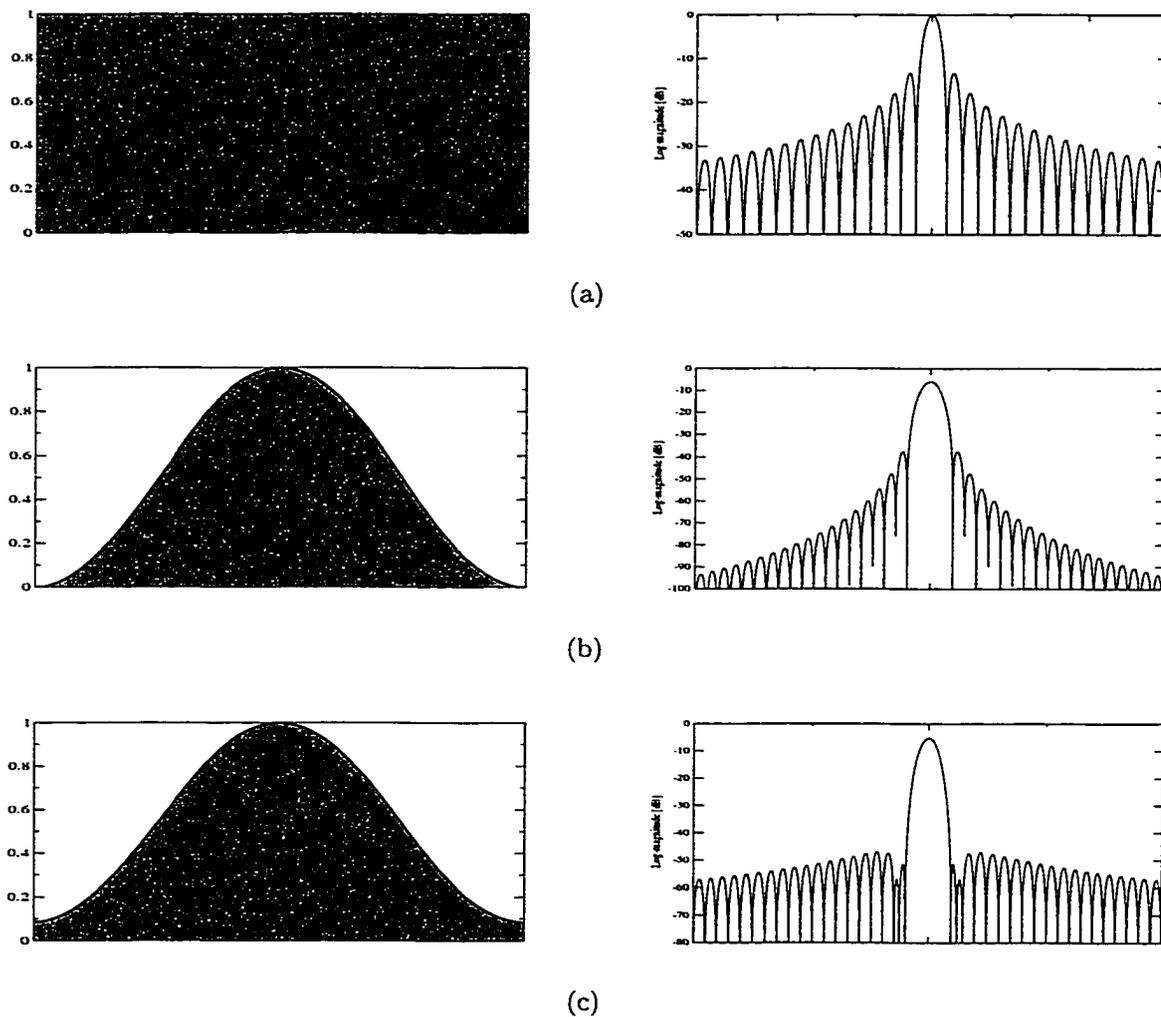


Figure 3.10: Apodizer transmittance and spectrum of (a) rectangle, (b) hamming, (c) raised cosine.

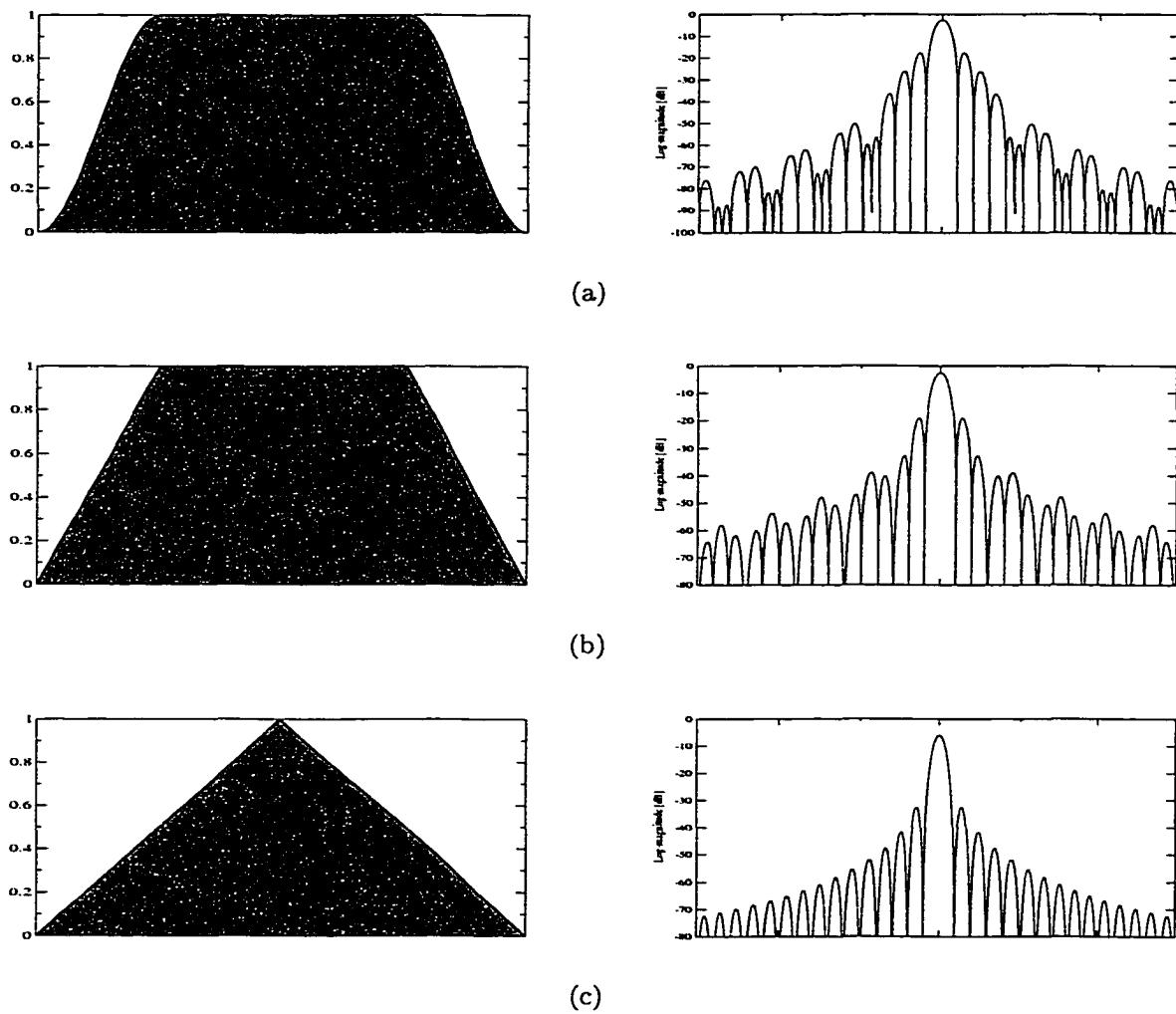


Figure 3.11: Apodizer transmittance and spectrum of (a) tukey, (b) trapezoid, (c) triangle.

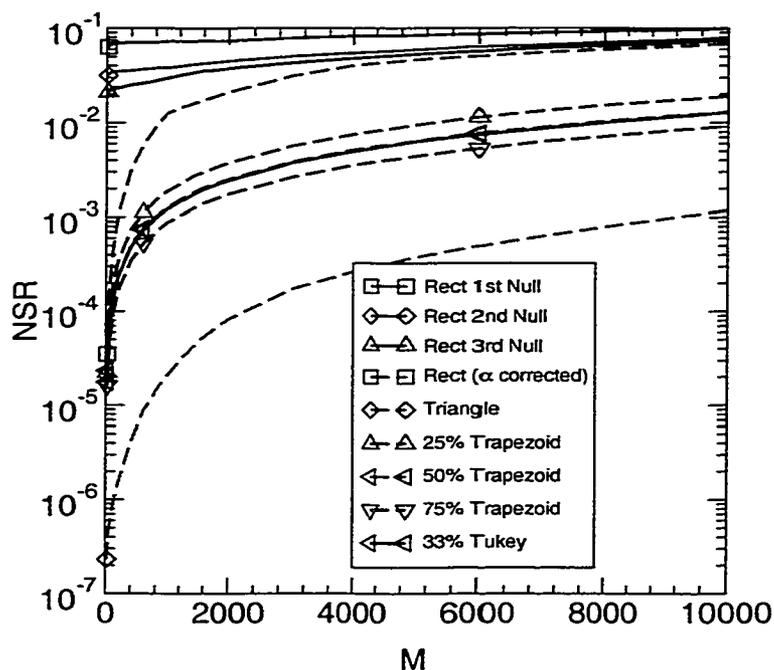


Figure 3.12: NSR (random page phase) for 1 cm crystal with $\alpha = 0.55 \text{ cm}^{-1}$.

memory, M for a variety of apodizers in addition to first null, second null, and third null sinc storage. Note that all the non-rectangular apodizers are implicitly absorption-corrected. We see from the default first, second, and third null storage that the NSR can be reduced by decreasing the multiplexing density. But clearly, if the NSR must be lowered, it is far more powerful to correct the absorption and use first null storage, or use one of the other windowing functions, such as the 50% trapezoidal apodizer.

When the pages add coherently (fig. 3.13), we see the NSR grows much faster than with the random phase. This result is expected as the random phase term serves to average out much of the cross talk contributions.

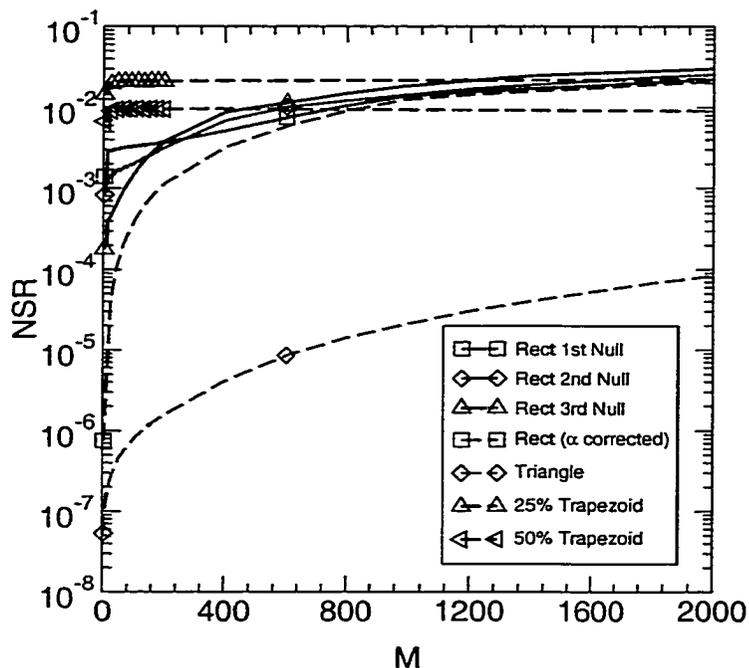


Figure 3.13: NSR (coherent page phase) for 1 cm crystal with $\alpha = 0.55 \text{ cm}^{-1}$.

When there is an angular pointing error associated with achieving the recording and reconstruction angles, then the cross talk from neighboring holograms establishes a large noise floor. Fig. 3.14 shows the histogram from an experiment using a New Focus picomotor to repeatably point at the same angle after random movements away from the desired angle. The standard deviation of the pointing accuracy was measured to be $7.6 \mu\text{rad}$. This corresponds to 3.3% of a Bragg null for a 0.5 cm crystal. The irradiance diffraction efficiency is plotted and labeled on the right axis. The histogram count is labeled on the left axis.

The jitter dominated IPI is shown in fig. 3.15 where the pointing error is zero mean and bounded below $\pm 3.3\%$ of the null spacing. We see in this case, the

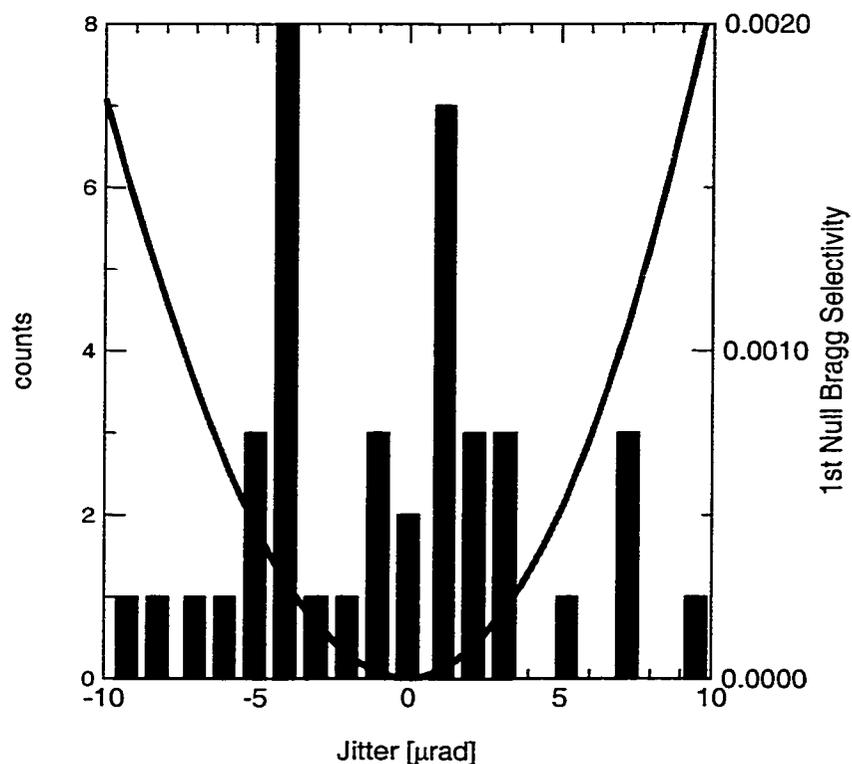


Figure 3.14: Experimental jitter histogram from repeated measurements on pointing accuracy and repeatability of New Focus picomotor. The standard deviation is $7.6 \mu\text{rad}$, which corresponds to 3.3% of a Bragg null spacing for a 0.5 cm crystal. The quadratic curve and the right axis refer to the Bragg selectivity function centered at the first null.

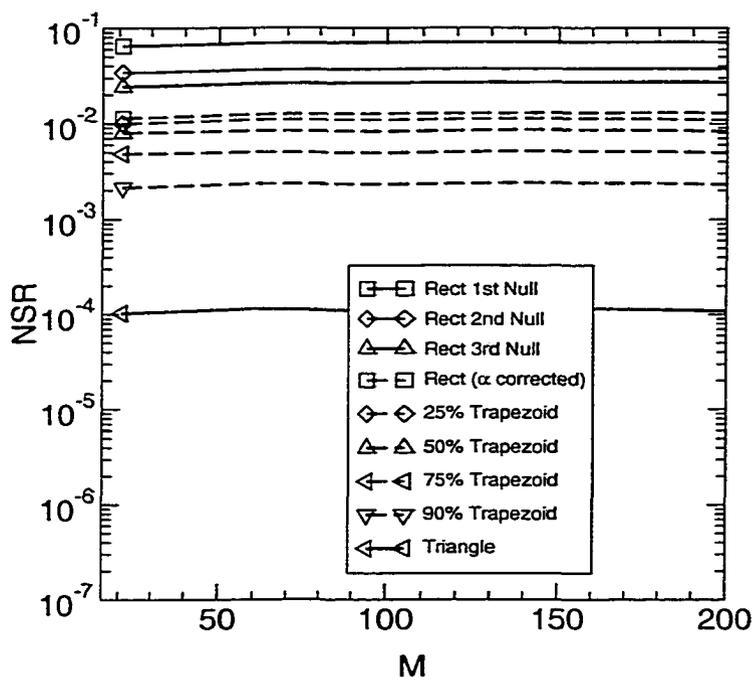


Figure 3.15: Jitter dominated IPI noise; $L = 0.45$ cm, $\alpha = 0.55$ cm $^{-1}$, 3.3% angular jitter.

cross talk does not scale significantly with M as the noise is saturated by the relative diffraction efficiency of the neighboring holograms.

3.6 Apodized $M/\#$

In this section we consider the impact of absorption on the recording and erasure dynamics of photorefractive volume holograms. α is a key component in the photorefractive effect driving the time rate of grating formation and erasure. When the absorption is very small, holographic gratings record slowly and erase slowly. When the absorption is larger, gratings record faster but subsequently erase faster. Somewhere between the two extreme cases exists an optimal absorption coefficient for a given set of material and system parameters. This relation

was derived from the underlying crystal physics and material parameters in reference [42]. The absorption is optimal in the sense that it maximizes the system metric $M/\#$.³¹ $M/\#$ provides the scaling constant relating the Bragg-matched page diffraction efficiency and the number of multiplexed pages M :

$$|\eta(\xi_z = 0)|^2 = \left(\frac{M/\#}{M} \right)^2, \quad (3.70)$$

where $\eta(\xi_z = 0)$ is the Bragg-matched amplitude diffraction efficiency considered in section 3.3. Following a similar approach to reference [42] we develop the $M/\#$ for the apodized system using Fe:LiNbO₃ as the photorefractive medium.

The reconstruction reference beam (eq. 3.23) incident on the crystal has a total power, P_{in} , of:

$$P_{in} \triangleq \int_0^H \int_0^L |R_0 R(z)|^2 dy dz \quad (3.71)$$

$$= HR_0^2 \int_0^L |R(z)|^2 dz \quad (3.72)$$

$$= HR_0^2 P_R \quad (3.73)$$

where for convenience we define the power in the apodizer shape as:

$$P_R \triangleq \int_0^L |R(z)|^2 dz \quad (3.74)$$

The diffracted power, P_{out} , can be found by integrating over the angular spectrum of the diffracted wavevectors:

$$P_{out} = \iint \left| A(\vec{k}_d, \vec{k}_i) \text{sinc}(k_{d,x}W) \text{sinc}(k_{d,y}H) \right|^2 dk_{d,x} dk_{d,y} \quad (3.75)$$

which by Parseval's theorem²⁷ can be reexpressed as the power in the reciprocal domain:

$$P_{out} = \int_0^H \int_0^W \left| \frac{E_d(0)}{WH} \text{rect}\left(\frac{x}{W}\right) \text{rect}\left(\frac{y}{H}\right) \right|^2 dx dy \quad (3.76)$$

$$= \frac{|E_d(0)|^2}{WH} \quad (3.77)$$

where $E_d(\xi_z)$ was defined in eq. 3.28.

The amplitude diffraction efficiency is the ratio of the square root of the input and diffracted powers, allows us to represent the Bragg-matched diffraction efficiency (BSF at $\xi_z = 0$) as:

$$\eta(0) = \sqrt{\frac{P_{out}}{P_{in}}} \quad (3.78)$$

$$= \frac{|E_d(0)|}{HR_0\sqrt{WP_R}} \quad (3.79)$$

$$= \frac{\epsilon_0 n_0^4 r_{13}}{H\sqrt{WP_R}} \left| \iiint E_1(\vec{r}) R(z) S_\alpha(x, z) e^{-i\vec{K}_g \cdot \vec{r}} d\vec{r} \right| \quad (3.80)$$

For small recording exposures the space charge field and the modulation depth are related as:

$$|E_1(\vec{r}; t)| \propto m(x, z) E_{sc}(\alpha) (1 - e^{-t/\tau_l(x, z)} e^{-i\omega_l(x, z)t}) \quad (3.81)$$

where the new parameters are defined in Burr's dissertation [42]. τ_l is the local recording time constant. ω_l is the imaginary component of the local recording time constant. E_{sc} summarizes a number of material properties. It can be thought of as a material scaling parameter. It is important to note that it depends on α among other things.

In terms of the saturation amplitude and recording time constant, the space charge field evolves according to:

$$|E_1(\vec{r}; t)| = A_0 [1 - e^{-t/\tau_r}] \quad (3.82)$$

Taking the time derivative at $t = 0$ for both forms allows us to express the recording slope (A_0/τ_r) as:

$$A_0/\tau_r = \frac{\epsilon_0 n_0^4 r_{13}}{\tau_x H \sqrt{W P_R}} |E_{sc}(\alpha) (1 + i\tau_l \omega_l)| \cdot \left| \int \int \int S_0 W_0 W(z) R(z) e^{-\frac{\alpha}{2}(x+L)} d\vec{r} \right| \quad (3.83)$$

$$= \frac{\epsilon_0 n_0^4 r_{13} S_0 W_0 e^{-\frac{\alpha}{2}L} (1 - e^{-\alpha W})}{\tau_x \sqrt{W P_R}} |E_{sc}(\alpha) (1 + i\tau_l \omega_l)| \cdot \left| \int_0^L W(z) R(z) dz \right| \quad (3.84)$$

During erasure, the space charge field decays as e^{-t/τ_e} . τ_e can be found by taking the ratio of η and the derivative at $t = 0$:

$$\tau_e = \frac{\tau_x \int_0^L R(z) W(z) dz}{S_0^2 \int_0^L R(z) W(z) e^{-\alpha z} dz + W_0^2 \frac{1+e^{-\alpha W}}{2} \int_0^L R(z) W(z) |W(z)|^2 dz} \quad (3.85)$$

Finally, the M/# can be computed as the product of the recording slope and the erasure time constant:

$$M/\# \triangleq (A_0/\tau_r) \tau_e \quad (3.86)$$

$$= \frac{1}{2} k_0 n_0^3 r_{13} |E_{sc}(\alpha)| \left(\frac{1 - e^{-\alpha W}}{\alpha} \right) \frac{S_0}{W_0} \sqrt{\frac{L}{W}} \frac{e^{-\alpha L/2}}{\sqrt{P_R}} \times \frac{\left(\int_0^L R(z) W(z) dz \right)^2}{\left(\frac{S_0}{W_0} \right)^2 \int_0^L R(z) W(z) e^{-\alpha z} dz + \frac{1 + e^{-\alpha W}}{2} \int_0^L R(z) W(z) |W(z)|^2 dz} \quad (3.87)$$

where $k_0 = 2\pi/\lambda_0$, n_o is the ordinary index of refraction, r_{13} is electro-optic coefficient, and $E_{sc}(\alpha)$ is the space-charge field amplitude which depends on a number of crystal parameters including α .

In figure 3.16 we plot the $M/\#$ for the three apodization cases versus the absorption coefficient. For each choice of α the beam ratio is selected to maximize the $M/\#$. Recording apodization restores the ideal Bragg nulls but suffers a 6.8% loss in $M/\#$. Apodization of the reconstruction beam loses 1.5% in $M/\#$ and nearly perfectly restores the ideal nulls.

An important result to notice is that the optimum absorption coefficient (that maximizes $M/\#$) has shifted due to apodization. For a $1 \times 1 \times 1 \text{ cm}^3$ LiNbO₃ crystal, the optimal α is 0.727, 0.697, 0.608 cm^{-1} for no apodization, reconstruction beam apodization, and recording beam apodization, respectively.

3.7 Summary

In this chapter we have presented some theoretical results about page-to-page cross talk in photorefractive media. We saw the the Bragg selectivity function can be related to the transverse electric profile through a Fourier transform. This allows modification of the reference beam to reduce cross talk and correct for deleterious material effects such as absorption. In addition, we developed the modifications necessary to describe the $M/\#$ of an apodized system. Simulations showed that there was only a minor cost in reduction of $M/\#$ when using apodization to correct the influence of absorption and restore the ideal Bragg nulls.

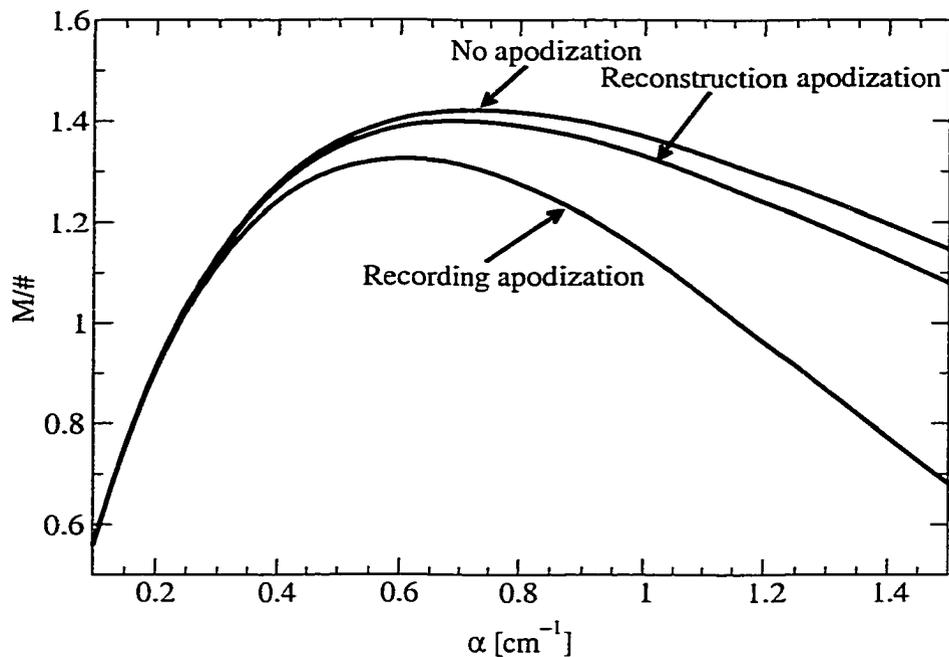


Figure 3.16: $M/\#$ versus absorption in a $1 \times 1 \times 1 \text{ cm}^3$ Fe:LiNbO₃ crystal for no apodization, reconstruction beam apodization, and recording beam apodization.

Notes Portions of this work have been published in *Proceedings-of-the-SPIE* as described in reference [47]. The basic derivation of $M/\#$ is due to Geoff Burr.⁴² Extensions to incorporate apodization in the derivation were originally considered by a former OCPL group member, Xue-Wen Chen. There are only a few minor differences between his and my derivation.

CHAPTER 4

BINARY ENUMERATION

4.1 Introduction

In this chapter we focus on increasing the total user information capacity of a VHM with binary-valued pixels by exploiting the non-traditional ($\text{snr} \propto 1/M^2$) nature of the channel. Until very recently, data pages contained, on average, an equal number of “on” and “off” pixels. This choice maximizes the user information per page. As we understand from earlier chapters, as more pages are stored in the memory, the pixel diffraction efficiency decreases as the square of the number of pages, eventually setting a limit on the number of pages that can be stored and reliably retrieved. In a similar manner, as fewer “on” pixels occur in a data page, the diffraction efficiency per pixel increases. By reducing the number of “on” pixels (N_{on}) per page of N pixels we can thus store more pages before the diffraction efficiency once again limits the reconstruction fidelity. We refer to the concept of reducing the number of “on” pixels per data page as page sparsity. In order to simplify comparison, we will call pages containing comparable numbers of “on” and “off” pixels “dense” pages. Increasing sparsity will facilitate the storage of more pages but at the expense of reducing the user information per page. We investigate this tradeoff in this chapter. We will show that in the presence of a fixed electrical detection noise floor, adjusting the data pages such that $N_{on}/N \approx 1/4$ will increase the volume storage capacity by 15%. In practice, VHMs suffer from additional noise sources such as cross-erasure, optical blur,

and photovoltaic damage. Sparse data pages are more tolerant to these noise sources than dense pages, indicating a further potential capacity improvement. In chapter 2 we defined the number of pixels per page as N^2 . However, in this and subsequent chapters it will be more convenient to refer to the number of pixels per page as N instead.

In section 4.2, the dynamics of VHM recording and erasure are briefly reviewed to extend the system metric of $M/\#$ (M-number)^{31,42} to include page sparsity. Once an additive thermal noise model is introduced in section 4.3, simple arguments can be used to relate diffraction efficiency to the detection bit-error-rate (BER) performance. Section 4.4 predicts the sparse capacity gain using the BER limited capacity. A more thorough and precise development of the capacity gain is made in section 4.5 using information-theoretic capacity bounds.

With respect to the practical details of achieving this proposed capacity improvement, section 4.6 introduces the binary enumeration code as an effective and powerful modulation code to encode the sparse pages. Encoding and decoding algorithms are presented as well. Finally, to verify the theory, experimental results are described in section 4.7 which demonstrate a capacity gain of 16% subject to a page retrieval BER of 10^{-4} .

4.2 Recording Dynamics in Holographic Storage

Data pages are typically stored in a VHM by following a recording exposure schedule^{31,37,49,50} which upon completion, results in equal diffraction efficiency per page. This schedule comes as a result of the well-known scaling law³⁵ for photorefractive materials which states that the intensity diffraction efficiency per

page, η_{page} , falls off as the square of the number of stored pages, M . The concept of the $M/\#$ (M-number) as a system metric allows the scaling to be characterized as

$$\eta_{\text{page}} = \left(\frac{M/\#}{M} \right)^2. \quad (4.1)$$

$M/\#$ succinctly describes the total efficiency of the complete storage/retrieval system (in terms of diffraction efficiency). In order to understand the effect of sparse data pages (*i.e.*, much fewer “on” pixels than “off” pixels), we need a more detailed analysis of the physical processes that underlie the $M/\#$.

Although the following results hold for a variety of page-oriented holographic recording environments, we consider the specific example of a Fourier-based holographic setup. Each “on” pixel in the data page is relayed through a Fourier lens to produce an approximation to a plane wave propagating at an angle determined by the position of the SLM pixel. The coherent superposition of the object and reference waves in the holographic material produces an interference pattern. This spatially varying intensity $I^j(\vec{x})$, is given by:

$$I^j(\vec{x}) = I_r + I_o + 2\text{Re} \left\{ E_o E_r^* \exp \left[i(\vec{k}_o^j - \vec{k}_r) \cdot \vec{x} \right] \right\}, \quad (4.2)$$

where $I_r = |E_r|^2$ is the peak reference intensity, $I_o = |E_o|^2$ is the data pixel peak intensity, \vec{k}_r is the reference wave vector, and \vec{k}_o^j is the wave vector of the j -th object pixel. The coherent superposition yields two intensity terms: a spatially constant bias term and a sinusoidal term. The user data is encoded in the sinusoidal term as a phase hologram. Note that E_o refers to the electric field amplitude associated with a single object pixel and not the composite data page. Because multiple object pixels have nearly identical wave vectors, the coherent

interference between them can be ignored. The next two subsections consider two important parameters characterizing the time evolution/decay of the holographic gratings: the recording slope and the erasure time constant.

4.2.1 Recording time constant

During recording, the amplitude of the holographic grating grows as

$$A_G(t) = A_0(1 - \exp[-t/\tau_R]), \quad (4.3)$$

where A_0 is the material-dependent saturation amplitude.³⁵ When the recording time is short compared to τ_R , then a first order series expansion around $t = 0$ is appropriate and yields

$$A_G(t) \doteq \frac{A_0}{\tau_R}t. \quad (4.4)$$

The quantity A_0/τ_R is commonly referred to as the recording slope. τ_R is the recording time constant whose value depends not only on the material parameters but also on the system parameters including beam ratio among others. For short recording times, we expect the grating amplitude to grow in proportion to the magnitude of the intensity modulation (eq. 4.2). We can thus express the recording slope in terms of the per pixel beam ratio (PBR):

$$\text{PBR} \triangleq |E_r|^2/|E_o|^2 = I_r/I_o \quad (4.5)$$

$$A_0/\tau_R \propto \text{Re}[E_o E_r^*] = I_r/\sqrt{\text{PBR}}. \quad (4.6)$$

Note that I_o refers to the intensity associated with a single pixel on a data page and not the whole data page.

4.2.2 Erasure time constant

As later holograms are recorded the previously recorded gratings erase according to the relation:

$$A_G(t) = A_G(0) \exp[-t/\tau_e] \quad (4.7)$$

where τ_e is the erasure time constant. τ_e is determined by the bulk exposure fluence to which the recording medium is subjected. τ_e is linearly proportional to the total intensity delivered to the crystal:

$$\tau_e \propto 1/(I_r + N_{on}I_o), \quad (4.8)$$

where N_{on} is the number of “on” pixels in each data page responsible for the erasure.

4.2.3 M/#

In order to characterize dependencies in the recording slope and erasure time constant on the number of pixels per page we compute $\tilde{M}/\#$, a modified form of the conventional M/#. The $\tilde{M}/\#$ [ref. 42] is defined as the product of the recording slope and erasure time constant in the short exposure limit. It is a function of the pixel beam ratio and the number of “on” pixels:

$$\tilde{M}/\# \propto \frac{\sqrt{\text{PBR}}}{\text{PBR} + N_{on}}. \quad (4.9)$$

If we reexpress the pixel beam ratio in terms of the bulk beam ratio (BR= $I_r/N_{on}I_o$) then we arrive at:

$$\tilde{M}/\# \propto \frac{\sqrt{\text{BR}}}{\text{BR} + 1} \frac{1}{\sqrt{N_{on}}} \quad (4.10)$$

$$= \text{M}/\# \frac{1}{\sqrt{N_{on}}}. \quad (4.11)$$

Notice that $\bar{M}/\#$ can be expressed as a product of two terms: the first depends only on the bulk recording modulation depth and the second term depends on the number of “on” pixels. Note that the first term is the conventionally defined $M/\#$, establishing that our modified metric $\bar{M}/\#$ is equivalent to the standard definition normalized by the number of “on” pixels per page.

From eq. 4.1 and eq. 4.11 we see that the per pixel diffraction efficiency η , is inversely proportional to N_{on} :

$$\eta \propto \left(\frac{\bar{M}/\#}{M} \right)^2 \quad (4.12)$$

$$= \left(\frac{M/\#}{M} \right)^2 \frac{1}{N_{on}} \quad (4.13)$$

Intuitively this agrees with the expectation that the page diffraction efficiency divides equally among all “on” pixels. The actual pixel diffraction efficiency fluctuates due to its dependence on the specific number of “on” pixels per page, a stochastic random variable. If we assume the binary pixels are chosen independently with *a priori* probabilities of π_1 and π_0 , then the number of “on” pixels per page is a statistical realization of a Bernoulli sequence; the probability of having n “on” pixels in a page of N pixels is

$$Pr [N_{on} = n] = \binom{N}{n} \pi_1^n \pi_0^{N-n}. \quad (4.14)$$

Computing the expected pixel diffraction efficiency, where expectation is performed over the random pixel data, results in

$$E[\eta] = \left(\frac{M/\#}{M} \right)^2 \frac{1}{\pi_1 N}, \quad (4.15)$$

under the condition that $\pi_1 N \gg 1$; that is, the average number of “on” pixels per page is much greater than one, a condition that will be satisfied in practice.

The percentage deviation of the pixel diffraction efficiency from the average value given in equation 4.15 is less than $2/\sqrt{\pi_1 N}$. Because most cases of interest will be in the range of $\pi_1 N \gg 10^3$, we can safely neglect the variation in η and hence consider the diffraction efficiency effectively as a deterministic constant:

$$\eta \doteq \left(\frac{M/\#}{M} \right)^2 \frac{1}{\pi_1 N} . \quad (4.16)$$

With the diffraction efficiency scaling understood, we next consider the noise that corrupts the measurement of a retrieved holographic data page.

4.3 Noise-limited detection

For the sake of this chapter, we consider that the memory is operating in an environment where the thermal noise of the CCD readout electronics dominates other stochastic variations in the retrieved data page. Thermal noise is appropriately modeled as an additive white Gaussian noise superimposed on the electronic CCD readout signal. The noise level establishes a constant noise floor since it is not influenced by system parameters such as the number of pixels on a page or the number of pages. One way to view the effect of the noise is to consider the limit it imposes upon the number of pages that can be stored and retrieved from the memory while maintaining a desired level of fidelity criterion. In our analysis we use the raw bit-error-rate (BER) metric as our measure of fidelity. The desired raw BER establishes a minimum electronic signal-to-noise (SNR) ratio. The noise floor sets the minimum acceptable pixel diffraction efficiency. Therefore the SNR is related to the system $M/\#$ at the same BER by

$$\text{SNR} \propto \left(\frac{M/\#}{M} \right)^2 \frac{1}{\pi_1 N} . \quad (4.17)$$

By decreasing the number of “on” pixels (decreasing π_1) in the data page we increase the SNR of each pixel thus allowing more reliable detection at a fixed noise level. Alternatively, a reduction in π_1 can be compensated by an increase in M while maintaining a constant SNR. The advantage of this second approach being that more data can be stored and retrieved at the same BER. Sparsity corresponds to decreasing π_1 from the dense page case of $\pi_1 = 1/2$. Sparsity adds redundancy to the data page which results in an associated loss of user information per page. This suggests a tradeoff between storing more pages while reducing the information per page. As we will see the overall effect will be a net gain in memory capacity for the correct choice of π_1 . The pixelwise correlations introduced by reducing π_1 can be thought of as a form of channel modulation coding appropriate for a holographic memory system.

4.4 Sparse data pages

Consider a VHM that contains M dense pages ($\pi_1 = 1/2$) with N pixels per page retrievable at a raw page BER goal of BER^* . The pixel diffraction efficiency is given by:

$$\eta^* = \left(\frac{M/\#}{M} \right)^2 \cdot \frac{1}{N/2}. \quad (4.18)$$

This system has a user capacity of $C = MN$ bits. From eq. 4.17, we know that if π_1 is decreased then the pixel SNR can be held constant if M is increased to M^* where

$$M^* = \frac{M/\#}{\sqrt{\eta^* \pi_1 N}}. \quad (4.19)$$

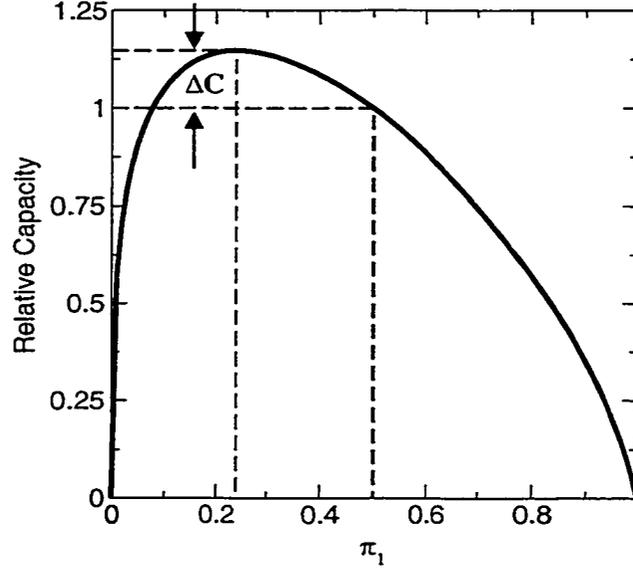


Figure 4.1: Capacity of a VHM with “on” pixel *a priori* probability π_1 . Capacity is relative to an equal priors ($\pi_1 = 1/2$) memory.

This represents an increase in the number of pages by a factor of $M^*/M = 1/\sqrt{2\pi_1}$. With the increased page count comes the associated loss in user information per pixel, $I_p(\pi_1)$. $I_p(\pi_1)$ can be bounded from above by the entropy $R_0(\pi_1)$ of an ideal source encoder. Entropy represents the largest amount of information conveyed by data sequences which have an average probability π_1 of selecting an “on” pixel:

$$I_p(\pi_1) \leq R_0(\pi_1) = -\pi_1 \log_2(\pi_1) - \pi_0 \log_2(\pi_0), \quad (4.20)$$

where $\pi_0 = 1 - \pi_1$ is the “off” pixel prior probability. The capacity of the sparse page memory ($\pi_1 < 1/2$) is:

$$C(\pi_1) = M^* I_p(\pi_1) N \quad (4.21)$$

$$\leq M^* R_0(\pi_1) N \quad (4.22)$$

Fig. 4.1 shows the upper bound on memory capacity of the sparse page memory relative to the dense page capacity. We see that if π_1 is approximately 1/4 we can achieve up to a 15% increase in the total user capacity. The constant SNR argument provided above is intuitively correct; however a more rigorous derivation of the true capacity gain is presented next.

4.5 Information-theoretic capacity

Information theory defines channel capacity as the largest amount of user information that can be stored and retrieved at an arbitrarily small bit-error rate.⁵¹ Any larger amount of information is guaranteed to be retrieved with a BER bounded away from zero.

For a specific choice of M and π_1 , the number of user bits stored in the memory, C , is given by:

$$C(M, \pi_1) = MN I_p(M, \pi_1) , \quad (4.23)$$

where I_p measures the impact of noise on the user information represented by a pixel, N is the number of pixels per page, and M is the number of pages in the memory. We then define the information-theoretic (IT) capacity C_{IT} as the peak value of C :

$$C_{IT} = \max_{M, \pi_1} C(M, \pi_1) . \quad (4.24)$$

I_p can be computed from the noise statistics of the independent and identically distributed Gaussian random variables and is given by:

$$I_p = -\frac{1}{2} \log_2 2\pi e - \int_{-\infty}^{\infty} p(y) \log_2 p(y) dy , \quad (4.25)$$

where $p(y) = \pi_1 g(y; \eta, \sigma_n^2) + \pi_0 g(y; 0, \sigma_n^2)$. $g(x; \mu, \sigma_n^2)$ is the Gaussian probability density function on the variable x with mean μ and variance σ_n^2 . η is the pixel diffraction efficiency given by eq. 4.16.

Fig. 4.2 shows C as a function of M . The solid curve denotes the capacity when π_1 is chosen to maximize C and the dotted curve shows the capacity when π_1 is restricted to be $1/2$. In addition, the circle symbol curve together with the right vertical axis shows the maximizing value of π_1 for each choice of M . In this example the system $M/\#$ was chosen from ref. [30] which represents a thermal noise-dominated VHM with 10^6 pixels per page operating at a page rate of 1 kHz. In this case, $C_{\text{IT}} = 3.22$ Gbits and $\pi_1^* = 0.253$. A different system $M/\#$ would scale the capacity but would not affect π_1^* . Comparing the peak capacity of the $\pi_1 = 1/2$ and $\pi_1 = \pi_1^*$ curves shows the 15% relative capacity advantage obtained through the use of sparse data pages.

4.6 Enumeration Code

In order to achieve this proposed capacity improvement, we need to find a simple, efficient, high rate modulation code that allows convenient adjustment of the pixel prior probability π_1 . We begin by considering block binary codes. In these codes, blocks of $\pi_1 = 1/2$ binary user data are mapped into blocks of binary codewords which are then composed onto a data page. The probability π_1 , can be computed as the average number of 1 bits per codeword. The code we will consider maps k -pixel data words into n -pixel codewords such that each codeword has weight (number of “on” pixels) exactly $m = \lfloor n\pi_1 \rfloor$. Note that this

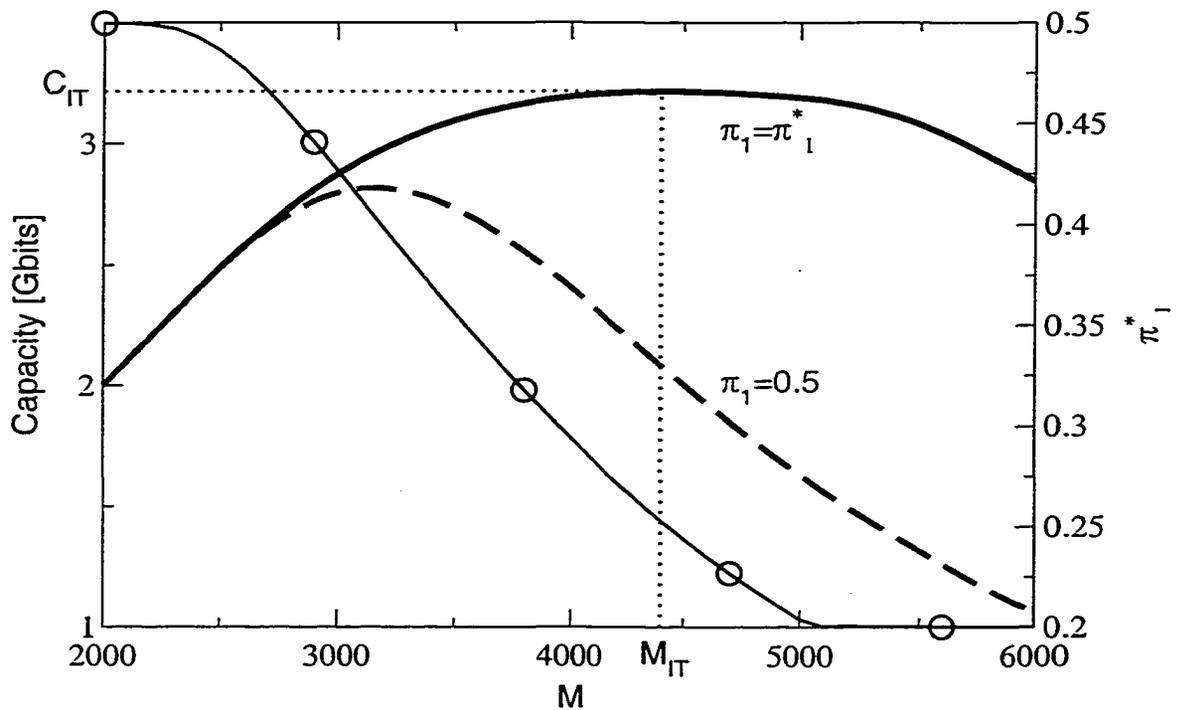


Figure 4.2: Information-theoretic (IT) capacity of a VHM with M pages. The solid curve shows the memory capacity with an “on” pixel *a priori* probability π_1 that maximizes the capacity and the dashed curve represents the memory capacity when π_1 is forced to be 0.5. The circle symbol curve along with the right vertical axis shows the maximizing value of π_1 as a function of M .

code does more than achieve the desired average value of π_1 ; it guarantees every codeword has the correct weight.

Enumerative source encoding⁵² was introduced by Cover in 1973 as an efficient scheme for compressing binary sequences containing certain forms of redundancy. For example, consider binary sequences of length n with weight m (e.g., the sequence contains exactly m 1 bits and $n - m$ 0 bits). There are precisely $S = \binom{n}{m}$ such sequences. If these sequences are labeled by the integers 0 to $S - 1$, we can alternatively refer to any particular sequence by its label. Cover describes an

efficient algorithm to quickly compute the integer label from the sequence and the reverse operation of finding the sequence from the label. Because the integer label requires $\lceil \log_2 S \rceil$ bits and the sequence requires n bits, when $\log_2 S \ll n$ the label representation will be significantly shorter, resulting in compression.

We propose using this source encoding scheme in reverse. We begin with the “compressed source” and expand it to the redundant form. Input blocks of k bits are considered to be the integer labels $\{0, 1, \dots, 2^k - 1\}$. The enumeration code then lets us associate each label with an n -bit codeword of weight m . The enumeration code is a specific type of the more general class of permutation codes.⁵³ Our shorthand notation to refer to a specific enumeration code will be $E(n, k, m)$.

An $E(n, k, m)$ code has a code rate of $R_E = k/n$. It is a constant-weight code since every codeword has Hamming weight of exactly m . The probability of a 1 or “on” pixel is $\pi_1 = m/n$. If a data page contains an integer number of codewords (*i.e.*, no codewords span multiple pages) then every page has exactly $N_{on} = \pi_1 N$ “on” pixels. Information theory guarantees that R_E must be bounded from above by the biased source entropy: a binary data source producing “on” bits with probability π_1 conveys an amount of information less than or equal to $R_0(\pi_1)$ (see eq. 4.20). The enumeration code achieves this entropy bound asymptotically as the block length increases, ensuring that longer enumeration codes provide a better code rate than shorter codes provide.

From the n and m parameters of the code we compute $k = \lceil \log_2 \binom{n}{m} \rceil$. This is the highest rate enumeration code of length n with $\pi_1 = m/n$. Since in general

$\binom{n}{m} > 2^k$, there are exactly $N_u = \binom{n}{m} - 2^k$ unused codewords; a necessary condition to avoid variable length input blocks. For certain choices of code parameters N_u may be nearly as large as 2^k . The advantage of N_u large is that the unused codewords actually provide additional error correction ability (increased distance between codewords) for the code while imposing no substantial decoding penalty. However, the minimum distance will not be improved, leading to a preference for the largest possible code rate and the smallest convenient value of N_u .

There exist alternatives to using enumeration to provide the 1:1 mapping between the input data and the constant-weight codewords. For small block lengths, a look-up table can be an effective solution. But the exponential storage requirements for the table as the block length increases prevent it as a practical answer for longer codes. Compact finite state machines can be constructed to output the codeword from the input data (or in reverse as well) but, in general, these solutions only exist for some choices of the parameters m , n , and k . We consider only the enumerative approach because it allows arbitrary selection of all the parameters, asymptotically achieves the entropy bound and has a simple encoding and low-complexity maximum-likelihood decoding algorithm.

4.6.1 Encoding algorithm

The user data to be encoded is considered as a k -bit positive integer, d . We construct the n -bit codeword, \vec{c} , from bit n to bit 1 in one pass as described in the following pseudo-code:⁵²

$m' \leftarrow m$: number of 1 bits to be assigned

For $i := n$ to 1,

```

 $n_s \leftarrow \binom{i-1}{m'}$ 
If  $n_s \leq d$  Then
     $c_i \leftarrow 1$ 
     $d \leftarrow d - n_s$ 
     $m' \leftarrow m' - 1$ 
Else
     $c_i \leftarrow 0$ 
End If
End For

```

The encoding time is linear in the block length n . The required binomial coefficients can be stored in a table (or electronic read-only-memory) requiring mn locations each holding a k -bit integer. These same coefficients will also be used by the decoder. Specific to our application goal of $\pi_1 = 1/4$, we expect the lookup table to be approximately $0.2n^3$ bits since $m = n/4$ and $k \approx 0.8n$.

4.6.2 Decoding algorithm

Let the noisy received codeword be a length n vector, $\vec{r} = \vec{c} + \vec{z}$, where \vec{c} is the transmitted codeword and \vec{z} is the noise vector imposed by the holographic recording/reconstruction process. A detection scheme must estimate \vec{c} from the noisy measurement \vec{r} . An advantage of the enumeration code is that we can easily implement maximum likelihood (ML) detection to find the most probable transmitted codeword. The m largest components of \vec{r} are set to 1 and the remaining $n - m$ components are set to 0. It can be shown that this estimate,

\vec{C}_{ML} is the most probable transmitted codeword because it has the smallest Euclidean distance to the received noisy vector.⁵³

From this sort-based codeword, we decode to the original user data in the following manner. The codeword bits follow the same ordering convention as defined in the encoder where we treat the codeword as an n -bit integer.

```

 $d \leftarrow 0$  : the user data, a  $k$ -bit positive integer
 $m' \leftarrow m$  : number of 1 bits not yet encountered
For  $i := n$  to 1,
  If  $c_i := 1$  Then
     $d \leftarrow d + \binom{i-1}{m'}$ 
     $m' \leftarrow m' - 1$ 
  End If
End For

```

Upon completion of this algorithm the decoded user data resides in d . As with the encoder, the time complexity of the decoder is linear in the block length n .

There is one further complication associated with decoding. In the beginning of this section we showed that there were N_u unused codewords in the code. The sort-based ML detection scheme may pick one of these unused codewords as its estimate. The decoding algorithm above can be modified slightly to correctly choose the minimum distance *valid* codeword. Details and a proof of correctness of the modified algorithm are beyond the scope of this chapter; refer to chapter 5

for the generalized version of this algorithm for grayscale data as well as for a proof of correctness.

4.7 Experimental Results

To confirm our theoretical predictions, we implemented a constant-BER holographic recording schedule⁵⁴ for two sets of data pages: sparse data pages ($\pi_1 = 1/4$) and the more common approach of dense data pages ($\pi_1 = 1/2$). Fig. 4.3 shows the layout of the angular multiplexing experiment. An Argon ion laser operating at 514 nm delivered 80 mW to the reference arm and 0.3 mW to the object arm. Fresnel holograms were recorded in a $1 \times 1 \times 1 \text{ cm}^3$ LiNbO₃:Fe crystal. The crystal was located 6 cm before the Fourier plane of the 20 cm focal length Fourier transform lens. Each data page contained 10^4 $36 \times 36 \mu\text{m}^2$ pixels imaged from an etched chrome-on-glass mask onto a full fill factor thermoelectrically cooled CCD with $9 \times 9 \mu\text{m}^2$ pixels. In order to reduce the effects of interpixel blur, the mask pixels were chosen much larger than the limits imposed by the lenses and the photorefractive crystal dimension. Example reconstructed data pages are shown in fig. 4.4 for both the sparse and dense data page memories. The mask data pattern was 303×171 pixels of which only a 141×95 sub-image was illuminated during recording. A linear translation stage moved the mask between exposures to ensure that no two recorded pages were the same.

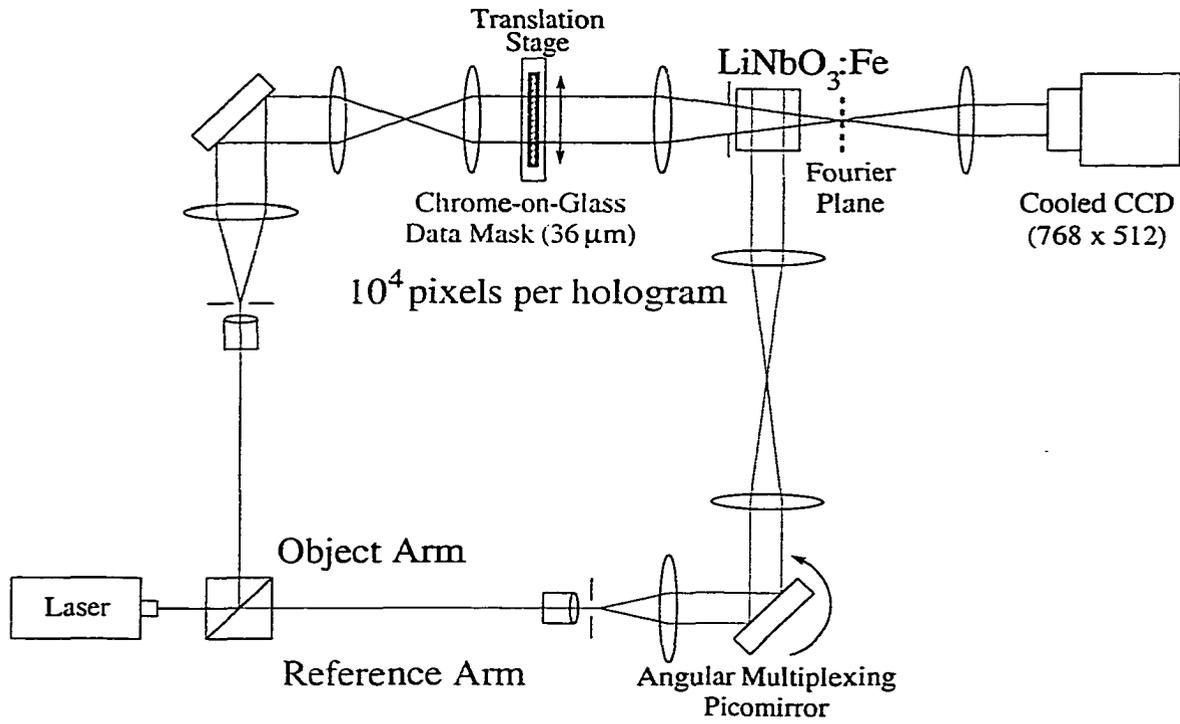
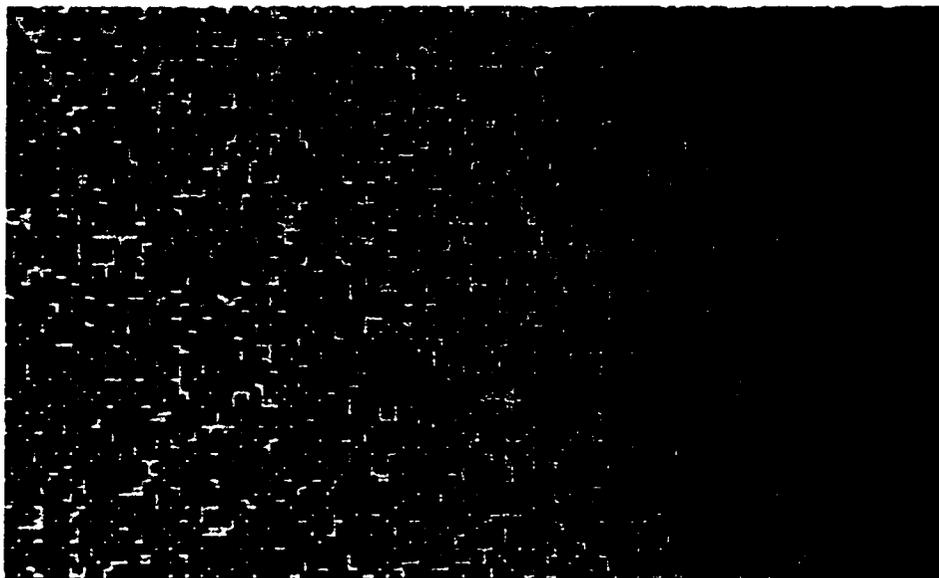
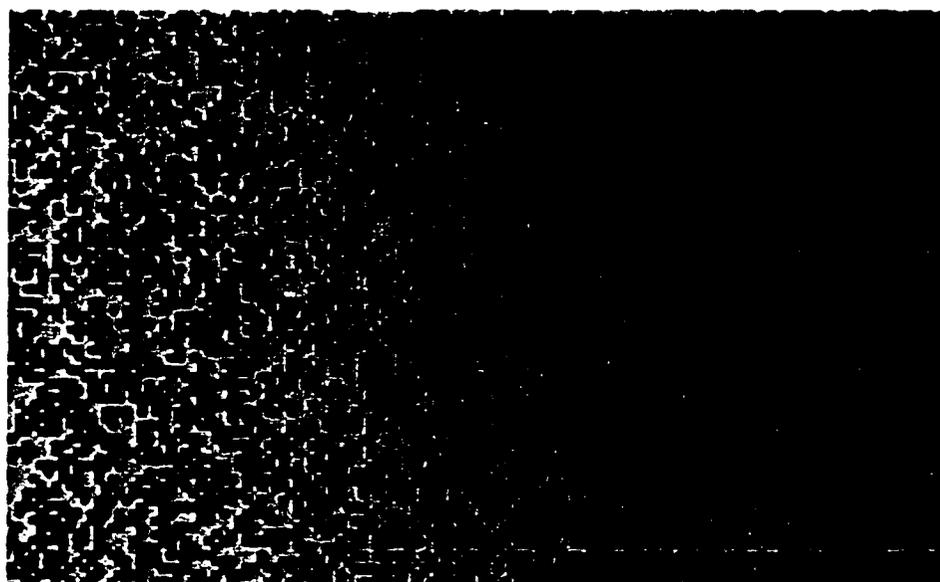


Figure 4.3: Block diagram of volume holographic memory experiment.

Our recording procedure consisted of iterative refinements on constant-BER schedules in order to produce a memory that met two requirements. First, the per page readout BER = 10^{-4} and second, the largest number of pages possible were stored that could still meet the BER requirement. To reduce the effects of photovoltaic buildup, the bulk beam ratio (total reference beam power divided by total object beam power) was kept around 250:1, guaranteeing a low grating modulation depth. Because the relative gain in capacity of sparse data pages to dense data pages is unaffected by beam ratio, we expect the experimental results to apply for other beam ratios.



(a) $\pi_1 = 1/4$



(b) $\pi_1 = 1/2$

Figure 4.4: Retrieved holographic data pages for (a) $\pi_1 = 1/4$ and (b) $\pi_1 = 1/2$.

In order to compare the two experimental memories in a fair manner, we considered exposure-limited readout. In a practical system, the readout integration time is limited by the inverse of the page rate, $T = 1/R$, as well as the maximum laser power, P . The product of the two defines a maximum readout exposure energy $E = P/R$ Joules. In our experiment the readout exposure was set to 500 mW at 33.4 ms or $E = 16.7$ mJ. By changing the integration time the readout exposure can be varied to provide detected pixels at a higher or lower BER as required by the specific system implementation.

Fig. 4.5 shows the average memory BER for the sparse and dense memories over a large range of readout exposures. It is clear that the two memories achieve the target BER goal of 10^{-4} at 16.7 mJ. In addition, notice that the two memories have very similar BER performance for all other readout exposures indicating that the memories are nearly identical in terms of raw BER. The error bars on each data point represent the 95% confidence interval of the measurement. For each data point, the average BER is computed by counting the number of bit errors out of either 3.7×10^5 retrieved data bits for the 3 mJ readout exposure up to 1.5×10^6 retrieved bits for the 18 mJ readout exposure.

Although the readout BERs for the two cases are similar, the number of multiplexed pages in each are not. The dense memory contained 97 pages with an

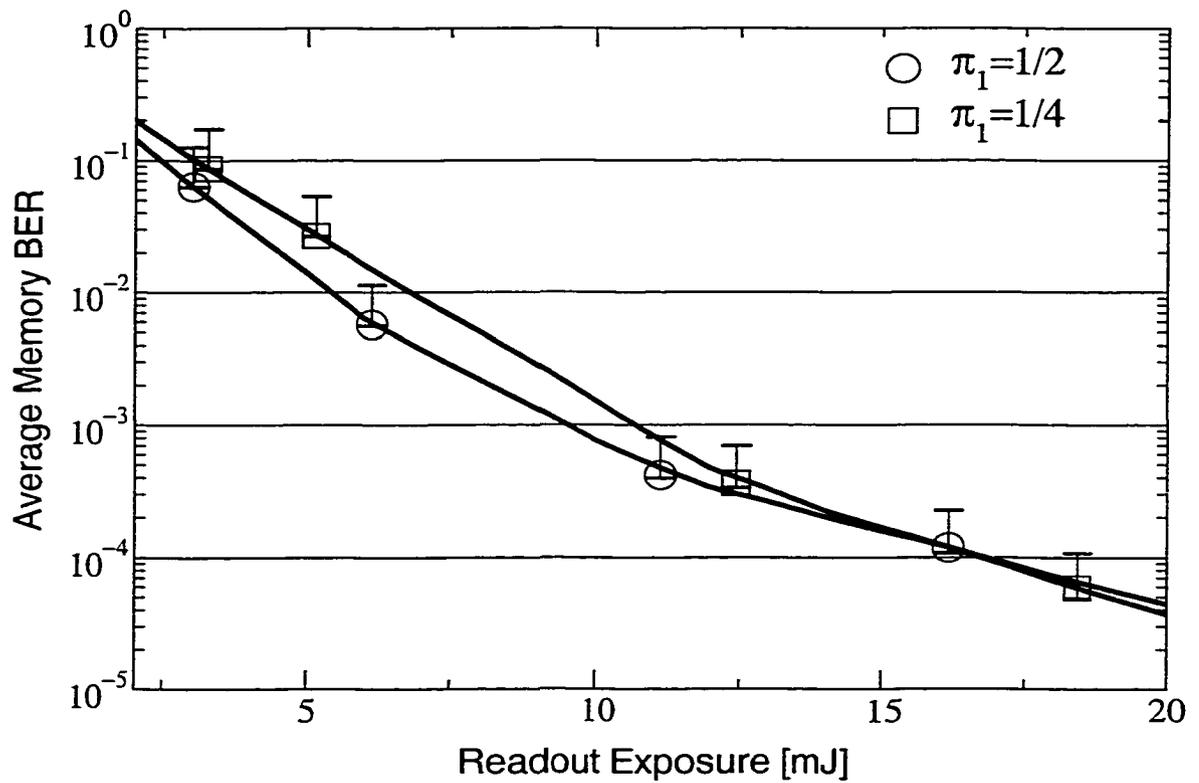


Figure 4.5: Bit error rate for different page readout exposures for the sparse and dense memories. The error bars on the data points indicate the 95% confidence interval.

average page diffraction efficiency of 4.5×10^{-9} , while the sparse memory contained 150 pages at an average page diffraction efficiency of 2.3×10^{-9} . The sparse pages were coded with a $E(52, 39, 13)$ enumeration code.

The user capacity of the dense memory was $C(\pi_1 = 1/2) = 0.97$ Mbits. For the sparse memory the enumeration code rate was 0.75 giving 7.5 kbits of information per page and a total user capacity of $C(\pi_1 = 1/4) = 1.125$ Mbits. The total increase in user capacity was experimentally seen to be 16%. This agrees well with our theoretical prediction of 15%.

4.8 Discussion

The above experiment focused on achieving a raw BER goal of 10^{-4} . In practice, a much more demanding user BER goal of around 10^{-12} must be met. In order to provide the much higher performance, the user data would be encoded using a strong error correction code, such as one of the Reed-Solomon (RS) family. The RS codewords would be input to the enumeration encoder and then composed into a data page. The code rate associated with the RS code would, in practice, lower the capacity further. However, the information-theoretic (IT) bounds derived in sec. 4.5 give the channel capacity (hence error-free communication) and are thus unaffected by the inclusion of the RS code rate.

A critical assumption in our analysis of sparse data pages has been the dominance of the electrical thermal detection noise. In general, a system that achieves the highest practical performance will balance all noise sources. Interpixel interference (IPI) or blur is one of the most common noise sources present in VHMs near the density limit.⁵⁵ By trying to maximize information per volumetric density, the IPI increases until it also limits the system capacity.⁵⁵ Although we have not analyzed this operating condition in detail, it seems clear that sparse pages offer an additional advantage over dense pages. Because sparse data pages have more distance on average between “on” pixels, they should be less affected by the blurring as compared to dense pages. A simple argument will support this idea. Let P_8 represent a 3×3 data pattern with a center “off” pixel and 8 neighboring “on” pixels. Due to the lowpass nature of the holographic channel P_8 is one of the most likely patterns to be decoded in error. The probability of the pattern P_8 occurring in a dense page is 170 times larger than for a sparse data page ($\pi_1 = 1/4$).

4.9 Summary

While dense data pages containing an equal number of “off” and “on” pixels convey the most user information per page, they suffer from enhanced interpixel interference and suboptimal use of the storage material dynamic range. Sparse

data pages which contain about 1/4 “on” pixels, leveraging the dynamic range advantage associated with storing “off” pixels in order to gain a user capacity increase of 15% in the thermal-detection-noise dominated case. A simple angular-multiplexing VHM experiment demonstrated a 16% user capacity gain at a raw page bit-error-rate of 10^{-4} .

We have extended the conventional definition of $M/\#$ to include the effect of data page sparsity. The adjusted $M/\#$ is simply normalized by the square root of the number of “on” pixels. Using $\tilde{M}/\#$ we found that the same detection error rate could be achieved for a dense page memory and a sparse page memory containing $M^*/M = 1/\sqrt{2\pi_1}$ more pages. A more precise information theoretic capacity was also derived for the sparse memory which predicts a 15% user capacity gain.

Enumeration coding was shown as an efficient way to construct sparse data pages. Due to the encoding and decoding algorithms that scale in time linearly in block length, the enumeration code is well suited for the very low latency necessary in a VHM with an aggregate data rate on the order of 1 Gb/sec. In addition, it can be shown that the decoding algorithm provides maximum-likelihood performance. The constant-weight nature of the code also provides invariance to unknown page-wise slow spatial intensity variations.

Notes Portions of this work have been published in *Applied Optics* as described in reference [56].

CHAPTER 5

GRAYSCALE ENUMERATION

5.1 Introduction

In this chapter we focus on extending the results of the previous chapter, which used a binary enumeration code. Here we consider using a grayscale shortened enumeration code. Slepian⁵³ defined this class of codes as a variant I permutation modulation code. The coding scheme produces modulation-encoded data pages that exploit the proposed capacity advantage. Subsequent details of the encoder/decoder implementation show that it admits a low-complexity solution and achieves maximum-likelihood (ML) decoding. Section 5.2 describes the history of the permutation code as well as the proposed shortened permutation code. In sections 5.3, 5.4 and 5.5 we introduce the grayscale (non-binary) shortened permutation code treating it as a channel modulation code of interest to holographic memories. Algorithmic details of the encoding and decoding schemes are presented in sections 5.6 and 5.7, respectively. We show that only a single pass through the codeword symbols is sufficient to decode a candidate codeword. In section 5.8 we describe a detection algorithm that produces the ML shortened codeword from the noisy retrieved data page. After introducing the mathematical

notation and some simple properties of the algorithm we proceed in section 5.9 to prove our claim that the algorithm achieves ML decoding. We conclude by considering the expected performance of the proposed class of codes in a thermal noise-limited VHM application.

5.2 Background

In 1965 Slepian first introduced the class of permutation codes.⁵³ The codebook of a variant I permutation code is defined by a single codeword and the rule that any unique permutation of the codeword symbols also constitutes an acceptable codeword. Slepian shows that although the code fails to achieve the band-limited channel capacity asymptotically in codeword length, it admits a very simple maximum-likelihood detection rule. It is the ML decoding which encourages further consideration of permutation codes. In 1973 Cover described a simple enumerative scheme⁵² for encoding/decoding user data into/from the permutation modulation codebook. Together the two techniques provide a low-complexity method for signaling user information through an additive noise channel by use of constant-weight codewords and ML decoding. Alternatively, enumerative coding and permutation codes can be used for a variety of other channel types as well. For example, a run-length-limited sequence can be thought of as composed of a number of primitive constituent phrases (sequences) concatenated in a specific

order. Permutation of the phrase ordering produces the set of run-length-limited sequences and enumeration of the ordering provides the sequence index. A number of authors have shown efficient run-length-limited encoding and decoding schemes based on this observation.⁵⁷⁻⁵⁹

In page-oriented optical memories we are interested in non-binary modulation codes with high code rate and unequal symbol distribution (*i.e.*, the elements of the signaling alphabet do not appear in codewords with equal probability). Enumerative coding of a variant I permutation code provides the necessary mechanism to achieve this goal. However, one disadvantage of a permutation code is that the size of these codebooks is typically not an integer power of two. This requires a more complicated scheme to select a codeword to transmit from input binary user data. Instead, we propose to expurgate the permutation codebook such that only the first 2^k codewords are retained. This permits a trivial memoryless block encoding scheme. Slepian's decoding algorithm is not maximum-likelihood for the shortened code because not all permutations are valid. But with a fairly simple extension to Slepian's approach we can still provide low-complexity ML decoding.

5.3 Permutation Codes

The codebook, \mathcal{C} , of a variant I permutation code consists of a basis codeword and all the distinct codewords produced by permuting the order of the symbols. Let the codeword symbols be elements of a finite alphabet $\mathcal{A} = \{s_j\}_{j=1}^L$ of cardinality $|\mathcal{A}| = L$. Furthermore, we assume there is an underlying ordering of the elements such that $s_1 < s_2 < \dots < s_L$.

Let the basis codeword contain m_j occurrences of symbol s_j . Because every codeword is a permutation of this basis codeword, all codewords must have the same symbol distribution $\{m_j\}_{j=1}^L$. Clearly the symbol distribution completely defines the codebook where the codeword length must be

$$n = \sum_{j=1}^L m_j. \quad (5.1)$$

For convenience, we will require that a L -ary permutation code must have $m_j > 0 \ \forall j = 1, \dots, L$. If this were not the case, we could represent the code using a smaller alphabet where the positivity requirement is satisfied.

Let the function $I_j(\underline{\mathcal{C}})$ count the number of occurrences of symbol s_j in a vector composed of symbols from \mathcal{A} . We can now explicitly define a codeword.

Definition 1 *A codeword $\underline{\mathcal{C}}$ is a length n vector with symbols from \mathcal{A} satisfying a given symbol distribution $\{m_j\}_{j=1}^L$:*

$$\underline{\mathcal{C}} = [c_1 c_2 \dots c_n], \quad I_j(\underline{\mathcal{C}}) = m_j \ \forall j. \quad (5.2)$$

The size of the codebook, $N = |\mathcal{C}|$, is easily computed by counting the number of unique possible permutations:

$$N = \frac{n!}{m_1!m_2!\cdots m_L!} \quad (5.3)$$

As an illustrative example of a permutation code, consider the codebook formed from the alphabet $\mathcal{A} = \{1, 2, 3\}$ and symbol distribution $m_1 = 2, m_2 = 1, m_3 = 1$. In this case $\mathcal{C} = \{ [1123], [1132], [1213], [1231], [1312], [1321], [2113], [2131], [2311], [3112], [3121], [3211] \}$.

Note that the permutation code only defines the collection of codewords. It does not establish any sense of sequence. We use Cover's technique of enumeration to index through the codebook.

5.4 Enumeration of Permutation Codes

As discussed in chap. 4 Cover's enumeration technique can be used in reverse to construct sequences with specific properties by encoding/decoding the label. Throughout the last decade, there have been a number of practical applications of binary enumeration codes as a modulation code for constrained channels, commonly magnetic or optical recording channels subject to a run-length-limited constraint.^{57,58,60-64} In many of these cases, the code properties involve limiting the minimum and/or maximum length of consecutive occurrences of the 0 symbol. The enumerative permutation code differs from these codes in that we consider

codewords subject to a constant-weight requirement as opposed to a run-length constraint.

By defining a lexicographical ordering on the codewords we can equivalently represent the codeword by its ordered index.

Definition 2 A codeword $\underline{C}^{(1)} = [c_1^{(1)} c_2^{(1)} \dots c_n^{(1)}]$ is defined to be less than a codeword $\underline{C}^{(2)} = [c_1^{(2)} c_2^{(2)} \dots c_n^{(2)}]$ iff for the first location in which they differ (say location t) $c_t^{(1)} < c_t^{(2)}$.

Lexicographical ordering in this case can be interpreted consistently with the arrangement of a conventional dictionary where we have assumed some fixed ordering of the composite symbol alphabet. We explicitly define the label, $u(\underline{C})$, of a codeword \underline{C} as its lexicographic index. Clearly, each codeword has a unique index such that $0 \leq u(\underline{C}) < N$. The role of the modulation encoder is thus to output the codeword corresponding to the label selected by the user data, b (e.g., $u(\underline{C}) = b$). Similarly, the decoder will perform the inverse operation.

5.5 Shortened Non-binary Permutation Codes

The user data b must select a codeword label, an integer in the range $[0, N)$. If the user data is binary and N is not an integer power of two then memoryless

block encoding is not possible without excluding some codewords. Finite memory encoding techniques can use all N codewords but incur both a complexity and error propagation cost. Error propagation occurs as decoding errors in one codeword may induce further errors in successively decoded codewords. Because it is common that N is not representable as $N = 2^k$ for an integer k we consider shortened enumerative permutation codes where only the first 2^k codewords are included in the codebook. k must satisfy:

$$0 < k \leq k_{\max} = \lfloor \log_2 N \rfloor. \quad (5.4)$$

Figure 5.1 shows the relevant portions of the communication channel under investigation, where we make two assumptions about the user data input to the modulation encoder. First, it is already encoded with an appropriate error-correction code (ECC) necessary to operate at a moderate channel bit-error-rate (BER), say roughly 10^{-3} , and provide an acceptable decoded end-user BER of 10^{-12} . Second, the input data is unbiased (balanced), that is, zeros and ones appear with equal probability. This is one reasonable system model we may expect to encounter in photorefractive volume holographic memories (VHMs).

In VHMs, a wide variety of noise sources can in theory corrupt the recording and subsequent page retrieval.^{13,17,55,65-69} We consider a limiting-case in which thermal electrical noise present in the readout electronics is the dominate corrupting influence of the channel. The thermal noise is represented by a zero-mean

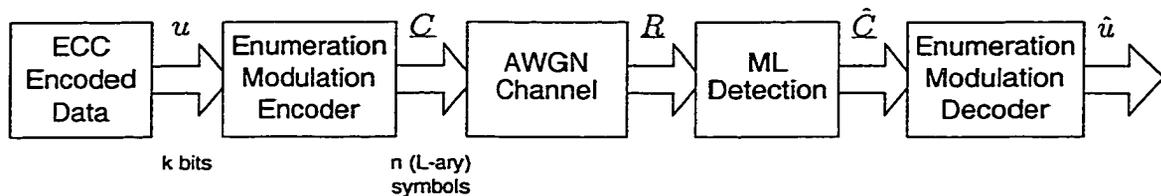


Figure 5.1: Block diagram describing the channel of interest. k bits of user data are encoded into a n symbol codeword in which each symbol can take one of L possible values. The codeword is stored and later retrieved as a real-valued vector. The detection scheme estimates the transmitted codeword which is then decoded to the estimated user data.

additive white Gaussian noise process. The measured codeword output from the channel is thus a real-valued length n vector embedded in white noise. We define the received vector and its components as $\underline{R} = [r_1 r_2 \cdots r_n]$. The detection scheme produces an estimate codeword $\hat{\underline{C}}$ given \underline{R} which is then decoded by computing the associated label $\hat{b} = \hat{u}(\hat{\underline{C}})$.

Slepian presented an implementationally simple ML detection scheme for such an additive white Gaussian noise (AWGN) channel when the complete permutation codebook is used.⁵³ The detection algorithm merely requires sorting the received vector of n real numbers and assigning the symbol s_1 to the m_1 symbols with the smallest received values, then s_2 to the next m_2 symbols, *etc.*

When only 2^k codewords are used, however, Slepian's approach does not take advantage of the reduced codebook and hence no longer achieves ML decoding for this shortened permutation code; it incurs a higher bit-error-rate at a given signal-to-noise-ratio (SNR). In section 5.8 we introduce an extension to

Slepian's approach that achieves ML detection for shortened enumerative permutation codes. To motivate consideration of the extended detection scheme, consider an example 4-ary length-189 permutation code with symbol distribution $m_1 = 113, m_2 = 48, m_3 = 20, m_4 = 8$. This code will be of interest later in the application to holographic memories. There are $2^{269.98}$ codewords yielding a code rate of $R = 269.98/189 = 1.4285$ bits per channel symbol. The entropy for this symbol distribution is $R_{\max} = 1.4818$ bits per channel symbol.

Figure 5.2 shows the BER versus the conventionally defined SNR (mean PDF separation divided by the noise standard deviation) for the example code using Slepian's decoding scheme (circle symbols) and the extended algorithm on two shortened codes $k = 215$ bits, $R = 1.1376$ (square symbols) and $k = 134$ bits, $R = 0.709$ (triangle symbols). We see for a target BER of 10^{-3} a SNR gain of 0.2961 dB and 0.6534 dB for the $k = 215$ and $k = 134$ shortened codes, respectively. Note that, as we expect, the SNR gain comes subject to a reduction in code rate.

Shortening the code results in an improvement in error correction ability (increased distance between codewords) in exchange for a reduction in code rate. A full analysis shows that in most cases reducing k is an inefficient strategy if the allowed system complexity permits a powerful outer algebraic ECC, such as

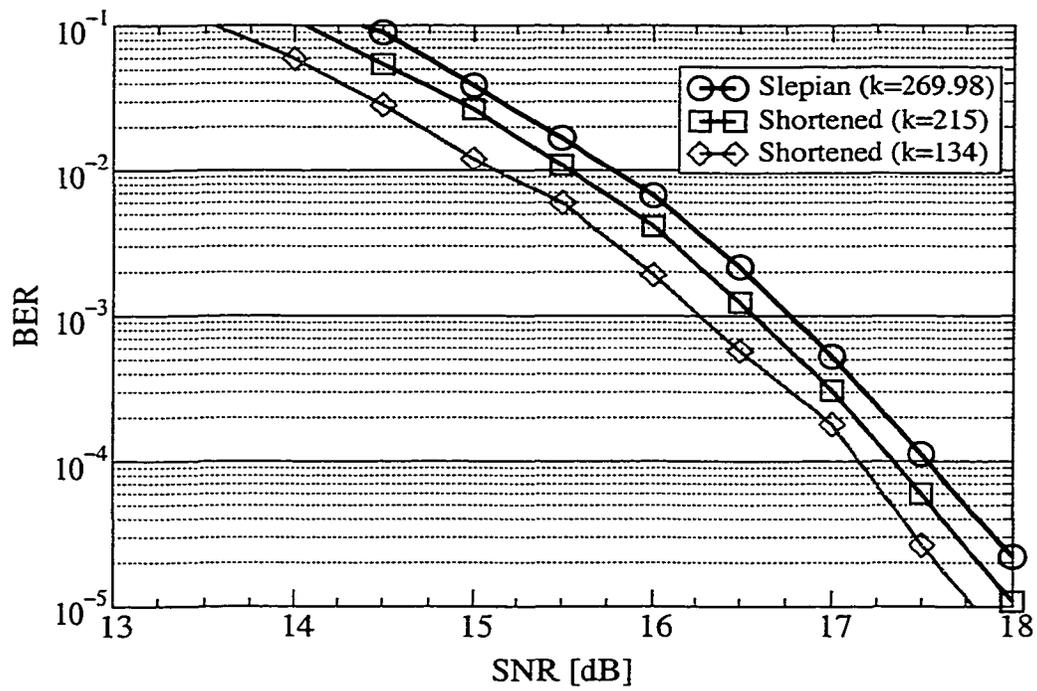


Figure 5.2: User bit-error-rate versus additive white Gaussian channel signal-to-noise-ratio for various detection/enumerative permutation codes.

a Reed-Solomon code. The outer code can typically provide the equivalent increased error correction at a higher code rate than the shortened permutation code; therefore, in those cases the practical best choice is to simply use Slepian's detection scheme on the shortened permutation code with $k = k_{\max}$. This provides the improved code rate for an acceptably small increase in BER.

5.6 Encoding Algorithm

A k -bit block of user data is provided as input to the encoder shown in Figure 5.1. The encoder maps each such block to the first 2^k enumerative permutation codewords. Reference [56] describes an encoder algorithm that maps k -bit user data into *binary* constant-weight enumeration codewords. Here we present the straightforward generalization of this algorithm (as suggested in Cover's 1973 paper) in which the codeword symbols are drawn from a general (*e.g.*, non-binary) finite alphabet \mathcal{A} .

The user data is treated as a k -bit non-negative integer label u . We construct the n -symbol codeword \underline{C} beginning with symbol 1 (leftmost) and finishing with symbol n (rightmost) in one pass as described in the following pseudo-code:

```

For  $j := 1$  to  $L$ ,
     $\tilde{m}_j = m_j$  (*  $\tilde{m}_j$  : number of unassigned  $s_j$  symbols *)
End For
For  $i := 1$  to  $n$ ,
     $j := 1$ 
    repeat := true

```

```

Do
   $z = n_s(c_1, c_2, \dots, c_{i-1}, s_j)$  (* see below for description *)
  If  $u < z$  Then
     $c_i = s_j$ 
     $\tilde{m}_j = \tilde{m}_j - 1$ 
    repeat := false
  Else
     $u = u - z$ 
  End If
  While (repeat = true)
End For

```

where $n_s(c_1, c_2, \dots, c_l)$ is the number of codewords for which the first l positions are $[c_1 c_2 \dots c_l]$. We see that for a fixed choice of L , the algorithm constructs the codeword sequentially beginning at the first position. In a practical implementation the function $n_s(\cdot)$ is stored in a look-up-table or read-only memory (ROM). Thus the encoder requires a maximum of nL k -bit integer addition and comparison operations, producing only linear complexity in block length n . We also see that this algorithm has linear complexity in alphabet size; however the space required to store $n_s(\cdot)$ increases exponentially in L .

5.7 Decoding Algorithm

Decoding a codeword \underline{C} to its associated index $u(\underline{C})$ involves counting the number of codewords below \underline{C} . This can be implemented as a simple sum as shown in the following pseudo-code:

```

 $u = 0$  (*  $u$  : the user data; a  $k$ -bit non-negative integer *)

```

```

For  $i := 1$  to  $n$ ,
  For  $s_j \in \{s : \forall s \in \mathcal{A} \quad s < c_i\}$ ,
     $u = u + n_s(c_1, c_2, \dots, c_{i-1}, s_j)$ 
  End For
End For

```

Similar to the encoding operation, we see that the decoding algorithm requires a maximum of nL k -bit addition operations.

5.8 Extended Detection Algorithm

The encoder transmits a codeword, $\underline{C}^{(j)}$ through the channel and a noise-corrupted observation \underline{R} is made by the receiver. The goal of the detection algorithm is to correctly estimate $\underline{C}^{(j)}$ from \underline{R} .

Here we consider operating on shortened enumerative permutation codes. Shortening the code can be thought of as dividing the complete variant I permutation codebook into two mutually-exclusive sets: valid and invalid. The valid set contains the first 2^k codewords of the general permutation code which comprise the shortened codebook. “First” is decided by enumerating the lexicographic index of the codeword. The remaining codewords are declared invalid. We refer to the set of valid codewords as \mathcal{C}_v .

Definition 3 *The maximum valid codeword, \underline{C}^{\max} , has the largest lexicographic index of all the valid codewords. It has a label of $u(\underline{C}^{\max}) = 2^k - 1$.*

Definition 4 *The maximum-likelihood (ML) codeword, \underline{C}^{ML} , is the codeword in \mathcal{C}_v that is closest (in squared Euclidean distance) to the received vector:*

$$\underline{C}^{ML} = \arg \min_{\underline{C} \in \mathcal{C}_v} d(\underline{C}), \quad (5.5)$$

where the squared Euclidean distance between two vectors is measured in the conventional manner as

$$d(\underline{C}) = \|\underline{C} - \underline{R}\|^2. \quad (5.6)$$

Definition 5 *The length- l prefix of a codeword is the first l coordinates of the codeword vector.*

The Slepian sort-based codeword, \underline{C}^S , is the minimum distance codeword in \mathcal{C} to the received vector. It is the ML codeword when codewords are selected from the codebook \mathcal{C} . See Reference [53] for a proof of this statement.

Refer to Figure 5.3 for a graphical example of Slepian's sort-based detection scheme showing a 3-ary code over $\mathcal{A} = \{1, 2, 3\}$ with symbol distribution $\{m_1 = 5, m_2 = 3, m_3 = 1\}$. The technique works by sorting the received vector \underline{R} and assigning the 1 symbol to the five locations with the smallest received values. The three locations with the next larger received values are labeled with the 2 symbol and the remaining location is assigned the one 3 symbol. There are two major advantages of sort-based decisions. First, the sort-based decision

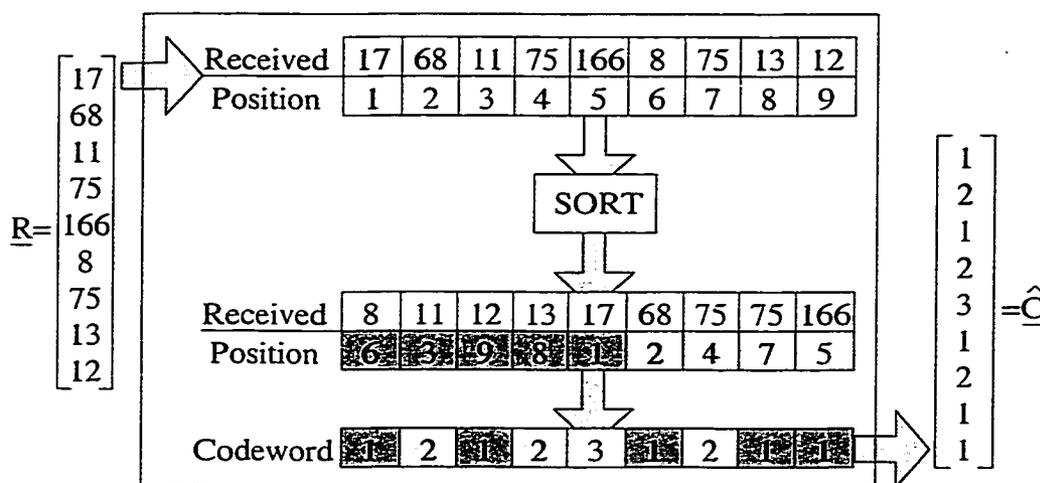


Figure 5.3: 3-level sort-based decision example. The “1” symbols are assigned to the locations with the 5 smallest received values. The “2” symbols are assigned to the 3 locations with the next largest received values. The “3” symbol is assigned to the location with the largest received value.

\underline{C}^s yields the maximum-likelihood permutation codeword with little computational overhead (when the code has not been shortened). A further advantage is its scale-invariance; there is no error introduced into the decision if the received measurement is subject to an arbitrary scaling. So channels which suffer from an unknown, slowly (relative to the codeword size) changing “gain” do not penalize sort-based decisions. This scale-invariance makes sort-detection a very robust detection scheme when the channel model deviates from the expected theoretical form. Also, note that the detection technique is threshold-less requiring no training data or preset calibration.

5.8.1 Algorithm Description

In general terms, the proposed algorithm works by extending the prefix of candidate codewords in such a way as to preserve important ordering properties as the prefix length increases. These properties will then guarantee that we find and recognize the ML codeword. We begin by defining notation for the length l prefix of a codeword \underline{C} as $F[\underline{C}, l] = [c_1 c_2 \cdots c_l]$. It is the vector of the first l symbols of \underline{C} .

The shortened permutation code (SPC) ML detection algorithm maintains a table containing P working codewords we label as $\underline{C}^{(i)}$ where $i = 1, 2, \dots, P$ along with their squared Euclidean distances to \underline{R} , $d^{(i)}$. The table is initialized to contain only one codeword: $\underline{C}^{(1)} = \underline{C}^s$. The algorithm attempts to construct $\underline{C}^{\text{ML}}$ one symbol per iteration, from left (position 1) to right (position n). At the beginning of iteration t of the main loop the codewords in the table represent the best candidate codewords, one of which will agree exactly with $\underline{C}^{\text{ML}}$ on a prefix of length $t - 1$. The goal of iteration t will be to extend this prefix by one position so that at the beginning of the next iteration, the table is guaranteed to contain a codeword $\underline{C}^{(j)}$ such that $F[\underline{C}^{(j)}, t] = F[\underline{C}^{\text{ML}}, t]$. This will be accomplished by adding new codewords to the table. During iteration t these new codewords will be obtained by swapping symbols within (not between) codewords that are already in the table. Specifically, a symbol in position t is swapped with a symbol

at position $q > t$ so that no change is made to the length $t - 1$ prefix of any codeword in the table.

Iteration t concludes with the deletion of certain codewords from the table. Codewords to be deleted are selected based on two conditions. Valid codewords are removed from the table because any further swapping can never decrease the Euclidean distance of these candidates. The best valid codeword found so far is retained in a special register \underline{B} , along with its distance d^* . Codewords with prefixes larger than $F[\underline{C}^{\max}, t]$ are also removed from the table because these codewords can never be valid. The algorithm requires a maximum of n passes through the table after which the table must be empty and $\underline{C}^{\text{ML}}$ will be \underline{B} . Commonly though, the table of working codewords will be empty well before n passes and the algorithm will terminate earlier. The algorithmic details are described by the following pseudo-code:

```

Sort  $\underline{R}$  to produce  $\underline{C}^s$ 
If  $\underline{C}^s$  is valid then
     $\underline{C}^{\text{ML}} = \underline{C}^s$ 
Else
     $d^* = \infty$ 
     $P = 1$ 
     $d^{(1)} = \|\underline{C}^s - \underline{R}\|^2$ 
    For  $t := 1$  to  $n$ ,
        For  $p := 1$  to  $P$ ,
            If  $c_t^{(p)} > s_1$  Then
                Add children codewords to table (* see below *)
            End If
            If  $c_t^{(p)} > c_t^{\max}$  Then
                Delete codeword  $\underline{C}^{(p)}$  from table

```

```

                End If
            End For
        End For
         $\underline{C}^{\text{nc}} = \underline{B}$ 
    End If

```

Procedure to add children codewords:

```

For  $b := s_j$  where  $j \in \{j : s_j < c_t^{(p)}\}$ ,
    Find position  $q = \arg \max_{q \in \Omega_q} r_q$  where  $\Omega_q = \{q : q > t, c_q^{(p)} = b\}$ 
    Construct  $\tilde{C} = \underline{C}^{(p)}$  with  $\tilde{c}_t = b$  and  $\tilde{c}_q = c_t^{(p)}$ 
    If  $\|\tilde{C} - \underline{R}\|^2 < d^*$  Then
        Add  $\underline{C}^{(P+1)} = \tilde{C}$  to table
         $P = P + 1$ 
         $d^{(P)} = d^{(p)} + 2(c_t^{(p)} - b)(r_t - r_q)$ 
        If  $\underline{C}^{(P)}$  is a valid codeword Then
            If  $d^{(P)} < d^*$  Then
                 $d^* = d^{(P)}$ 
                 $\underline{B} = \underline{C}^{(P)}$ 
            End If
            Remove  $\underline{C}^{(P)}$  from table
             $P = P - 1$ 
        End If
    End If
End For

```

Looking over the algorithm, it is clear that at the end of a main loop iteration there are only invalid codewords in the table. As soon as valid codewords are constructed they may be saved in \underline{B} but they are always discarded for reasons discussed in the previous paragraph.

5.9 Proof of ML Decoding of Shortened Permutation Codes

The basic operation used in our SPC detection algorithm is to swap two symbols in a candidate codeword. Lemma 1 determines the change in squared Euclidean distance (to the received vector).

Lemma 1 *If the symbols at locations q and t are swapped in a codeword \underline{C} to produce a new codeword $\underline{\tilde{C}}$ then $d(\underline{\tilde{C}}) = d(\underline{C}) + 2(c_t - c_q)(r_t - r_q)$.*

Proof.

$$\|\underline{\tilde{C}} - \underline{R}\|^2 = \sum_{j=1}^n (\tilde{c}_j - r_j)^2 \quad (5.7)$$

$$= \sum_{j \neq q,t} (c_j - r_j)^2 + (c_q - r_t)^2 + (c_t - r_q)^2 \quad (5.8)$$

$$= \sum_{j=1}^n (c_j - r_j)^2 - \sum_{j=q,t} (c_j - r_j)^2 + (c_q - r_t)^2 + (c_t - r_q)^2 \quad (5.9)$$

$$= \|\underline{C} - \underline{R}\|^2 + 2(c_q r_q + c_t r_t - c_t r_q - c_q r_t) \quad (5.10)$$

$$= \|\underline{C} - \underline{R}\|^2 + 2(c_t - c_q)(r_t - r_q) \quad \square \quad (5.11)$$

Theorem 1 *If \underline{C}^S is a valid codeword then $\underline{C}^{ML} = \underline{C}^S$.*

The proof of Theorem 1 follows directly from Reference [53], Appendix I. When \underline{C}^S is valid, the algorithm performs identically to Slepian's original scheme. It is when \underline{C}^S is invalid that the extension to his algorithm becomes necessary to find \underline{C}^{ML} .

In iteration t of the main loop, the algorithm constructs new codewords by taking already existing codewords in the table and swapping positions q and t . q is chosen such that $q = \arg \max_{q \in \Omega_q} r_q$ where $\Omega_q = \{q : q > t, c_q = b\}$ and \underline{C} is the specific candidate parent codeword. b is the value of the symbol to be swapped. Choosing q in this fashion is defined as “smart downstream swapping” (SDS). The SDS property guarantees that this choice minimizes the Euclidean distance of the newly created codeword to the received vector subject to the constraint that q is an element of Ω_q . The following lemma provides support for this statement.

Lemma 2 *If a codeword $\tilde{\underline{C}}$ is produced by swapping locations t and q in a codeword \underline{C} subject to the conditions that $c_t = a > c_q = b$ and $q > t$, then choosing $q = \arg \max_{q \in \Omega_q} r_q$ minimizes $d(\tilde{\underline{C}})$, where $\Omega_q = \{q : q > t, c_q = b\}$*

Proof. By Lemma 1 the squared distance of $\tilde{\underline{C}}$ is

$$d(\tilde{\underline{C}}) = d(\underline{C}) + 2(c_t - c_q)(r_t - r_q) \quad (5.12)$$

$$= d(\underline{C}) + 2(a - b)(r_t - r_q). \quad (5.13)$$

Because $a > b$ it is clear that $d(\tilde{\underline{C}})$ will be minimized by maximizing r_q . Hence position q corresponds to the largest received value from the set of allowed positions. \square

Lemma 3 *At the beginning of iteration t of the main loop, all codewords in the table represent the minimum distance codewords for their respective length*

$t - 1$ prefixes:

$$\underline{C}^{(i)} = \arg \min_{\underline{C} \in \Omega^{(i)}} d(\underline{C}) \quad \forall i = 1, 2, \dots, P \quad (5.14)$$

where $\Omega^{(i)} = \{\underline{C} : \forall \underline{C} \in \mathcal{C} \quad F[\underline{C}, t - 1] = F[\underline{C}^{(i)}, t - 1]\}$.

Proof. We prove this by induction on the loop index t . Clearly the property is true for the first iteration $t = 1$ because the table initially contains one codeword: $\underline{C}^{(1)} = \underline{C}^s$ which is by definition the minimum distance codeword with a prefix length of 0 (*i.e.*, over the set of all codewords). Assuming the lemma holds for iteration t we will show that it holds at the beginning of iteration $t + 1$ as well.

As a result of execution of iteration t , the table may contain the same codewords from the previous iteration (we define these as the parent codewords of this iteration) as well as a number of children codewords derived from the parents. We examine these two types of codewords separately.

First, consider a parent codeword, $\underline{C}^{(i)}$. Assume there is another codeword, say $\underline{\tilde{C}}$, with a smaller distance and the same length t prefix as $\underline{C}^{(i)}$.

$$\text{Assume } \exists \underline{\tilde{C}} \quad | \quad d(\underline{\tilde{C}}) < d(\underline{C}^{(i)}) \text{ and} \quad (5.15)$$

$$F[\underline{\tilde{C}}, t] = F[\underline{C}^{(i)}, t]. \quad (5.16)$$

But equation 5.16 must hold for length $t - 1$ as well:

$$F[\underline{\tilde{C}}, t - 1] = F[\underline{C}^{(i)}, t - 1], \quad (5.17)$$

which contradicts our assumption that $\underline{C}^{(i)}$ was the minimum distance codeword with prefix $F[\underline{\tilde{C}}, t-1]$. The contradiction establishes that $\underline{\tilde{C}}$ can not exist, insuring that $\underline{C}^{(i)}$ is the minimum distance codeword of prefix $F[\underline{C}^{(i)}, t]$ at the beginning of iteration $t + 1$.

Second, by lemma 2 we know that the choice of location q minimizes the distance of the child codeword $\underline{C}^{(P)}$. Because the child always differs from the parent at location t we know $\underline{C}^{(P)}$ is the minimum distance codeword with prefix $F[\underline{C}^{(P)}, t]$. This closes our inductive loop and completes the proof. \square

Theorem 2 *The proposed detection algorithm determines the maximum-likelihood codeword \underline{C}^{ML} for a shortened enumerative permutation code.*

Proof. Theorem 1 guarantees that if Slepian's sort-based decision \underline{C}^S yields a valid codeword it is the ML codeword. Otherwise, the algorithm will find \underline{C}^{ML} by successively modifying \underline{C}^S . We prove this as follows:

The algorithm begins with \underline{C}^S , an invalid codeword. \underline{C}^S and \underline{C}^{ML} share a common prefix of length $l - 1$, where l is the position of the first symbol for which the two codewords differ. $l = 1$ is a perfectly acceptable possibility indicating that they share no common prefix. At iteration l of the main algorithm loop \underline{C}^S will be a parent codeword which is by lemma 3 the minimum distance codeword with the prefix $F[\underline{C}^{ML}, l - 1]$. Because \underline{C}^S is an invalid codeword, $\underline{C}^S > \underline{C}^{ML}$.

During iteration l , \underline{C}^s parents children codewords with the same length $l - 1$ prefix and location l set to all possible lower-ordered symbols. One of these children must share a length l prefix with \underline{C}^{ML} . As long as the child codeword remains invalid, this process repeats with the next prefix symbol leading to a subsequent grandchild codeword which shares a length $l + 1$ prefix with \underline{C}^{ML} . Eventually, one of these offspring will be valid and thus represents the minimum distance valid codeword that agrees with \underline{C}^{ML} , hence it is \underline{C}^{ML} . If more than one codeword achieves exactly the same minimum distance, they are both considered \underline{C}^{ML} . \square

5.10 Application to VHM

Development of the SPC has been motivated by our specific application: volume holographic memories. Data is typically stored in a two-dimensional page parallel format containing on the order of 10^6 pixels. A pixel may take on one of many possible intensity values⁷⁰ but system considerations commonly restrict the choice to a few grayscale levels. There is a greater holographic material cost associated with storing bright pixels than dark pixels. In fact, there is no effective material cost in storing “off” pixels represented by the absence of diffracted light. In the binary-valued pixel case, previous work⁵⁶ has shown that the storage capacity of a VHM can be increased by 14% by adjusting the *a priori* probabilities

of the data page. The memory was assumed to be dominated by the thermal noise in the charge-coupled device (CCD) readout electronics. The optimal binary probabilities are such that “on” pixels occur with 25% probability and “off” pixels occur with 75% probability.

5.10.1 System Model

In the thermal noise-limited case, the noise-free electrical signal, x , from a detected pixel is

$$x = \frac{\Gamma}{M^2 N} a, \quad (5.18)$$

where N is the number of pixels in a page, M is the number of pages in the memory, a is the stored data pixel gray-level intensity. The constant Γ encompasses all the proportionality factors imposed by the specific system components (see Reference [30] for more discussion about Γ). The thermal noise corrupts the signal producing a measurement $y = x + n$ where n is drawn from an AWGN process with variance σ_n^2 . The probability density function of y , $p_Y(y)$ is thus Gaussian with variance σ_n^2 and mean $a\Gamma/M^2N$. Simulations use $N = 10^6$ and a 1 kHz page retrieval rate producing an aggregate data rate of 1 Gbit/sec.

The total number of user bits stored in the memory is $T = MI_pN$, where I_p is the number of user bits represented per pixel. The information-theoretic capacity of the memory, C_{\max} , using L -ary valued pixels is the maximum possible value of

T defined as

$$C_{\max} = \max_{M, \underline{\pi}} [M R_{\max}(M, \underline{\pi}) N], \quad (5.19)$$

where R_{\max} is the maximum amount of information represented by a pixel subject to the SNR set by M and the given symbol *a priori* distribution $\underline{\pi} = [\pi_1, \dots, \pi_L]$. R_{\max} is explicitly determined by the mutual information integral as

$$R_{\max}(M, \underline{\pi}) = - \sum_{l=1}^L \pi_l \int_{-\infty}^{\infty} p_Y(y) \log_2 \frac{p_{Y|A}(y|a)}{p_Y(y)} dy. \quad (5.20)$$

To avoid confusion, we will refer to the amount of user information stored in a sample memory as T bits for specific choices of code parameters, M , and a *priori* symbol distribution $\underline{\pi}$. C_{\max} will refer to the information-theoretic channel capacity when the grayscale signaling alphabet is constrained to a specific number of gray-levels.

5.10.2 Capacity Gain

Maximizing T over all possible $\underline{\pi}$ and M yields C_{\max} for each choice of L . As L increases, the capacity gain offered by gray-level signaling with optimal selection of $\underline{\pi}$ approaches 49%. Table 5.1 shows the gain associated with each graylevel compared to binary signaling with $\pi_1 = 0.5$.

We investigate using our proposed shortened permutation code as a practical method to achieve this capacity gain. Figure 5.4 plots the VHM storage capacity,

T , as a function of the number of stored pages. The solid curves show the maximum value of T when the memory is constrained to a given number of pages. The peak of these curves represents the VHM capacity C_{\max} . The solid curves represent graylevel signaling from $L = 3$ to $L = 8$. The two dashed curves show the number of user bits contained in the memory in the binary case ($L = 2$) when π_1 is constrained to be 0.5 ($\pi_1 = 0.5$) and when π_1 is chosen to maximize T ($\pi_1 = \pi_1^*$). The choice of the system proportionality coefficient Γ was taken from Reference [30] to be 3.8×10^{13} . Figure 5.4 establishes the upper bound on thermal noise-limited VHM performance using finite graylevel signaling.

We would also like to compare the following three practical codes: (1) binary pages with $\pi_1 = 0.5$ encoded with a Reed-Solomon (RS) code, (2) grayscale pages using $L = 4$ with $m_i = n/4$ encoded with a RS code, and (3) an inner $L = 4$ length 100 shortened enumerative permutation code with $k = 147$ bits, $m_1 = 58, m_2 = 24, m_3 = 12, m_4 = 6$ and an outer RS code. The RS code in all cases is a t symbol-error correcting length 255-symbol block code with 8-bit symbols. The performance of the various codes is plotted in Figure 5.5 along with the $L = 4$ capacity bound taken from Figure 5.4. We obtain the different points in each curve by varying the number of pages in the memory and adjusting the RS error correction ability, t to achieve a user BER of at most 10^{-12} for all

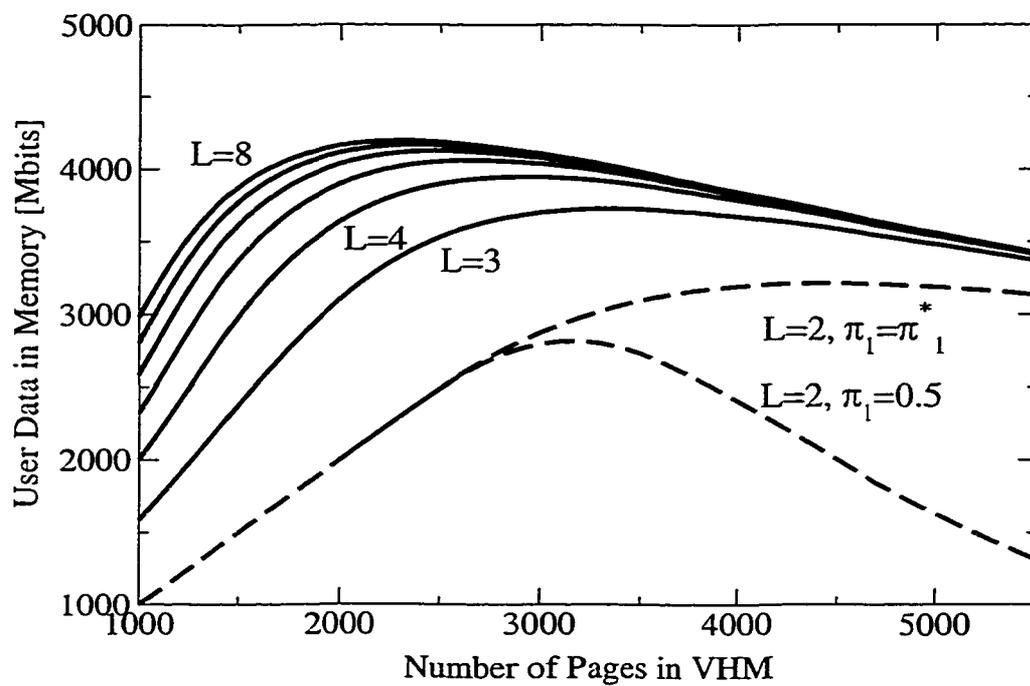


Figure 5.4: VHM storage capacity, T , for the case of L -ary pixel values. π_1 is the *a priori* probability of a pixel being “on”. π_1^* represents the choice of π_1 that maximizes the memory capacity.

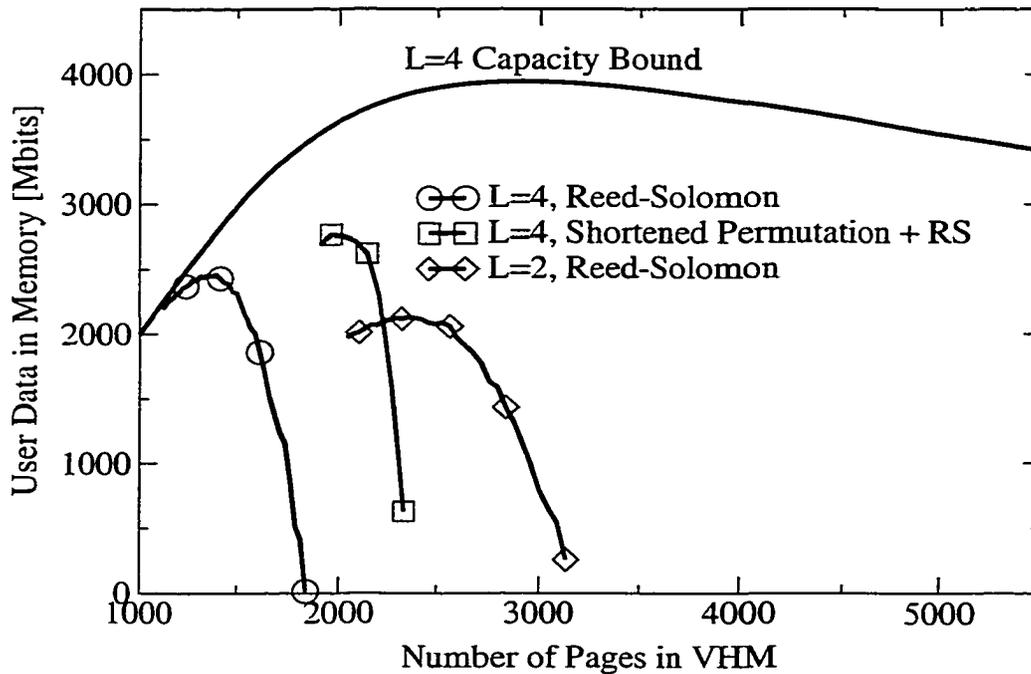


Figure 5.5: Memory capacity of various codes at a user BER of 10^{-12} .

cases. We observe a 12% improvement in the 4-ary enumeration approach over the 4-ary RS case, and a 30% increase in user capacity over the binary RS case.

Longer permutation code block sizes provide a higher code rate at the expense of decreased performance due to the phenomenon of error propagation. A single error in a modulation codeword symbol will upon decoding, magnify into a number of bit errors proportional to the codeword size. Researchers have suggested a number of approaches to address this error propagation problem.^{58,63,64,71} Initial consideration of this idea will be discussed at the end of the next chapter.

L	Capacity [Gbits]	Capacity Gain [%]
2	2.8179	0.0 ($\pi_1 = 0.5$)
2	3.2147	14.08 ($\pi_1 = \pi_1^*$)
3	3.7279	32.29
4	3.9464	40.04
5	4.0608	44.10
6	4.1273	46.46
7	4.1685	47.92
8	4.1949	48.86

Table 5.1: VHM capacity gain associated with graylevel signaling compared to binary $\pi_1 = 0.5$ signaling.

5.11 Summary

We have introduced a shortened enumerative permutation code that extended our earlier binary code from chap. 4 into a full non-binary constant-weight code. We proposed an associated low-complexity maximum-likelihood detection scheme. The detection algorithm is shown to be an extension to the procedure proposed by Slepian in Reference [53], and is proven to find the ML codeword from the shortened codebook. Because the shortened code contains an integer power of two codewords, block encoding of binary data becomes trivial. The fact that the code is parameterized by the *a priori* symbol distribution allows it to be effectively applied to produce grayscale modulated sparse data pages appropriate for volume holographic memories. Simulations predict a 30% capacity improvement obtained by using 2-bit grayscale pixels and a length 100 shortened enumerative

permutation code concatenated with a length 255 Reed-Solomon code. This improvement is relative to the storage capacity obtained by recording binary pages with only the Reed-Solomon code for error control. All pages are retrievable at an output BER of 10^{-12} . There is a theoretical 49% improvement in the VHM storage capacity when using a 3-bit (8 graylevel) signaling alphabet to compose sparse data pages.

Notes Portions of this work have been published in *IEEE Journal of Selected Areas in Communications* as described in reference [72].

CHAPTER 6

GRAYSCALE ENUMERATION EXPERIMENTS

6.1 Introduction

To verify the proposed capacity advantage of using a sparse non-binary enumerative permutation code, we conducted a number of experiments at IBM Almaden Research Center with the help of Geoff Burr, Hans Coufal, and the rest of the Holographic Data Storage crew. We integrated our enumeration encoder and decoder into the software control for the IBM DEMON³³ platform. Fig. 6.1 shows the block diagram of the DEMON system. It is the first VHM tester using a real-time SLM and a frame-rate CCD camera. The liquid crystal SLM is pixel matched to the 640×480 CCD camera. Holograms are recorded using the 90 degree angular multiplexing geometry. A Fe:LiNBO₃ crystal is the recording medium, placed 3 cm behind the Fourier plane of the 4F imaging system relaying the object beam to the CCD camera. An aperture in the Fourier plane can be changed to impose varying levels of blur. For this set of experiments the aperture was large introducing a minimal amount of blur.

Tables 6.1 and 6.2 describe the codes we evaluated experimentally. Code block lengths of 9, 12, 24, 48, 64, and 90 were used to completely cover the

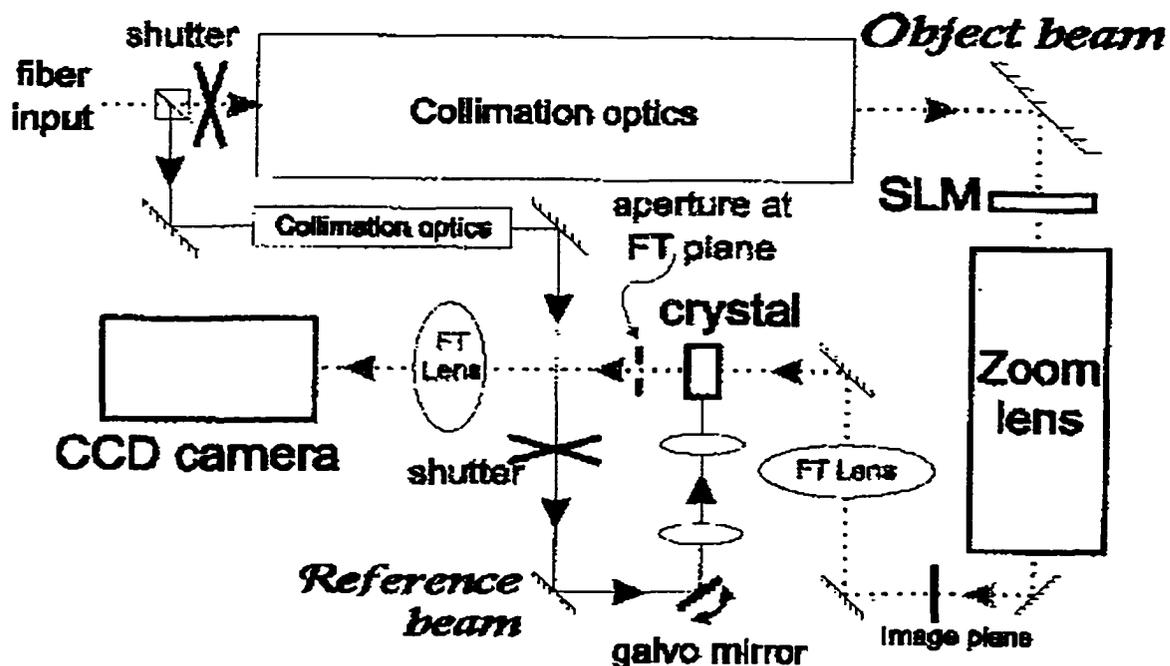


Figure 6.1: Block diagram of the IBM DEMON platform

practical usable range of lengths. Clearly, small block lengths such as 9 pixels have a terrible code rate and hence we expect them to not provide acceptable performance. Similarly, we expect a block length of 90 pixels to be too long. Longer blocks suffer a sharp BER performance loss due to the effect of error propagation; a single pixel error leads to a very large percentage of bit errors when the codeword is decoded. We expect a block length somewhere between these two extremes will provide the best usable capacity improvement. It has been discussed previously that using non-binary valued pixels will improve the system capacity. It was inferred in ref. [70] that due to practical system issues 3-ary valued pixels are the best choice. In this experiment we compare codes using binary, 3-ary,

and 4-ary valued pixels. As summarized in the results section, we concur with the previous finding that 3-ary valued pixels offer the best experimental capacity.

We begin by describing the code parameters and the testing procedure followed in the experiment. After that in sec. 6.3 we derive the relation between diffraction efficiency, pages in the VHM, and sparsity for the case of non-binary pixels. In sec. 6.4, we discuss how to compute the estimated system capacity based on the measurements from the experimental data. Finally in sec. 6.5 we compute, plot, and compare the results of the experiment.

6.2 Testing Procedure

The goal of the set of experiments was to characterize the retrieval BER performance of the proposed coding techniques, as well as the reference case of no coding. To achieve these results, we implemented the following experimental procedure for each set of code parameters under test.

1. Thermally erase the crystal
2. Do
 - (a) Move the reference beam to an unused angle
 - (b) Record the hologram using the predistortion technique
 - (c) For a range of CCD integration times: (typically about 19 exposures)
 - i. Save the captured 8-bit image
3. Repeat 3 times

Set	User bits	Block Length	# Pixels per graylevel				Rate
	k		n	m_0	m_1	m_2	m_3
A21	5	9	7	2			0.555556
A22	7	12	9	3			0.583333
A23	17	24	18	6			0.708333
A24	36	48	36	12			0.75
A25	48	64	48	16			0.75
A26	68	90	68	22			0.755556
A31	7	9	6	2	1		0.777778
A32	10	12	8	3	1		0.833333
A33	24	24	16	6	2		1
A34	54	48	31	12	5		1.125
A35	72	64	42	16	6		1.125
A36	104	90	59	22	9		1.15556
A42	13	12	7	3	1	1	1.08333
A43	30	24	14	6	3	1	1.25
A44	63	48	29	12	5	2	1.3125
A45	88	64	38	16	7	3	1.375
A46	124	90	54	23	9	4	1.37778
B31	8	9	5	3	1		0.88889
B32	11	12	7	4	1		0.916667
B33	27	24	14	7	3		1.125
B34	59	48	28	14	6		1.22917
B35	80	64	37	19	8		1.25
B36	115	90	52	27	11		1.27778
B41	10	9	5	2	1	1	1.11111
B42	15	12	6	3	2	1	1.25
B43	33	24	13	6	3	2	1.375
B44	71	48	25	13	7	3	1.47917
B45	97	64	34	17	8	5	1.51562
B46	140	90	48	23	12	7	1.55556

Table 6.1: IBM experimental set/parameter descriptions

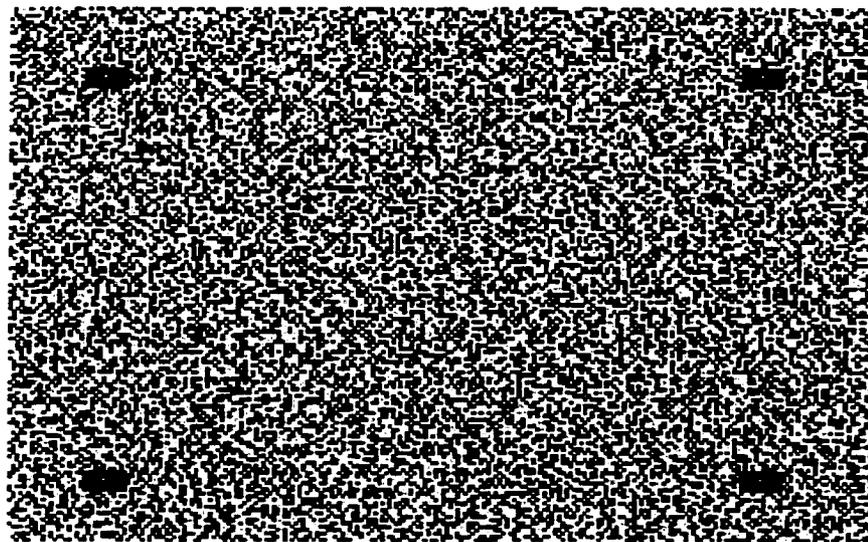
Set	Block Length	Pixel priors				% Entropy
		π_0	π_1	π_2	π_3	
	n					
A21	9	0.778	0.222			72.7
A22	12	0.750	0.250			71.9
A23	24	0.750	0.250			87.3
A24	48	0.750	0.250			92.4
A25	64	0.750	0.250			92.4
A26	90	0.755	0.245			94.2
A31	9	0.667	0.222	0.111		63.5
A32	12	0.668	0.250	0.082		70.1
A33	24	0.667	0.250	0.083		84.1
A34	48	0.646	0.250	0.104		90.2
A35	64	0.656	0.250	0.094		92.3
A36	90	0.656	0.244	0.100		94.1
A42	12	0.582	0.250	0.084	0.084	69.8
A43	24	0.583	0.250	0.125	0.042	82.3
A44	48	0.604	0.250	0.104	0.042	89.3
A45	64	0.594	0.250	0.109	0.047	91.5
A46	90	0.600	0.256	0.100	0.044	93.3
B31	9	0.556	0.333	0.111		65.8
B32	12	0.583	0.333	0.083		71.6
B33	24	0.583	0.292	0.125		83.5
B34	48	0.583	0.292	0.125		91.2
B35	64	0.578	0.300	0.125		92.4
B36	90	0.578	0.300	0.122		94.7
B41	9	0.556	0.222	0.112		67.0
B42	12	0.500	0.250	0.167	0.083	72.3
B43	24	0.542	0.250	0.125	0.083	81.3
B44	48	0.521	0.271	0.146	0.062	89.3
B45	64	0.531	0.266	0.125	0.078	91.6
B46	90	0.533	0.256	0.133	0.078	93.7

Table 6.2: IBM experimental set/parameter descriptions

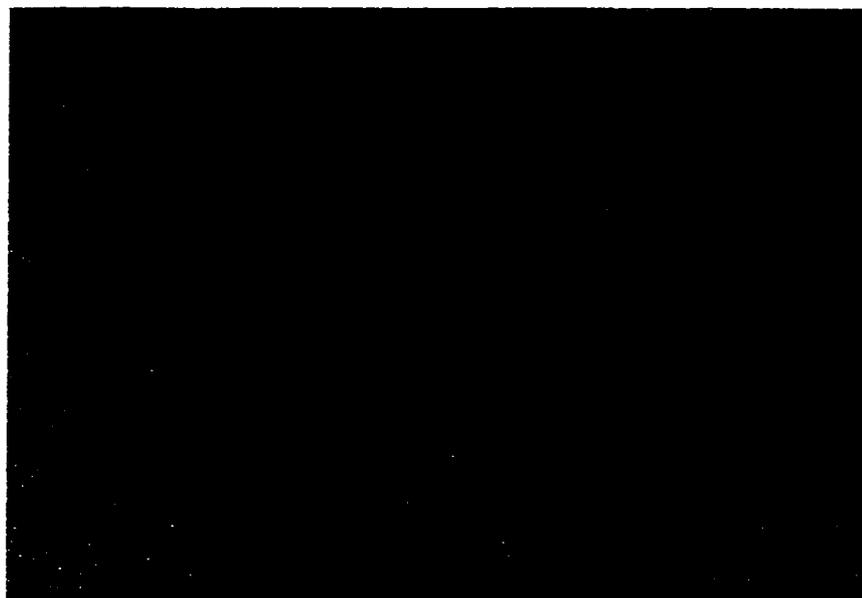
An example SLM image and corresponding CCD image of the reconstructed hologram are shown in fig. 6.2. The histogram of the hologram is presented in fig. 6.3.

The crystal was not always erased between sets. On average, it was erased approximately after every four sets of data were taken. Each hologram provides about 19 retrieved images over a variety of exposure times. These images are decoded using the code to provide an estimate of both the pixel-error-rate (PER) and the decoded user BER. In addition to decoded error rates, we compute the mean and variance of each of the L pixel levels, which we refer to as μ_i and σ_i , respectively, for $i \in \{0, 1, \dots, L - 1\}$.

We plot BER against the μ_{L-1} to determine the minimum required pixel value in order to achieve a desired BER. Fig. 6.4 shows an example curve from the A32 code. Once an operating raw BER requirement is established, say BER*, then from the BER curve fit for each code we determine μ_{L-1}^* , the minimum required pixel mean of the level $L - 1$ pixels. From the graph we see that for the XA32 set, $\mu_2^* = 26.4$ for $\text{BER}^* = 10^{-3}$.



(a)



(b)

Figure 6.2: (a) example SLM image and (b) reconstructed hologram for a 3-ary valued pixel case.

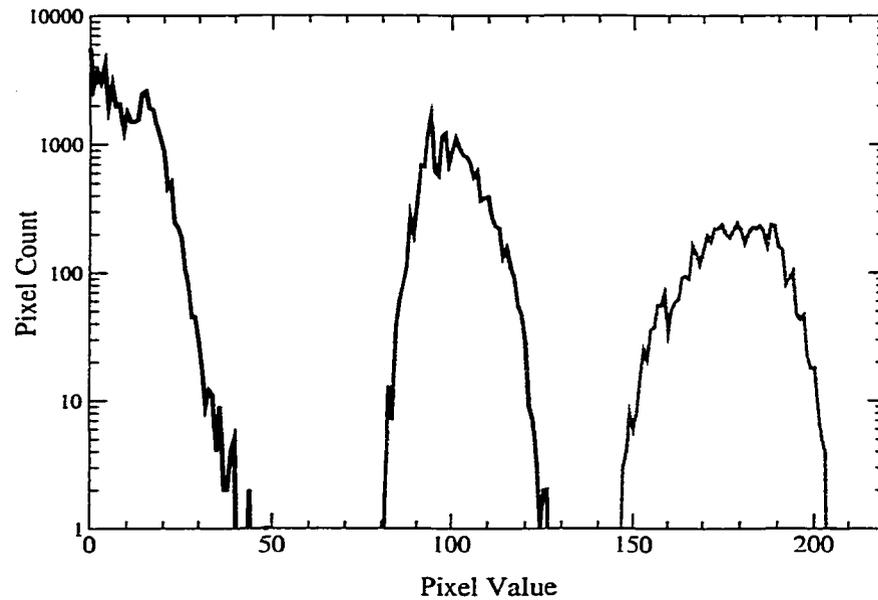


Figure 6.3: Histogram of 3-level hologram.

6.3 Grayscale $M/\#$

We know from chapter 4 that for binary sparse data pages the pixel diffraction efficiency, η , scales with pages, M , as:

$$\eta = \left(\frac{M/\#}{M} \right)^2 \frac{1}{\pi_1 N} \quad (6.1)$$

We wish to extend that result to establish the relation between pixel diffraction efficiency and pages for a sparse memory with L -ary valued pixels. Earlier results have also assumed that the pixel levels are equally spaced, *i.e.*, the level 2 pixel is twice as bright as the level 1 pixel. We wish to consider arbitrary level spacing because in the experiment, finite contrast will give the “off” pixel a non-zero diffraction efficiency in addition to the other pixel levels not being spaced equally.

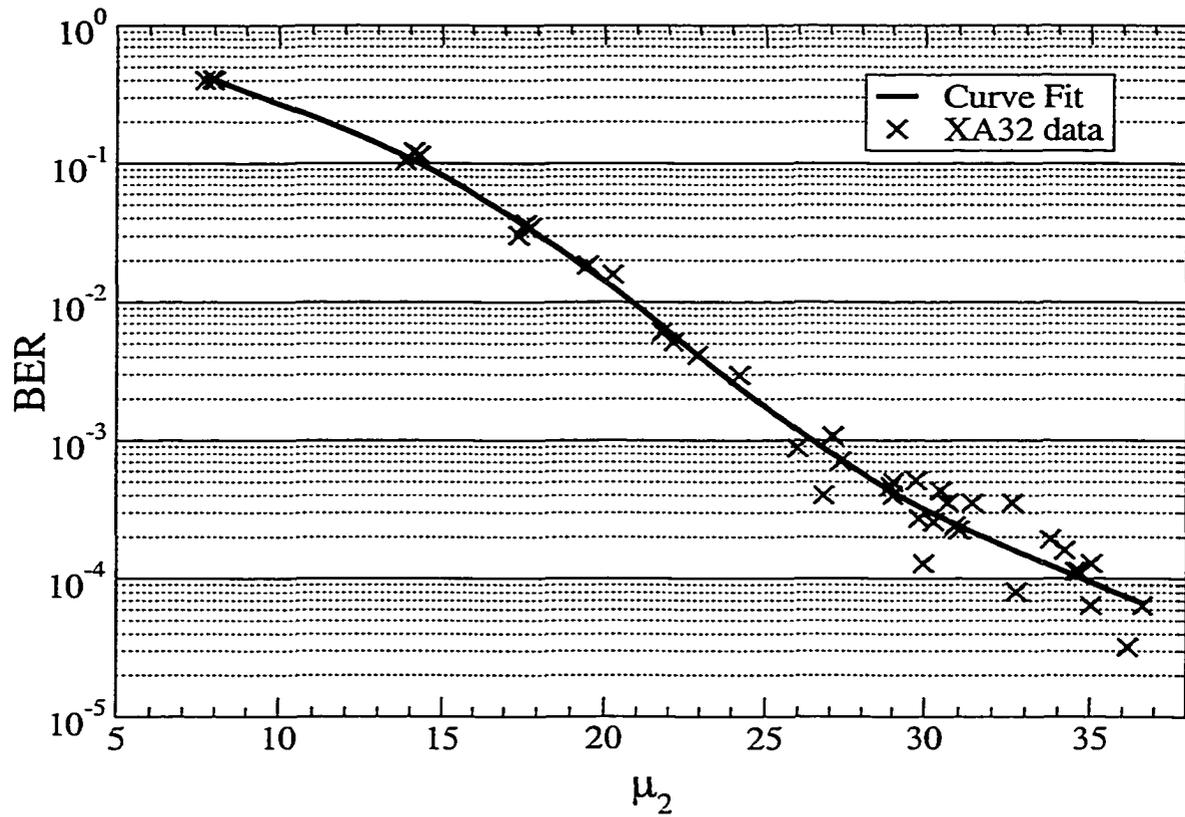


Figure 6.4: Experimental data and curve fit for XA32 data set

We derive the grayscale $M/\#$ by considering the recording and erasure constants for a specific graylevel, say level i , where $i = 0$ is the “off” pixel and $i = L - 1$ is the brightest pixel level. Similar to the development in sections 4.2.1-4.2.3, we consider the recording slope and the erasure time constant of the grating stored as the interference pattern between a single pixel in the object beam and the reference beam.

For pixel level i we wish to find the proportionality constant, $M/\#^{(i)}$, between level i pixel diffraction efficiency, η_i , and the number of pages in the memory, M , such that:

$$\eta_i = \left(\frac{M/\#^{(i)}}{M} \right)^2 \quad (6.2)$$

Let the total power in the object beam be I_o Watts, and the total power in the reference beam be I_r . The beam ratio is defined as:

$$\text{BR} = I_r/I_o \quad (6.3)$$

We can express the total object beam power as a sum of the beam power associated with all the pixels at each level. Let each level i pixel contain $I^{(i)}$ Watts such that:

$$I_o = \sum_{i=0}^{L-1} I^{(i)} N_i \quad (6.4)$$

where N_i is the number of pixels at level i .

To conveniently characterize the choice of pixel intensity levels, we define the ratio between the power in each pixel level and the brightest pixel level power as:

$$A_i \triangleq I^{(i)} / I^{(L-1)} \quad (6.5)$$

The recording slope goes as the amplitude of the interference pattern which for level i pixels is $\sqrt{I_r I^{(i)}}$. The erasure time constant is unchanged with respect to the earlier cases as it is the total illuminating power of $I_r + I_o$. The $M/\#$ for graylevel i can be computed as the product of the recording slope and the erasure time constant:

$$M/\#^{(i)} = \frac{\sqrt{I_r I^{(i)}}}{I_r + I_o} \quad (6.6)$$

$$= \frac{\sqrt{I_r I^{(L-1)} A_i}}{I_r + I_o} \quad (6.7)$$

$$= \frac{\sqrt{I_r I^{(L-1)}} \sqrt{A_i}}{\sqrt{I_o (BR + 1)} \sqrt{I_o}} \quad (6.8)$$

$$= \frac{\sqrt{BR} I^{(L-1)}}{BR + 1} \frac{\sqrt{A_i}}{\sqrt{I^{(L-1)} \sum_{j=0}^{L-1} A_j N_j}} \quad (6.9)$$

$$= \frac{\sqrt{BR}}{BR + 1} \frac{\sqrt{A_i}}{\sqrt{N \sum_{j=0}^{L-1} A_j \pi_j}} \quad (6.10)$$

$$= M/\# \frac{\sqrt{A_i}}{\sqrt{N \sum_{j=0}^{L-1} A_j \pi_j}} \quad (6.11)$$

The important result of the above sequence of equations is that the graylevel i pixel diffraction efficiency can be expressed in terms of the standard system-level definition of $M/\#$ if we just include the normalization factor associated with the relative power in each graylevel pixel. The overall relationship between level i pixel diffraction efficiency and the other VHM system parameters is thus:

$$\eta_i = \left(\frac{M/\#}{M} \right)^2 \frac{A_i}{N \sum_{j=0}^{L-1} A_j \pi_j} \quad (6.12)$$

The relationship between the minimum pixel mean and the total pages in the VHM can now be expressed by rearranging eq. 6.12:

$$M^* = \frac{M/\#}{\sqrt{N}} \frac{1}{\sqrt{\mu_{L-1}^*}} \frac{1}{\sqrt{\sum_{j=0}^{L-1} A_j \pi_j}} \quad (6.13)$$

where we have chosen μ_{L-1}^* as the minimum acceptable level $L - 1$ pixel mean that achieves a target raw BER of BER^* .

6.4 Experimental Results to Estimated Capacity

From the BER versus pixel mean experimental curve fits, we use eq. 6.13 to estimate the number of pages in the VHM and hence the capacity provided by each set of codes and associated parameters. The capacity estimate assumes the

diffraction efficiency scales as $1/M^2$ and that the storage of the additional pages in the memory does not degrade the fidelity of the already stored holograms; only the diffraction efficiency of the holograms is reduced. We know from earlier work⁵⁴ that this assumption is invalid and the consecutive storage of the data pages does affect the hologram fidelity to some degree. We recognize that the capacity estimates provided by this approach are therefore an upper-bound on the practical achievable storage capacity. With that caveat in mind, we convert our experimental data to a capacity estimate by the following procedure:

1. Curve fit the decoded user BER versus the mean $L - 1$ pixel value
2. Find μ_{L-1}^* that achieves a target raw BER goal of BER^*
3. Compute the estimated number of storage pages as:

$$M^* = \frac{M/\#}{\sqrt{N}} \frac{1}{\sqrt{\mu_{L-1}^*}} \frac{1}{\sqrt{\sum_{j=0}^{L-1} A_j \pi_j}} \quad (6.14)$$

where this equation was derived in the previous section.

4. Compute the estimated capacity as:

$$C = M^*NR \quad (6.15)$$

where R is the overall system code rate, or equivalently, the number of user bits represented per pixel.

In the experimental data, the noise variance on each pixel level is not the same, as is assumed in our thermal noise dominated model. As a result, it is better to choose unequally spaced pixel levels in order to reduce the BER. This

in turn reduces the idealized sparsity gain achieved by equally spaced levels. We keep track of this difference by considering two types of sparsity gain: the ideal sparsity gain using equally spaced levels, and the experimental sparsity gain using the level spacing achieved in the actual holograms. We refer to the ideal sparsity gain as S_I and the true gain as S . To compare the two, we define γ such that $S = \gamma S_I$, where γ must be in the range $[0, 1]$. The ideally spaced levels for L -ary valued pixels are:

$$B_j \triangleq \frac{j-1}{L-1} \quad (6.16)$$

The set of levels $\{A_j\}$ will refer to the level distribution of the experimental hologram. We can compute γ directly as:

$$\gamma = \sqrt{\frac{\sum_{j=0}^{L-1} B_j \pi_j}{\sum_{j=0}^{L-1} A_j \pi_j}} \quad (6.17)$$

To develop a better understanding of the many factors affecting the capacity estimate, C , we can re-express eq. 6.15 as a product of more insightful terms:

$$C = M^* N R \quad (6.18)$$

$$= \left(M / \# \sqrt{2N} \right) \frac{1}{\sqrt{\mu_{L-1}^*}} \frac{1}{\sqrt{2 \sum_{j=0}^{L-1} A_j \pi_j}} \quad (6.19)$$

$$= F_M F_B \gamma S_I R_E R_0 \quad (6.20)$$

F_M is the system factor which includes the $M/\#$ and the number of pixels per page. F_B is the BER factor which increases as the square root of the required pixel diffraction efficiency decreases. S_I is the ideal sparsity and γ is the constant describing how close to the maximum achievable sparsity the experimental data is. R_E is the code rate of the shortened enumerative permutation code that is directly responsible for providing the sparsity factor $\gamma S_I > 1$. R_0 is the code rate of any remaining error-correction or modulation codes. Remember that a detected pixel value of μ_{L-1}^* provides a decoded raw BER of BER^* which is likely on the order of 10^{-3} . An outer code of rate R_0 is necessary to reduce the raw BER to an output user BER goal on the order of 10^{-12} . Note the inclusion of the factor of $\sqrt{2}$ in F_M to account for the $\sqrt{2}$ factor placed in the denominator of the sparsity term. This normalizes the sparsity factor such that S_I for the binary uncoded case is identically unity, and for any sparse code S_I is larger than unity and represents the multiplicative factor on the increased number of storable pages relative to the uncoded equal prior case.

6.4.1 Local Threshold Detection

In order to provide a baseline comparison to the enumerative permutation code results, we also employ a local threshold detection scheme for the binary pixel data sets. This detection method works by processing the page in 8×8 pixel blocks

and computing the histogram of the 64 pixels. From that histogram and the true knowledge of the transmitted data page we exhaustively find the threshold that minimizes the BER of the block. This is an upper bound on the performance of a local thresholding scheme because it uses the (necessarily unavailable in practice) true data to choose the best possible local decision threshold. The conversion of the minimum required pixel value to a capacity estimate is identical to that of the enumeration-based codes. In fact, we consider the local threshold as another sort of “code” which has a code rate, R_E , of unity: 1 bit per pixel, and a sparsity factor of unity.

6.5 Results

The results of the various codes are summarized in table 6.3. Set XC20 is local threshold detection, and the other codes have all been previously defined in tables 6.1 and 6.2. Fig. 6.5 shows the results with the capacity plotted against the number of pages in the memory for the XC20, XA2, XA3, XA4, XB3, and XB4 sets.

For these results, $BER^* = 10^{-3}$ and an outer RS(255,t=17) is applied with a rate of $R_0 = 0.8667$, which provides an output user BER of 6.2×10^{-13} .

Notice that using these modulation codes causes a change in the readout rate of the VHM in addition to the change in the obvious change in capacity. The

Set	Capacity [Gbits]	Pages M^*	Ideal Sparsity S_I	Sparsity Factor γ	BER Factor F_B	Rate R_E
XC20	1.252	1445	1	0.9600	0.2441	1
XA21	1.208	2510	1.500	0.9352	0.2902	0.5556
XA22	1.083	2143	1.414	0.8997	0.2732	0.5055
XA23	1.360	2215	1.414	0.9297	0.2733	0.7083
XA24	1.312	2018	1.414	0.9223	0.2510	0.7500
XA25	1.328	2043	1.414	0.9369	0.2502	0.7500
XA26	1.340	2046	1.430	0.9353	0.2481	0.7556
XA31	1.122	1665	1.500	0.9134	0.1971	0.7778
XA32	1.212	1677	1.549	0.9031	0.1944	0.8333
XA33	1.358	1567	1.549	0.9008	0.1822	1.0000
XA34	1.389	1425	1.477	0.9103	0.1719	1.1250
XA35	1.323	1357	1.512	0.9018	0.1614	1.1250
XA36	1.084	1082	1.500	0.8979	0.1303	1.1560
XA42	1.107	1179	1.500	0.8881	0.1436	1.0830
XA43	1.183	1092	1.549	0.9285	0.1231	1.2500
XA44	1.212	1065	1.604	0.9052	0.1191	1.3120
XA45	1.169	981	1.569	0.9123	0.1112	1.3750
XA46	1.027	860	1.596	0.9103	0.0961	1.3780
XB31	1.130	1467	1.342	0.9280	0.1911	0.8889
XB32	1.194	1503	1.414	0.9207	0.1872	0.9167
XB33	1.270	1302	1.359	0.9252	0.1680	1.1250
XB34	1.351	1268	1.359	0.9184	0.1649	1.2290
XB35	1.268	1170	1.352	0.9294	0.1511	1.2500
XB36	1.324	1196	1.355	0.9306	0.1538	1.2780
XB41	1.071	1112	1.389	0.9208	0.1411	1.1110
XB42	1.044	964	1.342	0.9297	0.1254	1.2500
XB43	1.104	926	1.414	0.9277	0.1145	1.3750
XB44	1.073	837	1.414	0.9046	0.1062	1.4790
XB45	1.033	786	1.414	0.9021	0.1000	1.5160
XB46	0.845	627	1.409	0.9191	0.0785	1.5560

Table 6.3: Experimental capacity parameters

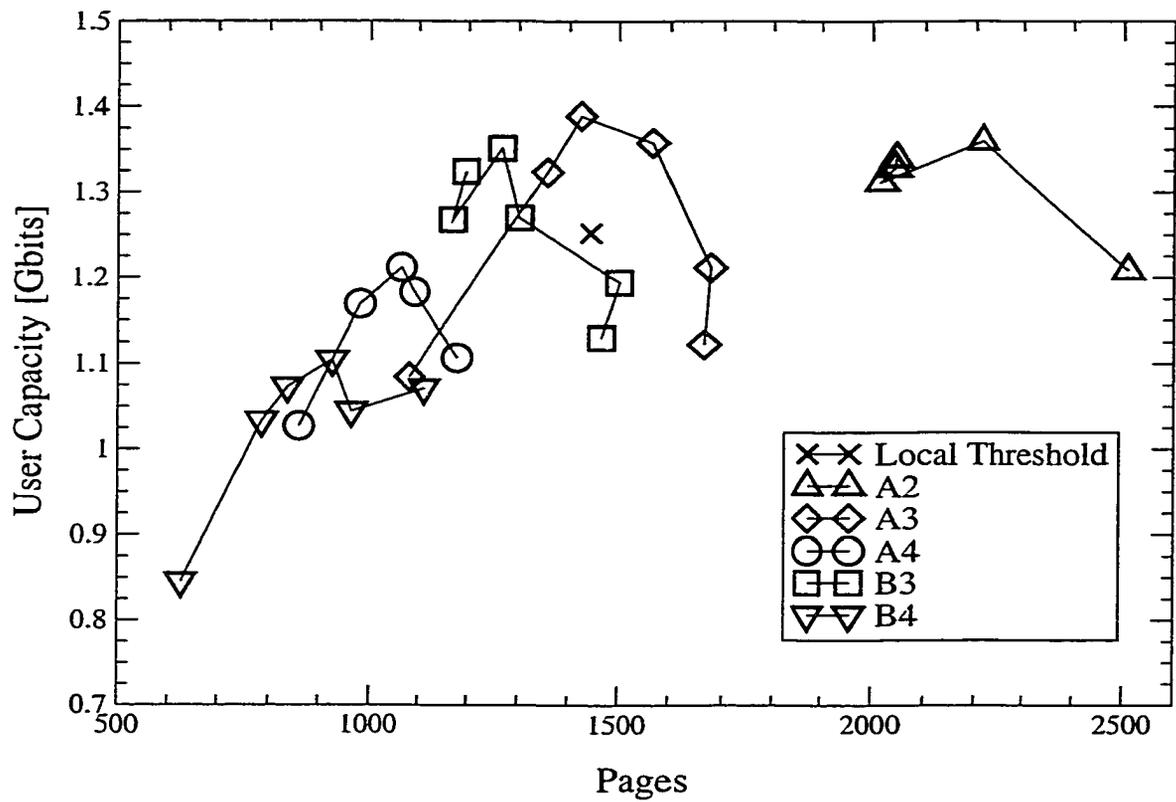


Figure 6.5: Experimental estimated VHM capacity

system $M/\#$ value of 4358.9 taken from ref. [30] assumes a page readout rate of 1 kHz with 10^6 pixels per page. So threshold detection operates at an aggregate transfer rate of 1 Gbit/sec. In table 6.4, we summarize the relative change in capacity and readout rate achieved for each experimental data set.

We see that the best performing code is the A34 code: 48 pixels per codeword, 31 “0” pixels, 12 “1” pixels, and 5 “2” pixels. Each pixel represents 1.125 bits of information. This code achieves a 10.93% increase in capacity and a 12.5% improvement in the readout rate as compared to no modulation coding and a local threshold detection scheme.

6.6 Error Propagation

The results in the previous section showed that sparse non-binary enumerative permutation codes can improve the VHM capacity. However, the phenomenon of error propagation degraded those results. We initially described error propagation in sec. 5.10.2, but we briefly review here the fundamental cause.

In order to achieve an efficient enumeration code rate (approaching entropy), long blocks must be used. The non-distance preserving mapping between the input and output of the enumeration code results in a small number of symbol errors present in the codeword decoding to a much larger number of bit errors in the output data. The severity of error propagation can be characterized by

Set	Capacity Change [%]	Readout Rate Change [%]
XA21	-3.53	-44.44
XA22	-13.49	-41.67
XA23	8.58	-29.17
XA24	4.77	-25.00
XA25	6.09	-25.00
XA26	6.99	-24.44
XA31	-10.38	-22.22
XA32	-3.25	-16.67
XA33	8.44	0
XA34	10.93	12.50
XA35	5.67	12.50
XA36	6.99	15.60
XA42	-11.63	8.30
XA43	-5.54	25.00
XA44	-3.25	31.20
XA45	-6.64	37.50
XA46	-17.99	37.80
XB31	-9.76	-11.11
XB32	-4.64	-8.33
XB33	1.38	12.50
XB34	7.89	22.90
XB35	1.25	25.00
XB36	1.25	27.80
XB41	-14.46	11.10
XB42	-16.61	25.00
XB43	-11.83	37.50
XB44	-14.33	47.90
XB45	-17.51	51.60
XB46	-32.50	55.60

Table 6.4: Change in capacity and readout rate relative to case of binary pixels and no modulation coding (local threshold detection).

looking at the error rate magnification upon decoding. Ideally, it would be a small constant, but unfortunately instead, it scales with the code block length, n .

To visualize this effect consider the following simulation. We take random codewords selected from the A46 code ($n = 90$ pixels, $L = 4$). We corrupt the codeword with a random weight 2 error pattern. This corresponds to the most likely error event we will encounter upon retrieval of the holographic data page. The corrupted codeword is decoded and the hamming distance is computed between the true and decoded 124 bit sequences. This process is repeated many times and we form a histogram of the number of bit errors incurred by a weight 2 codeword error. Fig. 6.6 shows the histogram. A nearest neighbor codeword error generates on average about 40 bit errors (30% of the user data block).

To combat this problem we propose a concatenated coding scheme as suggested in references [71, 58, 63, 64]. We will refer to the method as “reverse coding” due to the reversed application order of an algebraic error-correction code. Typically, user data is encoded with an outer strong ECC followed by a modulation code. We propose adding an inner weaker high-rate ECC after the modulation code in order to protect the modulation codeword from errors. Because, if we can correct the codeword errors incurred in the channel, then there is no error propagation. When we fail to correct all the errors, the error magnification penalizes the decoder.

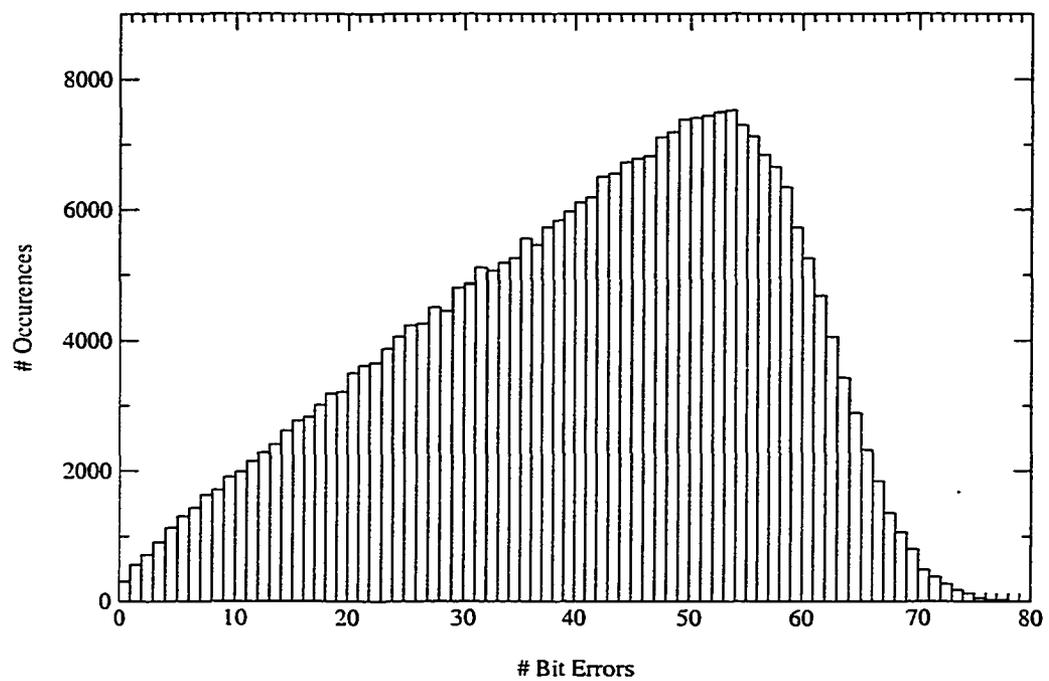


Figure 6.6: Histogram of the number of bit errors incurred when decoding random codewords from code A46 with weight 2 error patterns.

There are a number of optimizations that can be done to implement the reverse coding scheme. We will only consider a basic approach here.

Let the L -ary pixels be expressed in the form $L = q^b$; hence there are b q -ary symbols expressed in a pixel. Furthermore, we restrict our attention to only the cases where q is prime. For the inner code we choose a systematic Reed-Solomon code defined over $\text{GF}(q^m)$. The parameters of the codewords are defined in table 6.5. a codewords are concatenated to form the information symbols of the RS code: $\text{RS}(\tilde{N}, \tilde{K}, t)$. a is chosen as large as possible to fit as many codewords as possible into the $\tilde{K} = (q^m - 1) - 2t$ RS symbols. However, there will be a small leftover number of unused symbols because $n \nmid [(q^m - 1 - 2t)m/b]$ in general. In order to improve the code performance, the unused symbols are discarded from the code yielding a shortened RS code: $\text{RS}(N, K, t)$. It is good to note that shortening the code preserves the maximum distance separation (MDS) property of the original Reed-Solomon code. This can be easily proven by considering the shortened codewords as a linear subspace formed by deleting the unused codeword coordinates. Because the higher dimensional space satisfied the MDS property so must a linear subspace of it.

The codeword is protected up to t symbol errors. Encoding is a simple procedure as described in fig. 6.7. The user data protected with a strong outer ECC code is encoded with the modulation code E_1 producing a codeword, \underline{C} , that is

		q^m -ary symbols	L -ary pixels
Input data	K	$\lceil anb/m \rceil$	an
Parity	P	$2t$	$\lceil 2tm/b \rceil$
Length	\tilde{N}	$q^m - 1,$	$\lceil (q^m - 1)m/b \rceil$
Shortened Length	N	$K + P$	

Table 6.5: Inner code parameters

a concatenation of a enumeration codewords. \underline{C} is treated as a K RS information symbol block input to the inner systematic RS code which outputs the $2t$ symbols of parity. The parity symbols are expressed as pixels and appended to \underline{C} to produce a new protected codeword $\underline{C}' = [\underline{C} \mid \underline{P}]$. Protecting the codeword requires an additional $\lceil 2tm/b \rceil$ pixels. We consider two types of reverse codes: type I, and type II. In the type I approach (fig. 6.7a), the parity from the RS code is encoded directly into pixels which do not satisfy any modulation constraint (sparse or otherwise). Type II codes (fig. 6.7b) encode the RS parity with another enumeration code, E_2 , to produce a longer sequence of parity pixels that do satisfy the sparse modulation constraint.

It is clear that the parity in the type II code may suffer from error propagation as did our original unencoded codeword. We expect t to be small and hence the parity sequence to be significantly shorter than the enumeration codeword. In this case, the error propagation incurred on the parity pixels will hopefully be small.

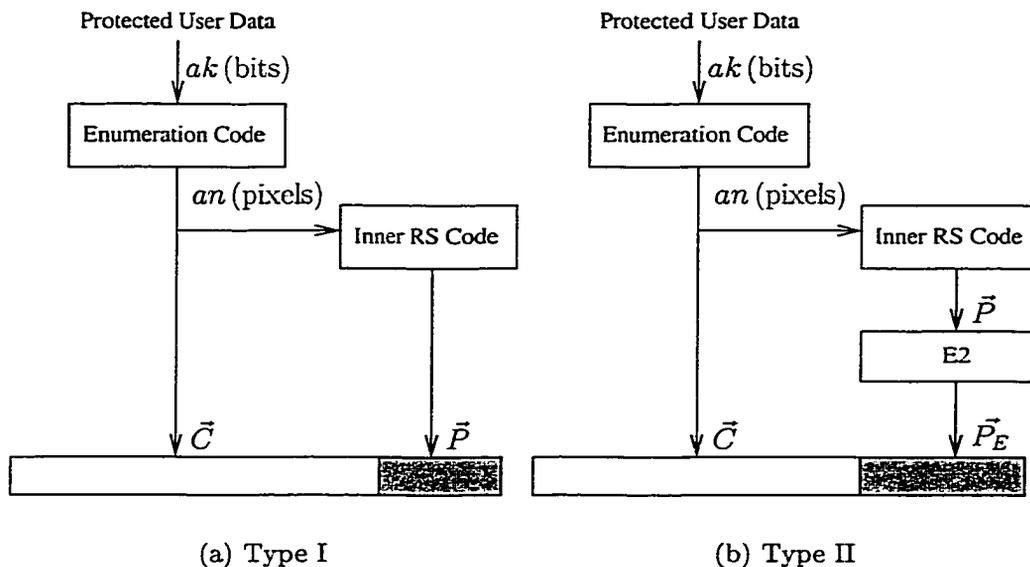


Figure 6.7: Block diagram of reverse encoding scheme. (a) Type I: parity pixels violate sparse constraint. (b) Type II: parity pixels meet sparse constraint due to encoding with enumeration code E2.

Another obvious extension of the type II code, is to repeat the protection and re-modulation coding of the parity recursively on the parity from the previous code stage. For example, a type III code could be defined by protecting the modulation-constrained parity from the type II code with a new (shorter) RS code. The new parity is then modulation encoded and appended to the sequence. This would help to combat the error propagation in the parity codeword. Clearly this process could be repeated until an implementational complexity issue is raised. We will only consider the type I and type II codes in this chapter.

6.6.1 Implementation Details

There are a number of minor details that have to be addressed when considering encoding the Reed-Solomon $GF(q^m)$ symbols in terms of L -ary (q^b) pixels. Thoughtful consideration will solve most of these at the expense of a partially utilized (wasted) pixel or two. In general, a block of r RS symbols represents mr q -ary digits of information, which can be expressed as $\lceil rm/b \rceil$ pixels. The number of pixels wasted (due to the ceiling function) will therefore be at most 1 pixel. Blocking the symbols into pixels in this manner results in a pixel possibly containing information associated with two symbols. The decoding complexity is not sabotaged by this fact because it is a trivial matter to (in hardware or software) treat each pixel as a serial concatenation of b q -ary digits. Pixels are appended to the digit stream until m digits are captured and a RS q^m symbol is output. Any remaining digits from the last pixel are shifted to the beginning of the conversion register and the process repeats for the next symbol.

However there is an issue concerning detection on the parity portion of the type I reverse code. By encoding the parity symbols directly into pixels, we do not meet the sparse modulation constraints imposed by the first enumeration code. This decreases the sparsity capacity advantage by increasing the priors associated with the brighter pixel values. It also prevents us from applying the sort-based ML detection algorithm. Simulations assume a thresholding scheme

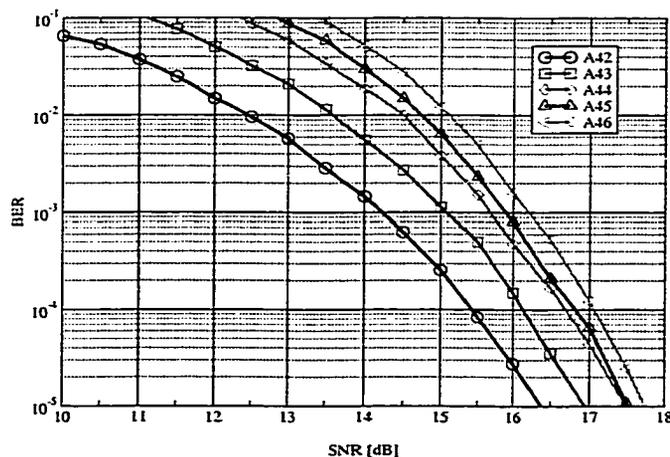


Figure 6.8: Unprotected BER performance of the A4 set of codes ($n = \{12, 24, 48, 90\}$, $L = 4$). Error propagation limits the performance of the long codes.

where the detection threshold is placed halfway between each possible graylevel. This is clearly not the best approach, but it is a reasonable solution. Some of the more obvious disadvantages include the unequal error rate for the end graylevel (levels: 0 and $L-1$) transition probabilities as well as the danger of local intensity variations in the retrieved page requiring some form of adaptive threshold tracking and estimation. Type II codes will use the ML detection scheme discussed in the last chapter.

6.6.2 Simulation Results

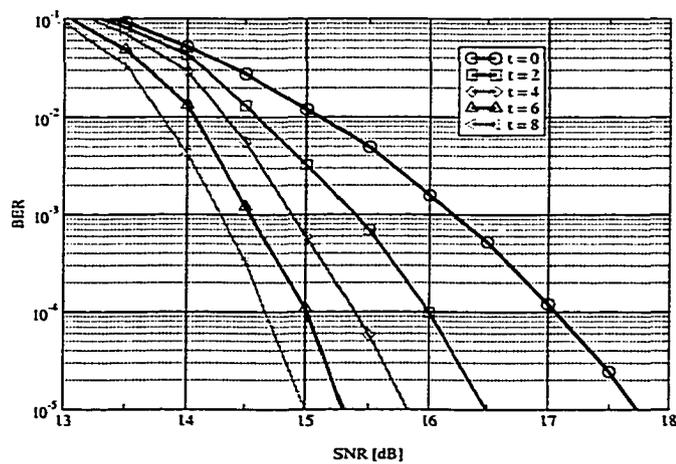
To see the effect of error propagation, look at fig. 6.8, which shows the BER performance on the A4 set of codes on the AWGN channel. All five codes have effectively the same prior symbol distribution and only differ in block length. A42

is 12 pixels long, while A46 is 90 pixels long. The loss in performance for longer codes is caused by the onset of error propagation.

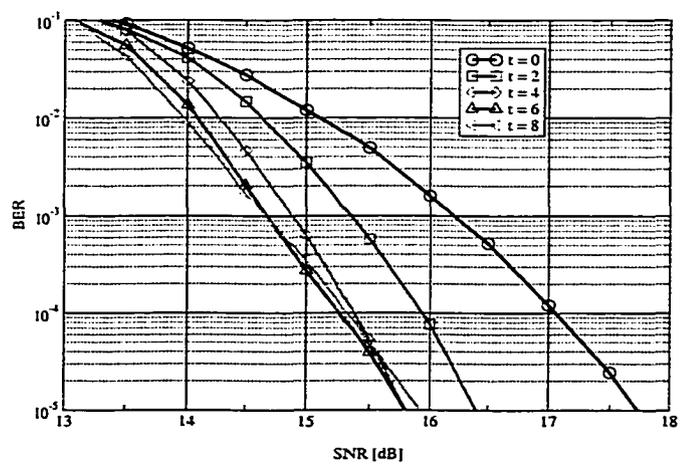
Fig. 6.9 shows the BER of the A46 enumeration code in the AWGN channel. The five curves range from $t = 0$ (no reverse coding) to $t = 8$. As an indicator to the power of the type I code, we see that the $t = 8$ code offers an SNR improvement of 2 dB at a raw BER of 10^{-3} . The cost in achieving the lower operating SNR is the 0.935 code rate. Fig. 6.9b shows the type II code performance. Note the unusual trend that increasing the error correction ability beyond $t = 4$ actually leads to an increase in the BER in the case of high SNR. This strange trend occurs due to the increased error propagation as the enumeration-encoded parity codeword is lengthened.

6.6.3 Reverse coding applied to the IBM experiment

In order to further improve the enumeration code results on the IBM experimental data, we propose estimating the improvement using reverse coding. Since the experiment did not actually include reverse coded pages, we must formulate a method to predict the expected performance change. In addition, we must justify that our prediction would match or bound the performance in an actual complete experiment.



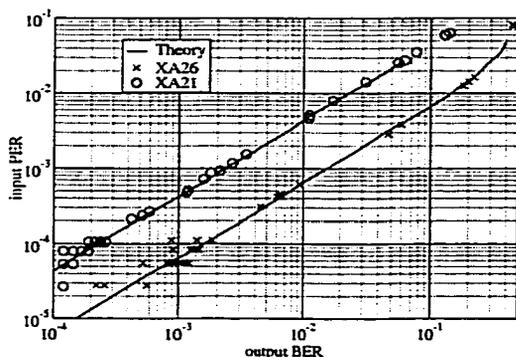
(a) Type I



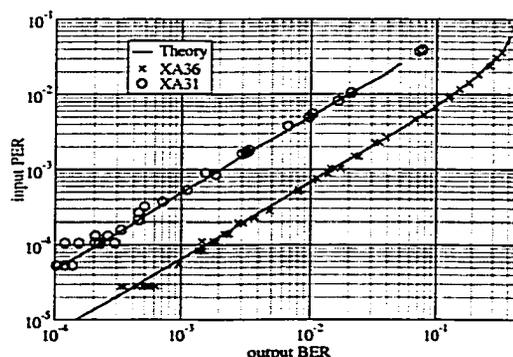
(b) Type II

Figure 6.9: Reverse coding on the A46 ($n = 90$, $L = 4$) code using an inner RS(255, K , t) code.

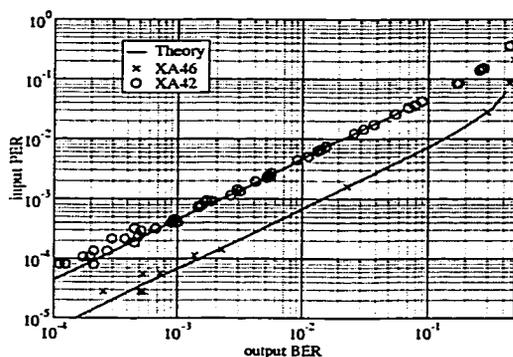
The procedure to apply the reverse coding is as follows. We begin with the target raw BER goal of the modulation code (typically 10^{-3}). From the AWGN channel simulations, We find the SNR necessary to achieve the target BER using the reverse code under test. Evaluating the theoretical BER of the enumeration code without reverse coding at this SNR yields the unprotected BER (UBER). Because the experimental data represents the same parameters and algorithm as the uncoded enumeration code (see noise model assumption discussed next), we expect the UBER to be representative of the raw error-rate we should observe in the experiment. Following the procedure developed earlier for estimating the capacity change due to the enumeration codes, we curve fit the experimental decoded user BER versus the mean $L - 1$ pixel value and find the mean level that achieves the UBER requirement (as opposed to BER* requirement without reverse coding). The reduction in the required operating pixel mean level represents the improvement in storage capability due to reverse coding through the previously introduced factor F_B (see eq. 6.20). Of course, the lower mean pixel values are achieved at a reduction in the code rate, reflected in the capacity factor R_0 . For the type I codes which do not meet the sparsity constraint, we also recompute the sparsity factor γ to include the adjustment to the prior graylevel probabilities. We expect that for small values of error-correction t , the capacity



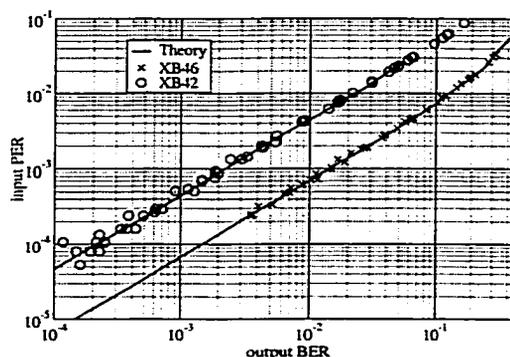
(a) XA2



(b) XA3



(c) XA4



(d) XB4

Figure 6.10: Comparison of theoretical and experimental input/output error-rates for data sets XA2, XA3, XA4, and XB4.

will be improved until a point is reached where increasing t does not provide enough of an increase in F_B to cover the loss in R_0 and γ .

To partially justify this capacity extrapolation based on theory and experiment, we must consider the possible large effect of the difference between the true experimental noise and the theoretically assumed thermal noise. Fig. 6.10

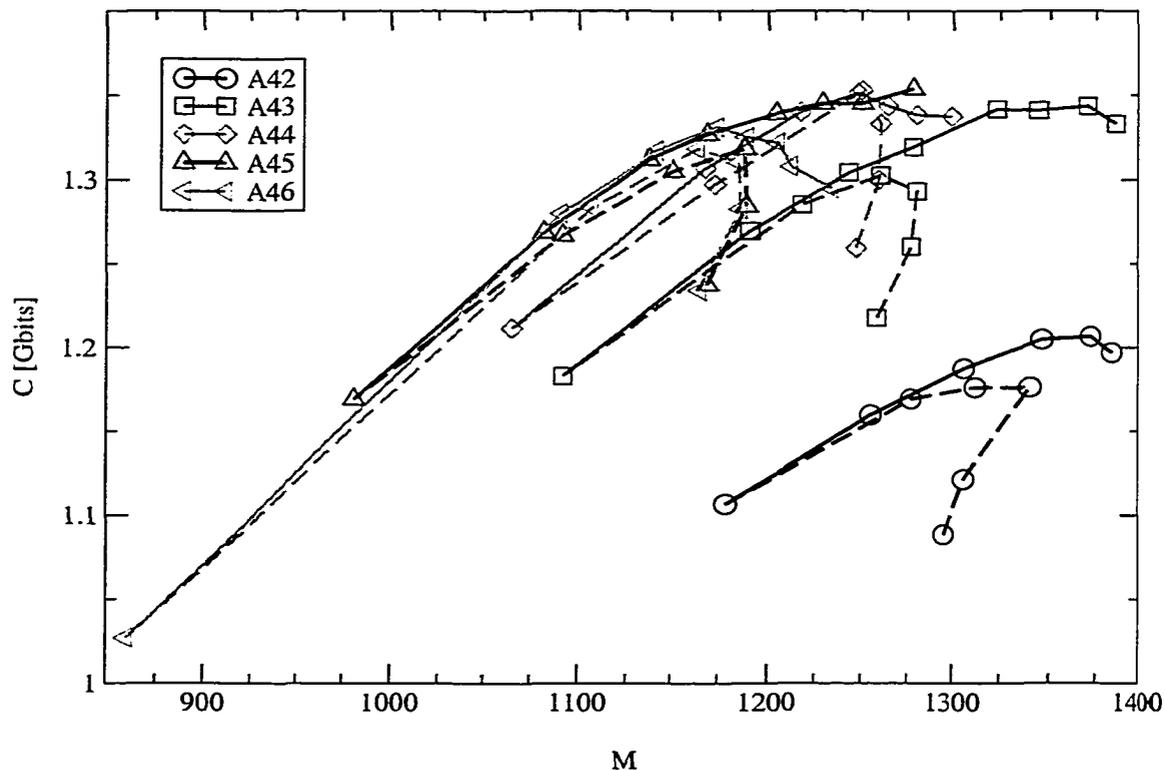


Figure 6.11: Extrapolated reverse code performance for XA4 sets. The solid curve represents the type I code and the dashed curve demarks the type II code.

plots the raw pixel-error rate (PER) versus the decoded user BER for both the theoretical and experimental results. The strong agreement between the output BER for all input PERs indicates that the theoretical simulations can predict, with high accuracy, the experimental output BER from the raw pixel-error rate. We will use this relation when estimating the extrapolated capacity of the reverse coding schemes applied to the experimental data sets.

For each specific data set we apply the type I or type II reverse code with varying levels of error correction (typically $t = 2$ up to $t = 10$), sweeping out a

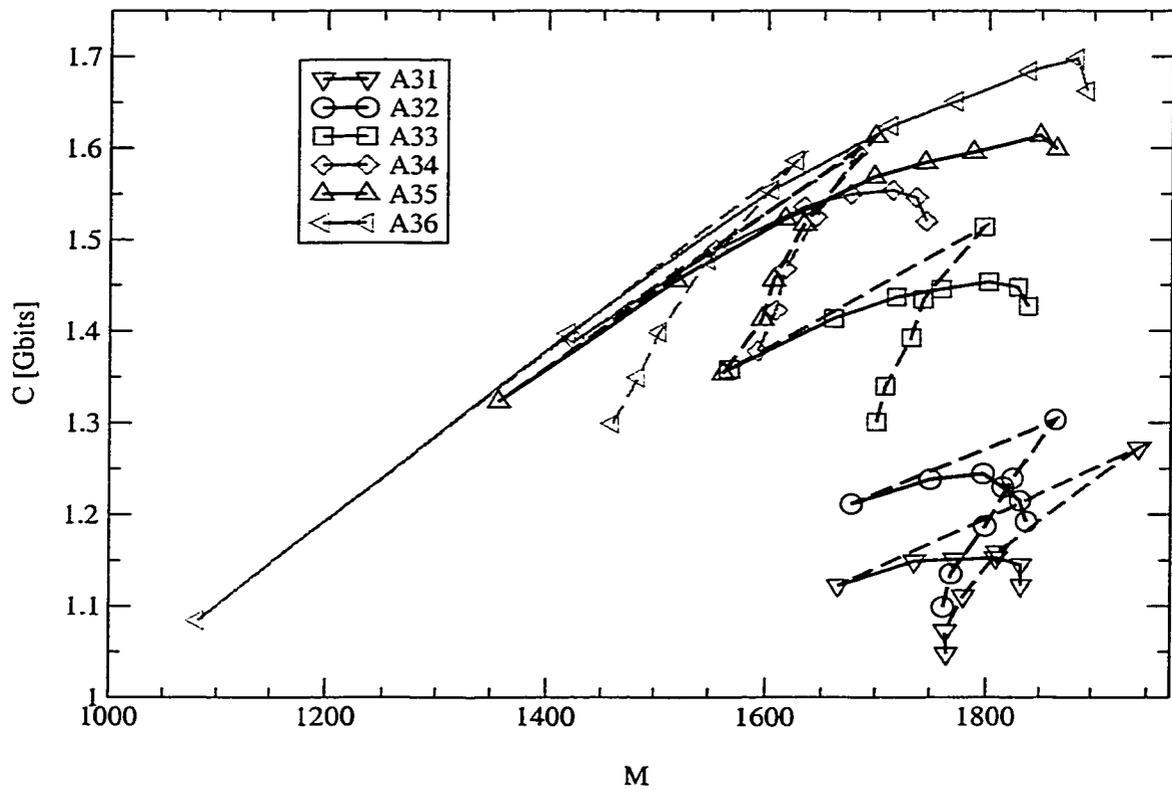


Figure 6.12: Extrapolated reverse code performance for XA3 sets. The solid curve represents the type I code and the dashed curve demarks the type II code.

curve on the capacity versus page count graph. As with all our previous graphs of this sort, the upper left corner is the best location, representing the highest storage capacity available in the fewest number of pages (faster data rate). Figures 6.11 and 6.12 shows the capacity for the XA4 and XA3 sets for both the type I and type II codes. The bottom left point of each curve represents the $t = 0$ (unprotected) case. The first few values of t quickly recover much of the capacity sacrificed to error propagation. Also notice that the type II codes perform in general worse than the type I codes. This indicates that sparse enumeration encoding of the parity, while gaining capacity due to the ML sort detection and the use of sparse priors, loses more capacity due to lower code rate (more parity pixels per codeword) and error propagation.

6.7 Summary

We see approximately a 11% improvement in the estimated experimental capacity using the 3-ary level shortened enumerative permutation code over local threshold detection on binary pages containing on average equal numbers of “on” and “off” pixels. If reverse coding is used, the effect of error propagation can be partially mitigated resulting in an estimated capacity increase of 35%. Note that 4-ary codes achieve a maximum gain of 8%. Theoretical results indicate that the 4-ary codes should outperform the 3-ary codes. Due to the increased difficulty

and the implicit broadening of the pixel distribution due to the recording process, 4-ary valued pixels offer less capacity than 3-ary pixels. This is in agreement with earlier work on non-binary pixels for use in VHMs.⁷⁰

In general, sparse enumerative permutation codes can increase the VHM storage capacity, though in practice, due to other important noise sources (holographic recording width, cross-erasure, photovoltaic) it may not necessarily be possible to achieve the predicted 50% improvement as suggested in chap. 5 operating in the thermal noise limit.

Error propagation reduces the effectiveness of the sparse enumeration codes, but we saw that applying an inner RS code to protect the modulation codeword could be used to recover a good amount of the lost performance at the expense of increased decoder complexity.

Notes This work has not yet been published.

CHAPTER 7

CONCLUSIONS

This dissertation has only touched upon a number of the important issues regarding volume holographic data storage. HDS promises an enormous storage capacity that can be randomly accessed at an extremely fast aggregate data rate. The system takes advantage of the inherent parallelism present in optics as well as the optical material properties allowing the recording of very high frequency spatial fringe patterns. The optoelectronic interfaces to the system pose significant technical challenges, some of which we concentrated on in this dissertation.

In Chapter 2 we designed a low-complexity CCD focal plane processor to perform local detection of optical data, correcting for the blur (interpixel interference) and a number of fixed-pattern alignment errors present in the imaging system. The design was based on a compact computation element at each pixel which communicated its current data estimation to the neighboring pixels. Because blur and optical misalignment errors are in nature a local phenomenon, we expect a local solution to be able to recover a significant amount of the performance that would be lost with a conventional design. The very blurred page scenario is appropriate in the high-density limit of optical storage.

Though photorefractive memories are a coherent optical system, we analyzed the PDFE performance in both the coherent and incoherent limits. We compared the results to other competitive decision techniques such as the Wiener filter, local thresholding, and a 1:2 modulation code. We found that the 3×3 Wiener filter offers a 107% increase in capacity as compared with simple threshold detection, when tested on the sinc^2 Sparrow resolved channel. We also found that the non-linear PDFE algorithm significantly improves on the Wiener filter estimate of the data page using the same 3×3 or 5×5 local connectivity. For the sinc^2 resolved channel, PDFE achieves an additional 9.7% capacity gain over the Wiener filter and on the most difficult ISI channel considered (Gaussian 25% beyond resolution channel), the 3×3 PDFE yields a 162% capacity improvement over the 3×3 Wiener filter for a SLM fill factor of 64%. We also examined the degradation of performance as systematic errors such as rotation, shift, and magnification were introduced and discovered that PDFE can tolerate up to a 0.48% magnification error, 0.28 degree rotation error, and 0.29 pixel shift error (tolerance being determined as the point where the performance drops to 90% of its value for no systematic errors). Tolerancing against prior channel knowledge was conducted at an operating point with 0.1% magnification error, 0.05 pixel shift, and 0.05 degrees of rotation. Notable results included a tolerance of up to 48% error in channel blur width, 0.29% magnification error, 0.17 degrees rotation

error, and a pixel shift of 0.17. Tolerance in this case is defined as the point where the performance drops to 90% of its value at the tolerance operating point.

When we considered the coherent optical system, it was necessary to use a quadratic version of the algorithm (QPDFE) in order to handle the quadratic ISI terms. In this case the performance of the QPDFE algorithm was demonstrated on the coherent channel and significant capacity improvements were found. For example, we recovered an additional 17.3% capacity by using the 7×7 QPDFE over the 7×7 Wiener filter. We found similar tolerances to systematic errors including a 0.38% tolerance to magnification error, 0.21 degree tolerance to rotation, and 0.2 pixel tolerance to shift.

A preliminary analysis of the VLSI resources necessary to realize the PDFE design showed that with current VLSI technology we could construct a 1024×1024 CCD read out row by row. The design would correct for a 3×3 pixel optical channel using three iterations of the PDFE algorithm and operate at 1000 frames per seconds (1 Mbit per second). A proof-of-concept fabrication in a MOSIS $2\text{-}\mu\text{m}$ analog process of a 3×3 PDFE design with 81 pixels was also described.

After our thorough consideration of the effect of the interference between pixels in a retrieved data page, we examined the case when the interference was between the holographic pages instead. Chapter 3 introduced the technique of reference beam apodization in order to control and/or eliminate the page-to-page cross talk

noise (IPI). Simulations showed that the crystal absorption will corrupt the Bragg selectivity function to produce a large amount of cross talk between neighboring pages. Apodization eliminated the absorption induced noise by correcting the modulation depth of the recorded gratings. In this case we saw the IPI noise was reduced providing a noise-to-signal ratio between 0.1 and 0.01. Jitter in the reference beam control was also examined and we found that even when the standard deviation of the jitter is around 4% of a Bragg null, a significant IPI noise is encountered. Apodization can be used to broaden the Bragg nulls and thus provide an increased tolerance to reference beam positioning errors.

In addition, we examined the relationship between apodization and the holographic system performance metric, $M/\#$. It was found that while apodization lowers the $M/\#$, it does so only slightly with the correct choice of apodizer. Absorption correction can be achieved at a very minor cost to the $M/\#$, improving high multiplexing-density storage.

Chapter 4 introduced the concept of data-page sparsity, which leveraged the asymmetric cost of storing an ON or bright pixel in a hologram as opposed to recording an OFF or dark pixel. Fewer ON pixels per page allowed more pages to be stored and retrieved at the same raw bit-error rate. For binary pixels, it was found that approximately 1/4 of a data page should be ON and 3/4 OFF. This case offered a information-theoretic capacity improvement of 15%. An experiment

verified the proposed scheme and showed a 16% improvement as compared to storage of normal pages which had a comparable number of OFF and ON pixels. An enumeration code solution was presented to act as a modulation encoder and decoder for the VHM system. A maximum-likelihood sort-based (thresholdless) decoding scheme was discussed and a very low-complexity encoding and decoding algorithm was presented.

The following chapter discussed extensions to the sparsity for arbitrary graylevel pixels. As more graylevels were considered, the capacity improvement of data page sparsity was seen to asymptotically approach 50%. In addition, we proved the ML property of the low-complexity modulation encoder and decoder.

Experiments conducted jointly with the IBM Almaden research center provided a large amount of data in which we could verify the expected performance of our enumeration-based codes. We saw that 3-ary sparse codes offered an 11% capacity improvement over uncoded binary pages. The practical performance suffered largely due to the effect of error propagation in the decoding of the enumeration codewords. Poor natural distance properties of the code led to a giant error magnification. A two-pixel error in the codeword would, upon decoding, result in a number of bit-errors proportional to the block length. We discussed reverse coding as a technique to counter error propagation. Applying it on the

experimental data showed excellent improvements. The estimated capacity of the memory increased by 35% using a reverse encoded 3-ary enumeration code.

In summary, we have addressed the volume holographic memory system along a number of different fronts. From the interpixel interference limit (blur) to the interpage interference limit. We examined closely the thermal noise limit and successfully applied channel coding to enhance the system performance. A new ML low-complexity encoding/decoding algorithm was developed to address some of the more practical issues.

REFERENCES

- [1] D. Gabor, "A new microscopic principle," *Nature*, vol. 161, p. 777, 1948.
- [2] E. N. Leith and J. Upatnieks, "Reconstructed wavefronts and communication theory," *J. Opt. Soc. Amer.*, vol. 52, p. 1123, 1962.
- [3] E. N. Leith and J. Upatnieks, "Wavefront reconstruction with continuous-tone objects," *J. Opt. Soc. Amer.*, vol. 53, p. 1377, 1963.
- [4] E. N. Leith and J. Upatnieks, "Wavefront reconstruction with diffused illumination and three-dimensional objects," *J. Opt. Soc. Amer.*, vol. 54, p. 1295, 1964.
- [5] P. J. van Heerden, "A new optical method of storing and retrieving information," *Appl. Opt.*, vol. 2, no. 4, pp. 387–392, 1963.
- [6] P. J. van Heerden, "Theory of optical information storage in solids," *Appl. Opt.*, vol. 2, no. 4, pp. 393–401, 1963.
- [7] D. Gabor and G. W. Stroke, "The theory of deep holograms," *Proc. Roy Soc. A.*, vol. 304, pp. 275–289, 1968.
- [8] H. Kogelnik, "Coupled wave theory for thick hologram gratings," *Bell Sys. Tech. J.*, vol. 48, no. 9, pp. 2909–2947, 1969.
- [9] R. M. Shelby, J. A. Hoffnagle, G. W. Burr, C. M. Jefferson, M.-P. Bernal, H. Coufal, R. K. Grygier, H. Günther, R. M. Macfarlane, and G. T. Sincerox, "Pixel-matched holographic data storage with megabit pages," *Opt. Lett.*, vol. 22, pp. 1509–1511, October 1997.
- [10] I. McMichael, W. Christian, D. Pletcher, T. Y. Chang, and J. H. Hong, "Compact holographic storage demonstrator with rapid access," *Appl. Opt.*, vol. 35, pp. 2375–2379, May 1996.
- [11] M. A. Neifeld and S. K. Sridharan, "Parallel error correction using spectral reed-solomon codes," *J. Opt. Comm.*, vol. 18, no. 4, pp. 144–150, 1997.

- [12] M. Aguilar, M. Carrascosa, and F. Agulló-López, "Optimization of selective erasure in photorefractive memories," *J. Opt. Soc. Amer. B*, vol. 14, pp. 110–115, January 1997.
- [13] M. A. Neifeld and M. McDonald, "Technique for controlling cross-talk noise in volume holography," *Opt. Lett.*, vol. 21, pp. 1298–1300, August 1996.
- [14] M. A. Neifeld and M. McDonald, "Optical design for page access to volume optical media," *Appl. Opt.*, vol. 35, pp. 2418–2430, May 1996.
- [15] G. W. Burr, F. H. Mok, and D. Psaltis, "Angle and space multiplexed holographic storage using the 90° geometry," *Opt. Commun.*, vol. 117, pp. 49–55, May 1995.
- [16] M. A. Neifeld and J. D. Hayes, "Error correction schemes for volume optical memories," *Appl. Opt.*, vol. 34, pp. 8183–8191, December 1995.
- [17] F. Dai and C. Gu, "Effect of Gaussian references on cross-talk noise reduction in volume holographic memory," *Opt. Lett.*, vol. 22, pp. 1802–1804, December 1997.
- [18] M. A. Neifeld and M. McDonald, "Error correction for increasing the usable capacity of photorefractive memories," *Opt. Lett.*, vol. 19, pp. 1483–1485, September 1994.
- [19] F. H. Mok, "Angle-multiplexed storage of 5000 holograms in lithium niobate," *Opt. Lett.*, vol. 18, pp. 915–917, 1993.
- [20] W.-C. Chou, *Coding applications in volume holographic memory systems*. PhD thesis, University of Arizona, 1999.
- [21] G. D. Forney, "The viterbi algorithm," in *Proc. IEEE*, vol. 61, pp. 268–78, 1973.
- [22] K. M. Chugg, "Performance of optimal digital page detection in a two-dimensional isi/AWGN channel," in *Conference Record of Thirtieth Asilomar Conference on Signals, Systems and Computers* (A. Singh, ed.), vol. 2, (Los Alamitos), pp. 958–962, IEEE Comput. Soc. Press, 1997.

- [23] J. F. Heanue, K. Gürkan, and L. Hesselink, "Signal detection for page-access optical memories with intersymbol interference," *Appl. Opt.*, vol. 35, pp. 2431–2438, May 1996.
- [24] C. L. Miller, B. R. Hunt, M. A. Neifeld, and M. W. Marcellin, "Bi-level image reconstructions via 2-d viterbi search," in *IEEE Intl. Conf. Image Processing*, (Santa Barbara, CA), October 1997.
- [25] X. Chen, K. M. Chugg, and M. A. Neifeld, "Near-optimal parallel distributed data detection for page-oriented optical memories," *IEEE. J. of Sel. Topics in Quantum Elec.*, vol. 4, pp. 866–879, 1997.
- [26] D. Messerschmitt, "A geometric theory of intersymbol interference: part i," *Bell Syst. Tech. Journal*, vol. 52, pp. 1483–1519, 1973.
- [27] J. G. Proakis, *Digital Communications*. McGraw-Hill, Inc., 3 ed., 1995.
- [28] M. A. Neifeld, K. M. Chugg, and B. M. King, "Parallel data detection in page-oriented optical memory," *Opt. Lett.*, vol. 21, pp. 1481–1483, September 1996.
- [29] B. M. King and M. A. Neifeld, "Parallel detection algorithm for page-oriented optical memories," *Appl. Opt.*, vol. 37, pp. 6275–6298, September 1998.
- [30] M. A. Neifeld and W.-C. Chou, "Information theoretic limits to the capacity of volume holographic optical memory," *Appl. Opt.*, vol. 36, pp. 514–517, January 1997.
- [31] F. H. Mok, G. W. Burr, and D. Psaltis, "System metric for holographic memory systems," *Opt. Lett.*, vol. 21, pp. 896–898, June 1996.
- [32] J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Channel codes for digital holographic data storage," *J. Opt. Soc. Amer. A*, vol. 12, pp. 2432–2439, November 1995.
- [33] G. W. Burr, J. Ashley, H. Coufal, R. K. Grygier, J. A. Hoffnagle, C. M. Jefferson, and B. Marcus, "Modulation coding for pixel-matched holographic data storage," *Opt. Lett.*, vol. 22, pp. 639–641, May 1997.

- [34] E. A. Lee and D. G. Messerschmitt, *Digital communications*. Kluwer Academic Publishers, 1988.
- [35] D. Brady and D. Psaltis, "Control of volume holograms," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1167–1182, July 1992.
- [36] S. S. Haykin, *Neural networks : a comprehensive foundation*. Macmillan, 1994.
- [37] S. Campbell, X. Yi, and P. Yeh, "Hybrid sparse-wavelength angle-multiplexed optical data storage system," *Opt. Lett.*, vol. 19, pp. 2161–2163, December 1994.
- [38] J. Heanue, M. Bashaw, and L. Hesselink, "Recall of linear combinations of stored data pages based on phase-code multiplexing in volume holography," *Opt. Lett.*, vol. 19, pp. 1079–1081, July 1994.
- [39] E. Chuang and D. Psaltis, "Storage of 1000 holograms with use of a dual-wavelength method," *Appl. Opt.*, vol. 36, pp. 8445–8454, November 1997.
- [40] G. A. Rakuljic, V. Leyva, and A. Yariv, "Optical data storage by using orthogonal wavelength-multiplexed volume holograms," *Opt. Lett.*, vol. 17, pp. 1471–1473, October 1992.
- [41] V. Markov, J. Millerd, J. Trolinger, M. Norrie, J. Dowaie, and D. Timucin, "Multilayer volume holographic optical memory," *Opt. Lett.*, vol. 24, pp. 265–267, February 1999.
- [42] G. W. Burr, *Volume Holographic Storage Using the 90° Geometry*. PhD thesis, California Institute of Technology, 1996.
- [43] H. J. Coufal, D. Psaltis, and G. T. Sincerbox, eds., *Holographic Data Storage*. Springer-Verlag, 2000.
- [44] P. Yeh and A. Yariv, *Optical waves in crystals*. John Wiley & Sons, Inc., 1993.

- [45] M. A. Neifeld, M. McDonald, and B. King, "Reference beam apodization in angularly multiplexed volume holography," in *Proc. SPIE*, vol. 3137, pp. 101–106, 1997.
- [46] M. E. McDonald, *Optical engineering for parallel data storage*. PhD thesis, University of Arizona, 1998.
- [47] B. M. King, M. A. Neifeld, and X.-W. Chen, "Improved multiplexing density in volume holographic memories through beam apodization," in *Optics in Computing 2000* (R. A. Lessard and T. V. Galstian, eds.), vol. 4089, pp. 132–137, June 2000.
- [48] C. Gu, J. Hong, I. McMichael, R. Saxena, and F. Mok, "Cross-talk-limited storage capacity of volume holographic memory," *J. Opt. Soc. Amer. A*, vol. 9, pp. 1978–1983, November 1992.
- [49] Y. Taketomi, J. E. Ford, H. Sasaki, J. Ma, Y. Fainman, and S. H. Lee, "Incremental recording for photorefractive hologram multiplexing," *Opt. Lett.*, vol. 16, pp. 1774–1776, November 1991.
- [50] A. Pu, K. Curtis, and D. Psaltis, "Exposure schedule for multiplexing holograms in photopolymer films," *Opt. Eng.*, vol. 35, pp. 2824–2829, October 1996.
- [51] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1 ed., 1991.
- [52] T. S. Cover, "Enumerative source encoding," *IEEE Trans. Inform. Theory*, vol. 19, pp. 73–77, January 1973.
- [53] D. Slepian, "Permutation modulation," *Proc. IEEE*, vol. 53, pp. 228–236, March 1965.
- [54] G. W. Burr, W.-C. Chou, M. A. Neifeld, H. Coufal, J. A. Hoffnagle, and C. M. Jefferson, "Experimental evaluation of user capacity in holographic data-storage systems," *Appl. Opt.*, vol. 37, pp. 5431–5443, August 1998.

- [55] M. Bernal, G. W. Burr, H. Coufal, and M. Quintanilla, "Balancing inter-pixel cross talk and detector noise to optimize areal density in holographic storage systems," *Appl. Opt.*, vol. 37, pp. 5377–5385, August 1998.
- [56] B. M. King and M. A. Neifeld, "Sparse modulation coding for increased capacity in volume holographic storage," *Appl. Opt.*, vol. 39, December 2000. In pre-production.
- [57] S. Datta and S. W. McLaughlin, "An enumerative method for runlength-limited codes: Permutation codes," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2199–2204, September 1999.
- [58] M. Mansuripur, "Enumerative modulation coding with arbitrary constraints and post-modulation error correction coding for data storage systems," *Proc. SPIE*, vol. 1499, pp. 72–86, 1991.
- [59] H. Hollman and K. A. Immink, "Runlength-limited sequences," *IEEE Trans. Inform. Theory*, vol. 37, no. 3, pp. 913–918, 1991.
- [60] K. Immink, "Error detecting runlength-limited sequences," *Proc. Eighth International Conference on Video, Audio, and Data Recording*, pp. 176–182, 1990.
- [61] K. A. Immink, "Runlength-limited sequences," *Proc. IEEE*, vol. 78, no. 11, pp. 1745–1759, 1990.
- [62] A. van Wijngaarden and K. Immink, "On the construction of constrained codes employing sequence replacement techniques," *Proc. IEEE International Symposium on Information Theory*, p. 144, 1997.
- [63] K. Immink and A. Janssen, "Error propagation assessment of enumerative coding schemes," *Proc. IEEE International Conference on Communications*, vol. 2, pp. 961–963, 1998.
- [64] J. L. Fan and A. Calderbank, "A modified concatenated coding scheme, with applications to magnetic data storage," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1565–1574, July 1998.

- [65] S. Campbell, S.-H. Lin, X. Yi, and P. Yeh, "Absorption effects in photorefractive volume-holographic memory systems. i beam depletion, ii material heating," *J. Opt. Soc. Amer. B*, vol. 13, pp. 2209–2228, October 1996.
- [66] A. A. Freschi, P. M. Garcia, I. Rasnik, and J. Frejlich, "Avoiding hologram bending in photorefractive crystals," *Opt. Lett.*, vol. 21, pp. 152–154, January 1996.
- [67] C. Gu and J. Hong, "Noise gratings formed during the multiple exposure schedule in photorefractive media," *Opt. Comm.*, vol. 93, pp. 213–218, October 1992.
- [68] E. S. Maniloff and K. M. Johnson, "Effects of scattering on the dynamics of holographic recording and erasure in photorefractive lithium niobate," *J. of Appl. Phys.*, vol. 73, pp. 541–547, January 1993.
- [69] F. Vachss, I. McMichael, and J. Hong, "Cross-erasure noise in high-density holographic-storage systems," *J. Opt. Soc. Amer. B*, vol. 14, pp. 1187–1198, May 1997.
- [70] G. W. Burr, G. Barking, H. Coufal, J. A. Hoffnagle, C. M. Jefferson, and M. A. Neifeld, "Gray-scale data pages for digital holographic data storage," *Opt. Lett.*, vol. 23, pp. 1218–1220, August 1998.
- [71] W. Bliss, "Circuitry for performing error correction calculations on baseband encoded data to eliminate error propagation," *IBM Tech. Discl. Bull.*, vol. 23, pp. 4633–4634, 1981.
- [72] B. M. King and M. A. Neifeld, "Low-complexity maximum-likelihood decoding of shortened enumerative permutation codes for holographic storage," *IEEE Jour. Sel. Areas Comm.*, 2001. Accepted, in pre-production.
- [73] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics*. John Wiley and Sons, 1991.
- [74] H. Coufal, "Solid information," *Nature*, vol. 393, pp. 628–629, June 1998.
- [75] D. Psaltis and G. W. Burr, "Holographic data storage," *Computer*, pp. 52–60, February 1998.

- [76] D. Psaltis and F. Mok, "Holographic memories," *Scientific American*, pp. 70–76, November 1995.
- [77] J. F. Heanue, M. C. Bashaw, and L. Hesselink *Science*, vol. 265, pp. 749–752, August 1994.
- [78] P. J. van Heerden, "A new optical method of storing and retrieving information," *Applied Optics*, vol. 2, pp. 387–392, April 1963.
- [79] D. Brady and D. Psaltis, "Information capacity of 3-d holographic data storage," *Optical and Quantum Electronics*, vol. 25, pp. S597–S610, 1993.
- [80] L. Hesselink and M. C. Bashaw, "Optical memories implemented with photorefractive media," *Optical and Quantum Electronics*, vol. 25, pp. S611–S661, 1993.
- [81] Y. N. Denisyuk, "Photographic reconstruction of the optical properties of an object in its own scattered radiation field," *Soviet Physics–Doklady*, vol. 7, no. 6, pp. 543–545, 1962.
- [82] L. Solymar and D. J. Cooke, *Volume Holography and Volume Gratings*. Academic Press, 1981.
- [83] P. S. J. Russell, "Optical volume holography," *Physics Reports*, vol. 71, no. 4, pp. 209–312, 1981.
- [84] P. Hariharan, *Optical Holography*. Cambridge University Press, 1984.
- [85] G. Saxby, *Practical Holography*. Cambridge University Press, second ed., 1994.
- [86] J. W. Goodman, *Introduction to Fourier Optics*. McGraw-Hill Companies, Inc., second ed., 1996.
- [87] J. W. Goodman, *Statistical Optics*. John Wiley & Sons, Inc, 1985.
- [88] C. C. Davis, *Lasers and Electro-Optics: Fundamentals and Engineering*. Cambridge University Press, 1996.

- [89] P. Yeh, *Introduction to photorefractive nonlinear optics*. John Wiley & Sons, Inc., 1993.
- [90] K. Buse, A. Adibi, and D. Psaltis, "Non-volatile holographic storage in doubly doped lithium niobate crystals," *Nature*, vol. 393, pp. 665–668, June 1998.
- [91] K. Buse, "Light-induced charge transport processes in photorefractive crystals I: Models and experimental methods," *Applied Physics B*, vol. 64, pp. 273–291, 1997.
- [92] K. Buse, "Light-induced charge transport processes in photorefractive crystals II: Materials," *Applied Physics B*, vol. 64, pp. 391–407, 1997.
- [93] H. Guenther, G. Wittmann, R. M. Macfarlane, and R. R. Neurgaonkar, "Intensity dependence and white-light gating of two-color photorefractive gratings in $LiNbO_3$," *Optics Letters*, vol. 22, no. 17, pp. 1305–1307, 1997.
- [94] P. Günter, "Holography, coherent light amplification and optical phase conjugation with photorefractive materials," *Physics Reports*, vol. 93, no. 4, pp. 199–299, 1982.
- [95] M. M. Wang, S. C. Esener, F. B. McCormick, I. Çokgör, A. S. Dvornikov, and P. M. Rentzepis, "Experimental characterization of a two-photon memory," *Optics Letters*, vol. 22, pp. 558–560, April 1997.
- [96] L. Dhar, A. Hale, H. E. Katz, M. L. Schilling, M. G. Schnoes, and F. C. Schilling, "Recording media that exhibit high dynamic range for digital holographic data storage," *Optics Letters*, vol. 24, pp. 487–489, April 1999.
- [97] W. E. Moerner, A. Grunnet-Jepsen, C. L. Thompson, M. S. Bratcher, and R. J. Twieg, "Recent advances in photorefractive polymer materials," in *Nonlinear Optical Properties of Organic Materials X* (M. G. Kuzyk, ed.), vol. 3147, pp. 84–94, SPIE, 30 July–1 August 1997.
- [98] J. D. Downie, D. A. Timuçin, D. T. Smithey, and M. Crew, "Long holographic lifetimes in bacteriorhodopsin films," *Optics Letters*, vol. 23, pp. 730–732, May 1998.

- [99] H.-Y. S. Li and D. Psaltis, "Three-dimensional holographic disks," *Applied Optics*, vol. 33, pp. 3764–3774, June 1994.
- [100] J. H. Hong, I. McMichael, T. Y. Chang, W. Christian, and E. G. Paek, "Volume holographic memory systems: Techniques and architectures," *Applied Optics*, vol. 34, pp. 2193–2203, August 1995.
- [101] K. Curtis, A. Pu, and D. Psaltis, "Method for holographic storage using peristrophic multiplexing," *Optics Letters*, vol. 19, pp. 993–994, July 1994.
- [102] D. Psaltis, M. Levene, A. Pu, and G. Barbastathis, "Holographic storage using shift multiplexing," *Optics Letters*, vol. 20, pp. 782–784, April 1995.
- [103] J. T. LaMacchia and D. L. White, "Coded multiple exposure holograms," *Appl. Opt.*, vol. 7, no. 1, pp. 91–94, 1968.
- [104] T. F. Krile, M. O. Hagler, W. D. Redus, and J. F. Walkup, "Multiplex holography with chirp-modulated binary phase-coded reference-beam masks," *Applied Optics*, vol. 18, pp. 52–56, January 1979.
- [105] C. Denz, G. Pauliat, and G. Roosen, "Volume hologram multiplexing using a deterministic phase encoding method," *Optics Communications*, vol. 85, pp. 171–176, September 1991.
- [106] C. Denz, G. Pauliat, G. Roosen, and T. Tschudi, "Potentialities and limitations of hologram multiplexing by using the phase-encoding technique," *Appl. Opt.*, vol. 31, pp. 5700–5705, September 1992.
- [107] K. Kamra, A. Kumar, and K. Singh, "Novel optical photorefractive storage-retrieval system using speckle coding technique in beam-fanning geometry," *J. of Mod. Opt.*, vol. 43, pp. 365–371, February 1996.
- [108] I. Lorgeré, F. Grelet, M. Ratsep, M. Tian, and J. L. L. Gouët, "Spectral phase encoding for data storage and addressing," *Journal of the Optical Society of America B: Optical Physics*, vol. 13, pp. 2229–2234, October 1996.

- [109] S. Tao, Z. H. Song, D. R. Selviah, and J. E. Midwinter, "Spatioangular-multiplexing scheme for dense holographic storage," *Applied Optics*, vol. 34, pp. 6729–6737, October 1995.
- [110] L. d'Auria, J. P. Huignard, C. Slezak, and E. Spitz, "Experimental holographic read-write memory using 3-d storage," *Applied Optics*, vol. 13, pp. 808–818, April 1974.
- [111] E. S. Maniloff, S. B. Altner, S. Bernet, F. R. Graf, A. Renn, and U. P. Wild, "Recording of 6000 holograms by use of spectral hole burning," *Applied Optics*, vol. 34, pp. 4140–4148, July 1995.
- [112] M.-P. Bernal, H. Coufal, R. K. Grygier, J. A. Hoffnagle, C. M. Jefferson, R. M. Macfarlane, R. M. Shelby, G. T. Sincerbox, P. Wimmer, and G. Wittmann, "A precision tester for studies of holographic optical storage materials and recording physics," *Applied Optics*, vol. 35, pp. 2360–2374, May 1996.
- [113] J.-J. P. Drolet, E. Chuang, G. Barbastathis, and D. Psaltis, "Compact, integrated dynamic holographic memory with refreshed holograms," *Optics Letters*, vol. 22, pp. 552–554, April 1997.
- [114] D. Lande, S. S. Orlov, A. Akella, L. Hesselink, and R. R. Neurgaonkar, "Digital holographic storage system incorporating optical fixing," *Optics Letters*, vol. 22, pp. 1722–1724, November 1997.
- [115] L. Dhar, K. Curtis, M. Tackitt, M. Schilling, S. Campbell, W. Wilson, A. Hill, C. Boyd, N. Levinos, and A. Harris, "Holographic storage of multiple high-capacity digital data pages in thick photopolymer systems," *Optics Letters*, vol. 23, pp. 1710–1712, November 1998.
- [116] C. Denz, K.-O. Müller, T. Heimann, and T. Tschudi, "Volume holographic storage demonstrator based on phased-coded multiplexing," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 4, pp. 832–839, September 1998.
- [117] H. J. Eichler, P. Kuemmel, S. Orlic, and A. Wappelt, "High-density disk storage by multiplexed microholograms," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 4, pp. 840–848, September 1998.

- [118] G. W. Burr, *Documentation for the DEMON System*. IBM Almaden Research Center, 0.3 ed., May 1997.
- [119] R. E. Blahut, *Theory and Practice of Error Control Codes*. Addison-Wesley Publishing, 1983.
- [120] S. Lin and C. Daniel J. Jr., *Error Control Coding: Fundamentals and Applications*. Prentice-Hall Inc., 1983.
- [121] S. A. Domrovskii, "Effectiveness of using error-correcting codes in holographic storage systems," *Avtometriya*, no. 2, pp. 57–61, 1989.
- [122] B. J. Goertzen and P. A. Mitkas, "Error-correcting code for volume holographic storage of a relational database," *Opt. Lett.*, vol. 20, pp. 1655–1657, August 1995.
- [123] G. A. Betzos, J. F. Hutton, M. Porter, and P. A. Mitkas, "Evaluation of array codes for page-oriented optical memories," in *Optics in Computing: 1997 Technical Digest Series (Post-conference Edition)*, vol. 8, pp. 204–206, 1997.
- [124] B. J. Goertzen and P. A. Mitkas, "Volume holographic storage for large relational databases," *Optical Engineering*, vol. 35, pp. 1847–1853, July 1996.
- [125] W. Kawakami and K.-I. Kitayama, "Optical error-correction coding encoder and decoder: Design considerations," *Applied Optics*, vol. 34, pp. 5064–5073, August 1995.
- [126] T. K. Matsushima, T. Matsushima, and S. Hirasawa, "Parallel encoder and decoder architecture for cyclic codes," *IEICE Transactions on Fundamentals of Electronics Communication and Computer Science*, vol. E79-A, pp. 1313–1323, September 1996.
- [127] M. E. Schaffer and P. A. Mitkas, "Requirements and constraints for the design of smart photodetector arrays for page-oriented optical memories," *IEEE J. of Sel. Topics in Quantum Elec.*, vol. 4, no. 5, pp. 856–865, 1998.

- [128] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *SIAM Journal on Applied Mathematics*, vol. 8, pp. 300–304, 1960.
- [129] M. Blaum, J. Bruck, and A. Vardy, "Interleaving schemes for multidimensional cluster errors," *IEEE Transactions on Information Theory*, vol. 44, pp. 730–742, March 1998.
- [130] W.-C. Chou and M. A. Neifeld, "Interleaving and error correction in volume holographic memory systems," *Applied Optics*, vol. 37, pp. 6951–6968, October 1998.
- [131] B. H. Olson and S. C. Esener, "Partial response precoding for parallel-readout optical memories," *Opt. Lett.*, vol. 19, pp. 661–663, May 1994.
- [132] A. Vardy, M. Blaum, P. H. Siegel, and G. T. Sincerbox, "Conservative arrays: Multidimensional modulation codes for holographic recording," *IEEE Transactions on Information Theory*, vol. 42, pp. 227–229, January 1996.
- [133] V. Vadde and B. V. K. V. Kumar, "Channel estimation and intra-page equalization for digital volume holographic data storage," in *Proceedings of the SPIE*, vol. 3109, pp. 245–250, 1997.
- [134] G. W. Burr, G. Barking, H. Coufal, J. A. Hoffnagle, C. M. Jefferson, and M. A. Neifeld, "Gray-scale data pages for digital holographic data storage," *Optics Letters*, vol. 23, pp. 1218–1220, August 1997.
- [135] S. Tao, B. Tang, Y. Zhou, and L. Shen, "Quantitative study of the gray-scale fidelity of volume holographic images," *Applied Optics*, vol. 38, pp. 3767–3777, June 1999.
- [136] K. M. Chugg, X. Chen, and M. A. Neifeld, "Two-dimensional equalization in coherent and incoherent page-oriented optical memory," *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, vol. 16, pp. 549–562, March 1999.
- [137] M.-P. Bernal, G. W. Burr, H. Coufal, R. K. Grygier, J. A. Hoffnagle, C. M. Jefferson, E. Oesterschulze, R. M. Shelby, G. T. Sincerbox, and M. Quintanilla, "Effects of multilevel phase masks on interpixel cross talk in digital holographic storage," *Applied Optics*, vol. 36, pp. 3107–3115, May 1997.

- [138] K. Bløtekjaer, "Limitations on holographic storage capacity of photochromic and photorefractive media," *Appl. Opt.*, vol. 18, pp. 57–67, January 1979.
- [139] K. Rastani, "Storage capacity and cross talk in angularly multiplexed holograms: two case studies," *Appl. Opt.*, vol. 32, pp. 3772–3778, July 1993.
- [140] H. Zhou, F. Zhao, and F. T. S. Yu, "Effects of recording-erasure dynamics of storage capacity of a wavelength-multiplexed reflection-type photorefractive hologram," *Appl. Opt.*, vol. 33, pp. 4339–4344, July 1994.
- [141] S. V. Miridonov, A. V. Khomenko, D. Tentori, and A. A. Kamshilin, "Information capacity of holograms in photorefractive crystals," *Opt. Lett.*, vol. 19, pp. 502–504, April 1994.
- [142] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.
- [143] C. E. Shannon, "Communication in the presence of noise," *Proceedings of the I.R.E.*, vol. 37, pp. 10–21, January 1949.
- [144] R. E. Blahut, *Principles and Practice of Information Theory*. Addison-Wesley, 1987.
- [145] C. Gu, F. Dai, and J. Hong, "Statistics of both optical and electrical noise in digital volume holographic data storage," *Electronic Letters*, vol. 32, no. 15, pp. 1400–1402, 1996.
- [146] D. Psaltis, D. Brady, and K. Wagner, "Adaptive optical networks using photorefractive crystals," *Applied Optics*, vol. 27, no. 9, pp. 1752–1759, 1988.
- [147] J. G. David Forney, "Coset codes—part i: Introduction and geometrical classification," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1123–1151, September 1988.
- [148] J. G. David Forney, "Coset codes—part ii: Binary lattices and related codes," *IEEE Trans. Inform. Theory*, vol. 34, pp. 1152–1187, September 1988.

- [149] T. R. Fischer and J. Pan, "Enumeration encoding and decoding algorithms for pyramid cubic lattice and trellis codes," *IEEE Trans. Inform. Theory*, vol. 41, pp. 2056–2061, November 1995.
- [150] S. B. Wicker and V. K. Bhargava, eds., *Reed-Solomon Codes and Their Applications*. IEEE Press, 1994.
- [151] G. Beenker and K. Immink, "A generalized method for encoding and decoding runlength-limited binary sequences," *IEEE Trans. Inform. Theory*, vol. 29, no. 5, pp. 1976–1984, 1983.
- [152] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 3rd ed., 1991.
- [153] J. Ashley and B. Marcus, "Two-dimensional lowpass filtering codes for holographic storage," *IEEE Trans. Comm.*, vol. 46, pp. 724–727, 1998.
- [154] V. Bhagavatula and M. Wong, "Crosstalk-limited signal-to-noise ratio resulting from phase errors in phase-multiplexed volume holographic storage," in *Optical society of America 1996 annual meeting*, p. 145, Optical Society of America, October 1996. session WZ6.
- [155] A. Vardy, M. Blaum, P. H. Siegel, and G. T. Sincerbox, "Conservative arrays: multidimensional modulation codes for holographic recording," *IEEE Trans. Inform. Theory*, vol. 42, pp. 227–230, January 1996.
- [156] A. A. Blok, "Effect of data coding methods in holographic memory on characteristics of reconstructed images of data pages," *Optoelectron. Instrum. Data Process.*, vol. 5, pp. 45–51, 1989.
- [157] C. B. Burckhardt, "Use of a random phase mask for the recording of fourier transform holograms of data masks," *Appl. Opt.*, vol. 9, pp. 695–700, March 1970.
- [158] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proc. IEEE*, vol. 66, pp. 51–83, January 1978.

- [159] H. C. Stankwitz, R. J. Dallaire, and J. R. Fienup, "Nonlinear apodization for sidelobe control in sar imagery," *IEEE Trans. Aero. and Elec. Sys*", vol. 31, no. 1, pp. 267–278, 1995.
- [160] G. W. Burr and B. Marcus, "Coding tradeoffs for high-density holographic data storage," in *Proc. SPIE*, vol. 3802, pp. 18–29, 1999.
- [161] T. Y. Chang, J. H. Hong, F. Vachss, and R. McGraw, "Studies of the dynamic range of photorefractive gratings in ferroelectric crystals," *J. Opt. Soc. Amer. B*, vol. 9, pp. 1744–1751, September 1992.
- [162] K. M. Chugg and X. Chen, "Efficient a-posteriori probability (app) and minimum sequence metric (msm) algorithms," *Submitted to IEEE Trans. Comm.*, August 1997.
- [163] T. N. Garrett and P. A. Mitkas, "Three-dimensional error correcting codes for volumetric optical memories," in *Proc. SPIE*, vol. 3468, pp. 116–124, 1998.
- [164] J. T. Gallo and C. M. Verber, "Model for the effects of material shrinkage on volume holograms," *Appl. Opt.*, vol. 33, pp. 6797–6804, October 1994.
- [165] J. W. Goodman, F. J. Leonberger, S. yuan Kung, and R. A. Athale, "optical interconnections for vlsi systems," *Proc. IEEE*, vol. 72, pp. 850–866, July 1984.
- [166] C. Gu and F. Dai, "Cross-talk noise reduction in volume holographic storage with an extended recording reference," *Opt. Lett.*, vol. 20, pp. 2336–2338, November 1995.
- [167] J. H. Hong, I. McMichael, T. Y. Chang, W. Christian, and E. G. Paek, "Volume holographic memory systems: techniques and architectures," *Opt. Eng.*, vol. 34, pp. 2193–2203, August 1995.
- [168] Y. Ichioka, T. Iwaki, and K. Matsuoka, "Optical information processing and beyond," *Proc. IEEE*, vol. 84, pp. 694–719, May 1996.
- [169] B. M. King and M. A. Neifeld, "Unequal a-priori probabilities for holographic storage," in *Proc. SPIE*, vol. 3802, pp. 40–45, July 1999.

- [170] T. F. Krile, R. J. M. II, J. F. Walkup, and M. O. Hagler, "Holographic representations of space-variant systems using phase-coded reference beams," *Appl. Opt.*, vol. 16, no. 12, pp. 3131–3135, 1977.
- [171] H. Lee, "Cross-talk effects in multiplexed volume holograms," *Opt. Lett.*, vol. 65, no. 6, pp. 874–876, 1989.
- [172] H.-Y. S. Li and D. Psaltis, "Alignment sensitivity of holographic tree-dimensional disks," *J. Opt. Soc. Amer. A*, vol. 12, pp. 1902–1912, September 1995.
- [173] M. Mansuripur and G. Sincerbox, "principles and techniques of optical data storage," *Proc. IEEE*, vol. 85, pp. 1780–1796, November 1997.
- [174] S. W. McLaughlin, A. R. Calderbank, R. Laroia, J. Gerpheide, and A. Jain, "Partial response modulation codes for electron trapping optical memory (etom)," in *Proc. SPIE*, vol. 2514, pp. 82–90, 1995.
- [175] D. Parsavand and M. K. Varanasi, "Rms bandwidth constrained signature waveforms that maximize the total capacity of pam-synchronous cdma channels," *IEEE Trans. Comm.*, vol. 44, pp. 65–75, January 1996.
- [176] J. D. Roberts, A. Ryley, D. M. Jones, and D. Burke, "Analysis of error-correction constraints in an optical disk," *Appl. Opt.*, vol. 35, pp. 3915–3924, July 1996.
- [177] S. Lin and D. J. Costello, *Error control coding: fundamentals and applications*. Prentice-Hall, 1983.
- [178] D. Z. Tsang, S. Dan L. Jr., A. Chu, and J. J. Lambert, "a technology for optical interconnections based on multichip integration," *Opt. Eng.*, vol. 25, pp. 1127–1131, October 1986.
- [179] A. Yariv, "Interpage and interpixel cross talk in orthogonal (wavelength-multiplexed) holograms," *Opt. Lett.*, vol. 18, pp. 652–654, April 1993.
- [180] X. Yi, P. Yeh, and C. Gu, "Cross-talk noise in volume holographic memory with spherical reference beams," *Opt. Lett.*, vol. 20, pp. 1812–1814, September 1995.