USC–SIPI REPORT #353

Multi-Dimensional Modulation Coding for Page-Oriented Optical Data Storage Systems

by

Dhawat Edward Pansatiankul

May 2002

Signal and Image Processing Institute
UNIVERSITY OF SOUTHERN CALIFORNIA
Department of Electrical Engineering-Systems
3740 McClintock Avenue, Room 400
Los Angeles, CA 90089-2564 U.S.A.

Acknowledgements

First and foremost, I would like to thank Prof. Alexander A. Sawchuk, my dissertation advisor, for giving me the opportunity to fulfill one of my goals. I am deeply grateful for his invaluable support and guidance not only in my research, but also in other aspects of my life. I have benefited greatly from his broad knowledge and experience. In addition, the wonderful environments that he provided have made my Ph.D. journey more enjoyable and memorable.

I would also like to thank Prof. B. Keith Jenkins and Prof. Wlodek Proskurowski for serving on both my dissertation and my qualifying examination committee, and Prof. Christos Kyriakakis and Prof. Robert M. Gagliardi for serving on my qualifying examination committee. Their suggestions and comments are very useful. I would in particular like to thank Prof. Wlodek Proskurowski for sharing his ideas in my first journal paper.

I am thankful to my former colleagues: Dr. Charles B. Kuznia, Dr. Jen-Ming Wu, Dr. Chih-Hao Chen, and Dr. Bogdan Hoanca for their assistance and encouragement during the early years of my Ph.D. study. A special thanks goes to Yunsong Huang, my officemate and encyclopedia, for his sincere friendship and helpful discussions in mathematics and physics. I also thank my current colleagues: Liping Zhang, Changki Min, Sunkwang Hong, Zahir Alpaslan, Nopparit Intharasombat, and Hyukjune Chung for their friendship and interesting conversation.

I gratefully thank Prof. P. Vijay Kumar, Prof. Antonio Ortega, and Prof. Shrikanth S. Narayanan for their valuable discussions and support. Thanks to Gloria Halfacre, Gerrielyn Ramos, and Diane Demetras for their administrative help. I acknowledge Dr. Allan G. Weber and Seth Scafani for their computer and network help. I would also like to thank Dr. Sanya Mitaim, Dr. Piyapong Thanyasrisung, Somkiat Kraikriangsri, Poonsuk Lohsoonthorn, and Wuttipong Kumwilaisak for their encouragement and discussions. Undoubtedly, there are other faculty, staff, and students of SIPI, IMSC, and EE-Systems at USC to whom I owe thanks, and I apologize for not mentioning them here.

Special debts are owed to my friends in Los Angeles, especially, Valailuk Limtanakool, Santi Jotibundhit, Ittipol Vatana, Thamapat Patram, and Watcharawadee Sawatdironnaphak; and also to my CUD'27 friends in Bangkok and elsewhere. Without them and their visits, my stay in Los Angeles would have been more difficult and less colorful.

Last, but certainly not least, I would like to thank my parents, Dr. Thawatchai and Dr. Boonchian Pansatiankul, and my lovely sister, Dr. Nalinee Pansatiankul, for their love and encouragement. I am deeply grateful for their total support during my years at USC. I may have never made it clear to them, but their endless support was crucial to my successful education.

Contents

List of Tables

List of Figures

Abstract

Page-oriented optical data storage (PODS) technology has great potential benefit for future computer applications that require tremendous data storage capacity and very high data transfer rates. As with any conventional data storage technology, PODS systems operating near their maximum capabilities need modulation coding to facilitate their recording and retrieving processes and, consequently, to improve their overall system performance. However, the modulation coding scheme designed for a specific data storage technology is generally inappropriate for others. This dissertation contains a comprehensive analysis of suitable modulation coding schemes and corresponding modulation codes for PODS systems that use two-photon absorption technology.

We present several two-dimensional (2-D) mathematical models for various two-photon PODS systems. Using these 2-D models, we describe the nature of intersymbol interference (ISI) in two-photon PODS systems and find that its characteristics are different from ISI in conventional data storage systems. To overcome the inherent ISI in two-photon PODS systems, we present a novel 2-D modulation coding scheme. We also present a number of fixed-length and variable-length 2-D modulation codes with diverse properties based on our 2-D modulation coding scheme. The bit-error-rate (BER) performance of these 2-D modulation codes is investigated and compared.

In addition to the ISI, the interpage interference (IPI) may also be significant in two-photon PODS systems. We present many three-dimensional (3-D) mathematical

models for a variety of two-photon PODS systems to examine IPI and its effects. In order to overcome the effects of ISI and the effects of IPI in two-photon PODS systems simultaneously, we present a new 3-D modulation coding scheme that is an extension of our 2-D modulation coding scheme. Examples of 3-D modulation codes derived from our 3-D modulation coding scheme are also presented. Finally, we discuss techniques to encode and extract clock information in two-photon PODS systems.

Chapter 1

Introduction

1.1 Motivation and Objective

New multimedia and Internet services such as high-definition television (HDTV), real-time archival digital video/audio, and future haptics (touch-related) applications increasingly demand enormous amounts of data storage capacity and massive data transfer rates. Improvements in conventional data storage technologies – magnetic hard disk, compact disc (CD), and digital versatile disc (DVD) – have managed to keep pace with such requirements. However, strong evidence indicates that these planar or quasi-planar data storage technologies are approaching fundamental limits that may be difficult to overcome [52]. An alternative candidate for next-generation data storage systems is page-oriented optical data storage (PODS) technology. By using this volumetric data storage technology, a system with very large data storage capacity (> 1 Tbits/cm^3) and very high data transfer rate (> 1 Gbits/s) is theoretically achievable [15,64]. Yet, before PODS systems can become commercial products, a large number of technical issues need to be solved.

One challenging issue is the reliability, in terms of bit-error-rate (BER), of PODS systems. Ideally, PODS systems are expected to have BERs to the users of 10^{-12} or lower before they can be released to the market [20,22]; unfortunately, real-world

systems operating near their maximum capabilities are generally subject to numerous sources of noise and interference. As with any other conventional data storage system, there are two basic types of data coding used in PODS systems to ensure the required user-end BERs. One is the modulation coding, whose main purpose is to facilitate the recording and retrieving processes. The other is the channel coding, which is used for error detection and error correction. Unlike conventional data storage systems, where the recording and retrieving processes are performed serially, the recording and retrieving processes of PODS systems are performed in parallel, implying that conventional modulation codes may be inappropriate for PODS systems.

Even though there exist many different error-correcting codes for channel coding, of particular interest to PODS systems is the Reed-Solomon codes and their extensions, which are also widely used in conventional data storage systems, because of their powerful abilities to correct both random and burst errors [55,68]. Nevertheless, in order to achieve the acceptable user-end BERs, Reed-Solomon coding schemes, or their extensions, need input BERs to satisfy specific upper bounds [20]. To maintain the input BERs of Reed-Solomon coding schemes under these specific constraints, suitable modulation coding schemes are necessary.

The goal of this research, therefore, is to explore possible modulation coding schemes that can be efficiently applied to PODS systems. To do so, the PODS systems must be properly modeled; the intrinsic noise and interference that occurs during the recording and retrieving processes must also be well examined. After developing such modulation coding schemes, we discuss the details and properties of corresponding

modulation codes. In addition, we analyze the performance of each modulation code in terms of BER.

1.2 Organization of the Dissertation

This dissertation is organized as follows. In chapter 2, the background relevant to this research is provided. It not only includes a brief overview of typical data storage systems, conventional modulation codes, and PODS systems, but also includes other significant related work. In the remaining chapters, we focus particularly on PODS systems that use two-photon absorption technology. Chapter 3 is a comprehensive discussion about the mathematical models and the inherent intersymbol interference (ISI) of two-photon PODS systems. We present a variety of models for two-photon PODS systems. Chapter 4 intuitively describes the appropriate modulation coding scheme for two-photon PODS systems and clearly illustrates its benefits. A number of fixed-length modulation codes based upon such modulation coding scheme and their general form are also described. In chapter 5, we roughly verify restrictions of fixed-length modulation codes. To surmount the shortcomings of fixed-length modulation codes, variable-length modulation codes for two-photon PODS systems are introduced. Chapter 6 presents numerous BER performance comparisons among various modulation codes. We also show how the optimal BER of a two-photon PODS system can be determined under specific conditions. In chapter 7, interpage interference (IPI) is taken into consideration, and more complicated models of two-photon PODS systems are

presented. To lessen the effects of IPI, we describe an extended modulation coding scheme and many examples of the corresponding modulation codes. Chapter 8 gives conclusions and discusses the possible future research work.

1.3 Contributions of the Research

The original research contributions contained in this dissertation can be summarized by the following:

(1) *Two-Dimensional Mathematical Models of Two-Photon PODS Systems.* We present several two-dimensional (2-D) models for different two-photon PODS systems, assuming that the distance between any two successive recorded data pages is large enough so that there is no IPI. Using these 2-D models, one can easily understand and study the nature of ISI and its effects in two-photon PODS systems.

(2) *Two-Dimensional Modulation Coding Scheme for Two-Photon PODS Systems.* We propose a novel 2-D modulation coding scheme that eliminates or relaxes the effects of ISI in two-photon PODS systems. A general form of 2-D modulation codes derived from the proposed 2-D modulation coding scheme is also presented. With this general form, one can distinguish and compare 2-D modulation codes.

(3) *Fixed-Length Two-Dimensional Modulation Codes for Two-Photon PODS Systems.* We present diverse examples of 2-D modulation codes for two-photon

PODS systems based on our 2-D modulation coding scheme. These 2-D modulation codes are classified as fixed-length 2-D modulation codes because their block sizes are fixed. Moreover, the pros and cons of each fixed-length 2-D modulation code are explained so that the most suitable 2-D modulation code for a particular PODS system can be selected.

(4) *Variable-Length Two-Dimensional Modulation Codes for Two-Photon PODS Systems.* By observing that fixed-length 2-D modulation codes normally have mediocre code rates due to their fixed block sizes, we propose new 2-D modulation codes that can overcome such obstacles. These alternative 2-D modulation codes are classified as variable-length 2-D modulation codes. Since the block sizes of variable-length 2-D modulation codes are not fixed, we gain additional degrees of freedom, leading to higher code rates. An example of a variable-length 2-D modulation code and its decoding procedure are presented. Additionally, many issues of a variable-length 2-D modulation code, such as the proper decoding procedure and the error propagation, are discussed.

(5) *BER Performance Analysis of Two-Dimensional Modulation Codes for Two-Photon PODS Systems.* We present a numerical procedure that is used to find the optimal threshold value for a binary threshold decision scheme applied at the detector array and the corresponding minimum BER of a noisy two-photon PODS system. We theoretically derive this procedure by assuming that the noise in a two-photon PODS system can be modeled as additive white Gaussian noise (AWGN). We also compare BER performance among 2-D modulation codes.

(6) *Three-Dimensional Mathematical Models of Two-Photon PODS Systems.* We present several three-dimensional (3-D) models for two-photon PODS systems, in which the distance between any two adjacent recorded data pages is small enough that the effects of IPI must be considered. Following these 3-D models, one can thoroughly investigate the characteristics of ISI and IPI in two-photon PODS systems.

(7) *Three-Dimensional Modulation Coding Scheme for Two-Photon PODS Systems.* In order to remove (or at least relieve) the effects of ISI and the effects of IPI simultaneously, we develop a novel 3-D modulation coding scheme that is an extension of our 2-D modulation coding scheme.

(8) *Three-Dimensional Modulation Codes for Two-Photon PODS Systems.* We present some examples of 3-D modulation codes based on our 3-D modulation coding scheme. All of them are derived directly from the examples of fixed-length 2-D modulation codes.

Chapter 2

Preliminaries

This chapter provides a brief overview of typical data storage systems, conventional modulation codes, and other background work related to page-oriented optical data storage (PODS) systems.

2.1 Data Storage Systems

A typical data storage system may be represented by the block diagram shown in Figure 2.1 [32,50]. In general, the input information to be stored in the data storage is in digital form. In the case of analog input information, an analog-to-digital (A/D) converter is required. For recording, the digital input information is first passed to the *channel encoder* (at point **A**). The purpose of the channel encoder is to add some redundancy to the digital input information in order to combat the noisy environment. Thus, the added redundancy is used to increase the reliability of the data storage system; in detail, this added redundancy is used for error detection and error correction. It should be noted that the output of the channel encoder is also in digital form.

The digital data at the output of the channel encoder is, then, passed to the *modulation encoder* (at point **B**). The primary purpose of the modulation encoder is to map the output data from the channel encoder into the data with more suitable properties

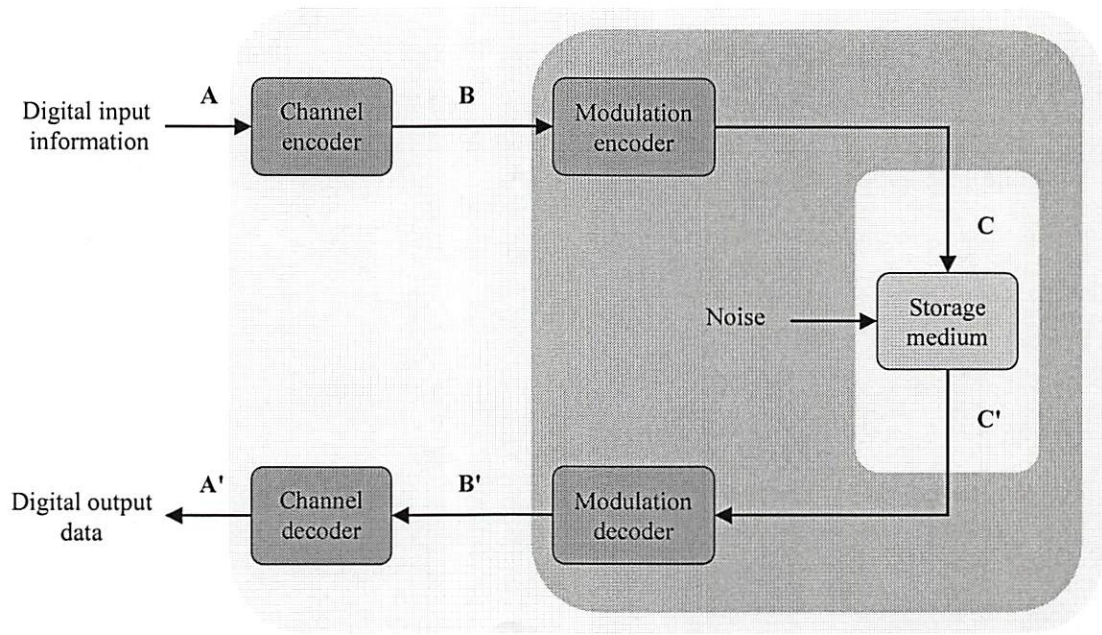


Figure 2.1 Block diagram of a typical data storage system.

for both recording and retrieval. In fact, the mapping of modulation encoder is used to facilitate the recording and retrieving processes in such a way that the characteristics of the stored data match the characteristics of the data storage system. The output of the modulation encoder is recorded in the *storage medium* (at point **C**) by the writing unit (not shown in Figure 2.1).

For data retrieval, the stored data is extracted from the storage medium by the reading unit (not shown in Figure 2.1), and is passed to the *modulation decoder* (at point **C′**). The stored data is subject to various sources of noise and interference, such as physical defects in the storage medium and thermal noise generated by electronic devices, and can be corrupted in a random or a deterministic manner. By taking advantage of the modulation coding properties, the digital data at the output of the

modulation decoder is expected to be a good estimate of the digital data originally input to the modulation encoder. The digital data at the output of the modulation decoder is, finally, passed to the *channel decoder* (at point **B′**). The channel decoder attempts to reconstruct the original input information by using the redundancy added at the channel encoder. Ideally, the output of the channel decoder (at point **A′**) should be a replica of the original input information, even though noise and interference may cause some decoding errors. If the original input information is in analog form, the digital-to-analog (D/A) converter is required after the channel decoder.

To measure the performance of a data storage system, several criteria can be used. One of them is the average probability of a bit-error, which is the frequency that errors occur at the output of the data storage system. Indeed, this probability of error is called bit-error-rate (BER). It should be emphasized that the BER of the data storage system, which is the BER that users see, i.e., the user-end BER, is generally not the same as the BER of the storage medium, i.e., the raw BER. Typically, the BER of the storage medium, measured between points **C** and **C′**, is very high, while the BER of the data storage system, measured between points **A** and **A′**, can be extremely low by utilizing the appropriate modulation coding and channel coding schemes. Other performance criteria include data transfer rate and useful data storage capacity. The data transfer rate often relies on the technology of the data storage system. Data storage systems having parallel access usually produce higher data transfer rate than those having serial access. The useful data storage capacity, on the other hand, depends on the code rates of both modulation code and channel code, where the code rate is defined as a fraction of the

useful information bits after modulation or channel coding. Apparently, the useful data storage capacity is different from the actual data storage capacity.

One familiar example of a data storage system is the Compact Disc (CD) system. In CD systems, the eight-to-fourteen modulation (EFM) code is used in modulation coding whereas the cross-interleaved Reed-Solomon code (CIRC) is used in channel coding [48]. Another example of a data storage system is the rotary-head digital audio tape (RDAT) system. The RDAT system uses the 8/10 group code for modulation coding and the Reed-Solomon product code, together with interleaving, for channel coding [67].

2.2 Modulation Codes in Conventional Data Storage Systems

Most conventional data storage systems need suitable modulation codes to enhance the system performance in a number of aspects [25,65]. For instance, modulation codes are employed in magnetic recording systems not only to reduce the temporal intersymbol interference (ISI) that arises during the retrieving process, but also to eliminate or minimize the DC content in the stored data [50]. Of particular importance to consider, however, is the fact that different data storage technologies may require different modulation codes, or even different modulation coding schemes.

Some well-known examples of modulation codes used in today's commercial data storage systems are: the (2,7) runlength-limited variable-length code used in many IBM disk storage systems [1,12], the EFM code in CD systems [48,66], and the

EFMPlus code in Digital Versatile Disc (DVD) systems [24]. Basically, these conventional modulation codes can be classified as runlength-limited codes [23,25]. A runlength-limited code is a binary sequence with a restriction on the number of 1s and 0s. Runlength-limited codes are normally described by two parameters, d and k, where d denotes the minimum number of consecutive 0s that separate any two 1s in a sequence and k denotes the maximum number of consecutive 0s that separate any two 1s in a sequence. The minimum constraint, d, is used to reduce the temporal ISI, while the maximum constraint, k, is used to ensure that the timing or clock information can be recovered. Note that runlength-limited codes can have either fixed-length or variable-length.

Input data	Output data
00000000	01001000100000
00000001	10000100000000
00000010	10010000100000
00000011	10001000100000
00000100	01000100000000
00000101	00000100010000
00000110	00010000100000
00000111	00100100000000
00001000	01001001000000
00001001	10000001000000
00001010	10010001000000

Table 2.1 Partial lookup table of the EFM code.

EFM and EFMPlus are two examples of runlength-limited fixed-length codes. Table 2.1 illustrates part of the lookup table for EFM code, which maps eight input bits

into fourteen output bits [48]. Using the EFM code and three merging bits, the runlength-limited fixed-length code with $d = 2$ and $k = 10$ can be obtained [23]. Since the EFM code totally maps eight input bits into seventeen output bits (including three merging bits), its code rate is, hence, equal to 8/17 (= 0.4706). Similarly, EFMPlus is a runlength-limited fixed-length code with the same runlength parameters as EFM, but has a slightly higher code rate of 8/16 (= 0.5) [24,25].

Input data	Output data
11	0100
10	1000
011	000100
010	001000
000	100100
0011	00100100
0010	00001000

Table 2.2 Lookup table of the (2,7) runlength-limited variable-length code.

Table 2.2 shows the lookup table of the (2,7) runlength-limited variable-length code [12]. With the mapping in Table 2.2, any two 1s in a sequence are separated by a run of at least two, but no more than seven, 0s. The code rate of the (2,7) runlength-limited variable-length code is equal to 0.5 because it maps two, three, or four input bits into four, six, or eight output bits, respectively. It is also worth pointing out that the error propagation of the (2,7) runlength-limited variable-length code due to a single erroneous bit can be limited if the proper decoding procedure is applied [23].

2.3 Page-Oriented Optical Data Storage (PODS) Systems and Other Related Work

In general, PODS systems are distinguishable from conventional data storage systems because the former systems fully utilize three-dimensional (3-D) volumetric storage media together with parallel recording/retrieving processes, whereas the latter systems only utilize two-dimensional (2-D) planar storage media, such as CDs, or the quasi-planar storage media (i.e., the 3-D volumetric storage media whose third dimension is used with coarse resolution), such as DVDs, together with serial recording/retrieving processes.

For PODS systems, several technologies exist including holography [15,18,31,51-54], spectral-hole burning [10,30,33,56,59], and two-photon absorption [11,21,35,47,64]. However, regardless of the technology, PODS systems generally suffer from a variety of noise and interference, when operated near their maximum capabilities. Many research groups have modeled and studied the effects of ISI in holographic PODS systems [2,19,61]. Instead of using a simple binary threshold decision scheme to digitize the retrieved data, a number of detection and equalization techniques have been proposed and investigated in order to mitigate the effects of ISI [7,9,17,27,28]. Various modulation coding schemes and partial response precoding approaches have also been discussed to further improve the performance, especially in terms of BER and data storage capacity, of holographic PODS systems [3,5,16,29,38,39,62,63]. One common goal of these detection/equalization techniques and modulation coding schemes is to guarantee that the BERs of holographic PODS

systems are at least on the order of 10^{-3}-10^{-4}. In addition, to reduce the BERs of holographic systems so that they satisfy the industrial requirement of 10^{-12} or lower, numerous error-correction, i.e., channel coding, schemes have been considered [8,13,22,36,37].

While there has been considerable study of holographic PODS systems, only a handful of researchers have concentrated on how to increase the reliability of the PODS systems that use spectral-hole burning and two-photon absorption technologies [26,69]. Nevertheless, it has been shown that the detection and equalization techniques described in Refs. 7, 9, and 28 can be applied not only to coherent imaging systems such as holographic PODS systems, but also to incoherent imaging systems such as two-photon PODS systems. Similarly, since the error-correction scheme is typically designed based on the format or layout of the recorded data, not on the technology of the data storage system, the error-correction schemes developed for holographic PODS systems may be applied directly, or with minor modifications, to PODS systems using other technologies that have the same recorded data format or layout. In contrast, the modulation coding scheme designed for a particular PODS technology is usually not appropriate for other PODS technologies, owing to the fact that the main purpose of modulation coding is to facilitate the technology-dependent recording/retrieving processes of each data storage system.

Chapter 3

Two-Dimensional Models and Intersymbol Interference of Two-Photon PODS Systems

While conventional data storage systems, such as the magnetic tape and Compact Disc (CD) systems, have been well modeled and widely studied by many groups of researchers for decades, page-oriented optical data storage (PODS) systems using two-photon absorption technology have not been investigated much. We describe in this chapter the detailed models and intersymbol interference (ISI) of two-photon PODS systems. The image formation process in two-photon PODS systems is described in the first section, leading to various two-dimensional (2-D) mathematical models. The second section discusses the performance deterioration due to ISI in these systems.

3.1 Two-Dimensional Models of Two-Photon PODS Systems

In two-photon PODS systems, data recording is performed using two propagating light beams. One beam, carrying the information to be stored, is imaged onto a desired plane in the volumetric medium while a second beam, specifying the location, is simultaneously focused throughout the same plane. The former is called the information beam whereas the latter is called the addressing beam. With sufficient photon energies,

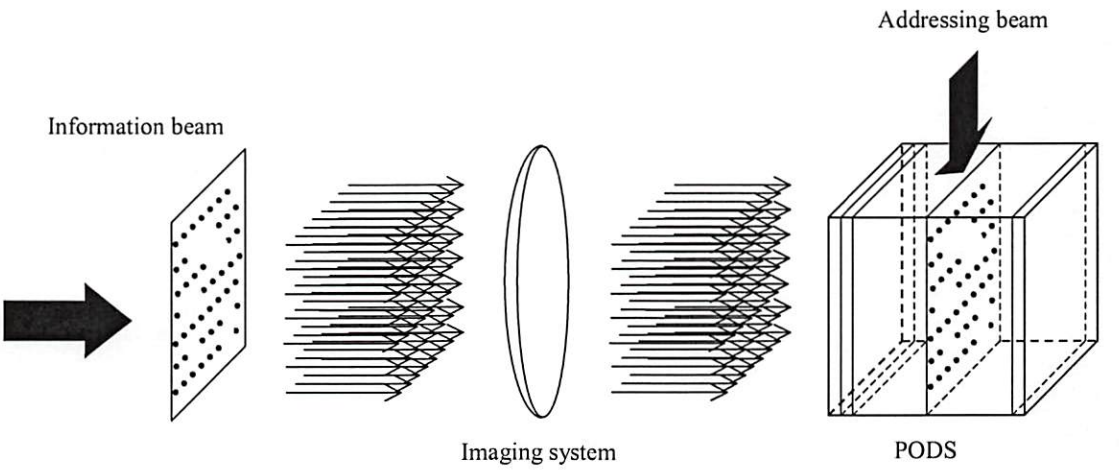


Figure 3.1 The recording of information in a two-photon PODS system.

a 2-D array of data marks (or spots) can be recorded on a plane (or page) where two beams intersect [11,35]. The typical number of data marks per page is 10^3 or more, perhaps as many as 10^6, and, generally, a very large number of data pages are stored in the medium. To retrieve the recorded data, the proper data page is illuminated by one readout beam, which is identical to the addressing beam used to record that page. This

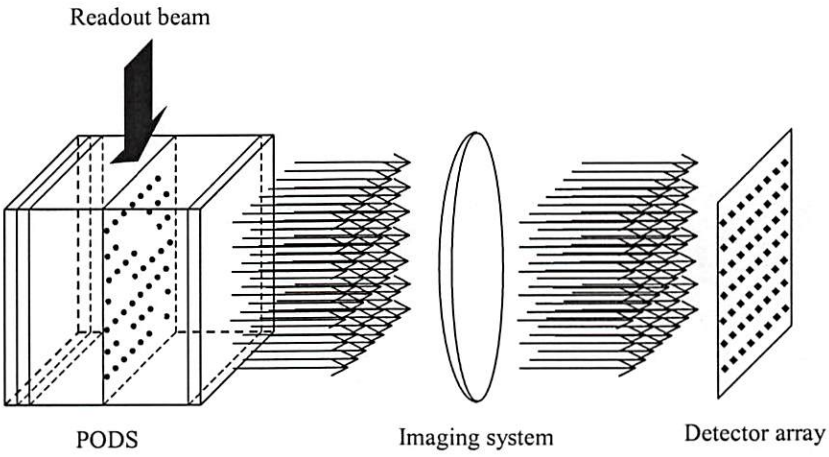


Figure 3.2 The retrieval of information from a two-photon PODS system.

illumination results in a fluorescence pattern that is imaged onto a charge-coupled device (CCD) camera or a 2-D integrated optoelectronic detector array. Figures 3.1 and 3.2 show the recording and retrieval of information in a two-photon PODS system, respectively.

Assuming that the recording process is perfect and that the distance between any two successive recorded data pages is large enough that there is no interaction between the data pages, then the mathematical model of a typical two-photon PODS system can be derived from the retrieving process as shown in Figure 3.3. In this mathematical model, $[i, j]$ and (x, y) denote an index of a 2-D discrete signal and an index of a 2-D continuous signal, respectively. Let $a[i, j]$, whose value is assumed to be either zero for a "0" bit or one for a "1" bit, represent the existence of a digital information bit (or data mark) recorded on a data page.

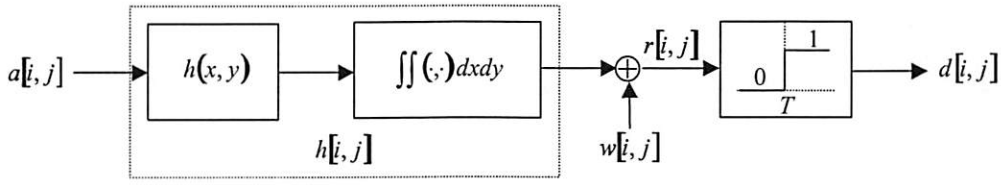


Figure 3.3 The 2-D mathematical model of a two-photon PODS system.

Since the readout intensity level of each data mark is composed of the superposition of light from a large number of statistically independent fluorescence elements, each radiating with a slightly different temporal frequency and a random phase, it is clear that the output of each data mark during the retrieving process is spatially incoherent. Therefore, ideally, the point-spread function (PSF), or the impulse

response, of this imaging system can be modeled mathematically as one of the following functions [14]:

$$h(x,y)=\frac{1}{b_s^2}\operatorname{sinc}^2\left(\frac{x}{b_s},\frac{y}{b_s}\right), \tag{3.1}$$

where $\operatorname{sinc}(x,y)=\frac{\sin(\pi x)}{\pi x}\frac{\sin(\pi y)}{\pi y}$ or

$$h(x,y)=\frac{1}{\pi b_c^2}\left[\frac{J_1\left(2\pi\sqrt{x^2+y^2}/b_c\right)}{\sqrt{x^2+y^2}/b_c}\right]^2, \tag{3.2}$$

where J_1 is a Bessel function of the first kind, order 1.

Equation (3.1) arises from a *diffraction-limited* imaging system containing a square aperture, where b_s, which is inversely proportional to the width of the aperture, represents the location of the first zero of the sinc^2 PSF. Similarly, Eq. (3.2) is obtained from a *diffraction-limited* imaging system containing a circular aperture, rather than a square, where b_c is a parameter inversely proportional to the radius of the aperture. Note that the J_1^2 PSF is circularly symmetric, but its zeroes are not equally spaced in radius. It should also be emphasized that the leading terms $1/b_s^2$ in Eq. (3.1) and $1/\pi b_c^2$ in Eq. (3.2) are the factors, simply derived using the Parseval's theorem, that normalize the volumes of the corresponding PSFs to one.

If an imaging system has a great amount of small *independent aberrations*, then, from the central limit theorem, the PSF can be described by a Gaussian function

$$h(x,y)=\frac{1}{2\pi b_g^2}\exp\left(-\frac{x^2+y^2}{2b_g^2}\right), \tag{3.3}$$

where b_g denotes the standard deviation [28,46]. Moreover, we note that other PSF models can also be used to describe the imaging optics of a particular system.

In our mathematical model, for simplicity, the magnification of the imaging system in the retrieving process is assumed to be one. Furthermore, the pitch between two nearest data marks, equivalently the pitch between two nearest detectors or the pixel pitch of the CCD camera, is set to one so that all other distances can be relatively expressed in units of this pitch. We also note that the size of the data mark is assumed to be very small compared to the extent of the PSF, implying that the data mark is effectively a delta function. From the assumptions, the received intensity at each detector, or each CCD pixel, of a *noiseless* retrieval system can be modeled by

$$r[i,j] = a[i,j] \otimes h[i,j], \tag{3.4}$$

where $\otimes$ represents a 2-D discrete convolution operator and $h[i,j]$ is the effective discrete PSF. Because the retrieved photons are integrated over the active area, assumed to be a square, of each detector or CCD pixel, the effective discrete PSF $h[i,j]$ in Eq. (3.4) is defined as

$$h[i,j] = \int_{j-\delta/2}^{j+\delta/2} \int_{i-\delta/2}^{i+\delta/2} h(x,y)\,dx\,dy, \tag{3.5}$$

where δ is a linear fill factor of a detector or a CCD pixel, whose value is between zero and one, and $h(x,y)$ is one of the continuous PSFs in Eqs. (3.1), (3.2), and (3.3).

However, in general, a retrieval system is not noiseless. It is assumed here that the retrieval system is noisy, mainly due to the electronics at the detector plane. Since the noise from electronic circuits, often called thermal noise, is typically modeled as

additive white Gaussian noise (AWGN) [4,57], the received intensity of a *noisy* retrieval system can be described by

$$r[i,j] = a[i,j] \otimes h[i,j] + w[i,j], \tag{3.6}$$

where $w[i,j]$ is an AWGN, whose mean and variance are zero and σ_w^2, respectively.

Finally, we assume that a simple binary threshold decision scheme is applied at the CCD or detector array, with a *priori* knowledge of a threshold value, T, in order to digitize the received intensities. Therefore, the detected data, i.e., the output of the retrieval system, can be modeled mathematically as

$$d[i,j] = \begin{cases} 1, \text{if } r[i,j] > T \\ 0, \text{if } r[i,j] < T \end{cases}. \tag{3.7}$$

It should also be noted that, for the purpose of analysis, it is often helpful to regard the 2-D discrete signals as matrices. For example, Eq. (3.6) can be expressed in matrix form by

$$\mathbf{R} = \mathbf{A} \otimes \mathbf{H} + \mathbf{W}, \tag{3.8}$$

where $\mathbf{R}$ is an $M \times N$ received intensity matrix whose entry in the p th row and q th column, R_{pq}, describes the real-valued intensity at the $[p,q]$-th detector or the $[p,q]$-th CCD pixel, $\mathbf{A}$ denotes the recorded $M \times N$ digital information data page in which each entry, A_{pq}, is either zero or one, $\mathbf{H}$ is the effective PSF matrix derived from $h[i,j]$ in Eq. (3.5) with minor modifications of indices so that $h[0,0]$ is the entry at the center of matrix $\mathbf{H}$, and $\mathbf{W}$ represents an $M \times N$ AWGN matrix whose entries are independent, identically distributed with mean and variance defined in Eq. (3.6).

3.2 Two-Dimensional Intersymbol Interference in Two-Photon PODS systems

From our mathematical model, a PSF defines the image intensity of each data mark at the detector plane. Figures 3.4, 3.5, and 3.6 illustrate the continuous PSFs in Eq. (3.1) with $b_s = 1.5$, Eq. (3.2) with $b_c = 2.5$, and Eq. (3.3) with $b_g = 0.6$, respectively. The x and y axes, as mentioned in the previous section, are in units of the data mark pitch or, equivalently, the detector or CCD pixel pitch. It is assumed in Figures 3.4, 3.5, and 3.6 that only one data mark, i.e., only one "1" bit, is recorded at $(x, y) = (0,0)$ (of the data plane) and, hence, the corresponding detector or CCD pixel is located at the same coordinate (of the detector plane). Suppose the linear fill factor δ is one, the active area of a detector or a CCD pixel is, then, a square, whose width is equal to one, centered at $(x, y) = (0,0)$ as shown in Figures 3.4, 3.5, and 3.6.

Note that the light originating from a data mark in Figures 3.4, 3.5, or 3.6 is not completely imaged onto the corresponding detector or CCD pixel. Some of the light is incident onto the local neighbor detectors or CCD pixels. This phenomenon is referred to as ISI or, more specifically, the 2-D ISI in which light from a data mark interferes *spatially* with its *local* neighbors. In fact, the 2-D ISI results from the inherent *lowpass* nature of the imaging system. Additionally, of particular importance to consider is the fact that the 2-D, or *spatial*, ISI in two-photon PODS systems is somewhat different from the one-dimensional (1-D), or *temporal*, ISI in today's commercial data storage systems, such as CD and Digital Versatile Disc (DVD).

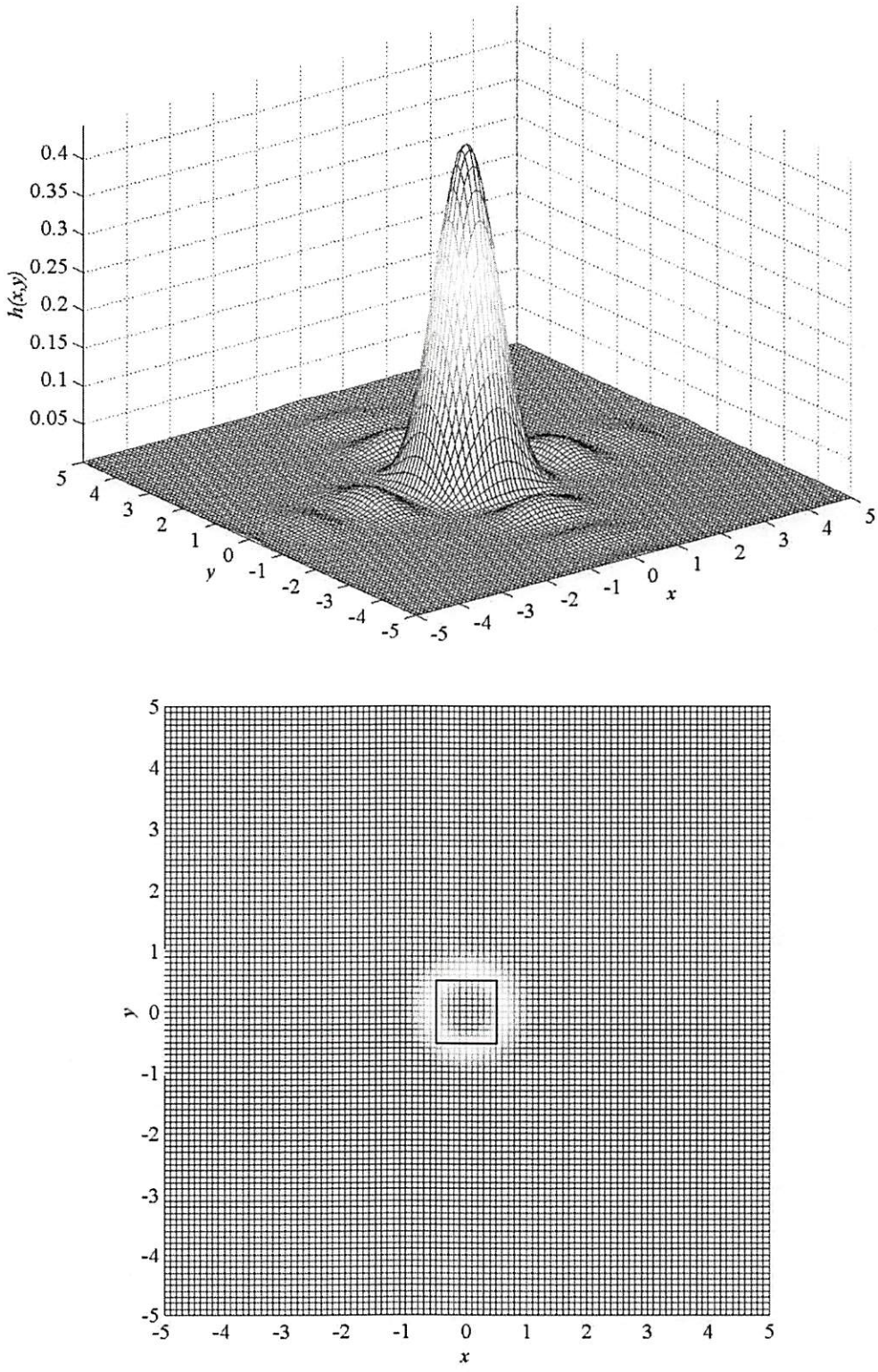


Figure 3.4 The sinc^2 PSF with $b_s = 1.5$.

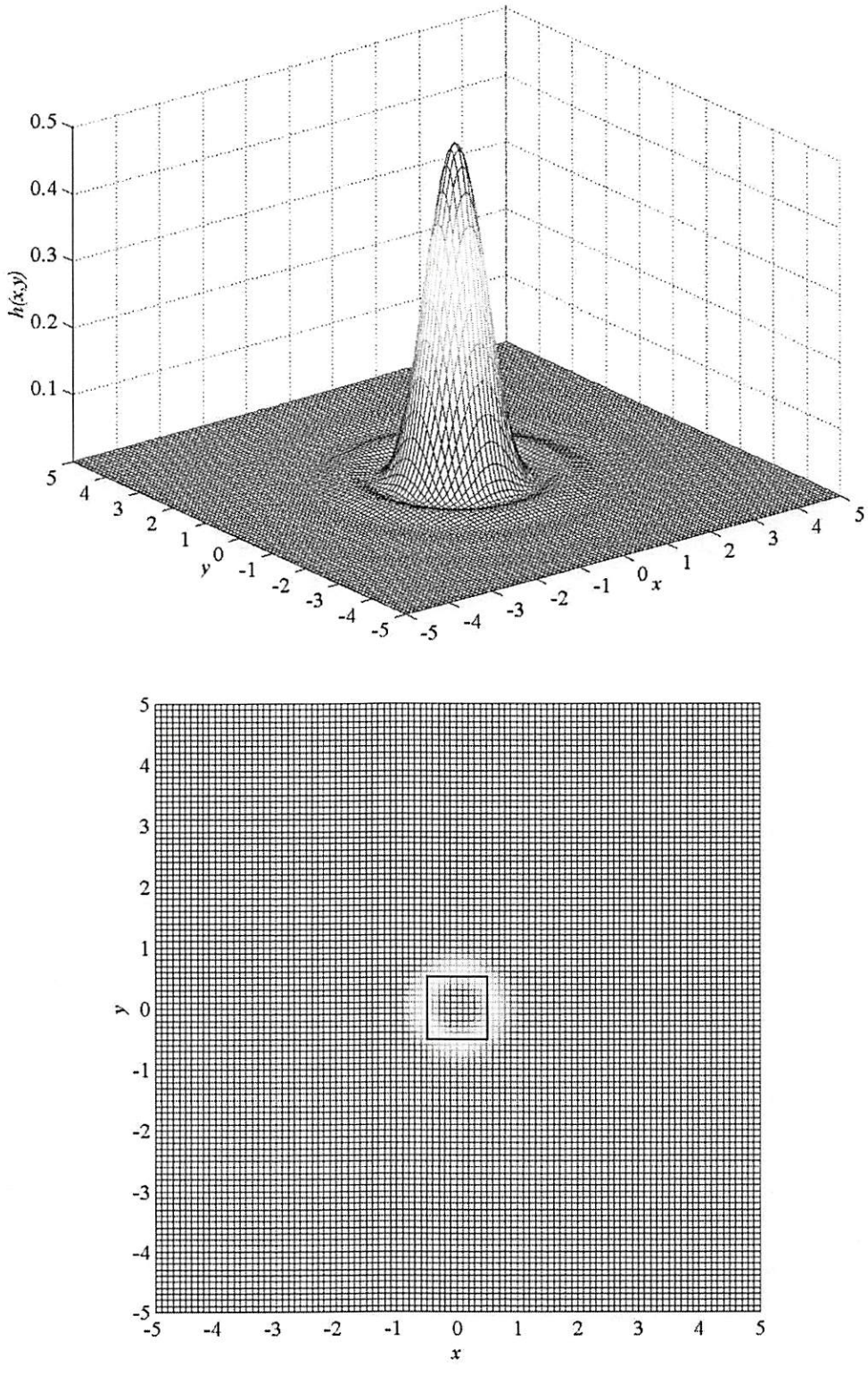


Figure 3.5 The J_1^2 PSF with $b_c = 2.5$.

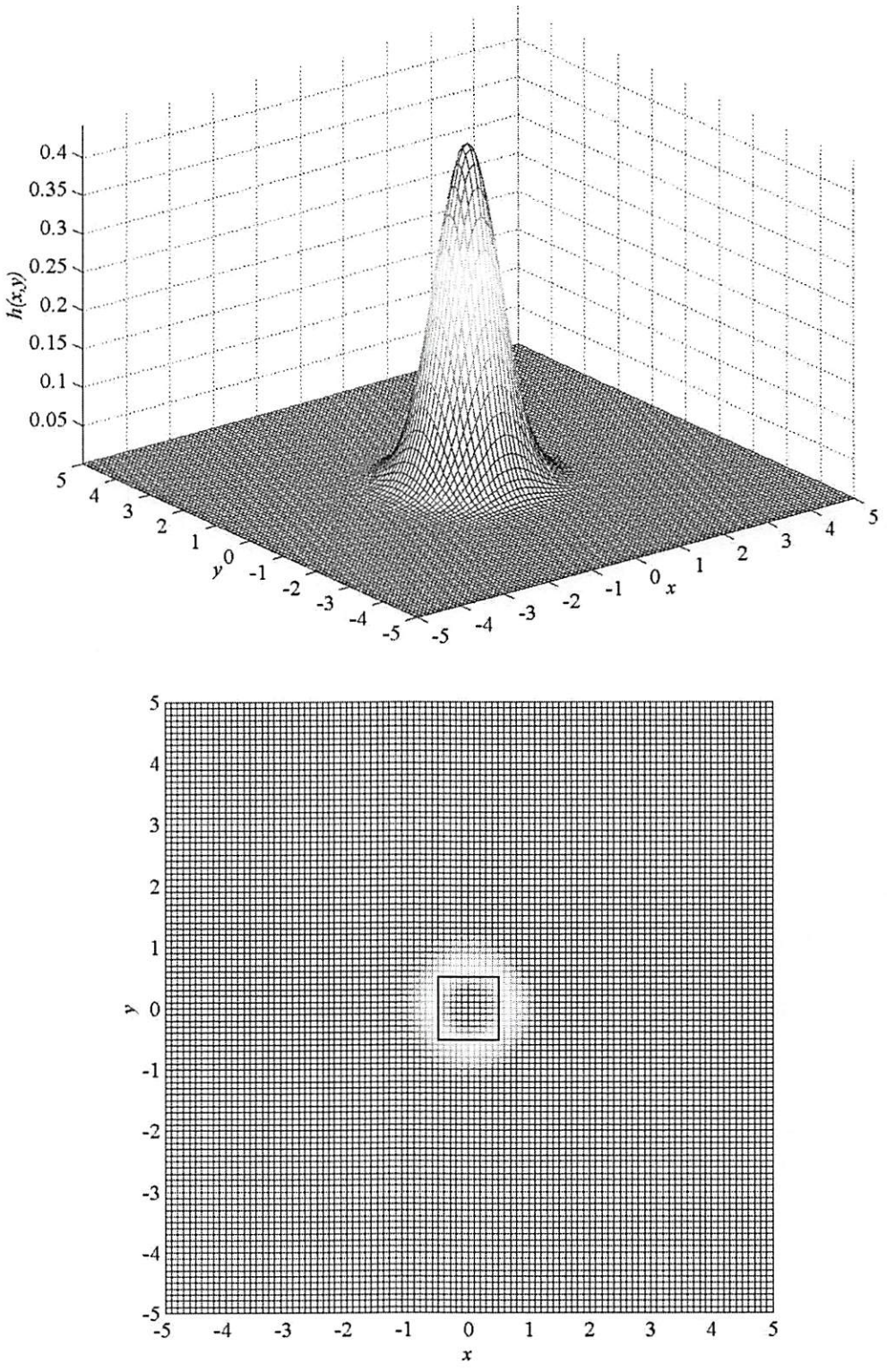


Figure 3.6 The Gaussian PSF with $b_g = 0.6$.

It should be clear from the mathematical model that the amount of light from a data mark that is incident onto the neighbor detectors or CCD pixels, i.e., the 2-D ISI, depends on both the PSF and the linear fill factor, δ, of the retrieval system. Indeed, for a given value of δ, the relative degree of the effects of 2-D ISI can be simply indicated by the value of b_s, b_c, or b_g. Figures 3.7, 3.8, and 3.9 show the 1-D slices through the x-axis of the sinc^2, J_1^2, and Gaussian PSFs at various values of b_s, b_c, and b_g, respectively. As b_s, b_c, or b_g increases, the amount of light from a data mark that is imaged onto the corresponding detector or CCD pixel decreases. Consequently, with the fact that the volume of each PSF is normalized to one, the remaining light from the same data mark that is incident onto the neighbor detectors or CCD pixels, i.e., the 2-D ISI, increases.

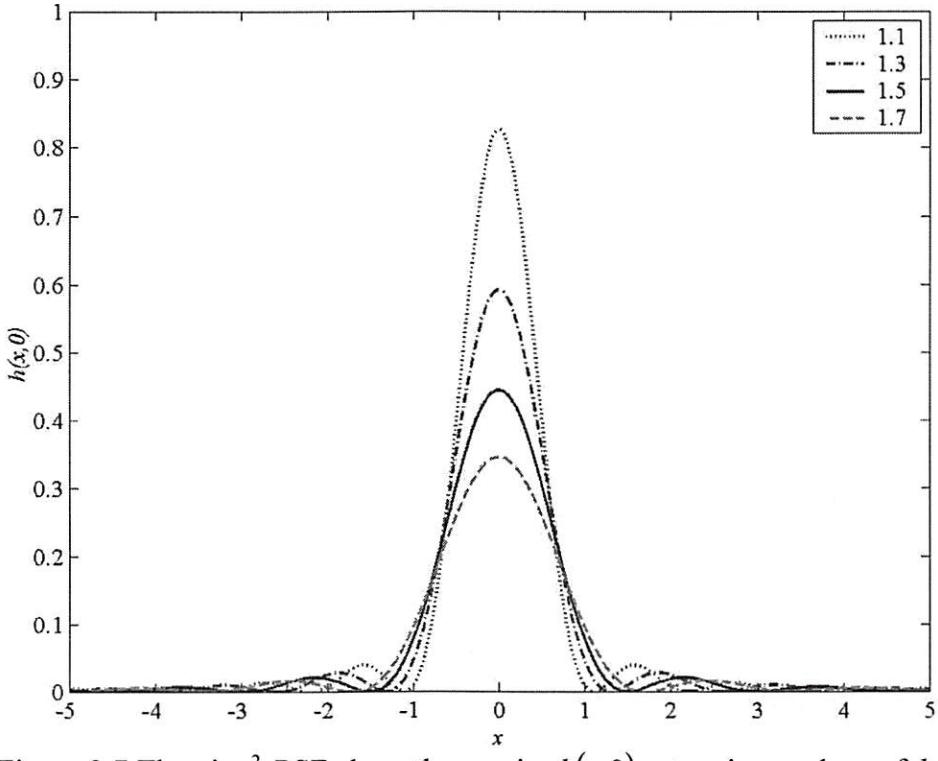


Figure 3.7 The sinc^2 PSF along the x-axis, $h(x,0)$, at various values of b_s.

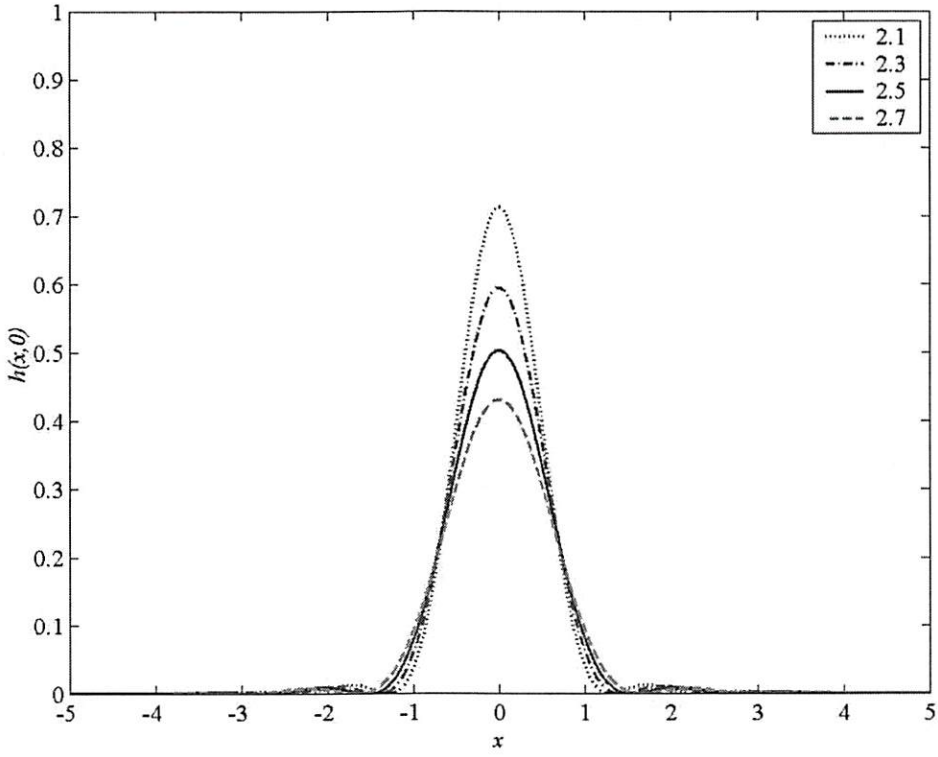


Figure 3.8 The J_1^2 PSF along the x-axis, $h(x,0)$, at various values of b_c.

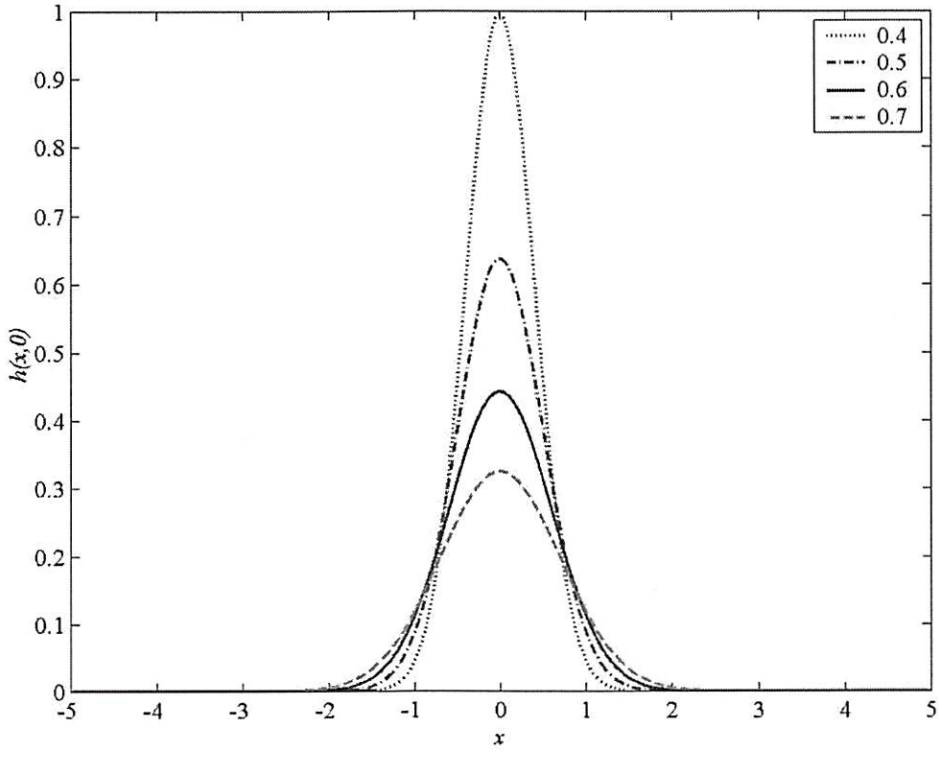


Figure 3.9 The Gaussian PSF along the x-axis, $h(x,0)$, at various values of b_g.

Figures 3.10, 3.11, and 3.12 depict more convenient ways to visualize the 2-D ISI. Assuming that a retrieval system is noiseless and its imaging system contains a square aperture, Figure 3.10(a) shows a 1-D slice through the x-axis of the image intensity that originates from two data marks recorded at $(x,y)=(-1,0)$ and $(x,y)=(1,0)$. In other words, Figure 3.10(a) shows the sum of $h(x+1,0)$ and $h(x-1,0)$ at the detector plane, where $h(x,y)$ is from Eq. (3.1) with $b_s=1.5$. Clearly, the tails of the image intensity from each data mark (each "1" bit) interfere with others (other "0" and "1" bits). In addition, Figure 3.10(b) shows the corresponding received intensity, which is determined from Eqs. (3.4) and (3.5) with $\delta=1$, of each detector or CCD pixel that lies on the x-axis, i.e., $r[i,0]$. Since only two data marks or two "1" bits are recorded, the digitized outputs of all detectors or CCD pixels, except the ones located at $[i,j]=[-1,0]$ and $[i,j]=[1,0]$, are ideally expected to be "0" bits. However, in Figure 3.10(b), the output of the detector or CCD pixel at $[i,j]=[0,0]$ is ambiguous because its intensity level is about half of the ("1" bit) intensity levels at $[i,j]=[-1,0]$ and $[i,j]=[1,0]$. Depending on a predetermined threshold value T, the detector or CCD pixel at $[i,j]=[0,0]$ may improperly make a decision leading to an incorrect output. It is important to emphasize that if the total number of data marks or "1" bits around the "0" bit at $(x,y)=(0,0)$ is more than two, then the effects of 2-D ISI on that "0" bit is even worse. Figures 3.11 and 3.12 illustrate the similar results for the cases of J_1^2 PSF with $b_c=2.5$ and Gaussian PSF with $b_g=0.6$, respectively.

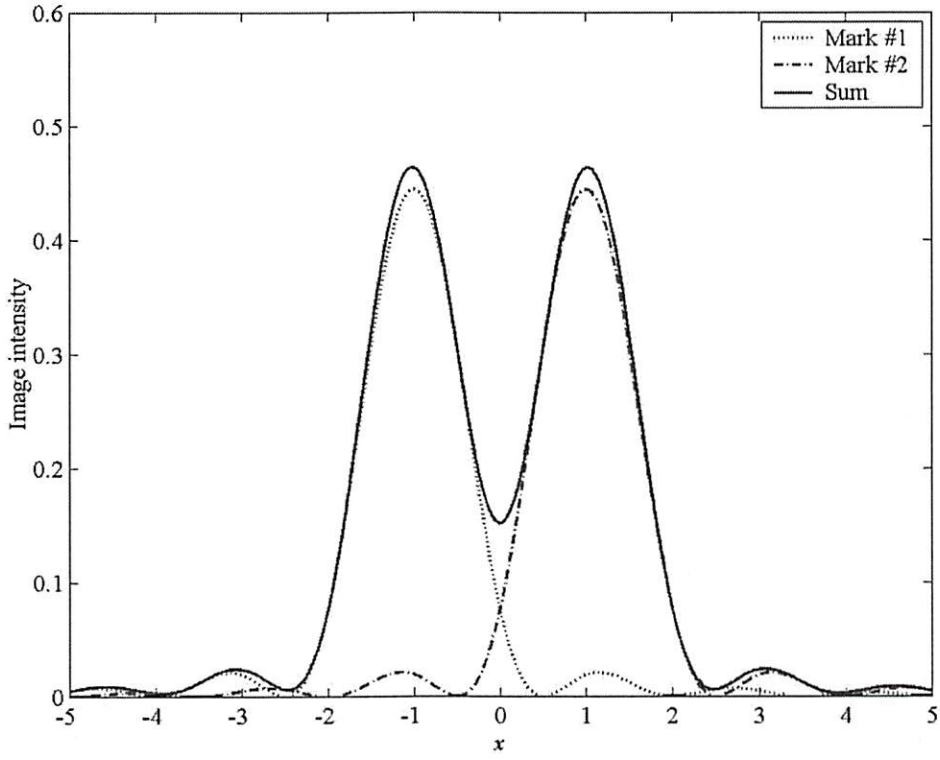


Figure 3.10(a) $h(x+1,0)$, $h(x-1,0)$, and their sum (sinc^2 PSF with $b_s = 1.5$).

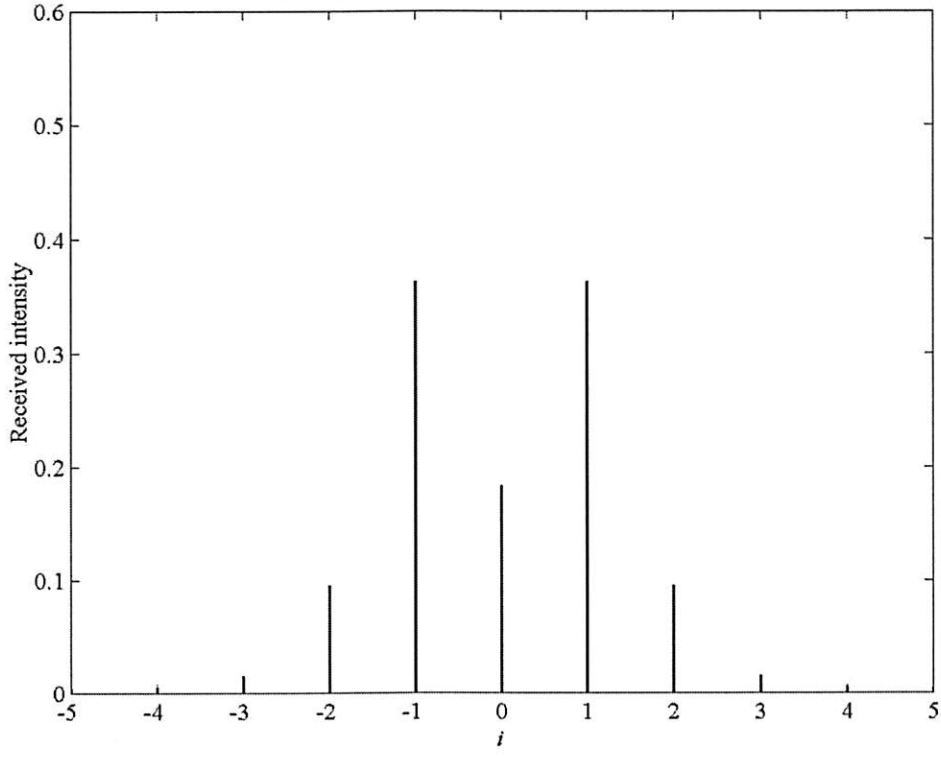


Figure 3.10(b) $r[i,0]$ (sinc^2 PSF with $b_s = 1.5$ and $\delta = 1$).

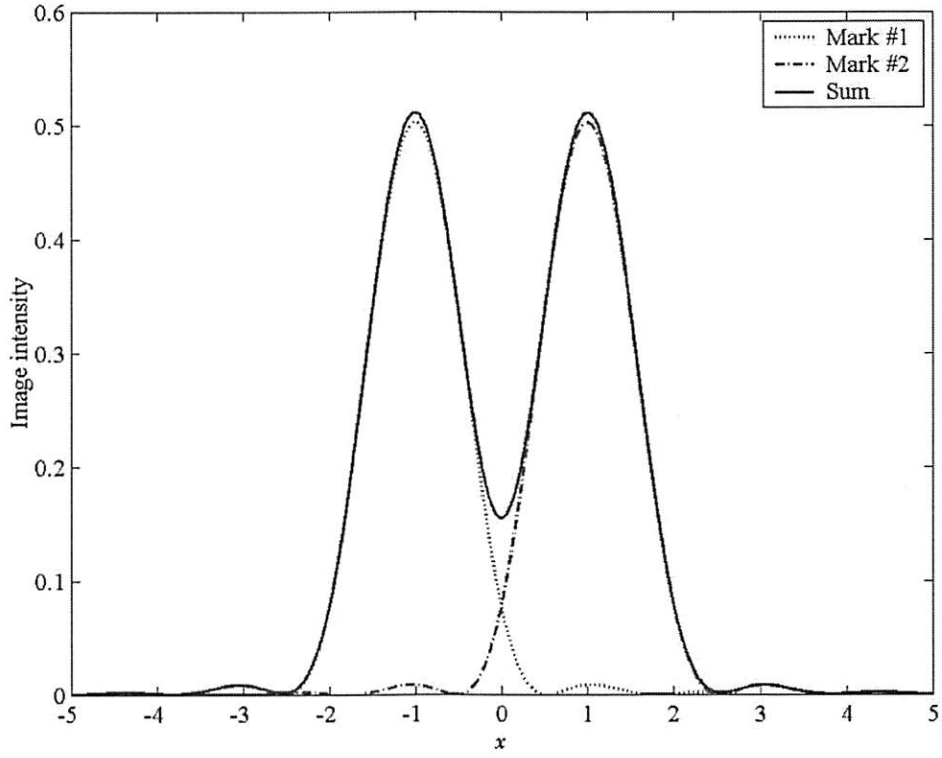


Figure 3.11(a) $h(x+1,0)$, $h(x-1,0)$, and their sum (J_1^2 PSF with $b_c = 2.5$).

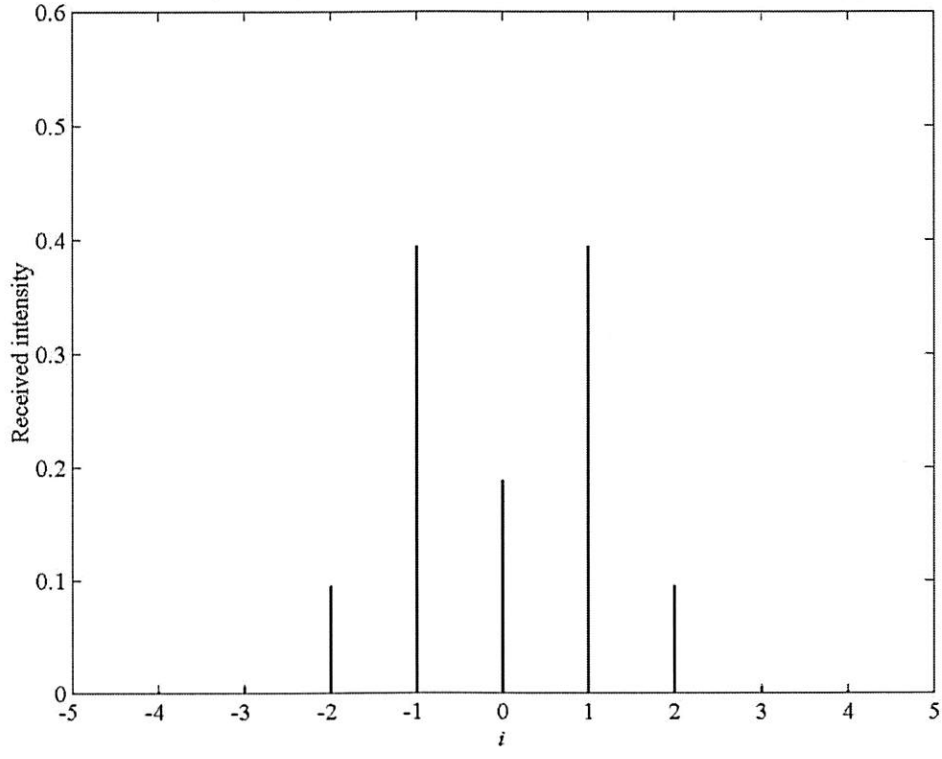


Figure 3.11(b) $r[i,0]$ (J_1^2 PSF with $b_c = 2.5$ and $\delta = 1$).

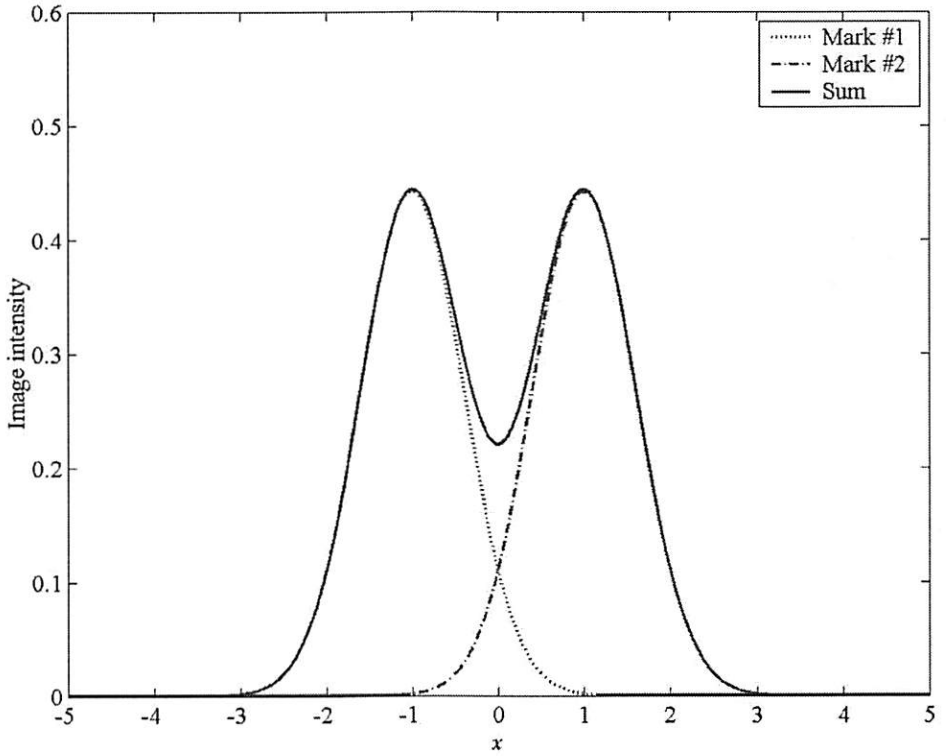


Figure 3.12(a) $h(x+1,0)$, $h(x-1,0)$, and their sum (Gaussian PSF with $b_g = 0.6$).

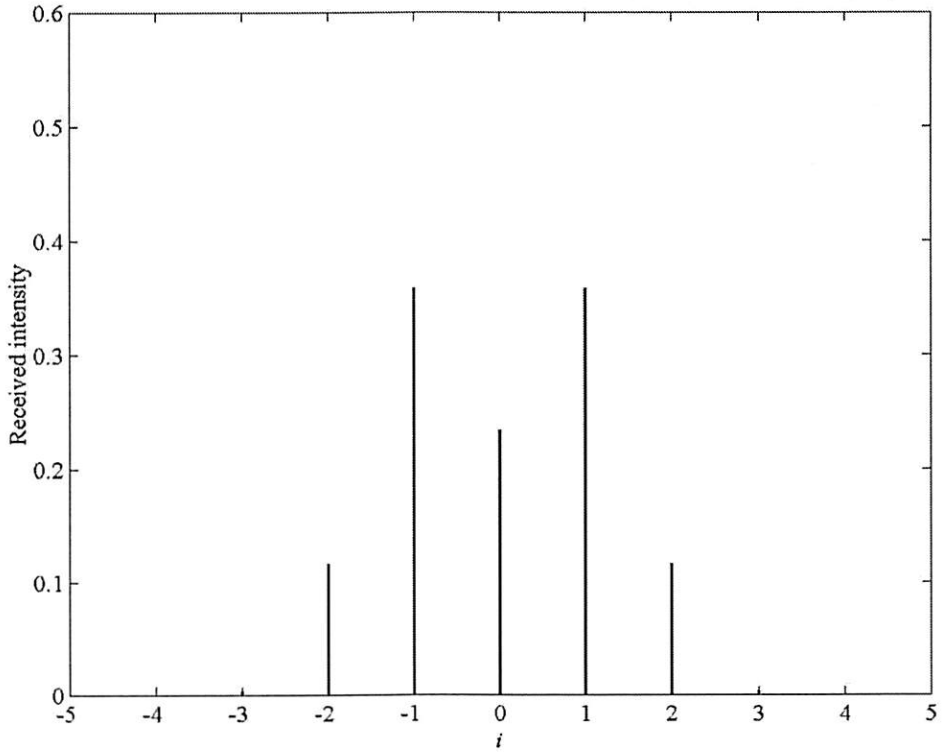


Figure 3.12(b) $r[i,0]$ (Gaussian PSF with $b_g = 0.6$ and $\delta = 1$).

To further support the fact that the relative degree of 2-D ISI depends on the value of b_s, b_c, or b_g, Figures 3.13, 3.14, and 3.15 show the 1-D slices through the x-axis of the image intensities, which originate from two data marks recorded at $(x,y)=(-1,0)$ and $(x,y)=(1,0)$, of the imaging systems with sinc^2, J_1^2, and Gaussian PSFs at various values of b_s, b_c, and b_g, respectively.

It is also important to note that another parameter that can affect the degree of 2-D ISI is the linear fill factor of the detector or CCD pixel. Although the amount of light from a data mark that is integrated by the corresponding detector or CCD pixel increases with increasing the linear fill factor, the amount of the remaining light from the same data mark that is integrated by the neighbor detectors or CCD pixels, i.e., the 2-D ISI, also increases.

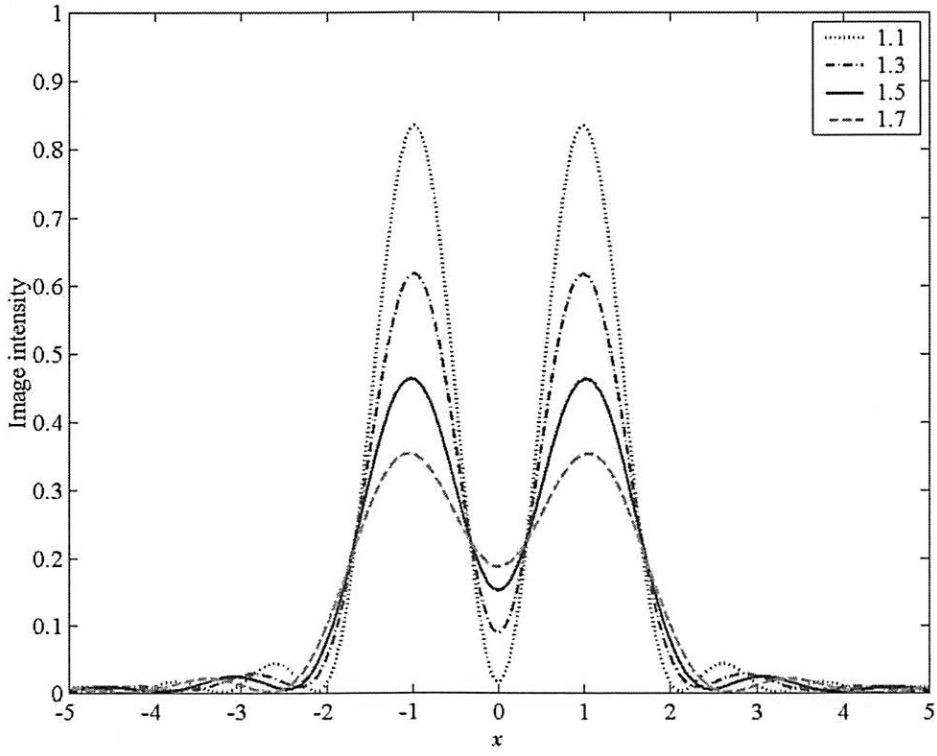


Figure 3.13 Sum of $h(x+1,0)$ and $h(x-1,0)$ at various values of b_s (sinc^2 PSF).

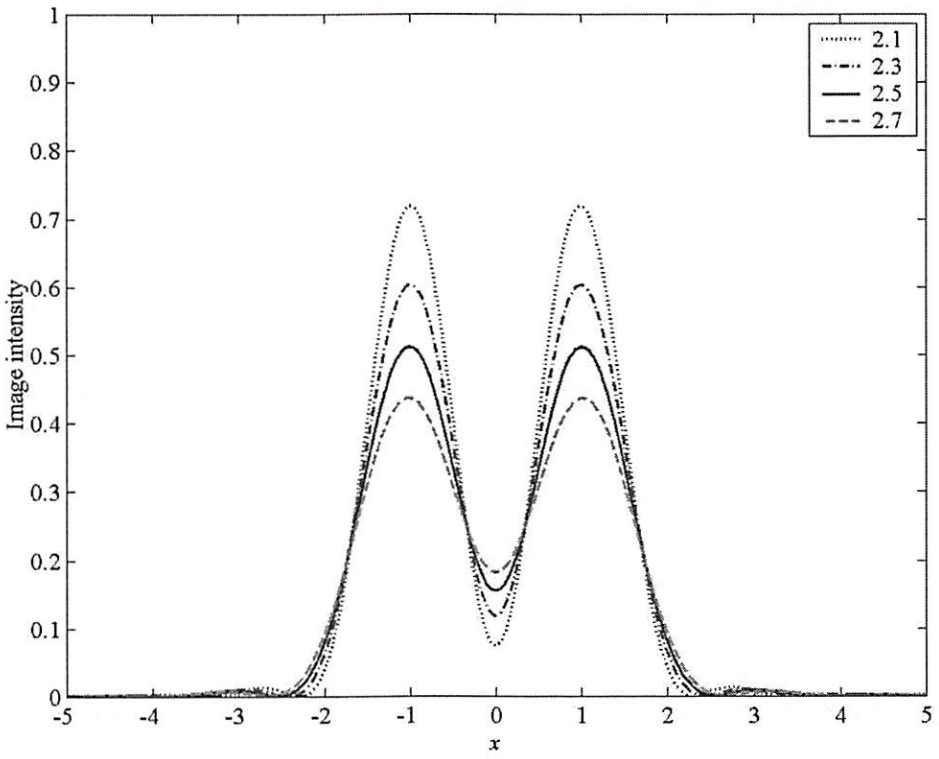


Figure 3.14 Sum of $h(x+1,0)$ and $h(x-1,0)$ at various values of b_c (J_1^2 PSF).

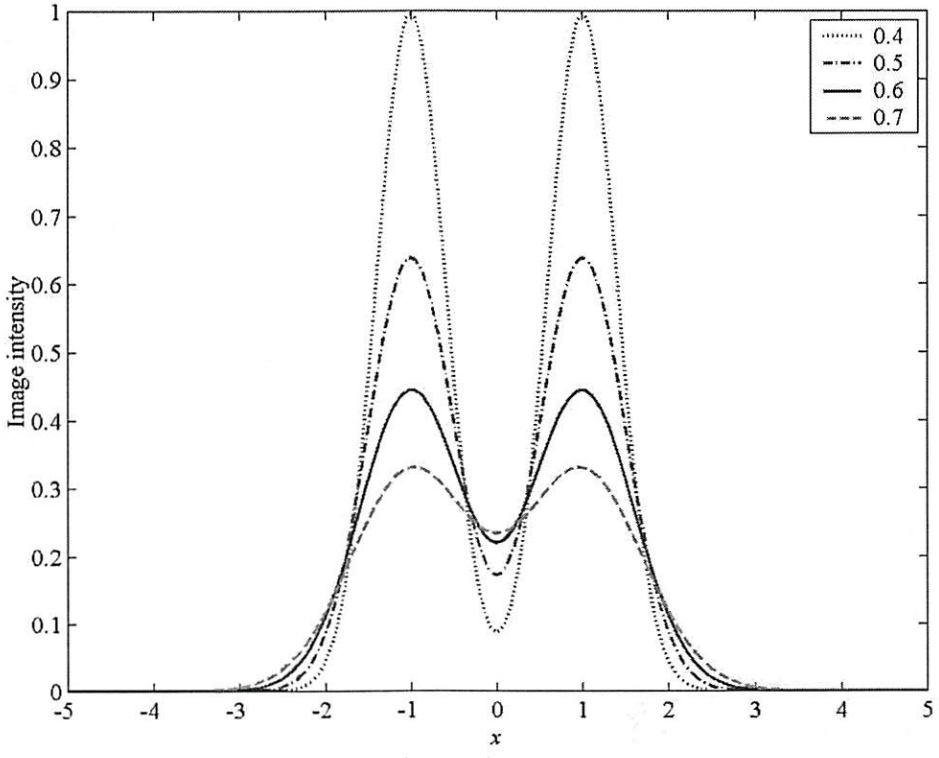


Figure 3.15 Sum of $h(x+1,0)$ and $h(x-1,0)$ at various values of b_g (Gaussian PSF).

Figure 3.16 illustrates the received intensity of each detector that lies on the x-axis, i.e., $r[i,0]$, at many values of the linear fill factor δ, assuming that a retrieval system is noiseless and its imaging system can be represented by a sinc^2 PSF with $b_s = 1.5$ and that only two data marks are recorded at $(x,y) = (-1,0)$ and $(x,y) = (1,0)$. It is clear that as δ increases from 0.8 to 1.0 the integrated intensities of every detector increase, implying that the effects of 2-D ISI also increase with δ. Similarly, Figures 3.17 and 3.18 show the results for the cases of the J_1^2 PSF with $b_c = 2.5$ and Gaussian PSF with $b_g = 0.6$, respectively. Again, if the total number of data marks or "1" bits around the "0" bit at $(x,y) = (0,0)$ is more than two, then the effects of 2-D ISI on that "0" bit is expected to be even worse.

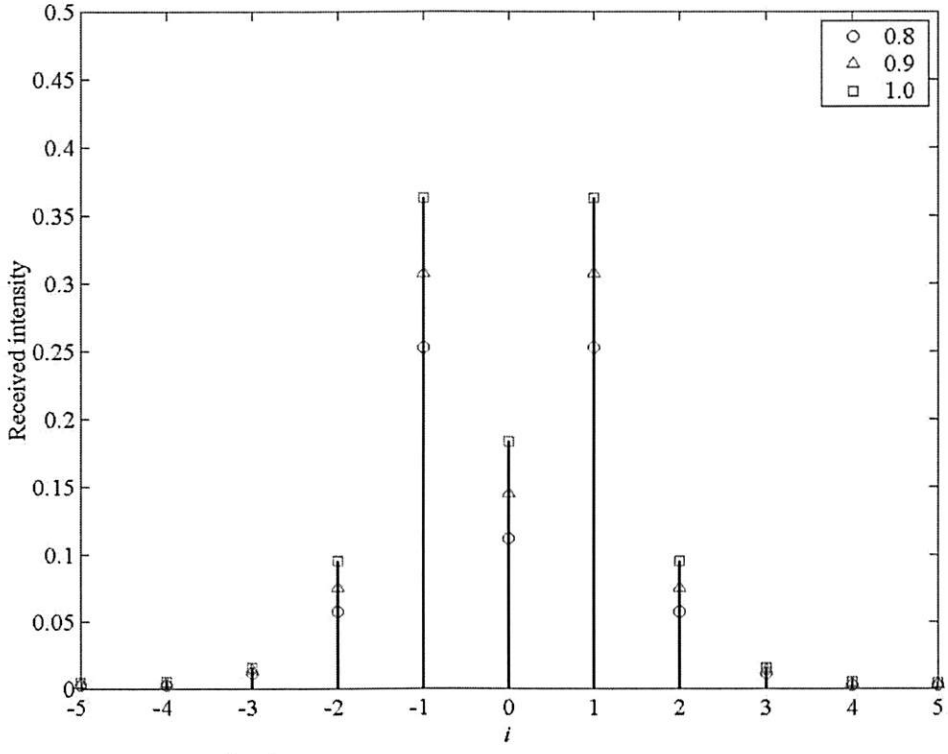


Figure 3.16 $r[i,0]$ at various values of δ (sinc^2 PSF with $b_s = 1.5$).

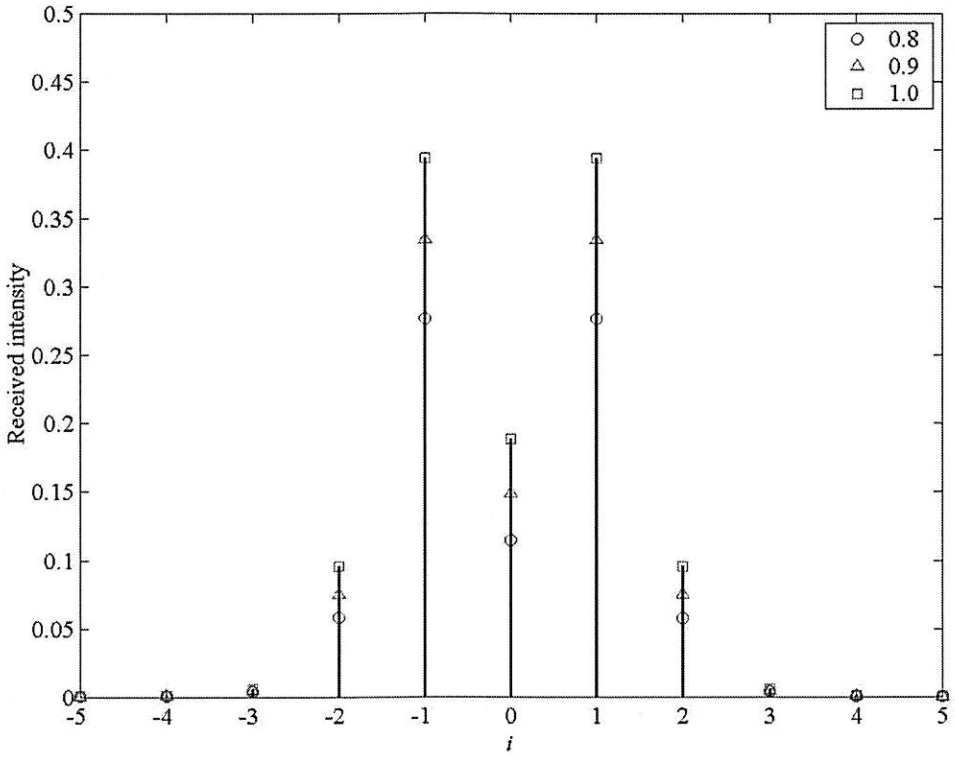


Figure 3.17 $r[i,0]$ at various values of δ (J_1^2 PSF with $b_c = 2.5$).

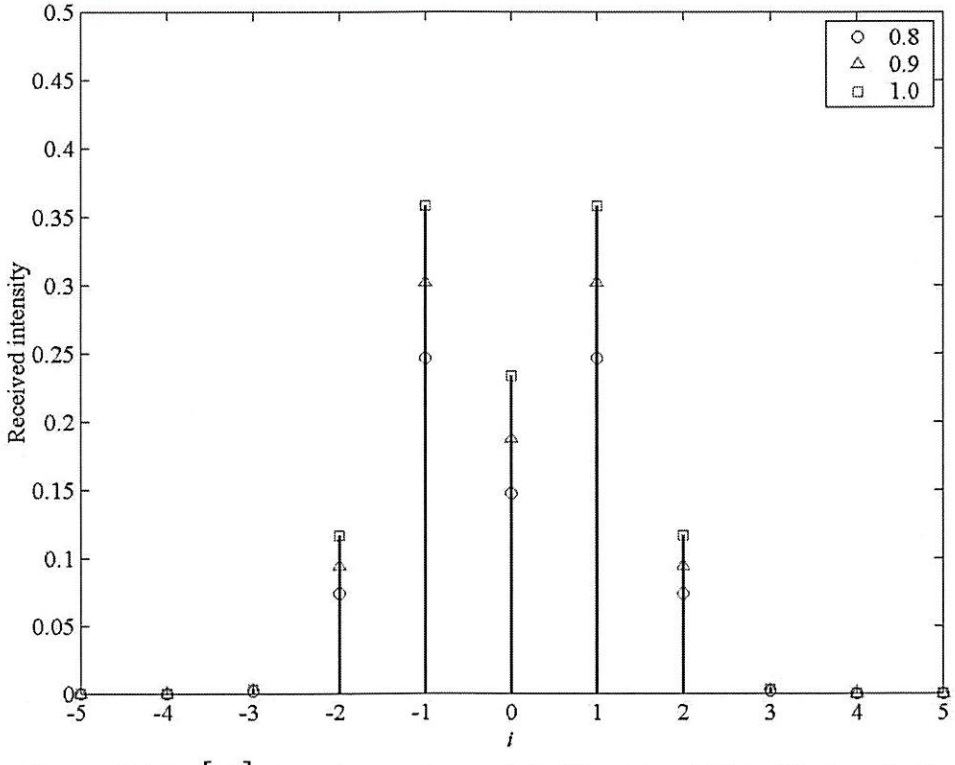


Figure 3.18 $r[i,0]$ at various values of δ (Gaussian PSF with $b_g = 0.6$).

Chapter 4

Two-Dimensional Modulation Coding for Two-Photon PODS Systems: Principles and Fixed-Length Codes

Modulation coding plays a significant role in most data storage systems. It is needed to reduce system errors, especially the effects of intersymbol interference (ISI), and also may increase the useful data storage capacity. Although existing modulation coding schemes are very much robust for today's state-of-the-art data storage systems, prospective data storage systems, such as two-photon page-oriented optical data storage (PODS) systems, require new modulation coding schemes to fulfill future system needs. This chapter is a comprehensive discussion about modulation coding for two-photon PODS systems. We propose in the first section a novel two-dimensional (2-D) modulation coding scheme to mitigate the effects of ISI in two-photon PODS systems. In the second section, we present an example of a 2-D modulation code, derived from the proposed 2-D modulation coding scheme, and its performance improvement. A general form of typical 2-D modulation codes is then described in the third section. With this general form, it is possible to categorize roughly the performance among 2-D modulation codes. Generally, 2-D modulation codes can have either fixed-length or variable-length. The last section of this chapter shows several fixed-length 2-D

modulation codes, while an example of variable-length 2-D modulation codes is presented in the next chapter.

4.1 Basic Principles of Two-Dimensional Modulation Coding Scheme

Obviously, from chapter 3, the simplest way to avoid 2-D ISI due to the imaging system in any two-photon PODS system is to separate the data marks and, consequently, the active areas of the detectors or charge-coupled device (CCD) pixels as far as needed. However, since the data mark pitch is increased, the utilization of the medium decreases. Moreover, since the detector pitch or the CCD pixel pitch is increased while the size of the active area is fixed, the linear fill factor decreases.

Instead of increasing the data mark pitch or detector pitch, it is suggested that modulation coding be applied to reduce the effects of 2-D ISI and, perhaps, increase the useful data storage capacity of a two-photon PODS system. As discussed in the previous chapter, 2-D ISI is regarded as the light from a data mark or a "1" bit that *spatially* interferes with other *local* neighbors, especially the "0" bits. Hence, intuitively, it is desirable to record the data in a two-photon PODS system such that any "0" bit is not surrounded by *too* many "1" bits. In contrast, any "1" bit is preferred to be surrounded by many "0" bits, although "1" bits can be recorded closely. As an example, with the same total number of "1" bits, Eq. (4.1) is the pattern of data that can relax the effects of 2-D ISI more than the one in Eq. (4.2).

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \tag{4.1}$$

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{matrix} \tag{4.2}$$

While the maximum number of "1" bits around any single "0" bit in Eq. (4.2) is eight, there are, at most, three "1" bits surrounding any single "0" bit in Eq. (4.1). The maximum value of the total amount of light from every "1" bits that interferes with any specific "0" bit in Eq. (4.1) thus should be less than that in Eq. (4.2). Note that the light from a "1" bit does not degrade the intensity signals of other neighbor "1" bits; in fact, it adds to the intensity signals of those "1" bits.

Alternatively, because the 2-D ISI can be viewed as a phenomenon that arises from the *lowpass* nature of the imaging system, the recorded data is expected to have low spatial frequency components as much as possible. In other words, it is preferable that the recorded data has a minimum of high spatial frequency components so that the lowpass characteristics of the imaging system do not significantly affect the recorded information [42]. In general, since it is known that any "1" bit *locally* interferes with other "0" bits in two-photon PODS systems, a 3×3 data block with high spatial

frequency content is defined as one having *a "0" bit in the middle surrounded by many "1" bits.* Similarly, a 3×3 data block with lower spatial frequency content is defined as one having *a "0" bit in the middle surrounded by fewer "1" bits.* For instance, the patterns of data in Eq. (4.3) are considered to have higher spatial frequency content compared with the ones in Eq. (4.4).

$$\begin{array}{cccc|cccc|cccc|cccc|cccc} 1&1&1&0&1&0&0&0&0&0&1&0&0&1&1&1&0&1&0&1\\ 1&0&1&0&0&0&1&1&1&1&0&0&0&1&0&1&1&0&1&0\\ 1&1&0&1'&0&1&0&1'&1&0&0&1'&0&1&1&1'&0&1&0&1\\ 0&0&1&0&0&0&1&1&1&1&1&0&0&0&0&0&1&0&1&0 \end{array} \qquad (4.3)$$

$$\begin{array}{cccc|cccc|cccc|cccc|cccc} 0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0&1&1&1&1\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&0&1&1&1&1\\ 0&0&0&0'&0&0&1&0'&0&1&0&0'&0&1&1&0'&1&1&1&1\\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1 \end{array} \qquad (4.4)$$

In conclusion, to reduce the effects of 2-D ISI in two-photon PODS systems, we propose a novel 2-D modulation coding scheme. Its main principle is to map a one-dimensional (1-D) input sequence of information into a 2-D output data in such a way that the spatial frequency content of the output pattern, i.e., the number of "1" bits around a "0" bit in any 3×3 block of data, satisfies the necessary low spatial frequency constraints in order to achieve the required system performance.

4.2 An Example of a Two-Dimensional Modulation Code and Its Performance Improvement

This section describes a particular example of a 2-D modulation code, derived from our 2-D modulation coding scheme, to illustrate possible performance improvements. This 2-D modulation code performs the simple, but non-trivial, input-output mapping [40,43]

$$\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 & a_6 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & a_1 & 0 & 0 \\ a_0 & X & a_3 & a_5 \\ 0 & a_2 & 0 & 0 \\ 0 & a_4 & 0 & a_6 \end{bmatrix}. \tag{4.5}$$

Suppose that there is a lengthy sequence of information bits so that each 1-D block of seven information bits is mapped into a 2-D block of 4×4 bits. If the total number of "1" bits in $\begin{bmatrix} a_0 & a_1 & a_2 & a_3 \end{bmatrix}$ is greater than two, then X is "1"; otherwise X is "0". In later sections, this mapping is called a $(4,4;7;2,4)$ 2-D modulation code according to a general form that we will describe in the next section.

Figure 4.1 shows histograms of the received intensities at the detector array or CCD camera when "0" bits and "1" bits are randomly page-wise recorded in the medium. It is assumed that each data page contains 120×120 bits, 100 pages of data are recorded, each detector or CCD pixel has a unit linear fill factor, and the retrieval system, containing a square aperture with $b_s = 1.5$, is noiseless, i.e., no additive white Gaussian noise (AWGN) is considered. Clearly, the histogram of the received intensities that originate from "0" bits overlaps with that from "1" bits; this, indeed, is a

consequence of the effects of 2-D ISI. Therefore, it is not possible to find an explicit threshold value for a binary threshold decision scheme that always gives correct results.

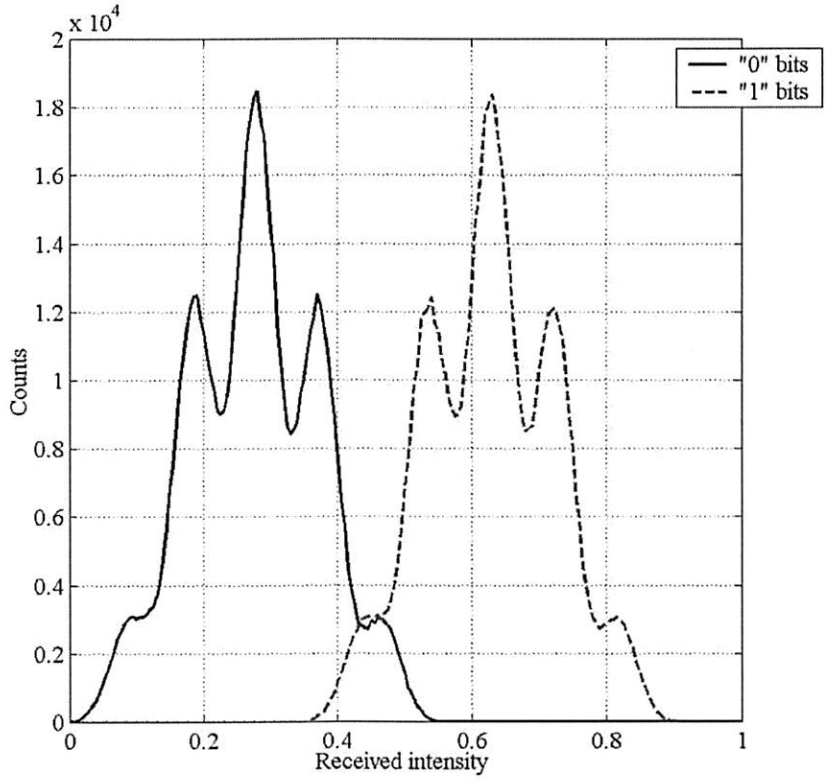


Figure 4.1 Histograms of the received intensities that originate from randomly page-wise recorded "0" and "1" bits, assuming that 100 data pages of 120×120 bits are recorded and that a noiseless retrieval system is characterized by the sinc^2 PSF with $b_s = 1.5$ and $\delta = 1$.

However, if the proposed 2-D modulation coding scheme is applied, the effects of 2-D ISI are expected to be reduced. Figure 4.2 shows the histograms of the received intensities at the detector array or CCD camera after the mapping in Eq. (4.5) is applied. As before, we assume that each data page contains 120×120 bits, 100 pages of data are recorded, each detector or CCD pixel has a unit linear fill factor, and the retrieval

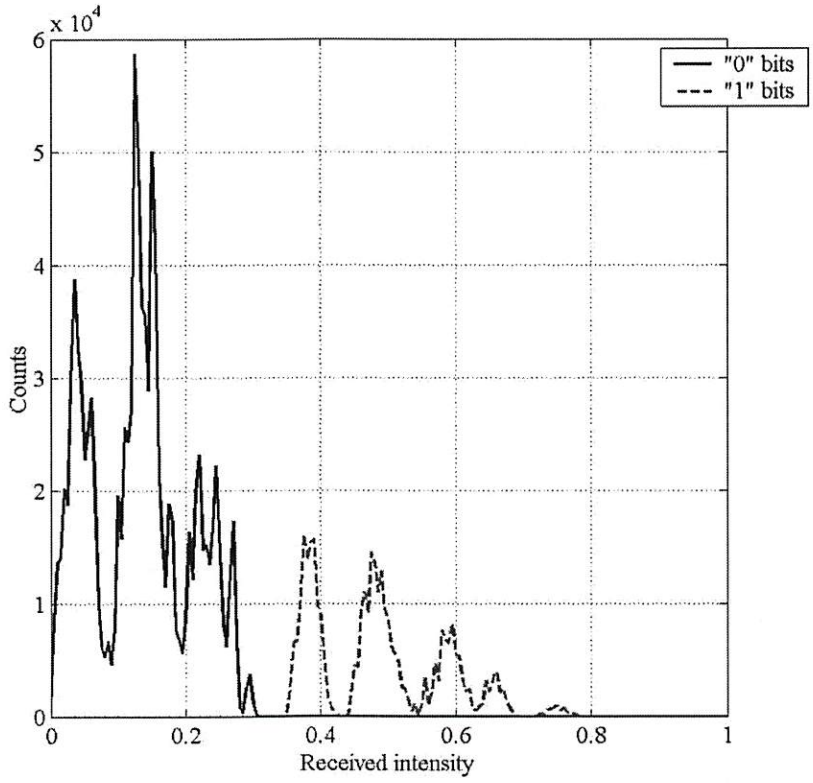


Figure 4.2 Histograms of the received intensities that originate from "0" and "1" bits after applying the mapping in Eq. (4.5), assuming that 100 data pages of 120×120 bits are recorded and that a noiseless retrieval system is characterized by the sinc^2 PSF with $b_s = 1.5$ and $\delta = 1$.

system, containing a square aperture with $b_s = 1.5$, is noiseless, i.e., the AWGN is not considered. It is apparent that, after applying the mapping, there is no overlap between the histogram of the received intensities that originate from "0" bits and that from "1" bits. Hence, it is ideally possible to find a threshold value for a binary threshold decision scheme such that the decisions are always correct.

There are several important aspects of Figure 4.2 that are worth discussing. First, note that the mapping in Eq. (4.5) must be applied to a randomly generated input sequence of 630,000 information bits to obtain 100 data pages of 120×120 bits to be recorded. Thus, the amount of the useful information is 43.75% of the total amount of the recorded data. On the other hand, since the mapping in Eq. (4.5) maps seven input bits into sixteen output bits, the fraction of useful information is 7/16 (= 0.4375). The overhead is the cost of lessening the effects of 2-D ISI.

In addition, it should be noted that the total number of counts in the histogram of the received intensities that originate from "0" bits, as expected from Eq. (4.5), is much higher than that in the histogram of the received intensities that originate from "1" bits. The extra "0" bits in Eq. (4.5) are used to mitigate the effects of 2-D ISI; to be more precise, they are padded in such a way that, in any block of 3×3 bits of each recorded data page, the maximum number of "1" bits around any single "0" bit is six. Two examples of such extreme cases are illustrated in Figure 4.3. The squares locate the 3×3 blocks of recorded data with the highest spatial frequency content, i.e., the maximum number of "1" bits around any single "0" bit. From Figure 4.3, the highest spatial frequency content may be contained in any 2-D block of 4×4 bits, i.e., any output block of the mapping, itself, or it may arise after any 2-D block of 4×4 bits is recorded with a specific set of neighbor 2-D blocks of 4×4 bits.

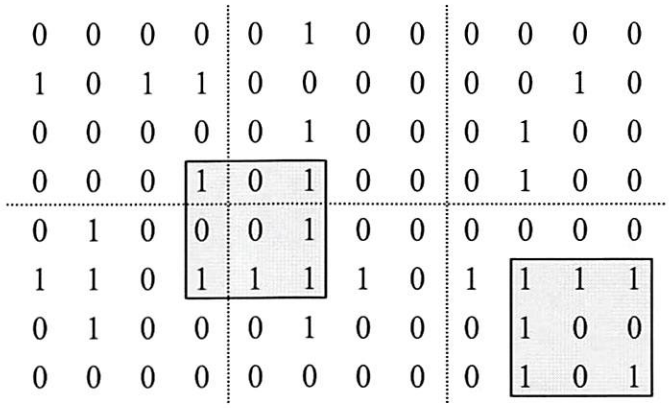


Figure 4.3 The examples of a 3×3 block of recorded data with the highest spatial frequency content after mapping by Eq. (4.5).

In practice, when the retrieval system is noisy, i.e., when the *random* effects of AWGN are included, the tails of the histograms always overlap, no matter whether the 2-D modulation coding scheme is applied or not, leading to an inexplicit threshold value. Nevertheless, the proposed 2-D modulation coding scheme is believed to significantly improve the system performance with the idea of lowering the *deterministic-like* effects of 2-D ISI.

Furthermore, we show in Figures 4.4 and 4.5 the histograms of the received intensities at the detector array or CCD camera before and after the mapping in Eq. (4.5) is applied, when the retrieval system is assumed to be characterized by the J_1^2 point-spread function (PSF) with $b_c = 2.5$, and other assumptions still remain the same as in Figures 4.1 and 4.2. The results in Figures 4.4 and 4.5 are similar to those in Figures 4.1 and 4.2, respectively. Figures 4.6 and 4.7, then, show the histograms of the received intensities before and after applying the mapping in Eq. (4.5), when the retrieval system

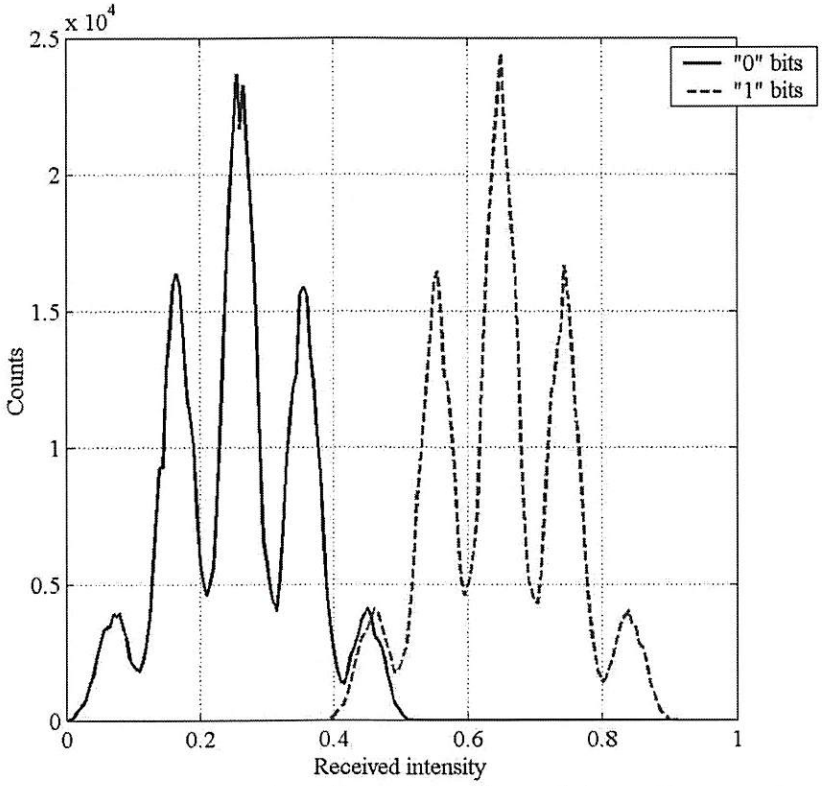


Figure 4.4 Histograms of the received intensities that originate from randomly page-wise recorded "0" and "1" bits, assuming that 100 data pages of 120×120 bits are recorded and that a noiseless retrieval system is characterized by the J_1^2 PSF with $b_c = 2.5$ and $\delta = 1$.

is assumed to be characterized by the Gaussian PSF with $b_g = 0.6$, and other assumptions remain the same as in Figures 4.1, 4.2, 4.4, and 4.5. In this particular example, the overlapping area of the histograms of the received intensities that originate from randomly page-wise recorded "0" and "1" bits in Figure 4.6 is very large, intuitively, because of a relatively large value of b_g, which implies a very high degree of 2-D ISI; hence, as shown in Figure 4.7, the mapping in Eq. (4.5) cannot overcome the effects of very high degree 2-D ISI, yet results in smaller overlapping area, i.e., less 2-D

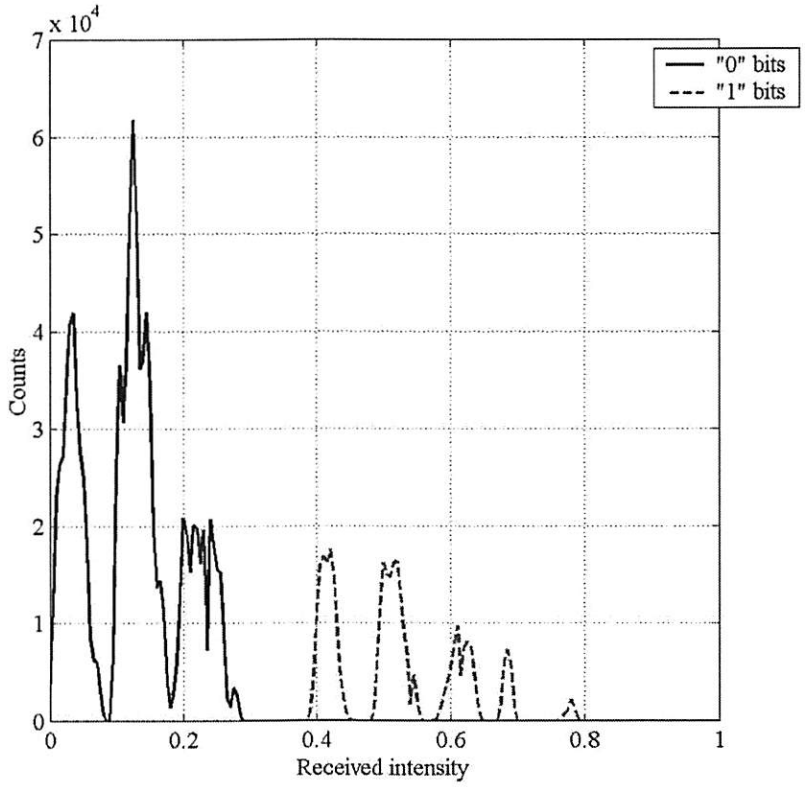


Figure 4.5 Histograms of the received intensities that originate from "0" and "1" bits after applying the mapping in Eq. (4.5), assuming that 100 data pages of 120×120 bits are recorded and that a noiseless retrieval system is characterized by the J_1^2 PSF with $b_c = 2.5$ and $\delta = 1$.

ISI effects. Based on our 2-D modulation coding scheme principles, in order to move the histograms in Figure 4.7 further apart, i.e., to further relieve the effects of 2-D ISI, the highest spatial frequency content allowed, i.e., the maximum number of "1" bits allowed around any single "0" bit in any 3×3 data block, in the recorded data pages, therefore, must be reduced; instead of Eq. (4.5), a new mapping is required.

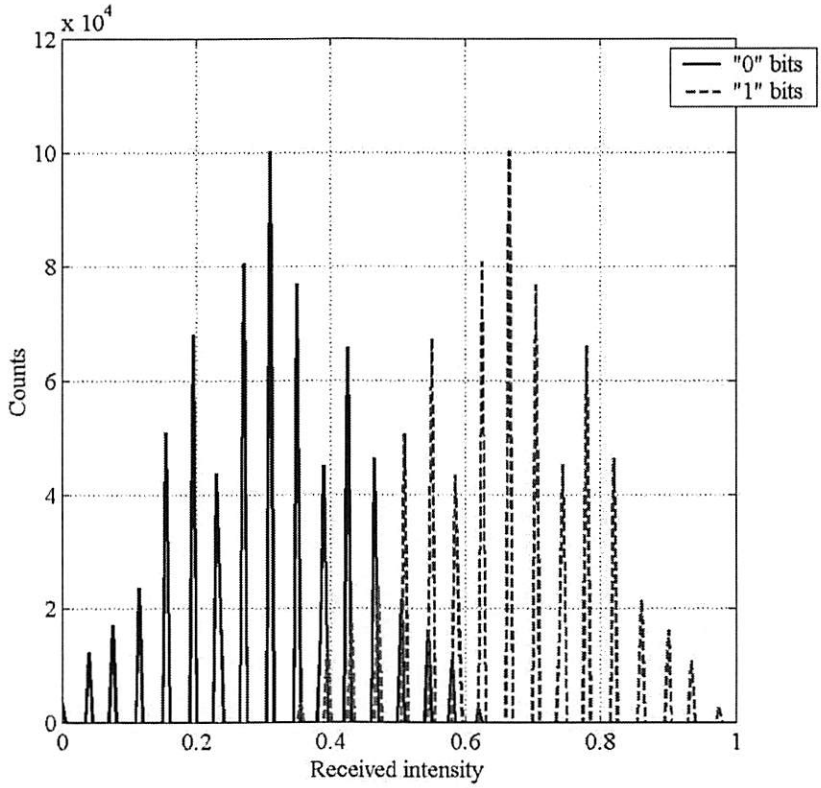


Figure 4.6 Histograms of the received intensities that originate from randomly page-wise recorded "0" and "1" bits, assuming that 100 data pages of 120×120 bits are recorded and that a noiseless retrieval system is characterized by the Gaussian PSF with $b_g = 0.6$ and $\delta = 1$.

Finally, to simulate the received intensities at the detector array or CCD camera, it should be emphasized that although the continuous PSFs, $h(x, y)$, in Eqs. (3.1), (3.2), and (3.3) spread infinitely, the effective PSF matrix, $\mathbf{H}$, in Eq. (3.8) is truncated to a finite-size matrix in the actual implementation to reduce the computational complexity in the simulation procedure. This truncation does not strongly affect the simulation results because it is known from chapter 3 that most of the light from a data mark is incident onto the corresponding detector or CCD pixel and its local neighborhood.

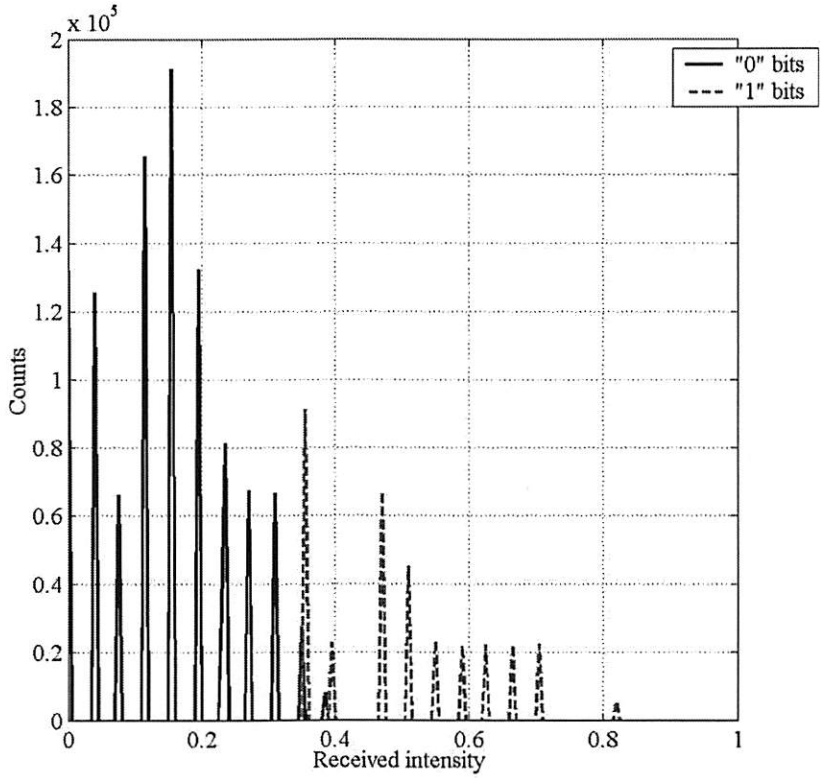


Figure 4.7 Histograms of the received intensities that originate from "0" and "1" bits after applying the mapping in Eq. (4.5), assuming that 100 data pages of 120×120 bits are recorded and that a noiseless retrieval system is characterized by the Gaussian PSF with $b_g = 0.6$ and $\delta = 1$.

In detail, we define the effective PSF matrix, $\mathbf{H}$, in a way such that it is the smallest possible $K \times K$ matrix, where K is an odd number, that contains *at least* 90% of the total light originating from a data mark. For instances, if a retrieval system is characterized by the sinc^2 PSF with $b_s = 1.5$ and $\delta = 1$, the corresponding effective PSF matrix is

$$\begin{bmatrix} 0.0000 & 0.0001 & 0.0008 & 0.0031 & 0.0008 & 0.0001 & 0.0000 \\ 0.0001 & 0.0004 & 0.0032 & 0.0120 & 0.0032 & 0.0004 & 0.0001 \\ 0.0008 & 0.0032 & 0.0241 & 0.0920 & 0.0241 & 0.0032 & 0.0008 \\ 0.0031 & 0.0120 & 0.0920 & 0.3507 & 0.0920 & 0.0120 & 0.0031 \\ 0.0008 & 0.0032 & 0.0241 & 0.0920 & 0.0241 & 0.0032 & 0.0008 \\ 0.0001 & 0.0004 & 0.0032 & 0.0120 & 0.0032 & 0.0004 & 0.0001 \\ 0.0000 & 0.0001 & 0.0008 & 0.0031 & 0.0008 & 0.0001 & 0.0000 \end{bmatrix}, \quad (4.6)$$

in which it contains 90.99% of the total light originating from a data mark. Noticeably, about 35.07% of the light is imaged onto the corresponding detector or CCD pixel (at the center of the matrix).

If a retrieval system is characterized by the J_1^2 PSF with $b_c = 2.5$ and $\delta = 1$, the effective PSF matrix is

$$\begin{bmatrix} 0.0014 & 0.0051 & 0.0054 & 0.0051 & 0.0014 \\ 0.0051 & 0.0194 & 0.0942 & 0.0194 & 0.0051 \\ 0.0054 & 0.0942 & 0.3888 & 0.0942 & 0.0054 \\ 0.0051 & 0.0194 & 0.0942 & 0.0194 & 0.0051 \\ 0.0014 & 0.0051 & 0.0054 & 0.0051 & 0.0014 \end{bmatrix}, \quad (4.7)$$

where it contains 91.11% of the total light originating from a data mark and the corresponding detector or CCD pixel receives about 38.88% of the total. Similarly, if a retrieval system is characterized by the Gaussian PSF with $b_g = 0.6$ and $\delta = 1$, the effective PSF matrix is

$$\begin{bmatrix} 0.0385 & 0.1168 & 0.0385 \\ 0.1168 & 0.3544 & 0.1168 \\ 0.0385 & 0.1168 & 0.0385 \end{bmatrix}, \quad (4.8)$$

in which it contains 97.53% of the total light originating from a data mark and the corresponding detector or CCD pixel receives about 35.44% of the total. Notice that the sizes of the effective PSF matrices are varied depending on the characteristics of the interested retrieval systems.

4.3 General Form of Two-Dimensional Modulation Codes

As discussed in sections 4.1 and 4.2, the recorded data pages in a two-photon PODS system should satisfy certain spatial frequency constraints in order to achieve the required system performance. Since, for a given two-photon PODS system, the highest spatial frequency content, i.e., the maximum number of "1" bits around a "0" bit in any data block of 3×3 bits in the recorded data page implies the degree of 2-D ISI effects, the general form of the 2-D modulation codes, derived from our 2-D modulation coding scheme, should include information about the highest spatial frequency content that is allowed in any recorded data page.

We also know from the properties of the PSFs in Eqs. (3.1), (3.2), and (3.3) that the amount of interference light originating from a data mark incident onto each of the corresponding four nearest neighbor detectors or CCD pixels is always greater than the amount of interference light originating from the same data mark incident onto each of the other neighbor detectors or CCD pixels (the effective PSF matrices in Eqs. (4.6), (4.7), and (4.8) also support this fact). Moreover, since the effects of 2-D ISI are considered to be local, we suggest that the general form of typical 2-D modulation codes

contain information about the maximum number of "1" bits in the four nearest neighbor positions around any "0" bit in the recorded data page and the maximum number of "1" bits in the next four nearest neighbor positions around the same "0" bit.

Let us label the neighbor positions around a "0" bit in a data block of 3×3 bits by the compass directions: north (N), northeast (NE), east (E), southeast (SE), south (S), southwest (SW), west (W), and northwest (NW), as shown in Eq. (4.9), where, similar to the terminology in morphological image processing, the four nearest neighbor positions, i.e., N, E, S, and W, are referred to as the four-connected neighbor positions, and the NE, SE, SW, and NW positions are referred to as the eight-connected neighbor positions [49].

$$\begin{matrix} \text{NW} & \text{N} & \text{NE} \\ \text{W} & 0 & \text{E} \\ \text{SW} & \text{S} & \text{SE} \end{matrix} \tag{4.9}$$

An arbitrary 2-D modulation code for two-photon PODS system that maps k bits of 1-D input information into a 2-D output block of $m\times n$ bits (shown in Figure 4.8), therefore, can be described by

$$(m,n;k;\alpha,\beta), \tag{4.10}$$

where α is the maximum number of "1" bits in the four-connected neighbor positions of any "0" bit and β is the maximum number of "1" bits in the eight-connected neighbor positions of the same "0" bit. It is necessary to note that α and β are not only the constraints in each 2-D $m\times n$ output block itself, but they are also the constraints after

the 2-D output blocks are placed side-by-side in a 2-D array that forms a tiling of the recorded data page.

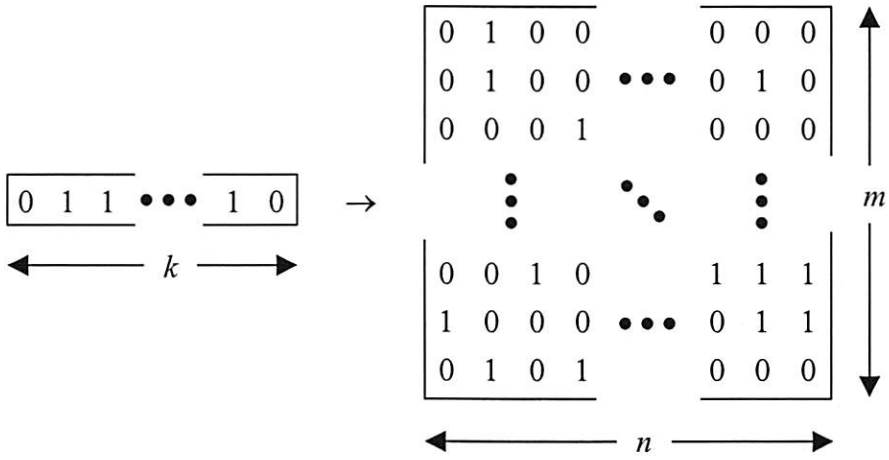


Figure 4.8 An $(m,n;k;\alpha,\beta)$ 2-D modulation code.

We also emphasize that since the effects from "1" bits in the four-connected neighbor positions of a "0" bit are stronger than those from "1" bits in the eight-connected neighbor positions, then the (α,β) in Eq. (4.10) is defined such that, after applying the $(m,n;k;\alpha,\beta)$ 2-D modulation code, it is possible for the recorded data page to contain a "0" bit that has more than β "1" bits in the eight-connected neighbor positions as long as the number of "1" bits in the four-connected neighbor positions is less than α. For instance, if the $(m,n;k;3,1)$ 2-D modulation code is applied, then a recorded data page that contains two "1" bits in the four-connected neighbor positions of a "0" bit and four "1" bits in the eight-connected neighbor positions of the same "0" bit still satisfies the given constraints: $\alpha = 3$ and $\beta = 1$.

Mathematically, the (α,β) of the $(m,n;k;\alpha,\beta)$ 2-D modulation code can be described as the following. Let $\hat{\alpha}$ and $\hat{\beta}$ be the number of "1" bits in the four-connected and eight-connected neighbor positions of any particular "0" bit, respectively. The $(\hat{\alpha},\hat{\beta})$ allowed in the recorded data page must satisfy

$$\begin{cases} \hat{\alpha} \le \alpha \\ \hat{\beta} \le (\beta-4)\hat{\alpha} + (4\alpha - \alpha\beta + \beta) \end{cases}, \tag{4.11}$$

where $\hat{\alpha},\hat{\beta} \in \{0,1,2,3,4\}$.

As an example, Figure 4.9 graphically depicts all possible combinations of $(\hat{\alpha},\hat{\beta})$, represented by the circles, and all combinations of $(\hat{\alpha},\hat{\beta})$ that satisfy the constraints of the $(m,n;k;3,1)$ 2-D modulation code, represented by the squares. Explicitly, any valid $(\hat{\alpha},\hat{\beta})$ agrees with

$$\begin{cases} \hat{\alpha} \le 3 \\ \hat{\beta} \le -3\hat{\alpha} + 10 \end{cases}. \tag{4.12}$$

We note that many essential issues of Eq. (4.10) should be discussed. First, the code rate of 2-D modulation code, which is defined as the fraction of the useful input information in the output, is equal to $k/(mn)$. The code rate is one of the criteria used to compare the performance of 2-D modulation codes. Second, in general, the size of each data page recorded in a two-photon PODS system, $M \times N$, is a positive integer multiple of m and n, i.e., $M = um$ and $N = vn$, where $u,v \in \{1,2,3,...\}$, although this is not a necessary condition. Third, the block size of a 2-D modulation code, specified by the general form in Eq. (4.10), is fixed. In other words, every 2-D output block of the

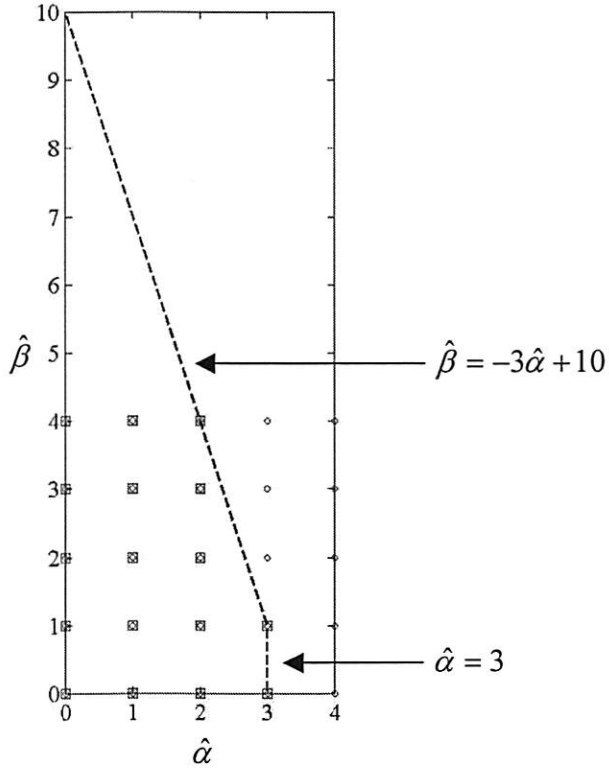


Figure 4.9 The points marked by a small square box are the possible combinations of $(\hat{\alpha}, \hat{\beta})$ that satisfy the constraints of the $(m, n; k; 3, 1)$ 2-D modulation code.

$(m, n; k; \alpha, \beta)$ 2-D modulation code has equal size, $m \times n$. Thus, 2-D modulation codes that can be described by Eq. (4.10) are generally called the *fixed-length* 2-D modulation codes. On the other hand, it is also possible to have 2-D modulation codes with variable block sizes. In such situation, the general form in Eq. (4.10) requires modifications.

4.4 Examples of Fixed-Length Two-Dimensional Modulation Codes

We present in this section many fixed-length 2-D modulation codes with different block sizes, code rates, and (α, β) constraints. Table 4.1 at the end of this chapter lists all the 2-D modulation codes described in this section.

4.4.1 (4,4;7;2,4) Two-Dimensional Modulation Code

In section 4.2, a 2-D modulation code that maps a 1-D block of seven information bits into a 2-D block of 4×4 bits is presented; it is indeed a fixed-length 2-D modulation code. Using the general form described in the previous section, it can be called the $(4,4;7;2,4)$ 2-D modulation code. We note that the code rate of this 2-D modulation code is equal to 7/16 (= 0.4375) and that a few examples of its highest spatial frequency content, which is $\alpha = 2$ and $\beta = 4$, are shown in section 4.2.

4.4.2 (3,3;4;2,3) Two-Dimensional Modulation Code

$$[a_0 \quad a_1 \quad a_2 \quad a_3] \rightarrow \begin{bmatrix} 0 & a_1 & 0 \\ a_0 & X & a_3 \\ 0 & a_2 & 0 \end{bmatrix} \tag{4.13}$$

The $(3,3;4;2,3)$ 2-D modulation code in Eq. (4.13) maps a 1-D block of four information bits into a 2-D block of 3×3 bits [42]. If the total number of "1" bits in $[a_0 \quad a_1 \quad a_2 \quad a_3]$ is greater than two, then X is "1"; otherwise X is "0". The entire lookup table of this 2-D modulation code is shown in Figure 4.10.

$$\begin{array}{cc} & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 & 0 \end{matrix} \rightarrow & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ & \begin{matrix} 0 & 0 & 0 \end{matrix} \end{array} \qquad \begin{array}{cc} & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 1 & 0 & 0 & 0 \end{matrix} \rightarrow & \begin{matrix} 1 & 0 & 0 \end{matrix} \\ & \begin{matrix} 0 & 0 & 0 \end{matrix} \end{array}$$

$$\begin{array}{cc} & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 0 & 1 \end{matrix} \rightarrow & \begin{matrix} 0 & 0 & 1 \end{matrix} \\ & \begin{matrix} 0 & 0 & 0 \end{matrix} \end{array} \qquad \begin{array}{cc} & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 1 & 0 & 0 & 1 \end{matrix} \rightarrow & \begin{matrix} 1 & 0 & 1 \end{matrix} \\ & \begin{matrix} 0 & 0 & 0 \end{matrix} \end{array}$$

$$\begin{array}{cc} & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 1 & 0 \end{matrix} \rightarrow & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ & \begin{matrix} 0 & 1 & 0 \end{matrix} \end{array} \qquad \begin{array}{cc} & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 1 & 0 & 1 & 0 \end{matrix} \rightarrow & \begin{matrix} 1 & 0 & 0 \end{matrix} \\ & \begin{matrix} 0 & 1 & 0 \end{matrix} \end{array}$$

$$\begin{array}{cc} & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 0 & 0 & 1 & 1 \end{matrix} \rightarrow & \begin{matrix} 0 & 0 & 1 \end{matrix} \\ & \begin{matrix} 0 & 1 & 0 \end{matrix} \end{array} \qquad \begin{array}{cc} & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ \begin{matrix} 1 & 0 & 1 & 1 \end{matrix} \rightarrow & \begin{matrix} 1 & 1 & 1 \end{matrix} \\ & \begin{matrix} 0 & 1 & 0 \end{matrix} \end{array}$$

$$\begin{array}{cc} & \begin{matrix} 0 & 1 & 0 \end{matrix} \\ \begin{matrix} 0 & 1 & 0 & 0 \end{matrix} \rightarrow & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ & \begin{matrix} 0 & 0 & 0 \end{matrix} \end{array} \qquad \begin{array}{cc} & \begin{matrix} 0 & 1 & 0 \end{matrix} \\ \begin{matrix} 1 & 1 & 0 & 0 \end{matrix} \rightarrow & \begin{matrix} 1 & 0 & 0 \end{matrix} \\ & \begin{matrix} 0 & 0 & 0 \end{matrix} \end{array}$$

$$\begin{array}{cc} & \begin{matrix} 0 & 1 & 0 \end{matrix} \\ \begin{matrix} 0 & 1 & 0 & 1 \end{matrix} \rightarrow & \begin{matrix} 0 & 0 & 1 \end{matrix} \\ & \begin{matrix} 0 & 0 & 0 \end{matrix} \end{array} \qquad \begin{array}{cc} & \begin{matrix} 0 & 1 & 0 \end{matrix} \\ \begin{matrix} 1 & 1 & 0 & 1 \end{matrix} \rightarrow & \begin{matrix} 1 & 1 & 1 \end{matrix} \\ & \begin{matrix} 0 & 0 & 0 \end{matrix} \end{array}$$

$$\begin{array}{cc} & \begin{matrix} 0 & 1 & 0 \end{matrix} \\ \begin{matrix} 0 & 1 & 1 & 0 \end{matrix} \rightarrow & \begin{matrix} 0 & 0 & 0 \end{matrix} \\ & \begin{matrix} 0 & 1 & 0 \end{matrix} \end{array} \qquad \begin{array}{cc} & \begin{matrix} 0 & 1 & 0 \end{matrix} \\ \begin{matrix} 1 & 1 & 1 & 0 \end{matrix} \rightarrow & \begin{matrix} 1 & 1 & 0 \end{matrix} \\ & \begin{matrix} 0 & 1 & 0 \end{matrix} \end{array}$$

$$\begin{array}{cc} & \begin{matrix} 0 & 1 & 0 \end{matrix} \\ \begin{matrix} 0 & 1 & 1 & 1 \end{matrix} \rightarrow & \begin{matrix} 0 & 1 & 1 \end{matrix} \\ & \begin{matrix} 0 & 1 & 0 \end{matrix} \end{array} \qquad \begin{array}{cc} & \begin{matrix} 0 & 1 & 0 \end{matrix} \\ \begin{matrix} 1 & 1 & 1 & 1 \end{matrix} \rightarrow & \begin{matrix} 1 & 1 & 1 \end{matrix} \\ & \begin{matrix} 0 & 1 & 0 \end{matrix} \end{array}$$

Figure 4.10 Lookup table of the $(3,3;4;2,3)$ 2-D modulation code.

Obviously, the code rate of the $(3,3;4;2,3)$ 2-D modulation code is 4/9 (= 0.4444). The values of α and β, i.e., the maximum numbers of "1" bits allowed in the four-connected and eight-connected neighbor positions of any specific "0" bit, are two and three, respectively. Figure 4.11 illustrates some examples of the possible highest spatial frequency content after applying the $(3,3;4;2,3)$ 2-D modulation code.

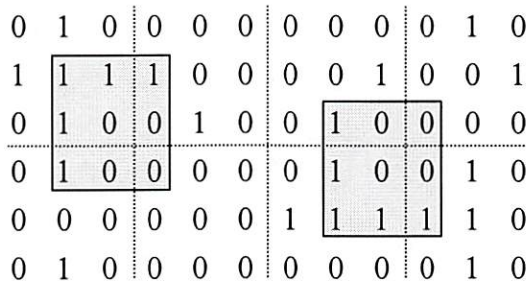


Figure 4.11 The examples of a 3×3 block of recorded data with the highest spatial frequency content after applying the $(3,3;4;2,3)$ 2-D modulation code.

Note that the $(3,3;4;2,3)$ 2-D modulation code should outperform the $(4,4;7;2,4)$ 2-D modulation code because the former has lower highest spatial frequency content and a higher code rate. Nevertheless, the latter might be preferred when the size of the recorded data page is restricted to a power of two, which is usually the case for typical PODS systems.

4.4.3 (1,2;1;2,4) Two-Dimensional Modulation Code

$$a \rightarrow [0 \quad a] \quad \text{or} \quad a \rightarrow [a \quad 0] \tag{4.14}$$

The $(1,2;1;2,4)$ 2-D modulation code in Eq. (4.14) maps one information bit into two bits by padding the extra "0" bit before or after the information bit. The code rate of this 2-D modulation code is, hence, equal to 0.5. Examples of the possible highest spatial frequency content, which is two "1" bits and four "1" bits in the four-connected and eight-connected neighbor positions of a "0" bit, respectively, after applying the $(1,2;1;2,4)$ 2-D modulation code are illustrated in Figure 4.12. Evidently, since the $(1,2;1;2,4)$ 2-D modulation code has higher code rate than the $(4,4;7;2,4)$ 2-D modulation code and their highest spatial frequency contents allowed are the same, the $(1,2;1;2,4)$ 2-D modulation code, thus, is expected to result in better overall performance.

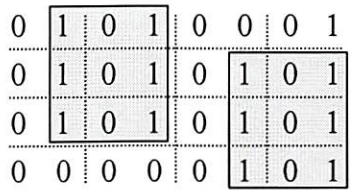


Figure 4.12 The examples of a 3×3 block of recorded data with the highest spatial frequency content after applying the $(1,2;1;2,4)$ 2-D modulation.

4.4.4 (4,4;8;2,2) Two-Dimensional Modulation Code

$$[a_0 \quad a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad a_7] \rightarrow \begin{bmatrix} a_0 & a_2 & 0 & 0 \\ a_1 & a_3 & 0 & 0 \\ 0 & 0 & a_4 & a_6 \\ 0 & 0 & a_5 & a_7 \end{bmatrix} \tag{4.15}$$

The $(4,4;8;2,2)$ 2-D modulation code in Eq. (4.15) maps a 1-D block of eight information bits into a 2-D block of 4×4 bits [44]. This 2-D modulation code has a code rate of 0.5 and its highest spatial frequency content contains two "1" bits in the four-connected neighbor positions of a "0" bit and the other two "1" bits in the eight-connected neighbor positions of the same "0" bit. Figure 4.13 shows several examples of the possible highest spatial frequency content after applying the $(4,4;8;2,2)$ 2-D modulation code.

0	1	0	0	0	0	0	0	1	0	0	0
0	1	0	0	0	1	0	0	0	1	0	0
0	0	1	1	0	0	0	0	0	0	0	1
0	0	1	0	0	0	0	1	0	0	0	1
0	0	0	0	0	1	0	0	0	1	0	0
0	0	0	0	0	1	0	0	1	1	0	0
0	0	0	1	0	0	0	0	0	0	0	0
0	0	0	0	0	0	1	1	0	0	0	1

Figure 4.13 The examples of a 3×3 block of recorded data with the highest spatial frequency content after applying the $(4,4;8;2,2)$ 2-D modulation code.

Despite the fact that the $(4,4;8;2,2)$ 2-D modulation code may give better performance, compared with the $(4,4;7;2,4)$, $(3,3;4;2,3)$, and $(1,2;1;2,4)$ 2-D modulation codes, in terms of code rate and highest spatial frequency content, another important criterion that should be taken into account for the overall performance is the complexity of the decoder, which is related to the size of the buffer needed to rearrange the 2-D detected data into the original 1-D information and the total delay (latency) during the decoding process.

4.4.5 (3,3;6;3,2) Two-Dimensional Modulation Code

$$\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & a_5 \end{bmatrix} \rightarrow \begin{bmatrix} a_0 & 0 & a_3 \\ a_1 & 0 & a_4 \\ a_2 & 0 & a_5 \end{bmatrix} \tag{4.16}$$

The $(3,3;6;3,2)$ 2-D modulation code in Eq. (4.16) maps a 1-D block of six information bits into a 2-D block of 3×3 bits. The code rate of this 2-D modulation code is 2/3 (= 0.6667). The maximum numbers of "1" bits allowed in the four-connected and eight-connected neighbor positions of a "0" bit are three and two, respectively. Examples of the possible highest spatial frequency content after applying the $(3,3;6;3,2)$ 2-D modulation code are shown in Figure 4.14.

It is clear that the $(3,3;6;3,2)$ 2-D modulation code should not be able to relax the effects of 2-D ISI as much as the $(4,4;7;2,4)$, $(3,3;4;2,3)$, $(1,2;1;2,4)$, or $(4,4;8;2,2)$ 2-D modulation code because of the higher spatial frequency content allowed. However, the $(3,3;6;3,2)$ 2-D modulation code has a higher code rate, i.e., a higher fraction of useful

information, therefore, implying that there exists a trade-off between the highest spatial frequency content allowed and the code rate.

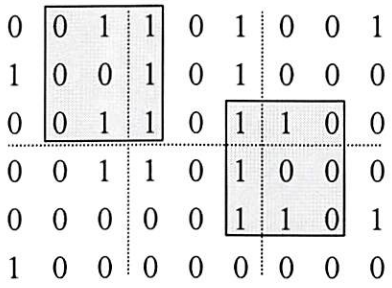


Figure 4.14 The examples of a 3×3 block of recorded data with the highest spatial frequency content after applying the $(3,3;6;3,2)$ 2-D modulation code.

Fixed-length code $(m,n;k;\alpha,\beta)$	Output block size (m,n)	Number of input bits k	Constraints (α,β)	Code rate $k/(mn)$
(4,4;7;2,4)	(4,4)	7	(2,4)	0.4375
(3,3;4;2,3)	(3,3)	4	(2,3)	0.4444
(1,2;1;2,4)	(1,2)	1	(2,4)	0.5
(4,4;8;2,2)	(4,4)	8	(2,2)	0.5
(3,3;6;3,2)	(3,3)	6	(3,2)	0.6667

Table 4.1 List of various fixed-length 2-D modulation codes.

Chapter 5

Two-Dimensional Modulation Coding for Two-Photon PODS Systems: Variable-Length Codes

It is obvious that the higher the code rate of a two-dimensional (2-D) modulation code, the higher the overall useful data storage capacity of a two-photon page-oriented optical data storage (PODS) system. We, however, observe that the fixed-length 2-D modulation codes presented in chapter 4 have rather moderate code rates. This is indeed due to the fact that they have fixed output block sizes, $m \times n$, which confine the total number of possible 2-D data blocks of $m \times n$ bits that satisfy the required constraints (i.e., the required highest spatial frequency contents) and, consequently, limit the code rates [45]. In this chapter, we first discuss in detail about the limitation of fixed-length 2-D modulation codes, in terms of code rate, by giving an example of the code rate upper bounds (at all possible values of spatial frequency constraints) of fixed-length codes with certain output block sizes. Also included in the first section is the procedure used to determine such empirical upper bounds. We later describe the alternative set of 2-D modulation codes, called the variable-length codes. Typically, variable-length codes can have higher code rates compared to fixed-length codes having the same spatial frequency constraints. An example of a variable-length 2-D modulation code, including its decoding procedure, is then presented.

5.1 An Example of the Code Rate Upper Bounds of Fixed-Length Two-Dimensional Modulation Codes

Ideally, the code rate limitation of a modulation code should be described by its maximum code rate, and the exhaustive search method, which search for all possible binary patterns that satisfy the required constraints, is the optimal way to determine such value. Following the generalized notation and basic principles of our 2-D modulation coding scheme discussed in chapter 4, the result of the exhaustive search, for any given m and n, is a collection of 2-D $m \times n$ binary patterns (output blocks) that satisfy the desired spatial frequency constraints α and β (not only within themselves, but also after they are placed side-by-side at all possible combinations in a 2-D array that forms a tiling of the recorded data page). Suppose the total number of such patterns is L. The maximum code rate is, thus, equal to

$$\left(\text{code rate}\right)_{\max} = \frac{\lfloor \log_2(L) \rfloor}{(mn)}, \tag{5.1}$$

where $\lfloor \lambda \rfloor$ represents the greatest integer not larger than λ.

We note, however, that the exhaustive search method is not practical for 2-D modulation codes because of its computational complexity (even when both m and n are as small as 3). To illustrate the code rate limitation of fixed-length 2-D modulation codes, instead of globally searching for the optimal solution, i.e., the maximum code rate, at any particular α and β, we simplify the problem by following the three steps

described below to reduce the size of the search space and, then, determining only the code rate upper bound.

Step 1: From the original search space, which contains all possible combinations of 2-D $m \times n$ binary patterns (for a total of 2^{mn} patterns), select only patterns that satisfy the desired spatial frequency constraints (α and β) within themselves, i.e., consider one pattern at a time and select only the patterns that satisfy the desired constraints.

Step 2: From the reduced search space obtained in Step 1, by considering one pattern at a time, select any pattern (X) that still satisfies the desired constraints (α and β) when four of them are placed as

X	X
X	X

Step 3: From the reduced search space obtained in Step 2, consider two patterns (X and Y) at a time and check whether they satisfy the desired constraints (α and β) if they are placed as

Y	Y	Y
Y	X	Y
Y	Y	Y

and

X	X	X
X	Y	X
X	X	X

Let L' be the total number of patterns in the reduced search space obtained from Step 2. The upper bound of the code rate of any fixed-length 2-D modulation code, then, can be calculated using the pseudocode in Figure 5.1. It is also necessary to note that although the upper bounds derived empirically from the above three steps are quite

loose, they are still useful enough to represent the significance of the code rate limitation of fixed-length 2-D modulation codes.

Furthermore, we show in Table 5.1 an example of the code rate upper bounds of the 2-D modulation codes with certain output block sizes of 3×3 (for all possible values of α and β). It should be emphasized that the code rates of the $(3,3;4;2,3)$ and $(3,3;6;3,2)$ 2-D modulation codes described in Eqs. (4.13) and (4.16), respectively, are very close to their corresponding upper bounds shown in Table 5.1. While the $(3,3;4;2,3)$ 2-D modulation code has a code rate of 0.4444 (= 4/9), its code rate upper bound is 0.5556 (= 5/9). On the other hand, the $(3,3;6;3,2)$ 2-D modulation code has a code rate of 0.6667 (= 6/9) whereas its code rate upper bound is 0.7778 (= 7/9).

begin initialize k' is the greatest integer not larger than $\log_2(L')$
  do
    $k' \leftarrow k'-1$
  until {the total number of patterns satisfying the conditions in Step 3 with at least $2^{k'}$ other patterns (pair-wise) is greater than or equal to $2^{k'}$ }
  code rate upper bound = $k'/(mn)$
end

Figure 5.1 Pseudocode used to determine the code rate upper bound following three steps of search space reduction.

(α,β)	0	1	2	3	4
1	0.1111	0.2222	0.2222	0.2222	0.2222
2	0.3333	0.4444	0.5556	0.5556	0.5556
3	0.6667	0.6667	0.7778	0.7778	0.8889
4	0.8889	0.8889	0.8889	0.8889	1

Table 5.1 The code rate upper bounds of the $(3,3;k;\alpha,\beta)$ 2-D modulation codes at all possible combinations of (α,β).

5.2 An Example of a Variable-Length Two-Dimensional Modulation Code and Its Decoding Procedure

As mentioned in the previous section, the code rates of fixed-length 2-D modulation codes are somewhat restricted. Thus, to have 2-D modulation codes with higher code rates at certain spatial frequency constraints, we propose that the sizes of 2-D output blocks be variable. In this discussion, we assume that only one of two dimensions of the 2-D output blocks is variable, i.e., either m or n is variable, while the other is fixed; the 2-D modulation codes with this property are classified as *variable-length* 2-D modulation codes. Indeed, our motivation for variable-length codes intuitively comes from the fact that the degrees of freedom to obtain 2-D data blocks, satisfying specific highest spatial frequency constraints, can be increased if the sizes of those data blocks are not fixed, i.e., the data block sizes are variable [45]. We also note that 2-D modulation codes with much higher code rates might be obtained by varying both

dimensions; however, the overhead needed for the coding and the complexity of the decoder might be excessive.

Remember that the general form in Eq. (4.10) is defined only for fixed-length 2-D modulation codes. For variable-length 2-D modulation codes, minor modifications are needed. One example of a variable-length 2-D modulation code is illustrated in Figure 5.2 [45]. It is called the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code because it maps a one-dimensional (1-D) block of two, four, or six information bits into a 2-D block of 4×1, 4×2, or 4×3 bits, respectively. The highest spatial frequency content allowed after applying the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code is $\alpha=2$ and $\beta=3$, meaning that in any recorded data page the maximum number of "1" bits in the four-connected neighbor positions of any "0" bit is two and the maximum number of "1" bits in the eight-connected neighbor positions of the same "0" bit is three.

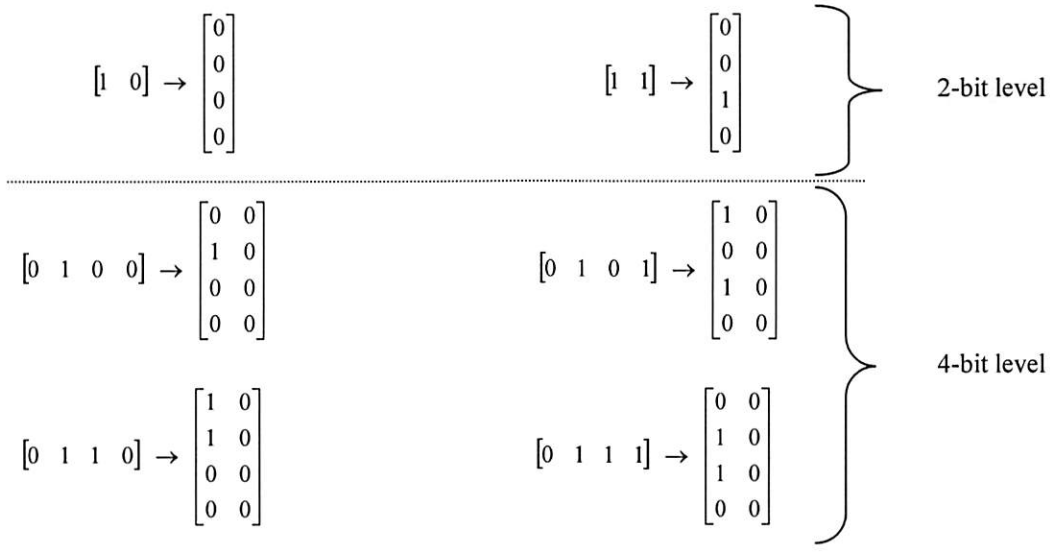


Figure 5.2 Lookup table of the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code.

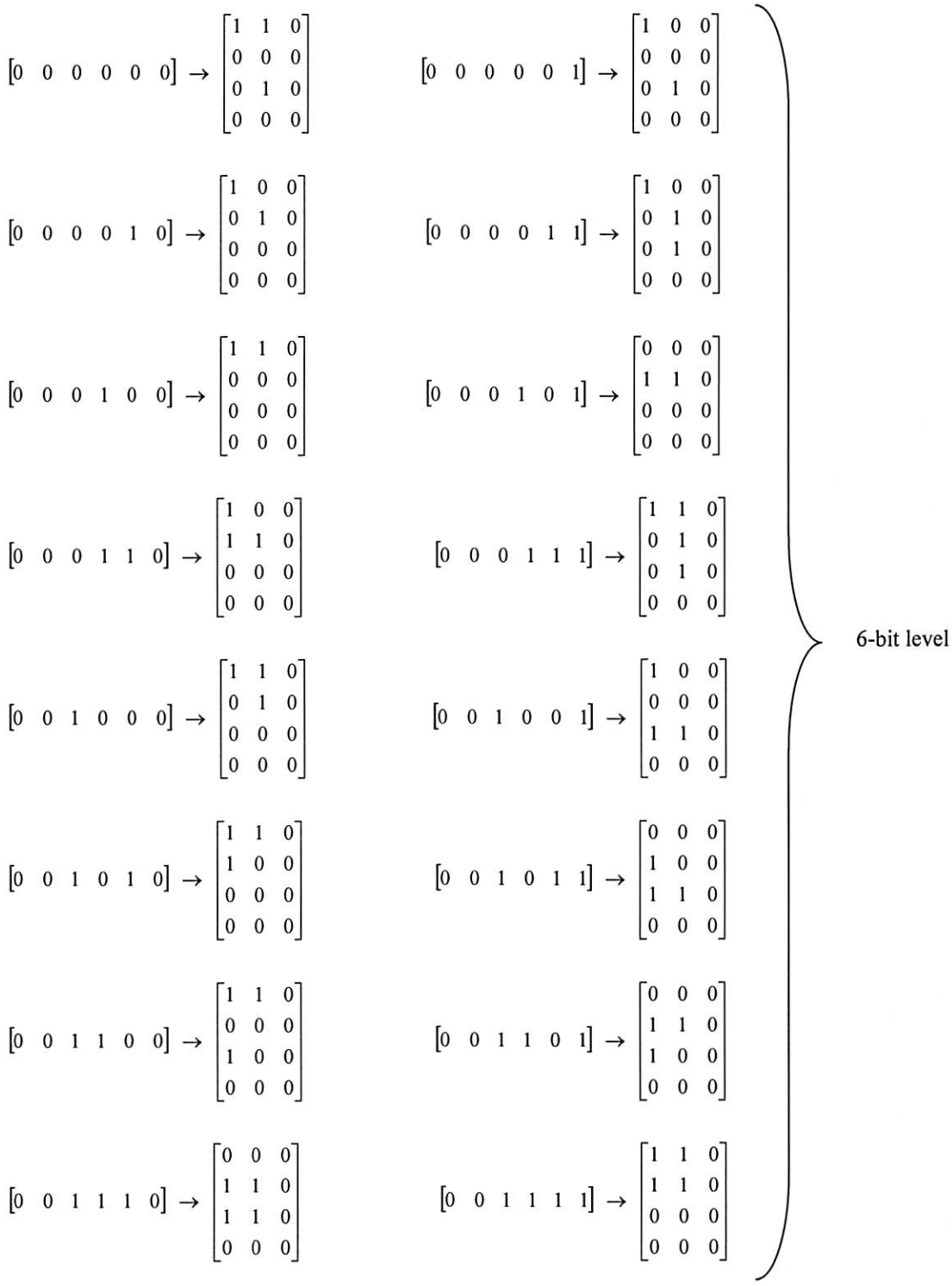


Figure 5.2(continued) Lookup table of the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code.

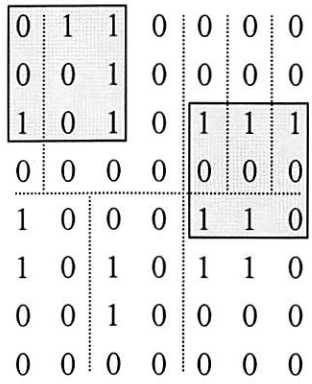


Figure 5.3 The examples of a 3×3 block of recorded data with the highest spatial frequency content after applying the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code.

Figure 5.3 depicts two examples of the possible highest spatial frequency content after applying the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code. There are several significant aspects of the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code that are worth discussing. First, the lookup table in Figure 5.2 is neither unique nor trivial. Second, observing that this variable-length 2-D modulation code maps two, four, or six bits into four, eight, or twelve bits, respectively, the code rate is then equal to 0.5. Hence, it is apparent that the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code should give better performance, in terms of the useful data storage capacity, than the $(3,3;4;2,3)$ fixed-length 2-D modulation code in section 4.4.2 because of the higher code rate at the same constraints of the highest spatial frequency content (α and β).

Furthermore, one of the most important issues about the variable-length codes is how to decode the information with least effects of the error propagation. Unlike the

simple decoding procedures for typical fixed-length 2-D modulation codes, the decoding procedures for variable-length 2-D modulation codes are usually more complicated. It is necessary to note that the error propagation due to an erroneous bit in a recorded data page of the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code is limited when the appropriate decoding procedure is applied. We show in Figure 5.4 the pseudocode of one such decoding procedure. Although the size of each recorded data page is $M \times N$ (as defined in chapter 3), we assume, without loss of generality, in the pseudocode that the size of 2-D array of data to be decoded is $4 \times (MN/4)$. With this decoding procedure, one erroneous bit would affect *at most* a 2-D data block of 4×5 bits, which is less than or equal to ten bits in the sequence of 1-D decoded output. One example of the worst scenarios is shown in Figure 5.5, where the decoding procedure proceeds from left to right and the square, with a "1" bit inside, locates the position of an erroneous bit. Nevertheless, it should be noted that it might be feasible to obtain a decoding procedure with higher efficiency for the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code.

begin initialize $l=1$, $[\underline{d}_1 \quad \underline{d}_2 \quad \cdots \quad \underline{d}_{MN/4}] \equiv 4\times(MN/4)$ array of data to be decoded

do

if $\underline{d}_l$ matches one in the 2-bit level of the lookup table

then obtain the corresponding 2-bit output, $l \leftarrow l+1$

else if $\underline{d}_{l+1} = [0 \quad 0 \quad 0 \quad 0]^t$

then if $[\underline{d}_l \quad \underline{d}_{l+1}]$ matches one in the 4-bit level of the lookup table

then obtain the corresponding 4-bit output, $l \leftarrow l+2$

else the output is $[0 \quad 0 \quad 0 \quad 0]$, $l \leftarrow l+2$, set the error flag

else if $\underline{d}_{l+2} = [0 \quad 0 \quad 0 \quad 0]^t$

then if $[\underline{d}_l \quad \underline{d}_{l+1} \quad \underline{d}_{l+2}]$ matches one in the 6-bit level of the lookup table

then obtain the corresponding 6-bit output, $l \leftarrow l+3$

else the output is $[0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$, $l \leftarrow l+3$, set the error flag

else the output is $[0 \quad 0 \quad 0 \quad 0 \quad 0 \quad 0]$, $l \leftarrow l+3$, set the error flag

until $l > (MN/4)-2$

do

if $l = (MN/4)-1$

then if $\underline{d}_l$ matches one in the 2-bit level of the lookup table

then obtain the corresponding 2-bit output, $l \leftarrow l+1$

else if $\underline{d}_{l+1} = [0 \quad 0 \quad 0 \quad 0]^t$

then if $[\underline{d}_l \quad \underline{d}_{l+1}]$ matches one in the 4-bit level of the lookup table

then obtain the corresponding 4-bit output, $l \leftarrow l+2$

else the output is $[0 \quad 0 \quad 0 \quad 0]$, $l \leftarrow l+2$, set the error flag

else the output is $[0 \quad 0 \quad 0 \quad 0]$, $l \leftarrow l+2$, set the error flag

else

then if $\underline{d}_l$ matches one in the 2-bit level of the lookup table

then obtain the corresponding 2-bit output, $l \leftarrow l+1$

else the output is $[0 \quad 0]$, $l \leftarrow l+1$, set the error flag

until $l > (MN/4)$

terminating conditions at the end of each recorded data page

end

Figure 5.4 Pseudocode of the decoding procedure for the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code.

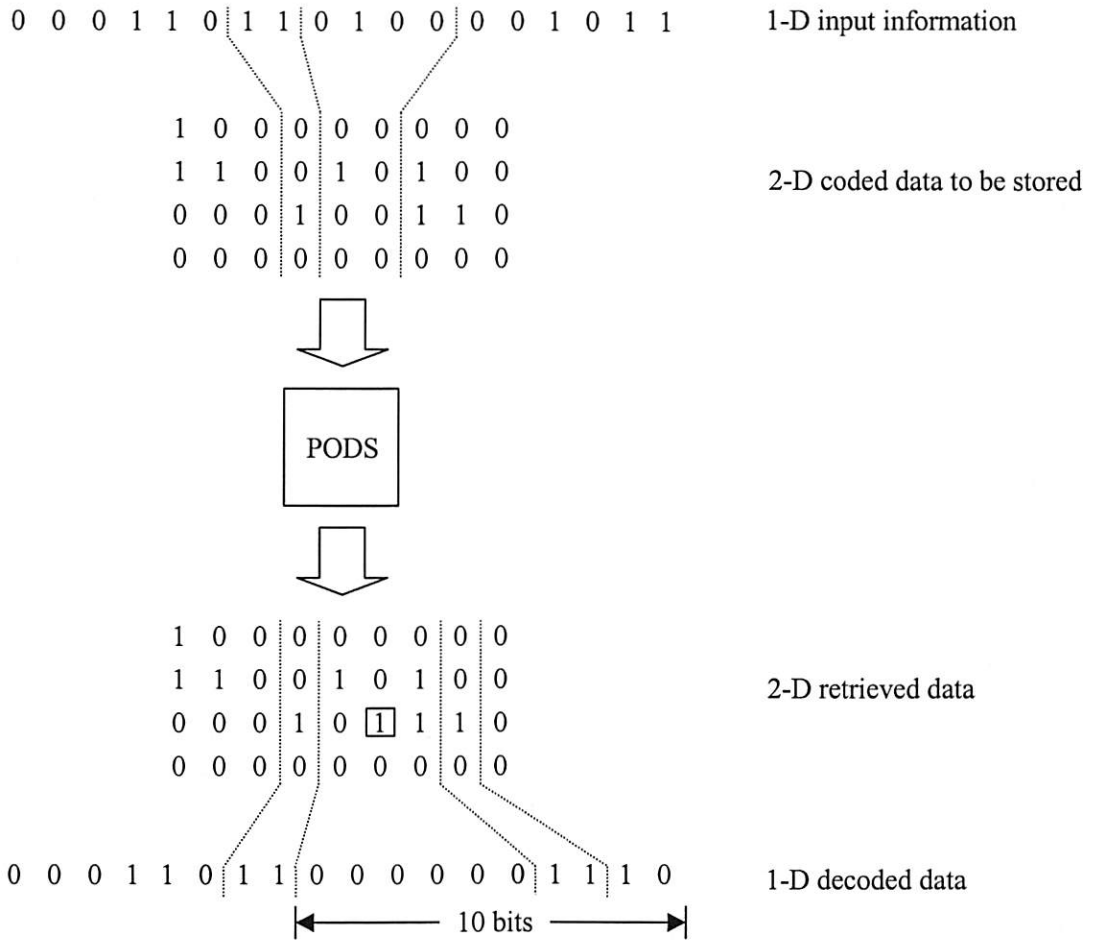


Figure 5.5 An example of the worst scenarios for the error propagation due to an erroneous bit after applying the decoding procedure in Figure 5.4, which proceeds from left to right.

Chapter 6

Bit-Error-Rate Performance Analysis of Two-Dimensional Modulation Codes for Two-photon PODS Systems

We recall from chapter 2 that a typical data storage system can be represented by the block diagram shown in Figure 2.1, in which there are two levels of coding: modulation coding and channel coding. To measure only the bit-error-rate (BER) performance of a modulation code, we must analyze the average probability of a bit-error between points B and B', rather than between points A and A'. In this chapter, we illustrate the BER performance of two-dimensional (2-D) modulation codes described in chapters 4 and 5. The first section presents the theoretical derivation of the minimum BER for a *noisy* two-photon page-oriented optical data storage (PODS) system. In this section, we model the noise as additive white Gaussian noise (AWGN) as discussed in chapter 3, and find the corresponding optimal threshold for a binary threshold decision scheme that is applied at the charge-coupled device (CCD) or detector array. However, in practice, it is often not possible to obtain such an optimal threshold, or a minimum BER, due to the fact that the exact probability density functions of noise in the system are usually unknown. We describe in the second section an alternative procedure that is used to determine a reliable BER of the practical two-photon PODS system. Based on this

practical procedure, the simulation results show that our proposed 2-D modulation codes, both fixed-length and variable-length codes, can significantly improve the system performance in terms of BER.

6.1 The Optimal Threshold and the Minimum Bit-Error-Rate of an Additive White Gaussian Noise Two-Photon PODS System

In chapter 4, we show many histograms of the received intensities, both before and after applying a 2-D modulation code, at the detector array or CCD camera of several *noiseless* two-photon PODS systems. Referring to Figures 4.2 and 4.5, a $(4,4;7;2,4)$ 2-D modulation code can totally overcome the effects of 2-D intersymbol interference (ISI), i.e., in each case, there is no overlap between the histogram of the received intensities that originate from "0" bits and that from "1" bits after applying such 2-D modulation code. Since there is no overlap, any intensity value that lies in the gap between the histograms can be used as a threshold value for the binary threshold decision scheme that is applied at the CCD or detector array. In fact, any intensity value that lies in that gap can be considered as an optimal threshold that always gives correct decisions and, consequently, results in a zero BER, i.e., the probability that an error will occur is equal to zero.

However, as discussed in chapter 3, a two-photon PODS system is generally not noiseless. Suppose that it is *noisy*, primarily, because of the electronics at the detector plane. In this case, the effects of AWGN are taken into account and, hence, the

histogram of the received intensities that originate from "0" bits always overlaps with that from "1" bits, regardless of any 2-D modulation coding. This, indeed, leads to a nontrivial optimal threshold value for the binary threshold decision scheme and a non-zero minimum BER. In this section, we describe analytically how to obtain such an inexplicit optimal threshold in order to achieve a minimum BER for a *noisy* (AWGN) two-photon PODS system.

Because the counts, i.e., the number of occurrences, of the received intensities in any histogram shown in chapter 4 are accumulated in a finite number of equally distributed intensity bins (in our simulations, the bin size is always set to 1/200 of the normalized intensity), the effects of AWGN in a *noisy* two-photon PODS system, then, can be simply included by convolving each intensity bin in the histogram with a Gaussian distribution function, whose mean and variance are zero and σ_w^2, respectively. Let us define $c_{0,l}$ as the number of counts in the l th bin, whose center is located at the intensity value $\mu_{0,l}$, of the histogram of the received intensities that originate from "0" bits. Analogously, $c_{1,l}$ is defined as the number of counts in the l th bin, whose center is located at the intensity value $\mu_{1,l}$, of the histogram of the received intensities that originate from "1" bits. Assuming that the total number of counts in the histogram of the received intensities that originate from "0" bits is C_0 and that the total number of counts in the histogram of the received intensities that originate from "1" bits is C_1, i.e.,

$C_0 = \sum_l c_{0,l}$ and $C_1 = \sum_l c_{1,l}$, we can derive the BER of an AWGN two-photon PODS system with intensity threshold value T as following:

$$\text{BER} \equiv \text{Prob}\{\text{error}\} = \text{Prob}\{\text{error}|\text{"0" bit is recorded}\}\text{Prob}\{\text{"0" bit is recorded}\} + \text{Prob}\{\text{error}|\text{"1" bit is recorded}\}\text{Prob}\{\text{"1" bit is recorded}\}. \quad (6.1)$$

It is straightforward from a binary threshold decision scheme, shown in Eq. (3.7), that the conditional probability of error given that a "0" bit is recorded is equivalent to the conditional probability that the received intensity, r, is greater than the threshold value, T, given that a "0" bit is recorded. Therefore,

$$\text{Prob}\{\text{error}|\text{"0" bit is recorded}\} = \text{Prob}\{r > T|\text{"0" bit is recorded}\} \quad (6.2)$$

$$= \frac{1}{C_0}\sum_l c_{0,l}\left[\int_T^\infty \frac{1}{\sqrt{2\pi}\sigma_w}\exp\left(-\frac{(r-\mu_{0,l})^2}{2\sigma_w^2}\right)dr\right] \quad (6.3)$$

$$= \frac{1}{2C_0}\sum_l c_{0,l}\,\text{erfc}\left(\frac{T-\mu_{0,l}}{\sqrt{2}\sigma_w}\right), \quad (6.4)$$

where erfc is the complementary error function defined as [34]

$$\text{erfc}(s) = \frac{2}{\sqrt{\pi}}\int_s^\infty \exp(-t^2)dt. \quad (6.5)$$

Similarly, the conditional probability of error given that a "1" bit is recorded is equivalent to the conditional probability that the received intensity, r, is less than the threshold value, T, given that a "1" bit is recorded. Thus,

$$\text{Prob}\{\text{error}|\text{"1" bit is recorded}\} = \text{Prob}\{r < T|\text{"1" bit is recorded}\} \tag{6.6}$$

$$= \frac{1}{C_1}\sum_l c_{1,l}\left[\int_{-\infty}^{T}\frac{1}{\sqrt{2\pi}\sigma_w}\exp\left(-\frac{(r-\mu_{1,l})^2}{2\sigma_w^2}\right)dr\right] \tag{6.7}$$

$$= \frac{1}{2C_1}\sum_l c_{1,l}\,\text{erfc}\left(\frac{\mu_{1,l}-T}{\sqrt{2}\sigma_w}\right). \tag{6.8}$$

Inserting Eqs. (6.4) and (6.8) into Eq. (6.1), we have

$$\text{BER} \equiv \text{Prob}\{\text{error}\} = \text{Prob}\{\text{"0" bit is recorded}\}\left[\frac{1}{2C_0}\sum_l c_{0,l}\,\text{erfc}\left(\frac{T-\mu_{0,l}}{\sqrt{2}\sigma_w}\right)\right] +$$

$$\text{Prob}\{\text{"1" bit is recorded}\}\left[\frac{1}{2C_1}\sum_l c_{1,l}\,\text{erfc}\left(\frac{\mu_{1,l}-T}{\sqrt{2}\sigma_w}\right)\right]. \tag{6.9}$$

Clearly, if the input information bits are recorded without applying any 2-D modulation code, then

$$\text{Prob}\{\text{"0" bit is recorded}\} = \text{Prob}\{\text{"1" bit is recorded}\} = 0.5. \tag{6.10}$$

However, if a 2-D modulation code is applied, then the probability that a "0" bit is recorded is generally not equal to the probability that a "1" bit is recorded. Based on our 2-D modulation coding scheme described in chapter 4, the probability that a "0" bit is recorded is often much higher than the probability that a "1" bit is recorded. Table 6.1 lists such probabilities for various 2-D modulation codes, both fixed-length and variable-length codes.

Finally, we evaluate Eq. (6.9) repeatedly for each threshold value T. A threshold value that results in a minimum BER is, hence, an optimal threshold for the binary threshold decision scheme of an AWGN two-photon PODS system.

2-D modulation code	Prob{"0" bit is recorded}	Prob{"1" bit is recorded}
(4,4;7;2,4)	195/256 (= 0.7617)	61/256 (= 0.2383)
(3,3;4;2,3)	107/144 (= 0.7431)	37/144 (= 0.2569)
(1,2;1;2,4)	3/4 (= 0.75)	1/4 (= 0.25)
(4,4;8;2,2)	3/4 (= 0.75)	1/4 (= 0.25)
(3,3;6;3,2)	2/3 (= 0.6667)	1/3 (= 0.3333)
(4,[1,2,3];[2,4,6];2,3)	631/768 (= 0.8216)	137/768 (= 0.1784)

Table 6.1 List of the probabilities that a "0" bit is recorded and the probabilities that a "1" bit is recorded for various 2-D modulation codes.

6.2 Bit-Error-Rate Performance of Two-Dimensional Modulation Codes

In section 6.1, we describe the numerical procedure that is used to find the optimal threshold value for a binary threshold decision scheme and the corresponding minimum BER of a noisy two-photon PODS system. It is necessary to emphasize that we theoretically derive such procedure based on the AWGN assumption. In other words, the optimal threshold and the minimum BER are derived by assuming that the probability density function of noise in a two-photon PODS system can be simply modeled as a Gaussian function, whose mean and variance are zero and σ_w^2, respectively. This simple assumption, however, is not valid for every practical two-photon PODS system (typically, the exact probability density functions of noise in practical two-photon PODS systems are very difficult to identify). Therefore, the

numerical procedure discussed in section 6.1, in general, may need modifications for practical two-photon PODS systems.

We present in this section an alternative procedure that is more practical and yet gives reliable values of BER for typical two-photon PODS systems. This practical procedure can be separated into two cases: the first assumes that no 2-D modulation code is applied, i.e., the raw or uncoded case, and the other assumes that a 2-D modulation code is applied, i.e., the coded case. For the former case, assuming that an enormous number of data pages are recorded without applying any 2-D modulation code, we retrieve as many recorded data pages as necessary to obtain at least 100 errors. We note that the number of errors is defined as the number of bits in the recorded data pages that are different from the corresponding bits in the detected data pages, which are obtained by digitizing the received intensities with a *priori* knowledge of a threshold value. We also note that the predetermined threshold value can be practically optimized, in favor of minimum BER, by a variety of ways. For examples, it can be estimated approximately from the histograms of the received intensities or we can use a trial-and-error approach. The BER for the raw case, thus, can be calculated by dividing the total number of errors, which is at least 100 errors, by the total number of detected data bits, which is the total number of detected data page resulting in that total amount of errors multiplied by the data page size.

For the latter case, suppose that a large number of data pages are recorded after applying a 2-D modulation code to a lengthy sequence of input information bits. We, again, retrieve the recorded data pages as many as necessary to obtain at least 100 errors.

Unlike the previous case, the number of errors in this case is defined as the number of bits decoded from the detected data pages, also obtained by digitizing the received intensities with a *priori* knowledge of a threshold value, that are different from the corresponding bits in the original input information sequence before applying a 2-D modulation code. It should be emphasized that the optimal predetermined threshold value in this case is generally different from that in the previous case, although it can be determined in a similar way. The BER of the coded case is, then, calculated by dividing the total number of errors, at least 100 errors, by the total number of original input information bits (before applying a 2-D modulation code), which results in that total amount of errors.

To illustrate the BER performance of our 2-D modulation codes, both fixed-length and variable-length codes, we simulate the BERs of several noisy two-photon PODS systems in this section. Our simulation procedure, basically, follows the practical procedure described above, except that we repeat the whole procedure ten times and then calculate the average BER in order to obtain a more reliable result. Figures 6.1 and 6.2 show pseudocodes of our simulation procedure for raw (uncoded) and coded cases, respectively. We note that every two-photon PODS system considered in this section is assumed to be noisy due to the electronics at the detector plane and can be modeled as discussed in chapter 3 by the AWGN, whose mean and variance are zero and σ_w^2, respectively. We also assume, for simplicity, in every simulation that each recorded data page always contains 120×120 bits; it is essential to emphasize that this assumption is possible because 120×120 is divisible by the output block sizes of all 2-D modulation

begin initialize {size of recorded data pages ($M \times N$), type of PSF and its parameter (sinc^2 PSF and b_s, J_1^2 PSF and b_c, or Gaussian PSF and b_g), linear fill factor of detectors (δ), standard deviation of AWGN (σ_w), predetermined intensity threshold (T), $l = 1$ }

do
  total number of errors = 0
  number of data pages = 0
  do
    randomly generate a digital data page
    number of data pages ← number of data pages + 1
    compute the received intensities at the detector array using Eq. (3.6) or (3.8)
    compute the detected data page using Eq. (3.7)
    compare the detected data page with the original digital data page and count the number of errors
    total number of errors ← total number of errors + number of errors
  until total number of errors ≥ 100

$$\text{BER}_l = \frac{\text{total number of errors}}{(\text{number of data pages}) \times MN}$$

$l \leftarrow l + 1$

until $l > 10$

$$\text{BER} = \frac{1}{10} \sum_{l=1}^{10} \text{BER}_l$$

end

Figure 6.1 Pseudocode of the BER simulation procedure for the raw (uncoded) case.

codes described in chapters 4 and 5. Additionally, in every simulation, the linear fill factor, δ, of each detector or CCD pixel is always set to one and, as defined in chapter 4, the effective point-spread function (PSF) matrix used in Eq. (3.6) or (3.8) contains *at least* 90% of the total light originating from a data mark. We note further that based on the AWGN assumption, all predetermined threshold values used in our simulations are obtained from the numerical procedure described in section 6.1, provided that 100 pages of data are recorded.

begin initialize { $(m, n; k; \alpha, \beta)$ 2-D modulation code, size of recorded data pages ($M \times N$), type of PSF and its parameter (sinc^2 PSF and b_s, J_1^2 PSF and b_c, or Gaussian PSF and b_g), linear fill factor of detectors (δ), standard deviation of AWGN (σ_w), predetermined intensity threshold (T), $l = 1$ }

do

total number of errors = 0

total number of input information bits = 0

do

randomly generate a digital input information sequence containing $(MN/mn)k$ bits

total number of input information bits ← total number of input information bits + $(MN/mn)k$

encode the digital input information sequence using the $(m, n; k; \alpha, \beta)$ 2-D modulation code

compute the received intensities at the detector array using Eq. (3.6) or (3.8)

compute the detected data page using Eq. (3.7)

decode the detected data page with proper decoding procedure

compare the decoded sequence with the original input sequence and count the number of errors

total number of errors ← total number of errors + number of errors

until total number of errors ≥ 100

$$\text{BER}_l = \frac{\text{total number of errors}}{\text{total number of input information bits}}$$

$l \leftarrow l + 1$

until $l > 10$

$$\text{BER} = \frac{1}{10} \sum_{l=1}^{10} \text{BER}_l$$

end

Figure 6.2 Pseudocode of the BER simulation procedure for the coded case.

Figures 6.3, 6.4, and 6.5 show several plots of the BER as a function of the standard deviation of the AWGN for many noisy two-photon PODS systems. First, the $(4,4;7;2,4)$, $(4,4;8;2,2)$, and $(3,3;6;3,2)$ 2-D modulation codes are of particular interest in Figure 6.3. In Figure 6.3(a), we assume that the retrieval system contains a square aperture and can be modeled by the sinc^2 PSF with $b_s = 1.5$. Apparently, if there is no modulation code applied (i.e., the raw case), the values of the BER are always very high, even with a very low value of the AWGN standard deviation. This, in fact, implies the

strong effects of 2-D ISI in a two-photon PODS system. On the contrary, when a 2-D modulation code is applied, it is clear that the values of the BER at the small values of the AWGN standard deviation improve significantly.

We note, however, that the BER performance improvement relies heavily on the characteristic of each 2-D modulation code. As illustrated in Figure 6.3(a), the $(3,3;6;3,2)$ 2-D modulation code cannot improve the BER performance as much as the $(4,4;7;2,4)$ and $(4,4;8;2,2)$ 2-D modulation codes can. It is, indeed, because the $(3,3;6;3,2)$ 2-D modulation code has higher spatial frequency constraints than the $(4,4;7;2,4)$ and $(4,4;8;2,2)$ 2-D modulation codes; more specifically, the $(3,3;6;3,2)$ 2-D modulation code has higher value of α than the $(4,4;7;2,4)$ and $(4,4;8;2,2)$ 2-D modulation codes. It should also be recalled from chapter 4 that the spatial frequency constraints of the $(4,4;8;2,2)$, $(4,4;7;2,4)$, and $(3,3;6;3,2)$ 2-D modulation codes are $(\alpha,\beta)=(2,2)$, $(\alpha,\beta)=(2,4)$, and $(\alpha,\beta)=(3,2)$, respectively, where we define α as the maximum number of "1" bits in the four-connected neighbor positions of any "0" bit in a recorded data page and β as the maximum number of "1" bits in the eight-connected neighbor positions of the same "0" bit.

Likewise, when we assume in Figure 6.3(b) that the retrieval system contains a circular aperture and can be represented by the J_1^2 PSF with $b_c = 2.5$, we obtain the similar results as in the case of Figure 6.3(a). In addition, in Figure 6.3(c), we consider the retrieval system that can be modeled by the Gaussian PSF with $b_g = 0.6$. It is shown in chapter 4 that, with this particular two-photon PODS system, the degree of 2-D ISI is

very high; thus, we expect the very low BER performance, if the effects of AWGN are taken into account. From Figure 6.3(c), it is obvious that only the $(4,4;8;2,2)$ 2-D modulation code can substantially improve the BER performance. We emphasize that this is because the $(4,4;8;2,2)$ 2-D modulation code has relatively low spatial frequency constraints, i.e., $(\alpha,\beta)=(2,2)$, compared with $(\alpha,\beta)=(2,4)$ and $(\alpha,\beta)=(3,2)$ of the $(4,4;7;2,4)$ and $(3,3;6;3,2)$ 2-D modulation codes, respectively.

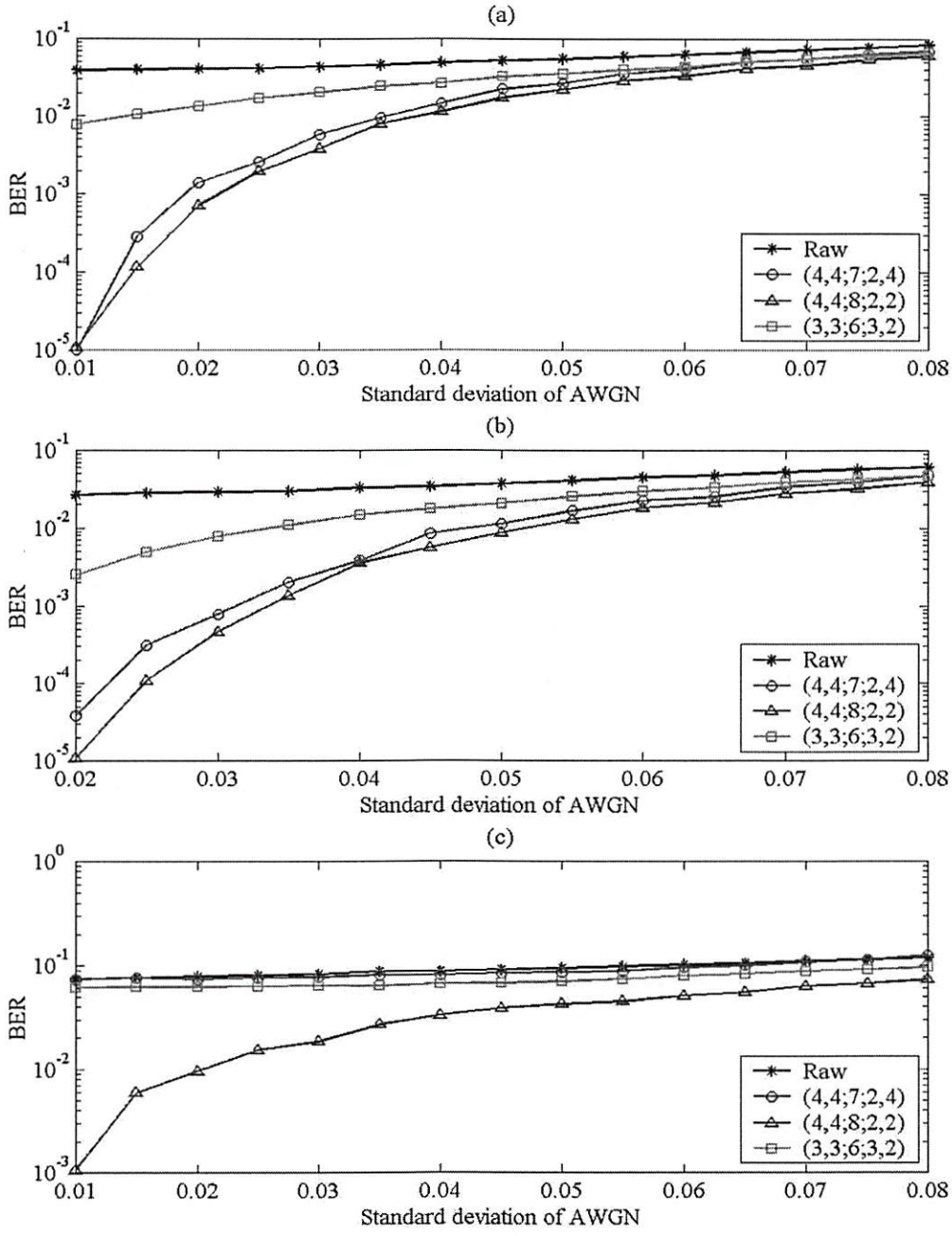


Figure 6.3 BER performance of $(4,4;7;2,4)$, $(4,4;8;2,2)$, and $(3,3;6;3,2)$ 2-D modulation codes in a noisy two-photon PODS system modeled by
(a) sinc^2 PSF with $b_s = 1.5$,
(b) J_1^2 PSF with $b_c = 2.5$, and
(c) Gaussian PSF with $b_g = 0.6$.

Secondly, we are interested in the $(3,3;4;2,3)$ fixed-length 2-D modulation code and the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code. The intention of Figure 6.4 is not only to examine the BER performance of a variable-length 2-D modulation code, but also to compare it with that of a fixed-length 2-D modulation code having the same spatial frequency constraints (α,β). Suppose in Figure 6.4(a) that the retrieval system is described by the sinc^2 PSF with $b_s = 1.5$. Noticeably, at the small values of the AWGN standard deviation, both $(3,3;4;2,3)$ fixed-length and $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation codes greatly improve the BER performance. Note, however, that the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code degrades the BER performance when the value of the AWGN standard deviation is excessively high. In other words, when the level of noise in a two-photon PODS system is very high, the BER performance of the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code is even worse than that of the raw case (i.e., when no 2-D modulation code is applied). This is due to the error propagation during the decoding process of the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code. Hence, although variable-length 2-D modulation codes may offer higher code rates and, as a result, higher overall useful data storage capacity, compared with fixed-length 2-D modulation codes having the same spatial frequency constraints (α,β), there are serious problems because of error propagation. Similar results are illustrated in Figure 6.4(b) for the case of a retrieval system that contains a circular aperture and can be modeled by the J_1^2 PSF with $b_c = 2.5$.

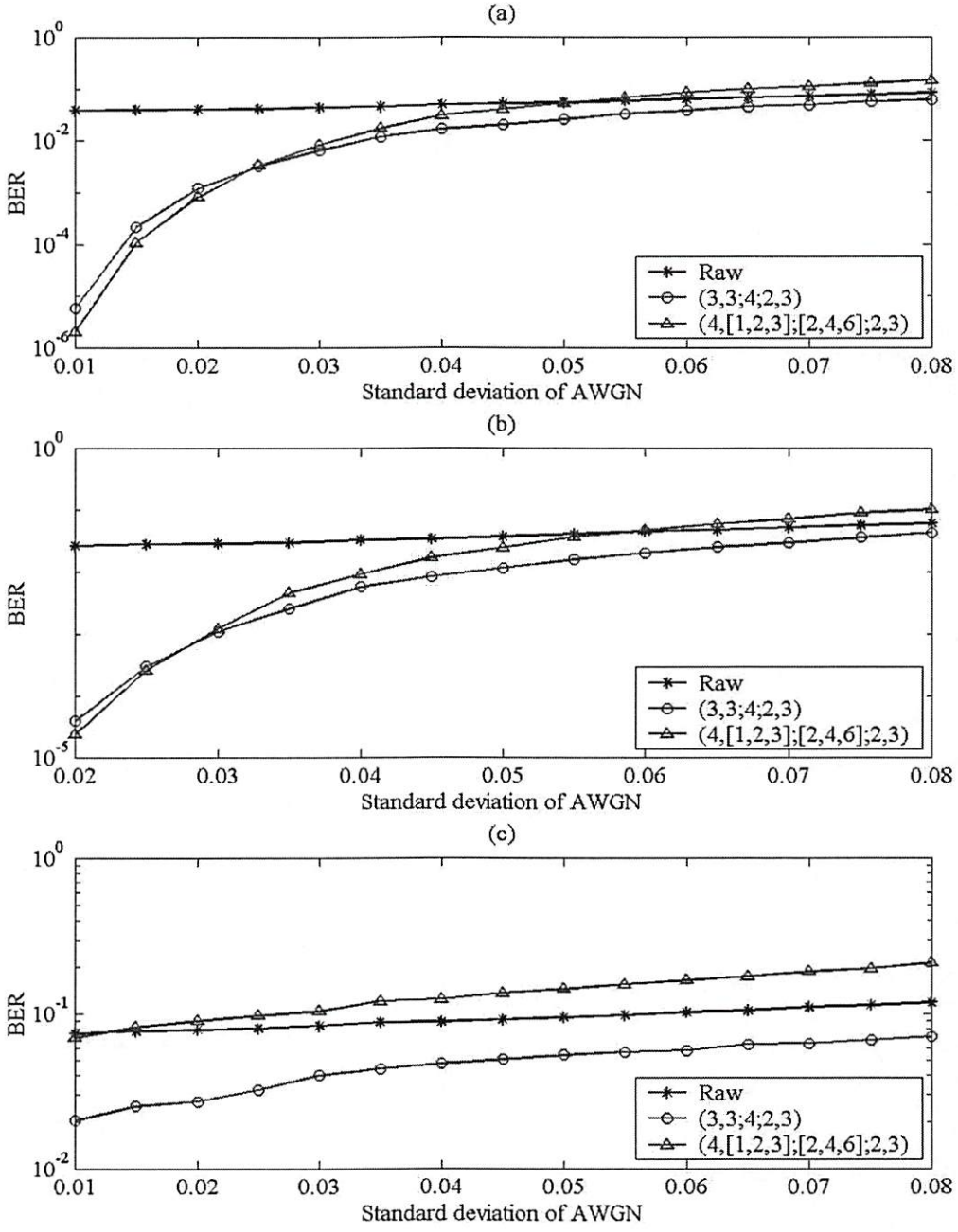


Figure 6.4 BER performance of $(3,3;4;2,3)$ fixed-length and $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation codes in a noisy two-photon PODS system modeled by

(a) sinc^2 PSF with $b_s = 1.5$,

(b) J_1^2 PSF with $b_c = 2.5$, and

(c) Gaussian PSF with $b_g = 0.6$.

Moreover, it is evident from Figure 6.4(c), in which the retrieval system is represented by the Gaussian PSF with $b_g = 0.6$, that when the degree of 2-D ISI is extremely high, the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code can be totally ineffective, regardless of the value of the AWGN standard deviation. Eventually, the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code weakens the BER performance of this specific two-photon PODS system, compared with the raw case. On the other hand, the $(3,3;4;2,3)$ fixed-length 2-D modulation code can slightly improve the BER performance of the same two-photon PODS system, although its spatial frequency constraints are identical to those of the $(4,[1,2,3];[2,4,6];2,3)$ variable-length 2-D modulation code. This example, again, supports the hurdles of error propagation in the variable-length 2-D modulation codes.

Lastly, we compare the BER performance of the fixed-length 2-D modulation codes that have the same spatial frequency constraints (α, β). In particular, we consider the $(4,4;7;2,4)$ and $(1,2;1;2,4)$ 2-D modulation codes in Figures 6.5(a), 6.5(b), and 6.5(c) for the cases of sinc^2 PSF with $b_s = 1.5$, J_1^2 PSF with $b_c = 2.5$, and Gaussian PSF with $b_g = 0.6$, respectively. It is apparent that both 2-D modulation codes give similar results for every case; however, to be more precise, the BER performance of the $(1,2;1;2,4)$ 2-D modulation code is somewhat better than that of the $(4,4;7;2,4)$ 2-D modulation code. Therefore, it should be noted that even though the 2-D modulation codes have the same spatial frequency constraints (α, β), their corresponding BER performance might be different.

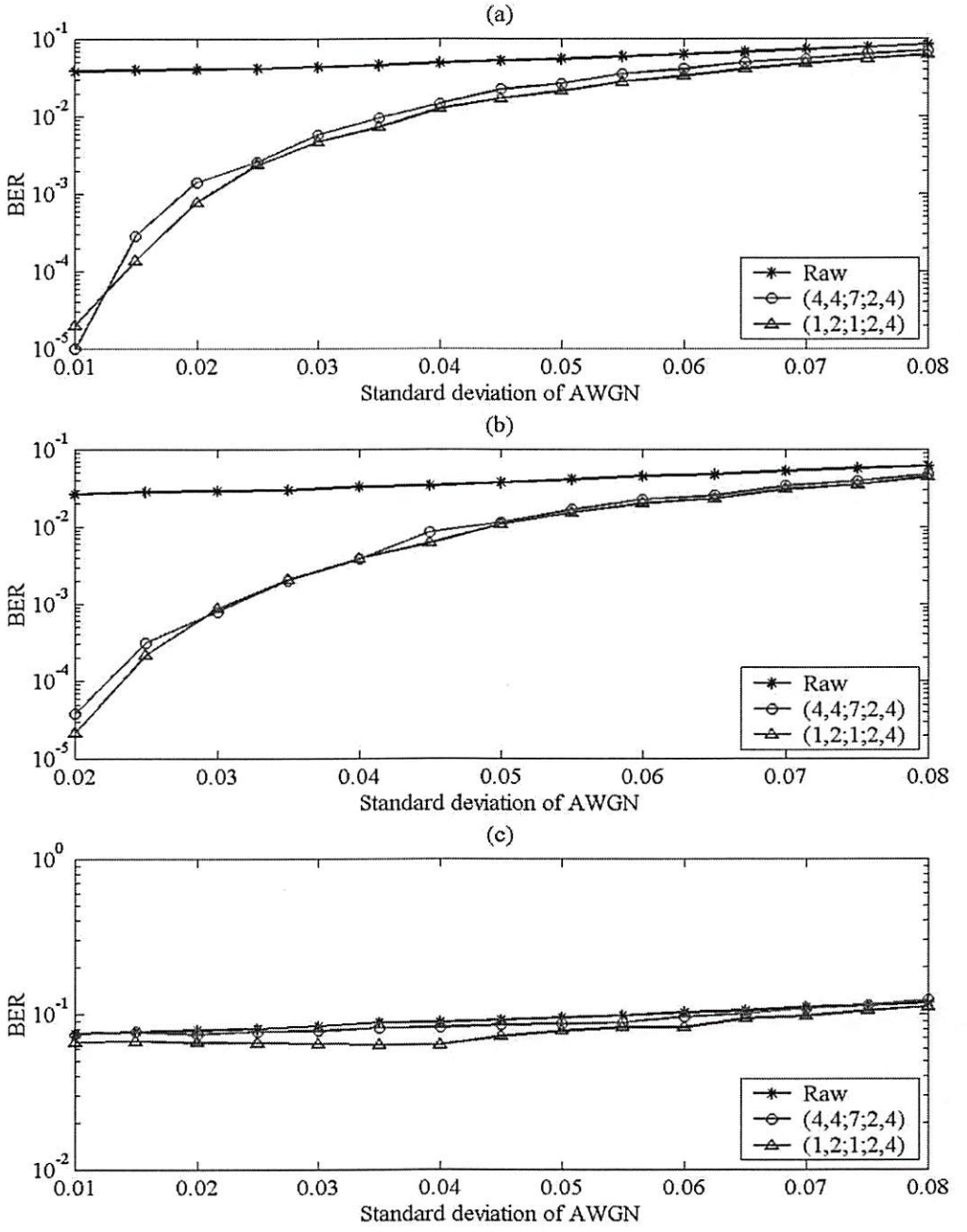


Figure 6.5 BER performance of $(4,4;7;2,4)$ and $(1,2;1;2,4)$ 2-D modulation codes in a noisy two-photon PODS system modeled by

(a) sinc^2 PSF with $b_s = 1.5$,

(b) J_1^2 PSF with $b_c = 2.5$, and

(c) Gaussian PSF with $b_g = 0.6$.

Chapter 7

An Extension to Three-Dimensional Modulation Coding for Two-Photon PODS Systems

In previous chapters, we have assumed that the distance between any two consecutive data pages recorded in a two-photon page-oriented optical data storage (PODS) system is large enough that there is no interaction or interference between the data pages. It is clear that the data storage capacity of the system can be increased if such distance can be reduced. However, if we record the data pages closer to each other, the bit-error-rate (BER) performance of the system may degrade due to the interference among closely recorded data pages. While the interference among bits within each data page discussed in earlier chapters is called the two-dimensional (2-D) intersymbol interference (ISI), the interference among data pages, which we discuss in this chapter, is called the interpage interference (IPI). The IPI, in fact, depends not only on the layout or format of the recorded data pages in the medium, but also on the arrangement of the data marks, i.e., the pattern of the information, in each recorded data page. In this chapter, we discuss three-dimensional (3-D) extensions of the 2-D system models described in chapter 3 and the 2-D modulation coding scheme described in chapter 4 in order to include the effects of IPI. The extended models of two-photon PODS systems are presented in the first

section. In the second section, we describe the basic principles of the 3-D modulation coding scheme and show a few examples of 3-D modulation codes.

7.1 Three-Dimensional Models and Interpage Interference of Two-Photon PODS Systems

Assuming that the recording process is perfect and that the distance between any two adjacent recorded data pages is small enough that they interfere with each other, then the mathematical model of a two-photon PODS system that includes the effects of interference among data pages, i.e., the effects of IPI, can be derived from the retrieving process as shown in Figure 7.1.

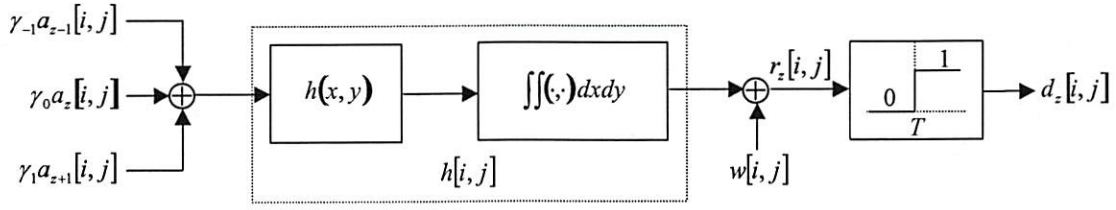


Figure 7.1 The 3-D mathematical model of a two-photon PODS system.

Let $[i,j]$ and (x,y) denote an index of a 2-D discrete signal and an index of a 2-D continuous signal, respectively. As described in chapter 3, the point-spread function (PSF) of the imaging system in the retrieving process, $h(x,y)$, can be modeled as either one of the functions in Eqs. (3.1), (3.2), and (3.3) or any other function depending on the

imaging optics and the characteristics of the retrieval system. For simplicity, we assume again that the magnification of the imaging system in the retrieving process is one. Moreover, the pitch between two nearest data marks, equivalently the pitch between two nearest detectors or the pixel pitch of the charge-coupled device (CCD) camera, is set to one so that all other distances can be relatively expressed in units of this pitch. It should also be noted that the size of the data mark is assumed to be very small compared to the extent of the PSF; hence, the data mark is effectively a delta function.

We assume that $a_z[i,j]$ represents the existence of a digital information bit (or data mark) recorded on the z th data page, where $z \in \{\cdots,-2,-1,0,1,2,\cdots\}$ and that it takes on value of either zero for a "0" bit or one for a "1" bit. We also assume that only the $(z-1)$th and $(z+1)$th data pages can interfere with the z th data page. Thus, the received intensity at each detector, or each CCD pixel, that arises from the z th data page in a *noiseless* retrieval system, then, can be modeled mathematically by

$$r_z[i,j] = \gamma_{-1}(a_{z-1}[i,j] \otimes h[i,j]) + \gamma_0(a_z[i,j] \otimes h[i,j]) + \gamma_1(a_{z+1}[i,j] \otimes h[i,j]) \quad (7.1)$$

$$= (\gamma_{-1}a_{z-1}[i,j] \otimes h[i,j]) + (\gamma_0 a_z[i,j] \otimes h[i,j]) + (\gamma_1 a_{z+1}[i,j] \otimes h[i,j]) \quad (7.2)$$

$$= (\gamma_{-1}a_{z-1}[i,j] + \gamma_0 a_z[i,j] + \gamma_1 a_{z+1}[i,j]) \otimes h[i,j]. \quad (7.3)$$

where γ_{-1}, γ_0, and γ_1 are nonnegative and $\gamma_{-1}+\gamma_0+\gamma_1 = 1$, $\otimes$ is a 2-D discrete convolution operator, and $h[i,j]$ represents the effective discrete PSF defined in Eq. (3.5).

Clearly, when the $(z-1)$th and $(z+1)$th data pages weakly interfere with the z th data page, the values of γ_{-1} and γ_1 should be much less than the value of γ_0. In

the extreme case, when the values of γ_{-1} and γ_1 are equal to zero and the value of γ_0 is one, i.e., when there is no IPI, Eq. (7.3) simply becomes Eq. (3.4). We note that because γ_{-1}, γ_0, and γ_1 indicate the relative degree of the effects of IPI, they are called the IPI coefficients. Furthermore, it is possible that other data pages, such as the $(z-2)$th and $(z+2)$th data pages, might interfere with the zth data page as well. For such an unlikely event, the additional terms representing the IPI effects from the $(z-2)$th and $(z+2)$th data pages must be included, where the IPI coefficients, again, are nonnegative and sum to one for normalization purposes.

If thermal noise effects from the electronic circuits at the detector plane are taken into account, the received intensity at each detector, arising from the zth data page, of a *noisy* retrieval system that also includes the effects of IPI can be described by

$$r_z[i,j] = (\gamma_{-1}a_{z-1}[i,j] + \gamma_0 a_z[i,j] + \gamma_1 a_{z+1}[i,j]) \otimes h[i,j] + w[i,j], \tag{7.4}$$

where $w[i,j]$ is an additive white Gaussian noise (AWGN), whose mean and variance are zero and σ_w^2, respectively.

Applying a simple binary threshold decision scheme with a *priori* knowledge of a threshold value, T, at the CCD or detector array to digitize the received intensities, the detected data, i.e., the output of the retrieval system, can be modeled as

$$d_z[i,j] = \begin{cases} 1, \text{if } r_z[i,j] > T \\ 0, \text{if } r_z[i,j] < T \end{cases}. \tag{7.5}$$

As discussed in chapter 3, it is often helpful to consider the 2-D discrete signals as matrices for the purpose of analysis. For example, Eq. (7.4) can be expressed in matrix form by

$$\mathbf{R}_z = (\gamma_{-1}\mathbf{A}_{z-1} + \gamma_0\mathbf{A}_z + \gamma_1\mathbf{A}_{z+1}) \otimes \mathbf{H} + \mathbf{W}, \tag{7.6}$$

where $\mathbf{R}_z$ is an $M \times N$ received intensity matrix derived from the z th recorded data page, whose entry in the p th row and q th column, $R_{z,pq}$, describes the real-valued intensity at the $[p,q]$-th detector or the $[p,q]$-th CCD pixel, $\mathbf{A}_{z-1}$, $\mathbf{A}_z$, and $\mathbf{A}_{z+1}$ denote the $(z-1)$th, z th, and $(z+1)$th recorded $M \times N$ digital information data page, respectively, in which each entry is either zero or one. Also, γ_{-1}, γ_0, and γ_1 represent the IPI coefficients that are always nonnegative and sum to one, $\mathbf{H}$ is the effective PSF matrix derived from $h[i,j]$ in Eq. (3.5) with minor modifications of indices so that $h[0,0]$ is the entry at the center of the matrix $\mathbf{H}$, and $\mathbf{W}$ represents an $M \times N$ AWGN matrix whose entries are independent, identically distributed with mean and variance defined in Eq. (7.4).

We exemplify in Figure 7.2 a convenient way to visualize the effects of IPI. Suppose that the retrieval system of a two-photon PODS system is noiseless and its imaging system can be modeled as a sinc^2 PSF with $b_s = 1.5$. We also assume that the $(z-1)$th recorded data page contains a data mark at $[i,j] = [0,0]$, the z th recorded data page contains two data marks at $[i,j] = [-1,0]$ and $[i,j] = [1,0]$, and the $(z+1)$th recorded data page contains a data mark at $[i,j] = [0,0]$. Assuming $\gamma_{-1} = 0.1$, $\gamma_0 = 0.8$, and

$\gamma_1 = 0.1$, the plots on the left side of Figure 7.2 from top to bottom show: a one-dimensional (1-D) slice through the x-axis of the IPI that comes from a data mark recorded in the $(z-1)$th data page; a 1-D slice through the x-axis of the image intensity originating from two data marks recorded in the z th data page; a 1-D slice through the x-axis of the IPI that arises from a data mark recorded in the $(z+1)$th data page; and a 1-D slice through the x-axis of the total image intensity, including the effects of IPI from two nearest neighboring data pages, at the detector plane when the z th data page is retrieved, respectively.

On the other hand, each plot shown on the right side of Figure 7.2 depicts the received intensity, corresponding to the image intensity shown on the left side, of the detectors or CCD pixels that lie on the x-axis, i.e., $r[i,0]$. Here, we assume that each detector has a unit linear fill factor ($\delta = 1$). Ideally, for the z th recorded data page, the digitized outputs of all detectors or CCD pixels should be "0" bits, except the ones at $[i,j]=[-1,0]$ and $[i,j]=[1,0]$. From the second plot on the right side of Figure 7.2, the effects of 2-D ISI from data marks at $[i,j]=[-1,0]$ and $[i,j]=[1,0]$ lead to the uncertainty of the output of the detector, or CCD pixel, at $[i,j]=[0,0]$ because the intensity level at $[i,j]=[0,0]$ is about half of the intensity level of the actual "1" bit at $[i,j]=[-1,0]$ or $[i,j]=[1,0]$. Indeed, the output of the detector, or CCD pixel, at $[i,j]=[0,0]$ in this particular case depends on the threshold value T of a simple binary threshold decision scheme.

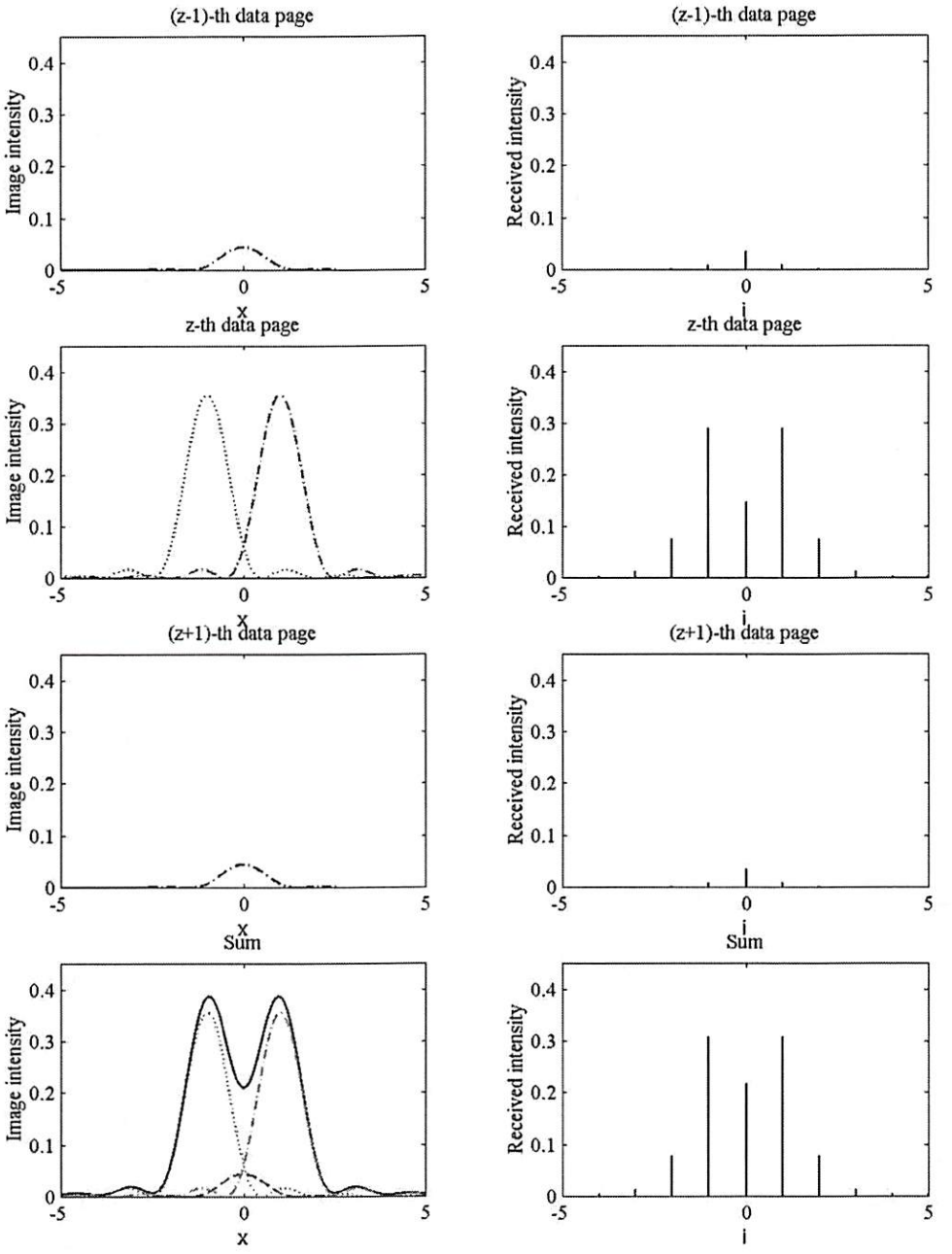


Figure 7.2 An example of the effects of IPI, where $\gamma_{-1} = 0.1$, $\gamma_0 = 0.8$, and $\gamma_1 = 0.1$, in a noiseless two-photon PODS system modeled by the sinc^2 PSF with $b_s = 1.5$ and $\delta = 1$.

However, it is evident from the bottom-right plot in Figure 7.2 that, adding the effects of IPI, the output of the detector or CCD pixel at $[i, j] = [0,0]$ is very likely to be a "1" bit, regardless of the value of a predetermined threshold, because the intensity level at $[i, j] = [0,0]$ is much more than half of the intensity level of the actual "1" bit at $[i, j] = [-1,0]$ or $[i, j] = [1,0]$. This implies that, in addition to the 2-D ISI, the IPI also further degrades the system performance. Figures 7.3 and 7.4 illustrate the similar results for the cases of J_1^2 PSF with $b_c = 2.5$ and Gaussian PSF with $b_g = 0.6$, respectively.

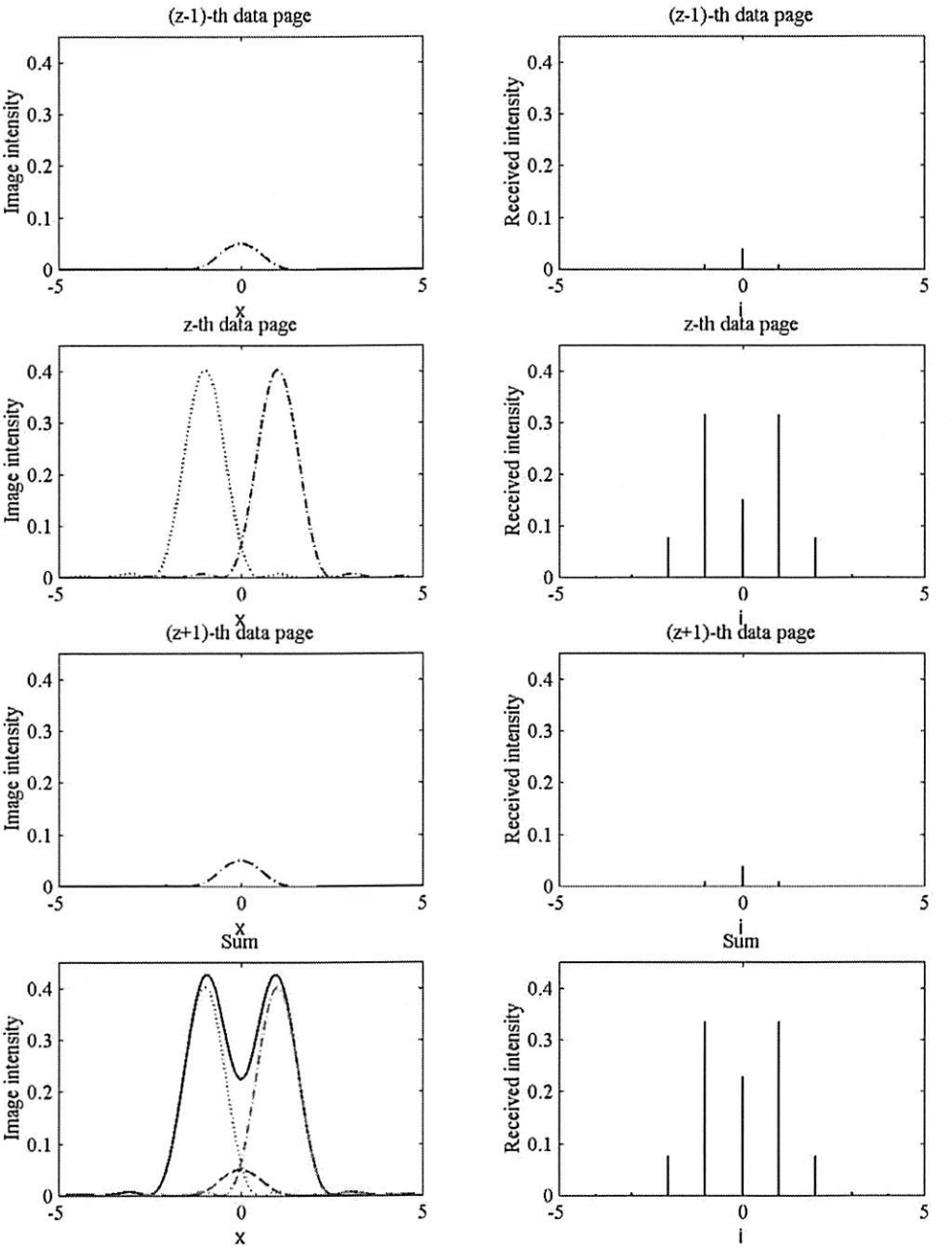


Figure 7.3 An example of the effects of IPI, where $\gamma_{-1} = 0.1$, $\gamma_0 = 0.8$, and $\gamma_1 = 0.1$, in a noiseless two-photon PODS system modeled by the J_1^2 PSF with $b_c = 2.5$ and $\delta = 1$.

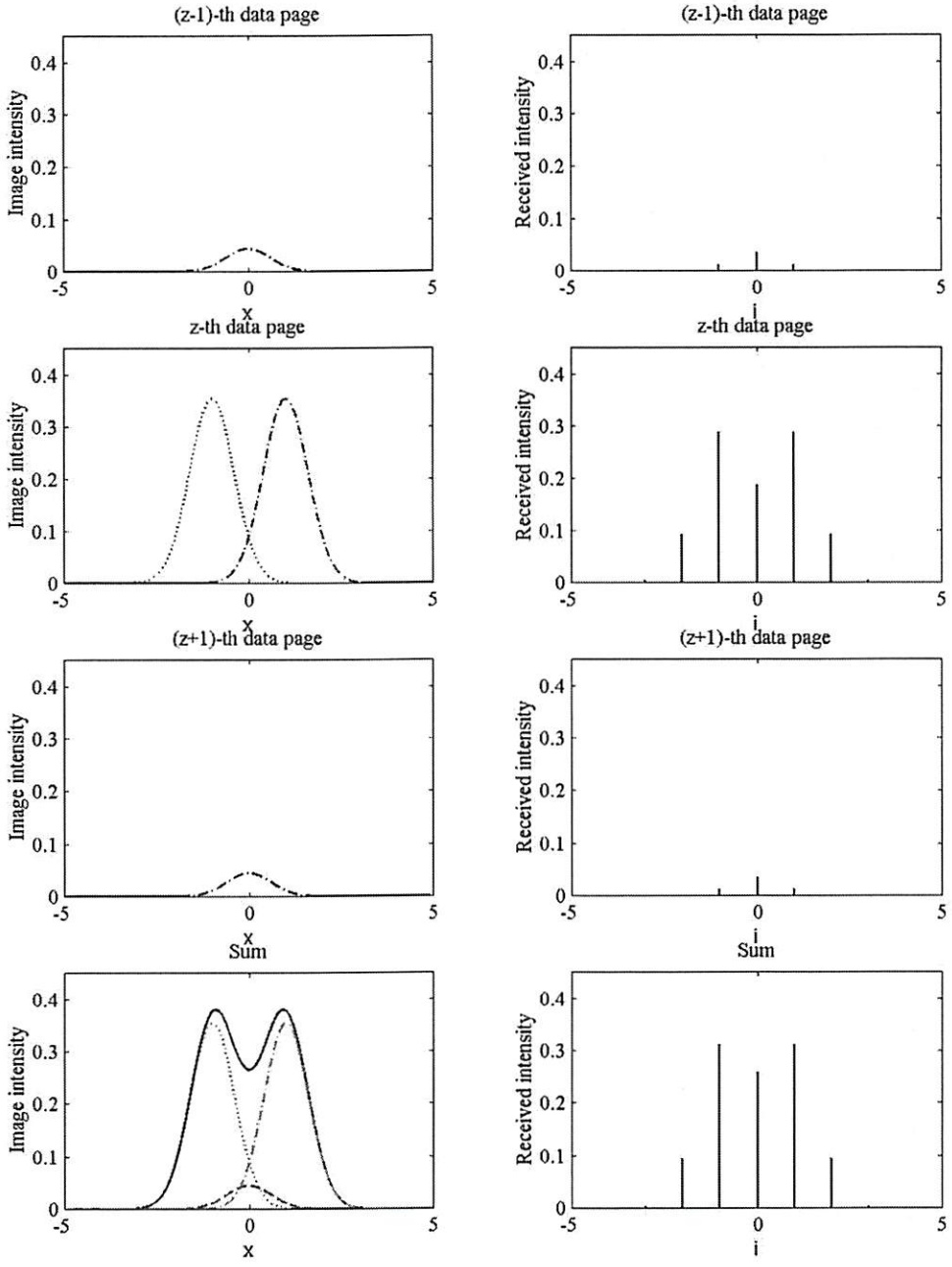


Figure 7.4 An example of the effects of IPI, where $\gamma_{-1} = 0.1$, $\gamma_0 = 0.8$, and $\gamma_1 = 0.1$, in a noiseless two-photon PODS system modeled by the Gaussian PSF with $b_g = 0.6$ and $\delta = 1$.

7.2 Basic Principles of Three-Dimensional Modulation Coding Scheme and Examples of Three-Dimensional Modulation Codes

Apparently, from section 7.1, the simplest way to avoid IPI in a two-photon PODS system is to separate the recorded data pages as far as needed. However, if the distances among recorded data pages are increased, the utilization of the medium, then, decreases. Therefore, rather than increasing the distances among recorded data pages, we propose a new 3-D modulation coding scheme that not only can reduce the effects of 2-D ISI, but also can reduce the effects of IPI.

As discussed in chapter 4, the 2-D ISI is regarded as the light from a "1" bit that interferes with other local neighbor "0" bits (of the same data page). The effects of 2-D ISI, thus, can be reduced by limiting the number of "1" bits around any "0" bit on each data page. On the other hand, as discussed in the previous section, the IPI can be regarded as the fraction of light from a "1" bit of the adjacent recorded data page that interferes with "0" bits of the retrieved data page. Hence, in order to mitigate the effects of 2-D ISI and the effects of IPI in a two-photon PODS system simultaneously, the main principle of our 3-D modulation coding scheme, intuitively, is to map a 1-D sequence of input information into a 3-D output data in such a way that any "0" bit recorded in the medium is not surrounded, volumetrically, by too many "1" bits [44]. We also note that any "1" bit in the medium is preferred to be volumetrically surrounded by many "0" bits, although "1" bits can be recorded closely.

Following the principles of the proposed 3-D modulation coding scheme, we present two examples of 3-D modulation codes. The first 3-D modulation code can be described mathematically as

$$[a_0 \quad a_1] \rightarrow \{[0 \quad a_0]_{2z}; [a_1 \quad 0]_{2z+1}\}, \tag{7.7}$$

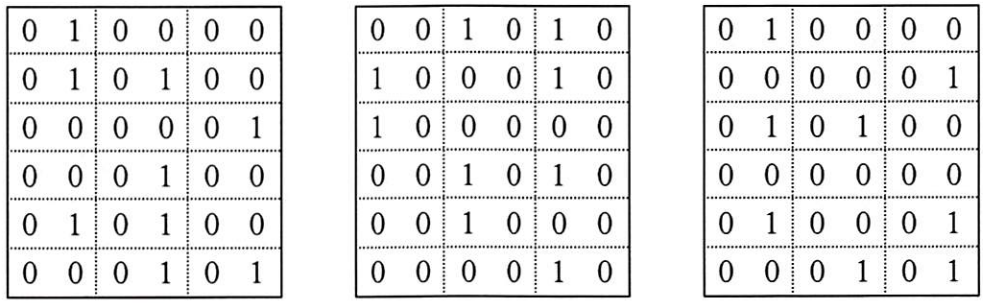
where $z \in \{\cdots, -2, -1, 0, 1, 2, \cdots\}$. Assuming that there is a lengthy sequence of input information bits, the 3-D modulation code in Eq. (7.7), then, maps each 1-D block of two input information bits into a 3-D block of $1 \times 2 \times 2$ bits. In other words, one of the two input information bits, a_0, is mapped into two bits, $[0 \quad a_0]$, on one data page, i.e., the $(2z)$th data page, while the other input information bit, a_1, is mapped into other two bits, $[a_1 \quad 0]$, on the following data page, i.e., the $(2z+1)$th data page. It is necessary to note that since this 3-D modulation code, totally, maps two bits into four bits, its code rate is equal to 0.5.

Notice that the 3-D modulation code in Eq. (7.7) can be derived from the $(1,2;1;2,4)$ 2-D modulation code described in chapter 4 and its mirror version; therefore, the effects of 2-D ISI in the retrieved data page can be reduced. Moreover, it should be emphasized that this 3-D modulation code ensures that there is always a "0" bit recorded on each adjacent recorded data page before and after any input information bit; thus, the effects of IPI among the recorded data pages can be reduced as well.

We note that the 3-D modulation code in Eq. (7.7) can also be described simply as

$$\begin{cases} a \rightarrow [0 \quad a], & \text{even data page} \\ a \rightarrow [a \quad 0], & \text{odd data page} \end{cases}, \tag{7.8}$$

where a is an input information bit. Figure 7.5 depicts an example of the recorded data pages in a two-photon PODS system after applying the 3-D modulation code in Eq. (7.7) or Eq. (7.8).



Even data page Odd data page Even data page

Figure 7.5 An example of the recorded data pages after applying a 3-D modulation code based on the $(1,2;1;2,4)$ 2-D modulation code.

Similarly, we can derive another example of a 3-D modulation code from a $(4,4;8;2,2)$ 2-D modulation code defined in chapter 4. This 3-D modulation code can be described mathematically as [44]

$$[a_0 \quad a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad a_7 \quad a_8 \quad a_9 \quad a_{10} \quad a_{11} \quad a_{12} \quad a_{13} \quad a_{14} \quad a_{15}] \rightarrow$$

$$\left\{ \begin{bmatrix} a_0 & a_2 & 0 & 0 \\ a_1 & a_3 & 0 & 0 \\ 0 & 0 & a_4 & a_6 \\ 0 & 0 & a_5 & a_7 \end{bmatrix}_{2z} ; \begin{bmatrix} 0 & 0 & a_{12} & a_{14} \\ 0 & 0 & a_{13} & a_{15} \\ a_8 & a_{10} & 0 & 0 \\ a_9 & a_{11} & 0 & 0 \end{bmatrix}_{2z+1} \right\}, \tag{7.9}$$

where $z \in \{\cdots, -2, -1, 0, 1, 2, \cdots\}$. The 3-D modulation code in Eq. (7.9) maps each 1-D block of sixteen input information bits into a 3-D block of $4 \times 4 \times 2$ bits. In fact, eight of

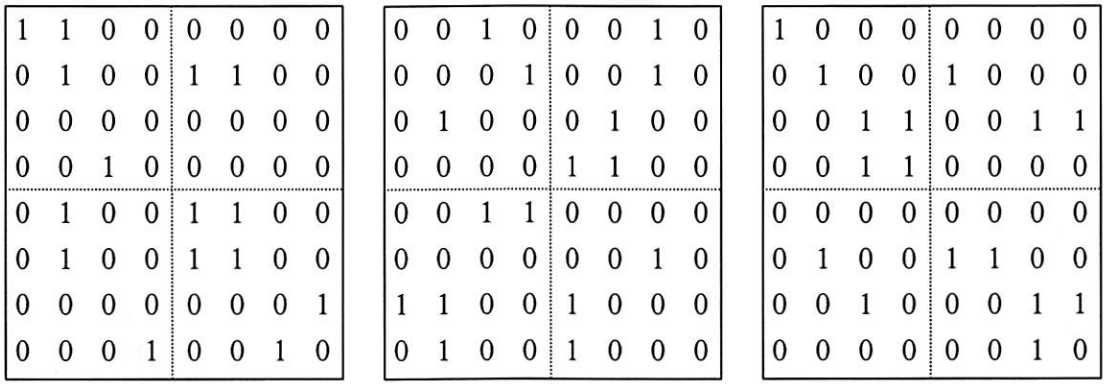
the sixteen input information bits are mapped into a 2-D block of 4×4 bits on one data page, i.e., the $(2z)$th data page, by using the $(4,4;8;2,2)$ 2-D modulation code, while the other eight input information bits are mapped into another 2-D block of 4×4 bits on the following data page, i.e., the $(2z+1)$th data page, by using the mirror version of the $(4,4;8;2,2)$ 2-D modulation code. We note that because this 3-D modulation code maps sixteen bits into thirty-two bits, its code rate is equal to 0.5.

Since the 3-D modulation code in Eq. (7.9) is derived from the $(4,4;8;2,2)$ 2-D modulation code and its mirror version, it can relax the effects of 2-D ISI in the retrieved data page. Additionally, it is clear that, with this 3-D modulation code, there is always a "0" bit recorded on each adjacent recorded data page before and after any input information bit. Hence, the effects of IPI among the recorded data pages can also be reduced.

Alternatively, the 3-D modulation code in Eq. (7.9) can be described as

$$\begin{cases} [a_0 \quad a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad a_7] \rightarrow \begin{bmatrix} a_0 & a_2 & 0 & 0 \\ a_1 & a_3 & 0 & 0 \\ 0 & 0 & a_4 & a_6 \\ 0 & 0 & a_5 & a_7 \end{bmatrix}, & \text{even data page} \\ [a_0 \quad a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad a_7] \rightarrow \begin{bmatrix} 0 & 0 & a_2 & a_0 \\ 0 & 0 & a_3 & a_1 \\ a_6 & a_4 & 0 & 0 \\ a_7 & a_5 & 0 & 0 \end{bmatrix}, & \text{odd data page} \end{cases}, \quad (7.10)$$

where $[a_0 \quad a_1 \quad a_2 \quad a_3 \quad a_4 \quad a_5 \quad a_6 \quad a_7]$ is a 1-D block of eight input information bits. We show in Figure 7.6 an example of the recorded data pages in a two-photon PODS system after applying the 3-D modulation code in Eq. (7.9) or Eq. (7.10).



Even data page Odd data page Even data page

Figure 7.6 An example of the recorded data pages after applying a 3-D modulation code based on the $(4,4;8;2,2)$ 2-D modulation code.

Chapter 8

Conclusions, Discussions, and Future Extensions

8.1 Conclusions and Discussions

This dissertation discusses the appropriate modulation coding for page-oriented optical data storage (PODS) systems that use two-photon absorption technology. In chapter 3, we have presented a number of two-dimensional (2-D) mathematical models for various two-photon PODS systems based on the assumption that the distance between any two successive recorded data pages is large enough such that there is no interaction between the data pages. With these 2-D mathematical models, we have examined the inherent intersymbol interference (ISI) of two-photon PODS systems and its effects. It is obvious that the ISI in two-photon PODS systems is different from the ISI in conventional data storage systems. The former is 2-D and spreads over the spatial domain, whereas the latter is one-dimensional (1-D) and spreads over the time domain.

We have proposed in chapter 4 a new 2-D modulation coding scheme that can relieve, or even eliminate, the effects of 2-D ISI in two-photon PODS systems. Since ISI in two-photon PODS systems is mainly due to the light from a data mark or a "1" bit that spatially interferes with other local neighbors, especially the "0" bits, on the same recorded data page, the fundamental principle of our 2-D modulation coding scheme is to limit the number of "1" bits around any "0" bit on each recorded data page. We have

also described a general form of 2-D modulation codes derived from the proposed 2-D modulation coding scheme. Using this general form, we can differentiate the 2-D modulation codes and approximately compare their performance. In addition, we have presented a variety of fixed-length 2-D modulation codes for two-photon PODS systems, including their advantages and disadvantages.

In chapter 5, we have discussed restrictions of fixed-length 2-D modulation codes in terms of code rate. We have also presented an example of the code rate upper bounds of fixed-length 2-D modulation codes to support this fact. To overcome the moderate code rates of fixed-length 2-D modulation codes, we have proposed variable-length 2-D modulation codes for two-photon PODS systems. Due to the higher degrees of freedom, variable-length 2-D modulation codes typically have higher code rates than fixed-length 2-D modulation codes (at the same constraints). However, one major concern about variable-length 2-D modulation codes is how to decode them with a minimum of error propagation. An example of a variable-length 2-D modulation code and its proper decoding procedure has been presented. We have shown in chapter 6 the bit-error-rate (BER) performance of these fixed-length and variable-length 2-D modulation codes. A numerical procedure that is used to find the optimal threshold value for a binary threshold decision scheme applied at the detector array of a noisy two-photon PODS system has also been described.

If the distance between any two consecutive data pages recorded in a two-photon PODS system is small enough that they interfere with each other, the interpage interference (IPI) must be taken into account. To investigate the effects of both 2-D ISI

and IPI, we have presented many three-dimensional (3-D) mathematical models for two-photon PODS systems in chapter 7. We have also shown that the IPI further degrades the system performance in addition to the 2-D ISI. Therefore, we have proposed a novel 3-D modulation coding scheme for two-photon PODS systems in order to mitigate the effects of 2-D ISI and the effects of IPI simultaneously. Because the 2-D ISI can be regarded as the light from a "1" bit of the retrieved data page that interferes with other local neighbor "0" bits on the same data page and the IPI can be regarded as the fraction of light from a "1" bit of the adjacent recorded data page that interferes with "0" bits of the retrieved data page, the basic idea of our 3-D modulation coding scheme is to have any "0" bit recorded in the storage medium surrounded, volumetrically, by a minimum number of "1" bits. We have also developed several examples of 3-D modulation codes based on examples of fixed-length 2-D modulation codes. With these 3-D modulation codes, we expect significant improvements, in terms of BER performance, in two-photon PODS systems.

Furthermore, it should be emphasized that while the primary objectives of conventional modulation coding are to reduce the effects of 1-D ISI and to recover the clock information from the recorded data, the main goals of our multi-dimensional modulation coding are to reduce the effects of 2-D ISI and to reduce the effects of IPI. Unlike conventional data storage systems, the clock information in two-photon PODS systems can be either embedded in the recorded data pages or supplied separately in the clock channels [41]. Figure 8.1 shows an example of the recorded data page format or layout in a two-photon PODS system with four clock channels (I, II, III, and IV). We

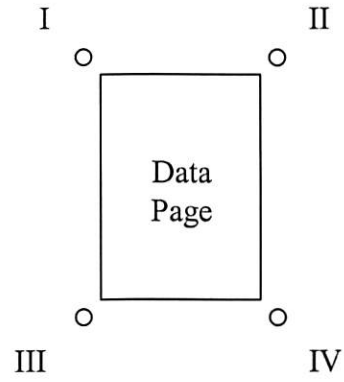


Figure 8.1 An example of the recorded data page format or layout in a two-photon PODS system with four clock channels at each corner of a data page (I, II, III, and IV).

note that although only one clock channel is necessary, the extra clock channels provide redundancy for error protection and are useful for data page tracking purposes. Figure 8.2 illustrates, for instance, how four clock channels can be used for data page tracking. On the odd data pages, clock channels I and IV carry the "1" bits (filled circles) whereas clock channels II and III carry the "0" bits. On the other hand, on the even data pages, clock channels I and IV carry the "0" bits whereas clock channels II and III carry the "1" bits (filled circles). It is clear that two "1" bits at a pair of clock channels across the diagonal are used not only to track the position of each data page, but also for redundancy of clock information.

Figure 8.2 An example of the data page tracking in a two-photon PODS system with four clock channels at each corner of a data page (each filled circle represents a data mark or a "1" bit).

8.2 Future Extensions

The work described in this dissertation provides many opportunities for future research. A number of interesting ideas can be outlined by the following:

- Further improve the 2-D and 3-D mathematical models of two-photon PODS systems. For example, consider more practical noise models than the additive white Gaussian noise (AWGN).
- Develop new fixed-length 2-D, variable-length 2-D, and nontrivial 3-D modulation codes with diverse properties for two-photon PODS systems.
- Analyze and simulate the BER performance of 3-D modulation codes for two-photon PODS systems.
- Study the optimization of the code rates of 2-D and 3-D modulation codes for two-photon PODS systems.

- Evaluate the trade-offs between the useful data storage capacity and the BER in two-photon PODS systems before and after applying our 2-D and 3-D modulation codes. Similar work has been investigated in holographic PODS systems [6].
- Develop novel channel coding schemes for two-photon PODS systems. An example of a 3-D interleaved error correcting coding scheme based on Reed-Solomon codes has been described in Refs. 40 and 43.
- Design a smart pixel array for retrieving (recording) process in two-photon PODS system that converts optical (electronic) signals into electrical (optical) signals, performs modulation decoding (encoding) based on our 2-D and 3-D modulation codes, and provides parallel-to-serial (serial-to-parallel) formatting [58].
- Study the feasibility of adaptive multi-dimensional modulation coding for two-photon PODS systems, in which the lookup table can be updated corresponding to the incoming input information. This idea has been initiated by the work in Ref. 60.

References

[1] K. G. Ashar, *Magnetic Disk Drive Technology: Heads, Media, Channel, Interfaces, and Integration* (IEEE Press, New Jersey, 1997).

[2] 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," Appl. Opt. **36,** 3107-3115 (1997).

[3] M.-P. Bernal, G. W. Burr, H. Coufal, and M. Quintanilla, "Balancing interpixel cross talk and detector noise to optimize areal density in holographic storage systems," Appl. Opt. **37,** 5377-5385 (1998).

[4] M. J. Buckingham, *Noise in Electronic Devices and Systems* (Halsted Press, New York, 1983).

[5] 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. **22,** 639-641 (1997).

[6] 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. **37,** 5431-5443 (1998).

[7] X. Chen, K. M. Chugg, and M. A. Neifeld, "Near-optimal parallel distributed data detection for page-oriented optical memories," IEEE J. Sel. Top. Quantum Electron. **4,** 866-879 (1998).

[8] W.-C. Chou and M. A. Neifeld, "Interleaving and error correction in volume holographic memory systems," Appl. Opt. **37,** 6951-6968 (1998).

[9] K. M. Chugg, X. Chen, and M. A. Neifeld, "Two-dimensional equalization in coherent and incoherent page-oriented optical memory," J. Opt. Soc. Am. A **16,** 549-562 (1999).

[10] C. De Caro, A. Renn, and U. P. Wild, "Hole burning, Stark effect, and data storage:2:holographic recording and detection of spectral holes," Appl. Opt. **30,** 2890-2898 (1991).

[11] A. S. Dvornikov, S. Esener, and P. M. Rentzepis, "Three-dimensional optical storage memory by means of two-photon interaction," in *Optical Computing Hardware*, J. Jahns and S. H. Lee, eds. (Academic Press, Massachusetts, 1994), pp.287-325.

[12] P. A. Franaszek, "Run-length-limited variable-length coding with error propagation limitation," U.S. Patent 3,689,899 (5 September 1972).

[13] B. J. Goertzen and P. A. Mitkas, "Error-correcting code for volume holographic storage of a relational database," Opt. Lett. **20,** 1655-1657 (1995).

[14] J. W. Goodman, *Introduction to Fourier Optics*, 2nd ed. (McGraw-Hill, New York, 1996).

[15] J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Volume holographic storage and retrieval of digital data," Science **265,** 749-752 (1994).

[16] J. F. Heanue, M. C. Bashaw, and L. Hesselink, "Channel codes for digital holographic data storage," J. Opt. Soc. Am. A **12,** 2432-2439 (1995).

[17] J. F. Heanue, K. Gurkan, and L. Hesselink, "Signal detection for page-access optical memories with intersymbol interference," Appl. Opt. **35,** 2431-2438 (1996).

[18] J. H. Hong, I. McMichael, T. Y. Chang, W. Christian, and E. G. Paek, "Volume holographic memory systems: techniques and architectures," Opt. Eng. **34,** 2193-2203 (1995).

[19] J. Hong, I. McMichael, and J. Ma, "Influence of phase masks on cross talk in holographic memory," Opt. Lett. **21,** 1694-1696 (1996).

[20] W.-F. Hsu, "Design of smart pixel interfaces for optical page-oriented memories," Ph.D. dissertation (University of Southern California, Los Angeles, California, 1996).

[21] S. Hunter, F. Kiamilev, S. Esener, D. A. Parthenopoulos, and P. M. Rentzepis, "Potentials of two-photon based 3-D optical memories for high performance computing," Appl. Opt. **29,** 2058-2066 (1990).

[22] J. F. Hutton, G. A. Betzos, M. Schaffer, and P. A. Mitkas, "Error correcting codes for page-oriented optical memories," in *Materials, Devices, and Processing for Optoelectronic Processing*, B. Javidi and J. A. Neff, eds., Proc. SPIE **2848,** 146-156 (1996).

[23] K. A. S. Immink, "Runlength-limited sequences," Proc. IEEE **78,** 1745-1759 (1990).

[24] K. A. S. Immink, "EFMPlus: The coding format of the multimedia compact disc," IEEE Trans. Consumer Electron. **41**(3), 491-497 (1995).

[25] K. A. S. Immink, "A survey of codes for optical disk recording," IEEE J. Sel. Areas Commun. **19,** 756-764 (2001).

[26] J. L. Kann and F. B. McCormick, "Numerical simulations of scattering in a two-photon optical data storage system," Appl. Opt. **37,** 4173-4182 (1998).

[27] M. Keskinoz and B. V. K. Vijaya Kumar, "Application of linear minimum mean-squared-error equalization for volume holographic data storage," Appl. Opt. **38,** 4387-4393 (1999).

[28] B. M. King and M. A. Neifeld, "Parallel detection algorithm for page-oriented optical memories," Appl. Opt. **37,** 6275-6298 (1998).

[29] B. M. King and M. A. Neifeld, "Sparse modulation coding for increased capacity in volume holographic storage," Appl. Opt. **39,** 6681-6688 (2000).

[30] B. Kohler, S. Bernet, A. Renn, and U. P. Wild, "Storage of 2000 holograms in a photochemical hole-burning system," Opt. Lett. **18,** 2144-2146 (1993).

[31] H.-Y. S. Li and D. Psaltis, "Three dimensional holographic disk," Appl. Opt. **33,** 3764-3774 (1994).

[32] S. Lin and D. J. Costello, Jr., *Error Control Coding: Fundamentals and Applications* (Prentice-Hall, New Jersey, 1983).

[33] 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," Appl. Opt. **34,** 4140-4148 (1995).

[34] MATLAB software (Math Works, Inc., Natick, Mass., 1994).

[35] F. B. McCormick, "2-photon optical storage technology," in *Optical Processing and Computing*, Y. Fainman, ed., SPIE International Technical Working Group Newsletter, Special Issue on Optical Data Storage, Vol. 9, No. 1, pp. 1, April 1998.

[36] M. A. Neifeld and J. D. Hayes, "Error-correction schemes for volume optical memories," Appl. Opt. **34,** 8183-8191 (1995).

[37] M. A. Neifeld and M. McDonald, "Error correction for increasing the usable capacity of photorefractive memories," Opt. Lett. **19,** 1483-1485 (1994).

[38] B. H. Olson and S. C. Esener, "Partial response precoding for parallel-readout optical memories," Opt. Lett. **19,** 661-663 (1994).

[39] B. H. Olsen and S. C. Esener, "One and two dimensional parallel partial response for parallel readout optical memories," in *Proceedings of IEEE International Symposium on Information Theory 1995*, pp. 141.

[40] D. E. Pansatiankul and A. A. Sawchuk, "Multi-dimensional modulation codes and error correction for page-oriented optical data storage," in *Optical Data Storage Topical Meeting*, SPIE Technical Digest (The International Society for Optical Engineering, Bellingham, Washington, 2001), pp. 94-96.

[41] D. E. Pansatiankul and A. A. Sawchuk, "Page-oriented optical data storage using multi-dimensional modulation codes," in *OSA Annual Meeting and Exhibit 2001*, Conference Program (Optical Society of America, Washington, DC, 2001), pp. 53.

[42] D. E. Pansatiankul and A. A. Sawchuk, "Two-dimensional modulation codes for page-oriented optical data storage systems," in *Optics in Computing*, OSA Technical Digest (Optical Society of America, Washington, DC, 2001), pp. 79-81.

[43] D. E. Pansatiankul and A. A. Sawchuk, "Multi-dimensional modulation codes and error correction for page-oriented optical data storage," in *Optical Data Storage 2001*, T. Hurst and S. Kobayashi, eds., Proc. SPIE **4342,** 393-400 (2002).

[44] D. E. Pansatiankul and A. A. Sawchuk, "Three-dimensional modulation codes for two-photon page-oriented optical data storage systems," submitted to *Joint International Symposium on Optical Memory and Optical Data Storage Topical Meeting 2002*, Waikoloa, Hawaii, July 2002.

[45] D. E. Pansatiankul and A. A. Sawchuk, "Variable-length two-dimensional modulation codes for two-photon page-oriented optical data storage systems," accepted for *2002 International Conference of Optics in Computing*, Taipei, Taiwan, ROC, April 2002.

[46] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. (McGraw-Hill, New York, 1991).

[47] D. A. Parthenopoulos and P. M. Rentzepis, "Three-dimensional optical storage memory," Science **245,** 843-845 (1989).

[48] K. C. Pohlmann, *The Compact Disc Handbook*, 2nd ed. (A-R Editions, Wisconsin, 1992).

[49] W. K. Pratt, *Digital Image Processing*, 2nd ed. (Wiley, New York, 1991).

[50] J. G. Proakis, *Digital Communications*, 3rd ed. (McGraw-Hill, New York, 1995).

[51] D. Psaltis, "Parallel optical memories," Byte **17**(9), 179-182 (1992).

[52] D. Psaltis and G. W. Burr, "Holographic data storage," Computer **31**(2), 52-60 (1998).

[53] D. Psaltis and F. Mok, "Holographic memories," Sci. Am. **273**(5), 70-76 (1995).

[54] A. Pu and D. Psaltis, "High-density recording in photopolymer-based holographic three-dimensional disks," Appl. Opt. **35,** 2389-2398 (1996).

[55] I. S. Reed and X. Chen, *Error-Control Coding for Data Networks* (Kluwer Academic Publishers, Massachusetts, 1999).

[56] A. Renn and U. P. Wild, "Spectral hole burning and hologram storage," Appl. Opt. **26,** 4040-4042 (1987).

[57] B. E. A. Saleh and M. C. Teich, *Fundamentals of Photonics* (Wiley, New York, 1991).

[58] A. A. Sawchuk, "Smart pixel devices and free-space digital optics applications," in *Proceedings of 1995 IEEE/LEOS Annual Meeting* (Institute of Electrical and Electronics Engineers, Piscataway, New Jersey, 1995), pp. 268-269.

[59] F. M. Schellenberg, W. Lenth, and G. C. Bjorkland, "Technological aspects of frequency domain data storage using persistent spectral hole burning," Appl. Opt. **25,** 3207-3216 (1986).

[60] L. Selavo, D. M. Chiarulli, and S. P. Levitan, "Real-time adaptive encoding for 3D optical memories," presented at SPIE Annual Meeting, San Diego, California, July 2001, paper 4459B-57.

[61] V. Vadde and B. V. K. Vijaya Kumar, "Channel modeling and estimation for intrapage equalization in pixel-matched volume holographic data storage," Appl. Opt. **38,** 4374-4386 (1999).

[62] V. Vadde and B. V. K. Vijaya Kumar, "Parity coding for page-oriented optical memories with intrapage intensity variations," Opt. Lett. **24,** 546-548 (1999).

[63] A. Vardy, M. Blaum, P. H. Siegel, and G. T. Sincerbox, "Conservative arrays: multidimensional modulation codes for holographic recording," IEEE Trans. Inf. Theory **42,** 227-230 (1996).

[64] M. M. Wang, S. C. Esener, F. B. McCormick, I. Cokgor, A. S. Dvornikov, and P. M. Rentzepis, "Experimental characterization of a two-photon memory," Opt. Lett. **22,** 558-560 (1997).

[65] J. Watkinson, *Coding for Digital Recording* (Focal Press, Massachusetts, 1990).

[66] J. Watkinson, *An Introduction to Digital Audio* (Focal Press, Massachusetts, 1994).

[67] J. Watkinson, *The Art of Data Recording* (Focal Press, Massachusetts, 1994).

[68] S. B. Wicker, *Error Control Systems for Digital Communication and Storage* (Prentice-Hall, New Jersey, 1995).

[69] L. Zhang and M. A. Neifeld, "Blockwise data detection for spectral hole-burning memories," Appl. Opt. **40,** 1832-1842 (2001).