J. Opt. A: Pure Appl. Opt. **10** (2008) 115305 (10pp) doi:10.1088/1464-4258/10/11/115305

Hybrid multinary modulation codes for page-oriented holographic data storage

G Berger, M Dietz and C Denz

Institut für Angewandte Physik, Westfälische Wilhelms-Universität Münster, Corrensstraße 2–4, D-48149 Münster, Germany

E-mail: gberger@uni-muenster.de

Received 4 June 2008, accepted for publication 8 September 2008
Published 29 September 2008
Online at stacks.iop.org/JOptA/10/115305

Abstract
Hybrid multinary block codes for implementation in page-oriented holographic storage systems are proposed. The codes utilize combined phase and amplitude modulations to encode input data. In comparison to pure amplitude-or pure phase-modulated block code designs hybrid multinary modulation coding allows us to augment the storage density at an unchanged error rate. Two different hybrid modulation code designs are introduced. Experimental implementation is thoroughly discussed, especially concentrating on readout concepts. Phase-resolved readout is accomplished by optical addition and subtraction, using an unmodulated reference page. Experimental results indicate that the overall error rate is usually dominated by errors related to amplitude detection. The study suggests that capacity gains of up to 31% or 47% are reasonable when utilizing phase modulations in conjunction with binary or ternary amplitude modulation.

Keywords: volume holographic data storage, sparse block code, phase retrieval

It is the goal of any data storage technology to enable high data densities and high data transfer rates. Besides the utilized multiplexing method, e.g. like angular, shift or phase-code multiplexing [1], in page-oriented volume holographic systems these figures of merit are crucially affected by the employed two-dimensional data encoding technique. In a given system configuration the achievable storage density depends on the data integrity, which can be controlled through the error management procedure employed, including suitable 2d channel encoding. This emphasizes the importance and accounts for the great interest in the research into new parallel coding designs and detection schemes during the last decade (e.g. [2–15]). Before recording data in a holographic device usually error-correcting codes (ECC) and interleaving are applied to the data [16]. The resulting data stream has to be transformed into a 2d pixellated structure that is to be displayed by a spatial light modulator (SLM) in the signal arm. When using a binary amplitude modulator such a data page simply consists of OFF and ON pixels. Subsequently, the data bearing signal wave and a coherent, indexed reference wave interfere in a storage medium to record a hologram. In this vein, in volume holographic systems usually many holograms are superimposed, e.g. by means of angular or phase-code multiplexing. In order to guarantee equal diffraction efficiencies of the holograms, a recording schedule has to be followed. To ease the process it is desirable to have signal waves that exhibit a constant intensity, independent of the input data.

During readout a camera provides digitized images of reconstructed analog two-dimensional signals, which have to be transformed back into the original binary input data streams, consisting of ‘ones’ and ‘zeros’. The easiest way to perform this task is to define a global threshold and to classify pixels as ‘ones’ if their brightness exceeds the threshold value. Otherwise pixels are interpreted as ‘zeros’. The problem of such a detection method is that it is very vulnerable with respect to intensity variations across the data page. Unfortunately, in real systems such variations are, to some extent, always present. Hence, when sticking to this simple threshold detection, one would usually have to tolerate poor user data densities, since a lot of capacity is consumed, for example, by ECC in order to provide reasonable error rates. With regard to page-oriented systems modulation codes provide a powerful coding technique that provides solutions to the emphasized demands. It guarantees constant signal wave intensities and significantly increases the achievable capacity of data pages at a constant bit error rate [6, 5].

It is the basic idea of modulation coding to divide data pages into blocks of equal size. Using a binary amplitude input SLM each of these blocks consists of a constant number of OFF and ON channels. User data is encoded in a way that each input data word of d bits is mapped to a unique assembly of OFF and ON channels that form a block. During the retrieval process one just has to identify the positions of the brightest (or darkest) channels within each block, in order to classify a read out block as to represent a certain code word. Thereby, the technique is someway similar to a local threshold detection method. By use of parity information about the block design, it implicitly finds the best threshold within a small region. Due to the constant number of m_1 ON channels per block, such codes are referred to as *constant-weight block codes*. If each block consists of an equal number of OFF and ON channels, the code is called *balanced block code*. Commonly a constant-weight modulation code can be symbolized by d:b (m_1), where d denotes the number of user bits that are encoded in each block of b channels. The associated code rate $r = d/b$ expresses the fraction of the theoretical storage capacity that can actually be utilized to encode user data (disregarding ECC, etc). In balanced block codes the weight m_1 is always equal to $b/2$. It is an important side effect of modulation coding that the total number of ON channels per page stays constant, no matter how the particular input user data looks like. This is especially important when using phase-code multiplexing, which relies on equal diffraction efficiencies of recorded holograms. In other words a constant number of ON channels per page contributes to a better data integrity. Assuming a nearly flat intensity profile and neglecting any non-uniform absorption effects, a constant number of ON channels implies a constant intensity of signal waves that are incident on the storage material. If the intensity of a signal wave that is composed just of ON pixels is normalized (=1), then the fraction $h = m_1/b$ is equivalent to the actual signal intensity when displaying data pages. The fraction h is usually referred to as sparseness.

The simplest modulation block code is a balanced 1:2 code, which produces data pages with blocks that consists of two channels. Each ‘zero’ or ‘one’ of a user bit stream is represented by a block that displays the channel sequence OFF–ON or ON–OFF, respectively. This coding design is usually referred to as *differential encoding* [6]. A 1:2 code provides a code rate of $r = 0.5$ and a sparseness of $h = 0.5$. In general, the code rate can be increased by increasing the size of each block [7]. For instance, the balanced 6:8 and the constant-weight 12:15 (7) code exhibit code rates of 0.75 and 0.8, respectively. Using binary encoding maximal a code rate of $r = 1$ can be reached when the block size goes to infinity. Any further increase of the data page capacity as well as enabling better control of the tradeoff between code rate, error rate and sparseness, can only be accomplished by utilizing more channel states. At first sight the utilization of additional gray values besides ON and OFF pixels suggests itself. Thorough experimental analyses of this approach suggest the best performance when employing ternary (3-ary) modulation [3, 8]. In this case a capacity gain of 11% can be accomplished. Using a special encoding scheme that extenuates error propagation, the estimated capacity improvement can be increased up to 35% [8]. The term ‘error propagation’ describes the fact that one misinterpreted channel results in an incorrect block symbol that may be associated with a completely different data word. Thereby one channel error can produce many user bit errors. Besides amplitude-modulated block codes, modulation codes for pure phase-modulated data pages have also been proposed [9].

In the following we present the implementation of constant-weight hybrid multinary modulation codes that are based on combined amplitude and phase modulations. In section 1, a fundamental construction rule and general expressions of code rates of different hybrid modulation codes are derived. When employing hybrid data pages the development of a readout procedure capable of detecting amplitudes and phases is a central issue. After briefly specifying the experimental set-up employed (section 2), in section 3 we present a robust readout method, whose implementation is straightforward in volume holographic storage systems that realize phase-code multiplexing. In section 4, the performance of several codes is experimentally studied and the pros and cons of different detection methods are discussed. Finally the practical value of hybrid modulation codes is estimated with respect to data capacity, error rate and data transfer rate.

1. Construction of phase-and amplitude-modulated block codes

The simplest hybrid encoding scheme, in which both amplitudes and phases are utilized, is known as hybrid ternary modulation (HTM) [17, 18]. This format is composed of three different channel states: one OFF state with zero amplitude and two ON states with an equal non-zero amplitude and diametrical phases (0 and π). Initially it has been proposed to reduce the dc component of Fourier-transformed data pages. Typically a focused purely amplitude-modulated data page gives rise to a very strong dc component that tends to saturate the storage medium while higher orders of the Fourier transform may not be sufficiently resolved. The idea of HTM is to smooth the intensity distribution in the Fourier plane, in order to allow effective recording of Fourier transform holograms. Here, this modulation scheme is utilized for data encoding. Corresponding modulation codes can be generated by specifying the frequency of occurrence of each state within each block. For instance, table 1(a) shows all code words that can be generated using HTM, where the block size comprises three channels ($b = 3$). Since data pages typically suffer more from undesired amplitude variations rather than relative phase variations of data channels (see below), a second approach also seems to be worthwhile. Codes may be designed as constant-weight modulation codes with respect to amplitude modulations, but without imposing any constraint on the utilization of predefined phase levels. In terms of hybrid ternary modulation table 1(b) shows all code words that can be generated for a block size of three channels ($b = 3$), if the sparsity of each phase state per block is variable. In the style of the above code notation, hybrid block codes can be symbolized

Table 1. List of code words that can be generated by means of hybrid ternary modulation (HTM) with constant (a) or variable (b) phase sparsity for a block size of 3 channels. '0' and '1' denote OFF and ON channels. The superscript denotes the relative phase (either 0 or π) of the corresponding ON channel.

No.	Code word s_i = constant				User word
(a)					
1	0	1^0	1^π	$\longleftrightarrow$	00
2	0	1^π	1^0	$\longleftrightarrow$	01
3	1^0	0	1^π	$\longleftrightarrow$	10
4	1^π	0	1^0	$\longleftrightarrow$	11
5	1^0	1^π	0		
6	1^π	1^0	0		
(b)					
1	0	1^0	1^0	$\longleftrightarrow$	000
2	0	1^0	1^π	$\longleftrightarrow$	001
3	0	1^π	1^0	$\longleftrightarrow$	010
4	0	1^π	1^π	$\longleftrightarrow$	011
5	1^0	0	1^0	$\longleftrightarrow$	100
6	1^0	0	1^π	$\longleftrightarrow$	101
7	1^π	0	1^0	$\longleftrightarrow$	110
8	1^π	0	1^π	$\longleftrightarrow$	111
9	1^0	1^0	0		
11	1^0	1^π	0		
11	1^π	1^0	0		
12	1^π	1^π	0		

by d:b $(m \times p)^{c,v}$, where m and p denote the numbers of different amplitude and phase states, respectively. For binary amplitude modulation m is equal to 2. b denotes the number of channels per block and d is the length of each user data word that can be encoded per block. The superscripts c and v indicate a constant or variable phase sparsity. Accordingly, employing four code words of table 1(a) enables the realization of a 2:3$(2 \times 2)^c$ code. Utilizing 8 of the code words shown in table 1(b) allows us to implement a 3:3$(2 \times 2)^v$ code. In these two cases code rates of $r = 0.67$ and 1 are obtained. Please note that the code rates presented are computed as the ratio of user bits d and the number of channels per block b. Therefore, when extending conventional binary data pages by means of additional phase modulations (or grayscale values) code rates greater than 1 can be achieved.

1.1. Number of distinguishable code words

Table 1 indicates that typically not all of the possible code words of a certain configuration are used to encode user data. The middle columns in tables 1(a) and (b) list all producible unique code words for the two cases of hybrid ternary modulation. The columns on the right side illustrate possible assignments of user data words. This incomplete utilization can give rise to difficulties during data detection. If all code words are utilized, every retrieved block symbol can unambiguously be assigned to a digital user word, even if channels have been falsely classified. But, as the utilization rate decreases, the possibility of identifying block symbols that are not part of the actually used set of code words increases. Instead of just sorting the channels of a block, more complicated decision schemes, e.g. based on the Viterbi algorithm, may be used for more accurate detection [10, 15].

In general, the number of distinguishable code words N needs to be larger than or equal to the number of possible user words. When employing binary modulation codes that are composed just of OFF and ON channels, this interrelationship can be expressed by $2^d \leqslant N$, where the number of code words N is given as $N^{\text{binary}} = \binom{b}{m_1}$. For instance, in order to encode 4-bit user words ($d = 4$), a suitable modulation code needs to be capable of distinguishing between at least $2^4 = 16$ different user words. This requirement is fulfilled by a code of block size $b = 6$ and weight $m_1 = 3$, since (6 *choose* 3) = 20 > 16, giving rise to a 4:6(3) code. When employing additional grayscale values the number of code words can be determined by means of the multinomial coefficient [7]

$$N^{\text{gray}} = \frac{b!}{\prod_i m_i!}, \tag{1}$$

where m_i is the absolute frequency of occurrence of the ith grayscale value. m_0 is the number of OFF channels, with $m_0 + \sum_{i>0}^{(m-1)} m_i = b$ (where m denotes the number of different amplitude states, e.g. $m = 2$ for binary encoding). For instance, if 6-bit user words are to be encoded per block, at least $2^6 = 64$ code words need to be provided. This can be accomplished by blocks of $b = 6$ channels when utilizing three grayscale values, e.g. OFF, 0.5· ON and ON. Evaluating (1) indicates that 90 code words can be generated if $m_i = 2$ for these three states.

When incorporating phase modulations, the two above discussed cases, i.e. constant or variable phase sparsity, have to be distinguished. If the frequency of occurrence of each combined amplitude–phase state is constant, the number of possible code words N^c is equal to N^{gray}, i.e. $N^c = b!/\prod_i m_i!$. Here, each m_i represents the number of channels per block that exhibit state i, corresponding to a unique amplitude–phase combination. If modulation coding is only implemented with respect to the distribution of amplitudes and no constraints are imposed on the utilization of phase states, higher code rates can be achieved due to greater numbers of allowed code words. It is discussed below that this capacity gain is accomplished at the expense of a weaker error robustness in comparison to the codes with constant phase sparsity. The number of producible code words N^v can be determined by modifying the coefficient given in (1). Let m_j denote the frequency of occurrence of each non-zero amplitude state j and p_j its number of possible phase states, then the number of producible code words can be calculated by

$$N^v = \frac{b!}{m_0} \prod_{j>0} \frac{p_j^{m_j}}{m_j!}, \tag{2}$$

where m_0 denotes the number of OFF channels per block.

Based on equations (1) and (2) it is straightforward to deduce the code rates related to hybrid multinary modulation codes. They can be written as

$$r^c = \lfloor \log_2 (N^c) \rfloor / b = \left\lfloor \log_2 \left(\frac{b!}{\prod_i m_i!} \right) \right\rfloor \Big/ b \tag{3}$$

Table 2. Sample hybrid modulation codes with constant phase sparsity (d = user bits, b = block length, m = no. of amplitude states, p = no. of phase states).

Code		No. of amplitude states			No. of phase states				Ch. states	Sparseness	Code rate
d:b	($m \times p$)	m_0	m_1	m_2	p_1	p_2	p_3	p_4	k	h	r
8:8	(2 × 2)	4	4	—	2	2	—	—	3		1.00
14:12	(2 × 2)	6	6	—	3	3	—	—	3		1.17
25:20	(2 × 2)	10	10	—	5	5	—	—	3		1.25
54:40	(2 × 2)	20	20	—	10	10	—	—	3	0.50	1.35
143:100	(2 × 2)	50	50	—	25	25	—	—	3		1.43
169:100	(2 × 3)	50	50	—	17	17	16	—	4		1.69
187:100	(2 × 4)	50	50	—	13	13	12	12	5		1.87
20:12	(2 × 4)	4	8	—	2	2	2	2	5	0.67	1.67
20:12	(3 × 2)	4	4	4	4	4	—	—	5	0.50	1.67
70:40	(3 × 2)	20	10	10	10	10	—	—	5	0.38	1.75
79:40	(3 × 2)	14	13	13	13	13	—	—	5	0.49	1.98
413:150	(3 × 4)	50	50	50	25	25	25	25	9	0.50	2.75
5:8	(2 × 2)	6	2	—	1	1	—	—	3		0.63
10:12	(2 × 3)	9	3	—	1	1	1	—	4		0.83
19:20	(2 × 4)	15	5	—	2	1	1	1	5	0.25	0.95
44:40	(2 × 4)	30	10	—	3	3	2	2	5		1.10
49:40	(3 × 4)	30	5	5	3	3	2	2	9		1.23

Table 3. Sample hybrid modulation codes with variable phase sparsity (d = user bits, b = block length, m = no. of amplitude states, p = no. of phase states).

Code		No. of amplitude states			Distinct ch. states	Sparseness	Code rate
d:b	($m \times p$)	m_0	m_1	m_2	k	h	r
10:8	(2 × 2)	4	4	—	3		1.25
15:12	(2 × 2)	6	6	—	3		1.25
27:20	(2 × 2)	10	10	—	3		1.35
57:40	(2 × 2)	20	20	—	3	0.50	1.43
146:100	(2 × 2)	50	50	—	3		1.46
175:100	(2 × 3)	50	50	—	4		1.75
196:100	(2 × 4)	50	50	—	5		1.96
24:12	(2 × 4)	4	8	—	5	0.67	2.00
23:12	(3 × 2)	4	4	4	5	0.50	1.92
74:40	(3 × 2)	20	10	10	5	0.38	1.85
83:40	(3 × 2)	14	13	13	5	0.49	2.08
430:150	(3 × 4)	50	50	50	9	0.50	2.87
6:8	(2 × 2)	6	2	—	3		0.75
12:12	(2 × 3)	9	3	—	4		1.00
23:20	(2 × 3)	15	5	—	5	0.25	1.15
49:40	(2 × 4)	30	10	—	5		1.23
57:40	(3 × 4)	30	5	5	9		1.43

for a constant phase sparsity and

$$r^{\mathrm{v}} = \left\lfloor \log_2 \left(\frac{b!}{m_0} \prod_{j>0} \frac{p_j^{m_j}}{m_j!} \right) \right\rfloor \Big/ b \qquad (4)$$

for a variable phase sparsity. $\lfloor\,\rfloor$ denotes the floor operator. In tables 2 and 3 several different sample codes are illustrated. The indicated amplitude sparseness h of each code is calculated assuming equidistant intensity values. k denotes the number of distinct channel states (i.e. of unique amplitude–phase combinations) that are utilized to generate code words. It is apparent that larger block sizes and more possible channel states enable higher code rates. In addition, as indicated by the form of the coefficients given in (1) and (2), it is obvious that equal frequency of occurrences of the different employed channel states suggest highest code rates in any given situation. However, code construction based solely on this criterion easily results in codes that exhibit a high sparseness. An example is the 20:12 $(2 \times 4)^{\mathrm{c}}$ code, which exhibits a high code rate of $r = 1.67$, but at a page sparseness of 0.67. When multiplexing many data pages a high sparseness is typically undesired, since corresponding pages sooner exhaust the usable dynamic range of the employed storage material. Hence, the actual number of holograms that can be multiplexed is reduced. Nevertheless, the codes listed in the tables emphasize that the utilization of amplitude and phase modulations for data encoding can significantly augment the page capacity in comparison to binary modulation coding. For instance, the balanced 9:12 binary modulation code ($r = 0.75$) can be improved by distinguishing two different phase states to form a 14:12 $(2\times2)^{\mathrm{c}}$ code and a 15:12 $(2\times2)^{\mathrm{v}}$ code, providing code rates of 1.17 or 1.25, respectively. When increasing the block size b, it becomes apparent that, on average, the relative gain in terms of the related code rate enabled by a variable phase sparsity over that of a constant phase sparsity becomes smaller. It should also be noted that in principle the same improvement with respect to the code rate can be achieved either by increasing the number of grayscale values or the number of phase states, when assuming constant phase sparsity. Concisely, the utilization of different phase states drastically improves the control over the tradeoff between amplitude sparseness and related code rates. Eventually, the actual capacity gain provided by any of these codes has to be experimentally estimated.

2. Experimental set-up

Experimental storage and readout of data pages modulated by hybrid block codes is investigated in the 90° set-up depicted in figure 1. An optically pumped semiconductor laser (OPSL) (λ = 488 nm, cw) serves as the light source. In

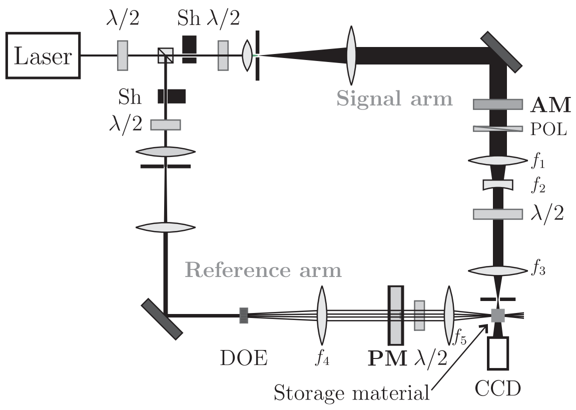

Figure 1. Volume holographic storage set-up employing phase-code multiplexing. (AM = amplitude modulator; PM = phase modulator; DOE = diffractive optical element; $\lambda/2$ = half − wave plate; Sh = shutter; POL = polarizer; CCD = camera for data page detection.)

the signal arm a transmissive twisted nematic liquid crystal display (TN LCD) with a resolution of 800 × 600 pixels (32 μm pitch, 85% fill factor) allows us to imprint amplitude-modulated data pages onto the expanded collimated beam. The signal wave propagates through a telescope arrangement ($f_1 = 80$ mm, $f_2 = -50$ mm, $f_3 = 80$ mm) with a focus point ≈2 mm in front of the storage media. A CCD camera with a 1280 × 1024 pixel resolution detects the imaged data page. In the reference arm a diffractive optical element (DOE) splits the incident collimated laser beam into 128 discrete beams that are focused ($f_4 = 160$ mm) on a phase-only modulator. By means of another lens behind the modulator ($f_5 = 100$ mm) 128 collimated reference beams are generated ($D \approx 1.5$ mm) that interfere in their point of intersection inside the storage medium with the signal wave. This configuration enables phase-code multiplexing. In the presented experiments typically 16 data pages are multiplexed, which allows us to group the reference beams into 16 sets of 8 beams that are distributed over the whole aperture. Thereby, unwanted intensity variations of the beams are kept below ±2.5%. The phase modulator allows us to control the relative phase shift of each reference beam with an accuracy of ≈1%. In the presented experiments iron-doped lithium niobate ($LiNbO_3$:Fe) is employed as storage material (5 × 7 × 7 mm^3). Recording data pages with a sparseness of 0.5, the configuration provides an M/# = 1.6.

3. Input and readout of hybrid data pages

In order to enable the utilization of hybrid modulation codes, two central demands have to be fulfilled. First of all, the associated phase and amplitude of each pixel of a data page need to be independently controlled. Second, during readout different phase states within reconstructed data pages have to be discriminated, in order to be able to retrieve all user information.

Usually, data modulators are operated in phase-only or amplitude-only modulation modes. Though it is, in principal, possible to perform phase and amplitude modulations at the same time, most devices would, in such cases, give rise to severe inhomogeneities. Typically, it is not possible to entirely avoid changes of the amplitude transmittance of a pixel as the phase delay is varied. The lack of a suitable modulator can be circumvented by using one modulator for pure phase modulation and another one for pure amplitude modulation. These two modulators could be stuck together or the input plane of the first modulator could be (pixel-matched) imaged onto the second one. In this way, it is possible to generate two-dimensional amplitude-and phase-modulated data pages. In addition, systems that are based on phase-code multiplexing allow a completely different approach. For this purpose the data pages to be stored are split into subpages, each one containing only data channels of equal phase. These subpages are consecutively recorded by the same phase-code, while adjusting the required overall phase offsets in the reference arm. Thereby, during readout with any phase-code, amplitude- and phase-modulated data pages are reconstructed. Since the phase adjustments are accomplished in the reference arm, rather than by a two-dimensional modulator in the signal arm, this method eliminates unwanted amplitude modulations related to signal wave phase modulation. However, the technique entails some constraints, as discussed in section 4.

In order to retrieve all information contained in reconstructed data pages, their two-dimensional phase and amplitude distributions have to be ascertained. It is, of course, straightforward to measure spatial intensity distributions using, for example, a CCD camera, but accomplishing phase resolution requires some additional effort. Again, when using a system based on phase-code multiplexing one readout approach suggests itself. Phase encoding inherently offers the opportunity to perform optical arithmetic operations like `AND`, `OR` and `XOR` [21, 22]. If a homogeneous reference page of sparseness 1 is recorded among the data pages, the phase distribution of any reconstructed data page can be revealed by simultaneous readout of that page and the reference page. At the same time the relative phase shift between the two reconstructed pages can be controlled by adjusting an offset in the readout beams. In this manner several phase states can easily be differentiated. Even though phase-code multiplexing obviously offers the easiest way to implement this readout method, it is not limited to this multiplexing technique. In systems based on pure angular multiplexing an additional reference beam could be used to record the additional reference page. Enabling phase control of one of the two reference beams allows us to follow the same approach. In the following the detection procedure is, in general, explained for such interferometric approaches.

Assume that various hybrid modulation-encoded data pages have been multiplexed. In order to retrieve the user information contained in one of the pages, in the first step, the desired data page is read out by addressing the medium with the appropriately indexed reference wave. The reconstructed page is detected by a CCD camera, revealing its two-dimensional intensity distribution. Figure 2(a) shows an example of this reconstruction step in the case of an HTM modulated data page. In this step no phase information

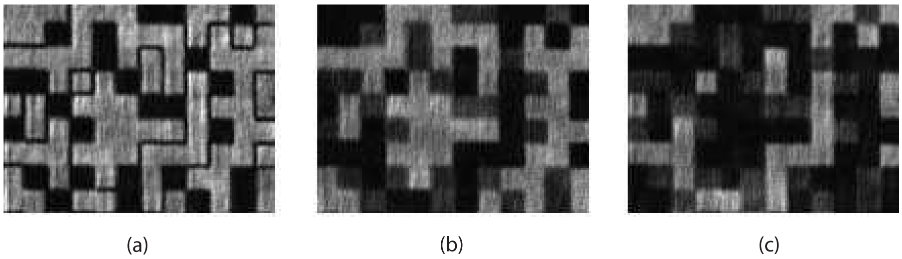

Figure 2. Magnified parts of images detected during readout of an HTM encoded data page. Each cutout displays 12 blocks. The data are encoded by a 15:12 (2×2) code, arranged in blocks of $b = 3 \times 4$ (y-dim $\times$ x-dim) data channels. (a) Reconstruction of the data page for OFF and ON channel determination. (b), (c) Simultaneous readout of the data page and the reference page in order to classify the phase states of the ON channels. The relative phase shifts of the reference wave with respect to the data page are adjusted to π (b) and 0 (c), respectively.

is retrieved. All further readout steps aim to resolve the phase states comprised in the addressed page. For that purpose the data page and the reference page are reconstructed simultaneously, while controlling the relative phase shift between the two pages to be 0 and π, respectively. In the first case the ON, 0 channels interfere constructively with the reference page, giving rise to strong channel intensities, while the ON, π channels exhibit very low intensities. When adjusting the relative phase difference to be π all ON, π channels receive strong intensities and all ON, 0 channels are attenuated. The phase detection process for a corresponding HTM data page is illustrated in the two magnified cutouts in figures 2(b) and (c). In both images the strong attenuation of parts of the ON channels is clearly discernible. In addition, it attracts attention that OFF channels exhibit considerable intensities. The light stems from the reference page, which is basically transmitted in these portions of the data page. Anyway, during phase detection, only intensity alterations of ON channels are considered, whose positions have already been ascertained in the first readout step. In the example shown in figure 2, it is straightforward to unambiguously differentiate the phase states. This readout procedure provides a high phase classification reliability, but requires several readout steps in order to access the whole data content of a page. If a hybrid ternary modulation code is utilized, three readout steps have to be performed. Each additional phase shift utilized in the employed hybrid code requires two additional reconstruction steps. In the following section results are presented that have been accomplished employing this readout method. In section 5 a readout method is discussed that reduces the required number of readout processes.

4. Performance of constant-weight hybrid block codes

Before discussing the actual experimental results it is necessary to define a measure that allows us to assess the capabilities of different coding scenarios. For that purpose, the commonly used bit error rate (BER) or symbol error rate (SER) definitions are modified in order to enable independent study of errors related to amplitude and phase detection. A data channel is accurately read out, if its phase, as well as its amplitude, are correctly classified. Let AER and PER denote the corresponding amplitude and phase error rates, respectively. Then, the probability of correct classification of data channels can be written as

$$P_{\text{correct}} = (1 - \text{AER}) \left(1 - \text{PER}_j \frac{m_j}{b}\right), \tag{5}$$

where m_j/b is the frequency of occurrence of the amplitude state j per block. When employing several phase states, based on (5), a data channel error rate (DCER) can be defined as

$$\begin{aligned} \text{DCER} &= 1 - (1 - \text{AER}) \left(1 - \sum_j \text{PER}(j) \frac{m_j}{b}\right) \\ &= \text{AER} + \sum_j \text{PER}(j) \frac{m_j}{b} \\ &\quad - \text{AER} \left(\sum_j \text{PER}(j) \frac{m_j}{b}\right), \end{aligned} \tag{6}$$

in which the last term can typically be neglected since $\text{AER} \cdot \text{PER} \ll 1$. Experimental results are in the following described in terms of amplitude, phase and data channel error rates. The error rates are determined by fitting the tails of corresponding histograms [23]. Figure 3 shows an example in which 7 HTM modulated data pages (with variable phase sparsity) and a reference page have been multiplexed. In this case the corresponding error rates can be computed to $\text{AER} \approx 10^{-5}$ and $\text{PER} \approx 3 \times 10^{-6}$. Typically, it is the aim to maximize the capacity as far as possible by adapting experimental parameters until an error rate of 10^{-3} is reached, which can effectively be improved down to 10^{-12} by suitable error-correcting codes (ECC) (e.g. Reed–Solomon codes).

4.1. *Effect of the number of utilized phase states on phase and amplitude error rates*

When employing simultaneous phase and amplitude modulation to realize block encoding, one of the first questions that crops up is 'What is the effect of an increased number of utilized phase states on the overall error rate'. To tackle this question, under the same conditions various data pages are recorded that have been modulated by different hybrid block codes at a constant block size of $b = 3 \times 4 = 12$ channels. The block codes comprise up to five equidistant phase states, while only two amplitude states (OFF and ON) are employed. The results are shown in figure 4. When utilizing two phase states,

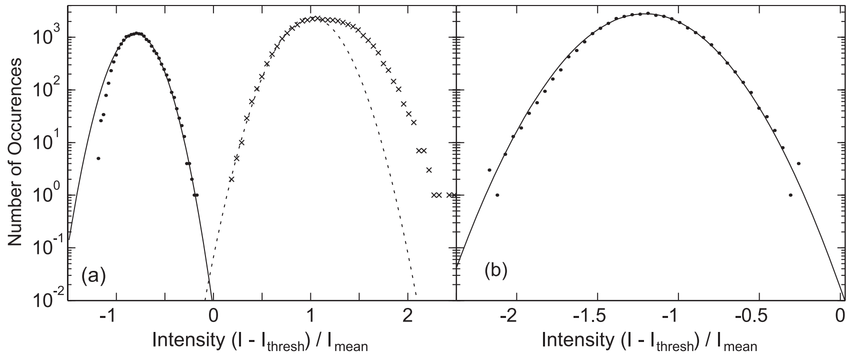

Figure 3. Error determination based on histograms that comprise data from 7 reconstructed HTM encoded data pages. The data pages and a reference page (sparseness 1) have previously been recorded by means of phase-code multiplexing. (a) OFF and ON channel histograms. (b) Phase channel histogram.

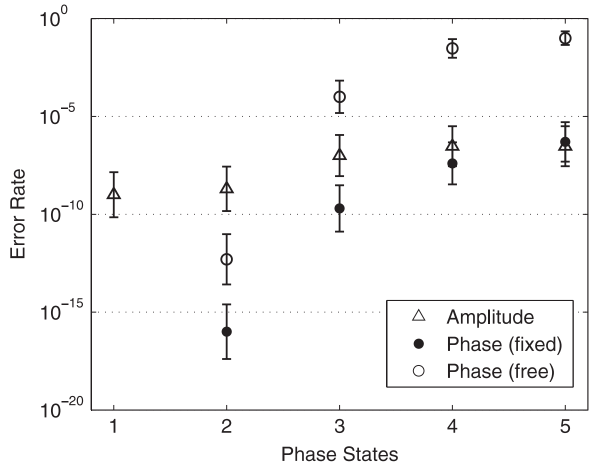

Figure 4. Amplitude and phase error rates for different numbers of phase states.

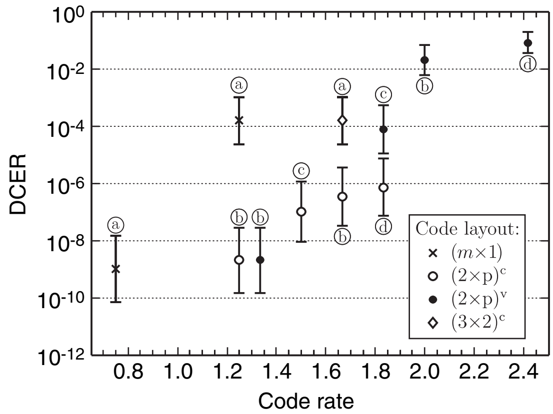

Figure 5. Data channel error rate for different data formats, in which the block size is set to $b = 3 \times 4 = 12$. The codes are composed by pure amplitude modulation $(m \times 1)$ with $m = 2, 3$, hybrid modulation with constant phase sparsity $(2 \times p)^{c}$ and variable phase sparsity $(2 \times p)^{v}$ with $p = 2, 3, 4, 5$. One hybrid code comprises three amplitude and two phase levels $(3 \times 2)^{c}$.

the amplitude error rate dominates the overall error rate. When increasing the number of possible phase states, it is clearly discernible that the phase error rate considerably deteriorates in the case of a variable phase sparsity. It adversely dominates the overall channel error rate if more than two different phase states are incorporated. In the case of a constant frequency of occurrence of the different phase states, the phase error rate remains below the amplitude error rate for up to 4 phase levels.

4.2. Enhancement of code rates

The above analysis suggests that the capacity of data pages can be considerably augmented by means of utilizing hybrid modulation codes. This conclusion is further illustrated in figure 5, where encountered data channel error rates of different codes are plotted versus the corresponding code rates. Note that the employed block codes in figure 5 are optimized with respect to the associated code rate. For that purpose the frequency of occurrences of each possible channel state (i.e. of each unique amplitude–phase combination) are equalized as far as possible. This approach results in sparseness values of $h_{(a)} = 0.5$, $h_{(b)} = 0.67$, $h_{(c)} = 0.75$ and $h_{(d)} = 0.83$. Such high sparseness values are only reasonable when multiplexing moderate numbers of holograms, since each page consumes a relatively large amount of the usable dynamic range during recording. As pointed out above, in such a case the overall storage density might be reduced, since fewer holograms can be multiplexed. If codes with a constant phase sparsity are used, the encountered error rate is significantly less than the target error rate of 10^{-3}. In terms of data encoding, in these cases the page capacity can be increased, e.g. by means of reducing the oversampling dimensions or by means of increasing the block lengths.

4.3. Estimation of actual capacity gains

In order to enable a reliable estimation of the achievable capacity gain when incorporating hybrid modulation codes the impact of the oversampling dimensions and the block lengths is analyzed. The results are shown in figures 6(a) and (b) for constant and variable phase sparsity, respectively. It is obvious that the probability of phase errors is always less than that of amplitude errors independent of the oversampling dimensions and the block length. The difference between the error rates is

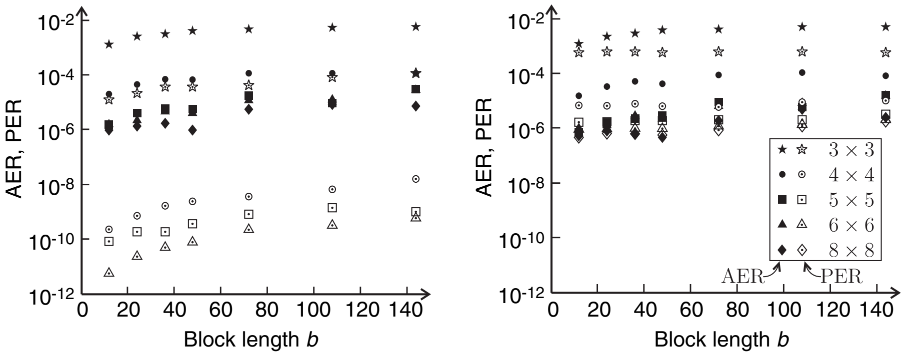

Figure 6. Amplitude and phase error rates of HTM encoded data for different oversampling dimensions and block lengths. (a) Constant phase sparsity. (b) Variable phase sparsity.

much more significant in the case of a constant phase sparsity than for a variable phase sparsity. In order to maximize the capacity one tends to reduce the oversampling dimensions as far as possible. Figure 6 indicates that an oversampling dimension of 3×3 pixels seems to be reasonable only for small block lengths in the present system. Further reduction of the oversampling is complicated when intending to use hybrid codes, since in reconstructed data pages adjacent channels are clearly separated if they display different phase states. The effect can be seen in figure 2(a). Due to this effect, in general a channel state has to be classified by evaluating fewer camera pixels, giving rise to a worse classification reliability, in comparison to the use of purely amplitude-modulated pages. Obviously, the effect is more pronounced when using smaller oversampling dimensions. In fact, the experiments suggest that 2×2 oversampling or no oversampling on the input modulator is only feasible if the camera used for detection possesses a much higher resolution or if the employed system enables pixel matching. Figure 6 suggests that maximal page capacities can be achieved by use of 3×3 oversampling at small block sizes or 4×4 oversampling at large block sizes. Further experiments of HTM block codes in the phase-encoded system reveal that the encountered characteristics persist, i.e. that the error rate due to erroneous phase detection stays below that related to amplitude detection.

Based on the above results, the actual capacity gain that can be achieved by means of hybrid modulation codes in the phase-encoded holographic storage system can be estimated. The page-related capacity can be controlled through the oversampling dimensions, the block lengths, the number of phase levels, the number of amplitude levels and the page sparseness. It is obvious that small oversampling dimensions are advantageous. When using 4×4 oversampling, block lengths up to $b = 150$ can be implemented while guaranteeing error rates less than 10^{-3}. However, although its implementation is complicated and block lengths have to be small, 3×3 oversampling can be used to increase the page capacity. Consider the use of 4×4 oversampling and block lengths $b = 150$ channels. Corresponding HTM data pages can encode 44 200 user bits. In contrast, HTM encoded data pages, where 3×3 oversampling and block lengths $b \leqslant 12$ are used, can encode maximal 66 495 user bits. In comparison, under the same conditions ($b = 12$, 3×3 oversampling) binary modulation codes would provide a capacity of just 39 987 bits. If no phase modulations are employed, block lengths of up to $b = 40$ can be utilized, resulting in a maximal page capacity for binary modulation coding of 49 025 bits. In order to augment the storage capacity, first of all, the oversampling dimensions should be decreased as far as possible. Thereafter the block length of the employed 2d channel code has to be adjusted in a way that the amplitude error rate stays just below 10^{-3}. Assume that the page sparseness is set to 0.5. If 3 or 4 distinct phase levels are employed to generate code words with constant phase sparsity, the capacity can be increased up to 70 928 bits and 75 361 bits, respectively. The latter capacity can also be realized by use of three grayscale values and two phase levels.

5. Practical value of hybrid modulation coding

According to the above investigation capacity gains of up to 54% can be realized in the developed phase-encoded holographic storage system. In this estimation it is assumed that data pages possess a sparseness of 0.5. In addition, error propagation has been neglected. Due to the increased number of possible channel states, the effect of error propagation can be more critical than for use of binary modulation codes. For instance, assume a case where an ON channel has falsely been interpreted as an OFF channel. In this case, one of the real OFF channels of the block will be classified as an ON channel, since the frequency of occurrence of each amplitude state is fixed. In addition, the 'phase state' of the false ON channel will be determined and another phase state of the same block is falsely interpreted. As a consequence, a completely different code word is detected, which is associated with a different data word, potentially resulting in many user bit errors. With respect to error propagation a variable phase sparsity is more advantageous. Since the frequency of occurrence of each phase state is variable, at least any false classification of a phase state does not result in another 'swapped' phase state. Thereby, the impact of phase errors is weaker.

In order to realize hybrid modulation coding the employed system configuration should be able to enable independent control over the 2d amplitude and phase distribution of signal waves. Due to the lack of a suitable modulator, here, another approach is followed by splitting the data pages into subpages that contain only ON channels of equal phase. The required phase delay is adjusted in the reference arm. However, this method is only feasible up to moderate numbers of multiplexed pages. In a real system the method will give rise to additional constraints when tending to exhaust the dynamic range of a material. Only in a perfect storage material would splitting the pages not shorten the maximal number of multiplexed pages [8, 24].

In the introduction it has been emphasized that it is an important goal of any 2d channel encoding technique to decrease the expectable error rate, while enabling high data transfer rates. Though hybrid codes impressively enhance the page capacity, data transfer rates are downgraded when following the readout procedure presented in section 3. The procedure is based on performing two readout attempts per employed phase level. In this manner, the technique achieves a high reliability in resolving all phase states. On the other hand, in cases where also the transfer rate during readout is a critical figure of merit, the effective number of necessary readout steps can be obstructive. This drawback can partly be circumvented. Assume that hybrid pages that exhibit two different phase levels are to be recorded. Instead of using phase shifts of 0 and π, the phases are adjusted to 0 and $\pi/2$. In this case all information can be read out in two instead of three readout steps. In the first step the reference page is controlled to interfere destructively with the ON, 0 channels. In the second step it is subtracted from the ON, $\pi/2$ states. Subsequently, after normalization, the derived intensities of the block channels are subtracted. The ON channels of the different phases exhibit high values of opposite sign, while the OFF channels appear with some smaller values in between. The latter effect causes a higher AER in comparison to the above conducted readout method. In our experiments the AER worsens by about 1.5 orders of magnitude when using the $\pi/2$ readout procedure, while the PER remains in both cases much lower. Though the $\pi/2$ readout technique improves the data transfer rate, it does not allow us to reliably differentiate more than two phase states that are displayed by channels of the same amplitude. In addition, in order to facilitate one further gray level, a high contrast and a rather homogeneous intensity distribution are necessary. If the system allows us to reliably discriminate an additional gray level, it also potentiates additional phase encoding without further constraints. Another method to improve the transfer rate during readout could be based on the use of two cameras. In this case the reconstructed data page has to be split to illuminate each of the detectors, where it interferes with separately calibrated reference pages.

6. Conclusion

We introduced hybrid multinary modulation codes for page-oriented holographic storage systems. Construction of two code designs and the computation of the number of producible code words and related code rates has been discussed. In order to implement hybrid modulation codes a suitable input configuration is necessary to independently modulate the 2d phase and amplitude distribution of a signal wave. In phase-encoded storage systems and when recording only moderate numbers of holograms input of hybrid data pages can be accomplished by means of splitting the pages to be stored into subpages. During readout the best reliability is obtained by performing one readout step for classification of amplitude states and two readout steps to resolve each employed phase level. Based on this method experimental investigations suggest that up to four phase levels can be discriminated at error rates that outperform the corresponding amplitude error rates. However, the procedure downgrades the achievable data transfer rate. An adapted method allows us to partly circumvent this drawback, while tolerating slightly higher error rates. It suggests reasonable performance for up to three amplitude states and two phase states. Hence, in practice the latter scenario is probably most attractive, since it provides the best tradeoff between achievable capacity gain and expected degradation of the data transfer rate. Using this coding design and assuming a sparseness of 0.5 the capacity of data pages can be augmented up to ≈47 % (including ECC) in comparison to binary modulation codes. Utilizing only two amplitude levels and two phase states a capacity gain of ≈31 % (including ECC) can be realized at a sparseness of 0.5. Concisely, the introduced novel hybrid block code design offers significant capacity enhancements over contemporary purely amplitude-modulated data pages in holographic storage systems.

References

[1] Coufal H J, Psaltis D and Sincerbox G T (ed) 2000 *Holographic Data Storage* (New York: Springer)

[2] Burr G W, Ashley J, Coufal H, Grygier R K, Hoffnagle J A, Jefferson C M and Marcus B 1997 *Opt. Lett.* **22** 639–41

[3] Burr G W, Barking G, Coufal H, Hoffnagle J A, Jefferson C M and Neifeld M A 1998 *Opt. Lett.* **23** 1218–20

[4] Burr G W, Chou W-c, Neifeld M A, Coufal H, Hoffnagle J A and Jefferson C M 1998 *Appl. Opt.* **37** 5431–43

[5] King B M and Neifeld M A 2000 *Appl. Opt.* **39** 6681–8

[6] Marcus B 2000 in *Holographic Data Storage* ed H J Coufal, D Psaltis and G T Sincerbox (New York: Springer) pp 283–91

[7] Berger G, Müller K-O, Denz C, Földvári I and Péter Á 2003 *Proc. SPIE* **4988** 104–11

[8] King B M, Burr G W and Neifeld M A 2003 *Appl. Opt.* **42** 2546–59

[9] John R, Joseph J and Singh K 2005 *J. Opt. A: Pure Appl. Opt.* **7** 391–5

[10] Heanue J F, Gurkan K and Hesselink L 1996 *Appl. Opt.* **35** 2431–38

[11] King B M and Neifeld M A 1998 *Appl. Opt.* **37** 6275–98

[12] Miller C, Hunt B R, Marcellin M W and Neifeld M A 2000 *J. Opt. Soc. Am.* A **17** 265–75

[13] Chou W-c and Neifeld M A 2001 *J. Opt. Soc. Am.* A **18** 185–94

[14] Neifeld M A and Wu Y 2002 *Appl. Opt.* **41** 4812–24

[15] Moinian A, Fagoonee L and Honary B 2005 *Opt. Eng.* **44** 105201

[16] Burr G W, Ashley J, Marcus B, Jefferson C M, Hoffnagle J A and Coufal H 1998 *Proc. SPIE* **3468** 64–75

[17] Jang J-S and Shin D-H 2001 *Opt. Lett.* **26** 1797–9

[18] Reményi J, Várhegyi P, Domján L, Koppa P and Lõrincz E 2003 *Appl. Opt.* **42** 3428–34

[19] Denz C, Pauliat G, Roosen G and Tschudi T 1991 *Opt. Commun.* **85** 171–6

[20] Denz C, Pauliat G, Roosen G and Tschudi T 1992 *Appl. Opt.* **31** 5700–5

[21] Heanue J F, Bashaw M C and Hesselink L 1994 *Opt. Lett.* **19** 1079–81
[22] Denz C, Dellwig T, Lembcke J and Tschudi T 1996 *Opt. Lett.* **21** 278–80
[23] Hoffnagle J A and Jefferson C M 2000 *Holographic Data Storage* ed H J Coufal, D Psaltis and G T Sincerbox (New York: Springer) pp 91–100
[24] Mok F H, Burr G W and Psaltis D 1996 *Opt. Lett.* **21** 896–8