

# 数字全息数据存储系统中的格雷映射等重调制码的设计

裴先登<sup>1</sup>, 陈端荣<sup>1,2</sup>, 谢长生<sup>1</sup>

(11 华中科技大学计算机科学与技术学院, 湖北武汉 430074; 21 湖北大学数学与计算机学院, 湖北武汉 430062)

**摘 要:** 在数字全息数据存储系统的记录通道中, 使用二进制映射等重调制码有利于信号的检测. 本文作者提出了 6:8 格雷映射等重调制码的编码方案, 将每 6 比特数据视为一个格雷码字, 而相应的等重码设计为具有 2 单位距离特性. 当检测到的 8 比特码字中发生一个 2 码元交换错误时, 该方案具有一定的容错能力. 误码性能的仿真结果表明, 在系统的信噪比较低时 6:8 格雷映射等重调制码优于 6:8 二进制映射等重调制码.

**关键词:** 全息存储; 格雷码; 等重码; 调制码

**中图分类号:** TN911 **文献标识码:** A **文章编号:** 0372-2112 (2003) 07-1012-03

## Design of Gray2Mapped Constant Weight Modulation Code for Digital Holographic Data Storage System

PEI Xian2deng<sup>1</sup>, CHEN Duan2rong<sup>1,2</sup>, XIE Chang2sheng<sup>1</sup>

(1. Dpt of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074 China;  
21 Dpt of Maths and Computer Science, Hubei University, Wuhan, Hubei 430062, China)

**Abstract:** Using a binary2mapped constant weight modulation code (BCWMC) for the recording channel of a digital holograph2ic data storage system can facilitate the decision on the detected signal. The authors propose a new modulation coding scheme, 6:8 Gray2mapped constant weight modulation code (GCWMC), in which each digital data having length 6 is viewed as a Gray codeword and the constant weight code is designed with a characteristic of binary 2unit distance. Thus the 6:8 GCWMC scheme can possess the ability of its all the 8bit codewords to bear some error when a two2digit swapping event occurs in an estimated codeword. Simulation results show that the 6:8 GCWMC scheme can outperform the 6:8 BCWMC one in the bi2error rate (BER) performance at low signal2noise ratio (SNR) for the system.

**Key words:** holographic storage; Gray code; constant weight code; modulation code

## 1 引言

数字全息数据存储是一种基于页的光存储技术, 由于同时具有高速的数据传输速率 (1Gbits/s)、巨大的存储容量 ( $V/K^3$ ,  $V$  是全息存储晶体的体积,  $K$  是光源的波长) 和短暂的访问响应时间 ( $< 100\text{fs}$ ), 极有可能成为下一代数字数据存储技术<sup>[1]</sup>, 由此可作为人们使用高速度、大容量的存储器的一种选择. 尽管早在 60 年代初期, Van Heerden 就估计到体全息存储的巨大容量<sup>[2]</sup>, 但是直到 90 年代初期, 当激光器、光电互换器件、记录材料等关键技术逐步达到数字全息数据存储的基本要求后<sup>[3]</sup>, 在 90 年代初期, 美国、日本、英国等发达国家投入巨资竞相研究, 力争该技术早日商品化; 我国几所重点大学和一些科研机构则在 90 年代末期, 从光电器件、记录材料和记

录通道信号处理等全方位展开了协同攻关.

由于数字全息数据存储系统 (DHDSS) 是一个有噪声的数据传输通道, 因此, 为满足商品化的要求, 必须采用纠错码和调制码技术<sup>[4~6]</sup>. 在 DHDSS 中, 使用等重调制码能使待存储的数据最大限度地适应光通道的传输特性, 并且有利于从检测出的光电信号中较准确地恢复为原数字数据. 例如 6:8 二进制映射等重调制码 (binary2mapped constant weight modulation code, BCWMC) 是将每 6 比特数据信息 (本文称之为一个数据字) 映射为一个相应的 8 比特等重码字.

仿真结果和实验<sup>[6,7]</sup>表明, DHDSS 记录通道中的各种各样的噪声源, 尤其是在系统的信噪比较低时, 极易使所检测到的等重码字 (与原来写入的码字相比较) 发生一个数字 1 和一个数字 0 的 2 码元交换错误, 由此导致其数据字中的平均误

比特数约为 3. 为此, 我们先利用深度优先算法将 64 个 8 比特等重码字设计为 2 码距编码, 然后, 将 64 个具有二进制单位距离的格雷码字映射到这 64 个 8 比特等重码字中(此时, 一个 6 比特二进制数据流形式上视为一个 6 比特格雷码字). 因此, 任何一个 8 比特等重码字发生一个 2 码元交换错误时, 至少都有 2 次机会仅产生 1 比特误码. 我们称这种编码为格雷映射等重调制码(Gray2mapped constant weight modulation code, GCWMC).

## 2 格雷映射等重调制码的编码

格雷码<sup>[8]</sup>(Gray Code)是一种单位距离码, 它表示任意两个相邻十进制整数所对应的两个二进制码字中仅有一个码元不同(包括首尾两个码字). 由于具有单位距离特性, 格雷码能够避免在码字转换的过渡区内产生瞬时误码. 因此, 格雷码广泛用于角度及位移编码的机械和数字通信系统中. 若一个分组码的每个码字中的 1 的个数都相等, 称为等重码; 此时若数字 0 和数字 1 的个数相等, 则称之为平衡码. 我们借助集合来描述 6:8 GCWMC 的编码方法如下:

$S = \{0, 1, 2, \dots\} = \{s | s \text{ 为非负的十进制整数}\};$

$B = \{b_i | S | 0 \leq b_i \leq 63, (000000)_2 \leq (b_i)_2 \leq (111111)_2, 0 \leq i \leq 63, 0 \leq j \leq 5\};$

$G = \{g_i | B | 0 \leq g_i \leq 63, g_{i,j} \text{ 是第 } i \text{ 个 6 位二进制反射格雷码字的第 } j \text{ 位}, g_{i,j} = b_{i,j+1} \oplus b_{i,j}, 0 \leq j \leq 5, b_{i,6} = 0, \oplus \text{ 为异或运算}\};$

$X = \{x_u | S | 15 \leq x_u \leq 240, 0 \leq u \leq 69, x_u \text{ 对应的 8 位二进制数中含 4 个 } 1/10, 4 \text{ 个 } 0\};$

$Y = \{y_i | X | y_i \text{ 的选取按文献[9], 以便使用硬件实现 Viterbi 算法的快速译码. } y_{i,k} \text{ 表示第 } i \text{ 个 8 位二进制码字的第 } k \text{ 位}, 0 \leq i \leq 63, 0 \leq k \leq 7\}.$

本文定义的格雷映射方式 F 如下:

$F: y_i = f(g_i), 0 \leq i \leq 63, g_i \in G, y_i \in Y$ , 其中, F 满足下列两个条件(包括首尾的  $y_0$  与  $y_{63}$ , 即此处的  $/+0$  和  $-0$  号是模 64 上的运算符):

$$(1) \{y_{i,k_1} \oplus y_{i-1,k_1} = 1\} \cap \{y_{i,k_2} \oplus y_{i-1,k_2} = 1\}$$

$$\cap \{y_{i,k_1} \oplus y_{i,k_2} = 1\} \cap \bigcap_{t=3}^8 \{y_{i,k_t} \oplus y_{i-1,k_t} = 0\}$$

$$(2) \{y_{i,k_{11}} \oplus y_{i+1,k_{11}} = 1\} \cap \{y_{i,k_{12}} \oplus y_{i+1,k_{12}} = 1\}$$

$$\cap \{y_{i,k_{11}} \oplus y_{i,k_{12}} = 1\} \cap \bigcap_{t=13}^{18} \{y_{i,k_t} \oplus y_{i+1,k_t} = 0\}$$

表 1 6:8 格雷映射等重调制码编码示意图

	$b_{i,j}$	$g_{i,j}$	$y_{i,k}$	$y_i$
0	000000	000000	10100110	166
1	000001	000001	10100101	165
2	000010	000011	10010101	149
3	000011	000010	10011100	156
...	...	...	...	...
60	111100	100010	10110100	180
61	111101	100011	10110001	177
62	111110	100001	10101001	169
63	111111	100000	10101100	172

对于位于集合 B 和集合 G 中的同一个 6 位二进制数据而言, 映射方式 F 隐含着一种复合映射:  $B \rightarrow Y \rightarrow G$ . 现用表 1 示意 6:8 GCWMC 码表(限于篇幅此表未全部列出, 本文的实验中使用的 6:8 GCWMC 码表则是利用了上述的复合映射关系).

6:8 GCWMC 与 6:8 BCWMC 的信号检测(译码)方法一样: 排序法、全相关法和 Viterbi 算法. 利用文献[9]中的 Viterbi 算法可使用硬件实现对 6:8 等重调制码的快速译码, 判决一个码字仅需要 13 次加和 13 次比较操作.

## 3 格雷映射等重调制码的性能分析

我们在计算机的仿真过程中使用了 Monte Carlo 方法, 以便较为准确地比较用这两种等重调制码编码的数据在 DHSS 的记录通道中的误码性能. 仿真过程是随机地产生一个大小为  $720 \times 720$  的二值化数据页, 分别用 6:8 GCWMC 与 6:8 BCWMC 编码后, 经过一个 2 维符号间串扰(使用文献[10]中的一个  $5 \times 5$  的串扰矩阵)并伴有不同方差的加性高斯白噪声的记录通道传输, 再利用文献[9]中的 Viterbi 检测器对恢复的数据进行检测, 最后对估计的码字解调后分别统计出它们各自的误比特数(共计 10 页), 由此画出它们的 BER(b2error rate)性能曲线如图 1 所示.

由图 1 可见, 当信噪比 SNR(signal noise ratio)较小时, 6:8 GCWMC 优于 6:8 BCWMC(注: 由于纵轴使用的是对数坐标, 即使误比特数相差很大, 也不易察觉), 而且在  $BER = 10^{-3}$  时, 6:8 GCWMC 比 6:8 BCWMC 能获取约 0.7 dB 的编码增益. 在仿真实验中曾使用过 RS(15, 11, 2)纠错码, 与文献[7]的实验结果一样能确保将 BER 从  $10^{-3}$  降至  $10^{-12}$  以下. 但是随着 SNR 的增加, 这两种编码的效果几乎一样.

因此, 我们用一个码字发生 2 码元交换错误后具有格雷映射关系的概率(简称为格雷率, GRAY2mapped Occurrence, GO)来说明 6:8 GCWMC 的误码性能在某种程度上优于 6:8 BCWMC 的原因. 对于 6:8 GCWMC 而言, 它的每一个码字的格雷率均大于或等于  $2/16$ (图 2 中的点画线表示  $GO = 1/8$ ). 数字 16 表示一个 8 位等重码字中数字 0 与数字 1 能发生 2 码元交换的最大组合数, 而数字 2 是此 16 种组合方式中的 2 种, 此 2 种组合方式所对应的 2 个 8 位等重码字的数据字与原来的 8 位等重码字的数据字相比仅有 1 比特错误发生. 例如表 1 中的 6:8 GCWMC, 当原码字为 10100101(000001)误判决为 10100110(000000)或 10010101(000011)时, 译码后的误比特数仅是 1.

图 2 是本文码表的 6:8 GCWMC 的格雷率与其码字数的分布关系, 其实际格雷率的大小是  $1/7 \sim 5/14$ (由于在集合 X 中存在 6 个禁用码字, 因此 64 个码字发生 2 码元交换的组合数有 14, 15, 16 三种), 平均格雷率为 0.233984; 而 6:8 BCWMC 中的部分码字隐含着格雷映射关系, 其实际格雷率的大小是  $0 \sim 5/16$ , 它的 24 个码字的格雷率小于  $1/8$ , 其中有 4 个码字的格雷率为 0, 平均格雷率为 0.132096.

利用我们的码表, 计算出 6:8 GCWMC 与 6:8 BCWMC 中的一个 8 比特码字发生 1 次 2 码元交换错误后的平均误比特

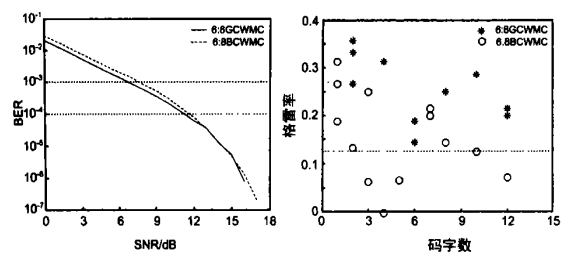


图 1 2 种 6:8 等重调制码的 BER 性能曲线

图 2 2 种 6:8 等重调制码的格雷率分布

数分别是 2.638132 和 2.764220 比特.并且用表 2 列出了这两种编码的各自 64 个码字发生有效的 2 码元交换时(由于 6 个禁用码字的存在,2 码元交换的有效总数为 948 次,无效总数是 76 次)所能导致的 1~6 比特误码的数目分布(表中的 6:8 G 表示 6:8 GCWMC,6:8 B 表示 6:8 BCWMC):

表 2 2 种 6:8 等重调制码的误比特数目分布						
	1bit	2bit	3bit	4bit	5bit	6bit
6:8 G	222	244	220	186	70	6
6:8 B	126	278	312	164	64	4

由表 2 可见,6:8 GCWMC 的 1 比特误码数几乎是 6:8 BCWMC 的一倍,由此也可以解释系统的信噪比较低时 6:8 GCWMC 具有/ 分布式容错0能力的原因,但这种编码增益是以 4-6 比特的误码数增加为代价换取的.当系统的 SNR 增加后,2 码元交换错误的发生率下降,所以此时 6:8 BCWMC 的 BER 性能几乎与 6:8 GCWMC 的一样.

4 结语

按照本文格雷映射方式 F 中的两个条件,首先将 64 个 8 比特等重码字构造为 2 单位距离码,然后将具有单位距离特性的 64 个 6 比特格雷码字(即为 6 比特数据字)映射到该 64 个 8 比特等重码字中,并利用计算机确定出这两者之间的最优映射关系(即 64 个 8 比特码字中那一个作为首码字与 6 比特格雷码的首码字 000000 相对应).当记录通道中的各种噪声源使得 6:8 GCWMC 中的任何一个 8 比特码字发生 2 码元交换错误后,所恢复的 6 比特数据字与原来写入的 6 比特数据字相比较,至少有 2 次机会使其仅发生 1 比特错误.误码性能的仿真结果表明在系统的信噪比 SNR 较低时 6:8 GCWMC 优于 6:8 BCWMC,而在信噪比 SNR 较高时它们的误码性能几乎一样,此外它们的译码方法相同.

参考文献:

[ 1 ] John F Heanue, Matthew C Bashaw, Lambertus Hesselink. Volume holographic storage and retrieval of digital data [ J ]. Science, 1994, 265: 749- 752.

[ 2 ] P J Van Heerden. Theory of optical information storage in solid [ J ]. Applied Optics, 1963, 2( 4 ): 393- 400.

[ 3 ] 谢长生, 陈端荣, 裴先登. 体全息数字存储系统中的几项关键技术[ J ]. 计算机研究与发展, 2002, 39( 9 ): 1113- 1119.

[ 4 ] Geoffrey W Burr, Jonathan Ashley, Hans Coufal, et al. Optimizing the holographic digital data storage channel [ A ]. Proceedings of the SPIE, Advanced Optical Memories and Interfaces to Computer Storage [ C ]. San Diego: CA, 1998.

[ 5 ] John F Heanue, Matthew C Bashaw, Lambertus Hesselink. Channel codes for digital holographic data storage [ J ]. Journal of the optical society of America A, 1995, 12 ( 11 ): 2432- 2439.

[ 6 ] Geoffrey W Burr, Jonathan Ashley, Hans Coufal, et al. Modulation coding for pixelmatched holographic data storage [ J ]. Optics Letters, 1997, 22( 9 ): 639- 641.

[ 7 ] Geoffrey W Burr, Wu2chun Chou, Mark A Neifeld, et al. Experimental evaluation of user capacity in holographic data2storage systems [ J ]. Applied Optics, 1998, 37( 23 ): 54325443.

[ 8 ] Gray F. Pulse code communication [ P ]. U. S. Patent 2632058, 1953.

[ 9 ] Jonathan James Ashley, Brian Harry Marcus. Encoding and detection of balanced codes [ P ]. U. S. Patent 6016330, 2000.

[ 10 ] Brian M king, Mark A Neifeld. Parallel detection algorithm for page2oriented optical memories [ J ]. Applied Optics, 1998, 37( 26 ): 6275- 6298.

作者简介:



**裴先登** 男, 1934 年出生, 教授, 博士导师, 从事硬盘存储、近场光存储和数字全息数据存储系统记录通道的研究.



**陈端荣** 男, 1960 年出生, 讲师, 博士生, 从事全息数据存储记录通道中的编解码技术的研究.



**谢长生** 男, 1957 年出生, 教授, 博士导师, 从事网络存储、计算机系统结构和数字全息数据存储系统记录通道的研究.