科學 专著丛书

组合编码原理及应用

COMBINATORIAL CODING THEORY AND ITS APPLICATIONS

靳 蕃 陈 志 编著

上海科学技术出版社

科學专著丛书

组合编码原理及应用

靳蕃　陈志　编著

上海科学技术出版社

科学专著丛书
组合编码原理及应用
靳 蕃 陈 志 编著
上海科学技术出版社出版、发行
（上海瑞金二路450号）
新华书店上海发行所经销　江苏如东印刷厂印刷
开本787×1092　小1/16　印张23　插页4　字数354,000
1995年6月第1版　1995年6月第1次印刷
印数 1—1,200
ISBN 7-5323-3657-3/TP·48
定价：36.00元

内 容 提 要

本书内容主要包括作者在完成国家自然科学基金项目——组合编码领域中的研究成果以及近年来国内外在组合码方面的最新成果。它将组合设计和编码技术这两个学科领域有机地结合起来，构成了数学基础理论和信息应用技术紧密相连的新体系。书中系统地阐明了组合编码的基本原理和基本方法，详细介绍了国内外在组合编码技术方面的最新成就。

全书共11章，内容翔实，资料充分。适合于国内外计算机、通信技术、信息工程、自控遥控、数字信息处理与测量等专业的师生，以及从事上述专业和试验设计、应用数学等方面的广大工程技术人员。

《科学专著丛书》序

如果说科学研究论文是创造性科学工作的发表性记录，那末科学技术学术专著则是创造性科学工作的总结性记录。前者注重的是优先权，后者注重的是系统化。

在大量科学研究的基础上，对一个专题或一个领域的研究成果，作系统的整理总结，著书立说，乃是科学研究工作不可少的一个组成部分。著书立说，既是丰富人类知识宝库的需要，也是探索未知领域、开拓人类知识新疆界的需要。特别是在科学各门类的那些基本问题上，一部优秀的学术专著常常成为本学科或相关学科取得突破性进展的基石。所以，科学技术学术专著的著述和出版是一项十分重要的工作。

近20年来，中国的科学事业有了迅速的发展，涌现了许多优秀的科学研究成果，为出版学术专著提供了坚实的基础。值此20世纪90年代，在出版学术专著方面，中国的科学界和出版界都在抓紧为本世纪再加些积累，为迎接新世纪多作些开拓。我高兴地看到，作为这种努力的一个部分，《科学》杂志的出版者——上海科学技术出版社推出了这套《科学专著丛书》。

上海科学技术出版社是科学技术界熟悉和信赖的一家出版社，历来注重科学技术学术专著的出版。《科学》杂志的编者组织编辑学术系列丛书，也不是第一次。在本世纪三四十年代，就曾推出过《科学丛书》，其中不乏佳作，对当时的学术研究起了很好的作用。

《科学》在中国是一份历史最长的综合性科学刊物，80年来与科学技术界建立了广泛的密切联系。现在推出的这套《科学专著丛书》正是这种

联系的产物。我相信，加强这种联系，著者与编者、出版者，科技界与出版界共同努力，精心选题，精心编辑，精心出版，一定能使这套专著丛书反映出中国科学技术研究的最新水平，为本世纪多留下几本中国学者的优秀专著，为迈向新世纪多铺下几块引路的基石！

周光召

（《科学》杂志编委会主编）

1994 年 8 月

前　言

为了保证数字通信的可靠性，建立在有限域代数基础上的差错控制编码技术，经历了40多年的发展，已经日臻完善。

随着以计算机为核心的数字信息处理技术迅猛发展，对差错控制编码技术在译码速度、功能的多样性、使用的灵活性以及编码译码器的简单化模块化等方面，都提出了传统编码方法所不曾具备的新要求。于是，以组合设计为基础的组合编码方法，近些年来开始受到国内外编码界的关注与重视。

本书以11章的篇幅，将组合设计和编码技术这两个学科领域有机地结合起来，构成数学基础理论和信息应用技术紧密相连的新体系。书中除较为系统地阐明组合编码的基本原理和方法外，还扼要地介绍了近年来国内外在组合编码技术方面的新成就，其中也包括了本书的作者及所担负的国家自然科学基金项目成员在国内外学术刊物和学术会议上发表的研究成果。

在绪论中简要地阐述差错控制编码和组合学的基本概念。第2章的数论基础和第3章的区组设计是以后各章编码方法的理论依据。从第4章至第10章，分别介绍了不同特色的组合编码方法和所构造出来的组合码。在第11章中介绍了若干实际应用示例。

在本书的撰写过程中，西南交通大学曹建猷教授和西安电子科技大学王新梅教授、美国Lehigh大学曾开明教授(Prof.K.M.Tzeng)提供了许多宝贵的建议和意见，范平志副教授、郝红鸣老师参加了部分研究工作。常季坤女士、蔡琳女士为整理书稿付出了辛勤的劳动。对此，作者一并表示诚挚的谢意。

组合编码是一门方兴未艾的新学科分支，它在差错控制编码、数据加密编码以及信源编码等领域中都显现出广阔的发展应用前景。在这个亟待进一步开发的科学宝库面前，作者不过起着探路尖兵的作用而已。

殷切期望读者对书中疏漏不妥之处提出批评指正，并盼更多有兴趣者共同一道来探究和攀登本学科领域中的新高峰。

作 者　1994年春

目　录

《科学专著丛书》序
前言
第 1 章　绪论 …… 1
§1.1　差错控制编码历史的回顾 …… 1
§1.2　差错控制编码的基本原理 …… 3
§1.3　基本组合法则 …… 7
§1.4　组合编码技术的发展与前景 …… 10
第 2 章　数论基础 …… 14
§2.1　数论的历史与地位 …… 14
§2.2　素数 …… 15
§2.3　约数与倍数 …… 20
§2.4　同余方法 …… 27
§2.5　威尔逊、费马和欧拉定理 …… 38
§2.6　本原根与指数 …… 46
§2.7　平方剩余的基本概念 …… 51
第 3 章　区组设计 …… 59
§3.1　区组设计的基本概念 …… 59
§3.2　区组设计的特性 …… 61
§3.3　三连系 …… 74
§3.4　构造区组设计的方法 …… 77
§3.5　循环差集 …… 83
§3.6　有限域生成 BIBD …… 98
§3.7　小区组设计组合成大区组设计 …… 101
§3.8　t 设计 …… 104
§3.9　部分均衡不完全区组设计 PBIBD …… 105
第 4 章　组合编码方法 …… 110
§4.1　组合编码的基本特点 …… 110

§4.2 关联矩阵 $s(u,v)$ …… 111
§4.3 线性组合码 …… 119
§4.4 非线性组合码 …… 136
§4.5 组合编码译码电路示例 …… 148
§4.6 神经网络与组合编码的关系 …… 150
第5章 幂次剩余码…… 154
§5.1 平方剩余与高次剩余 …… 154
§5.2 平方剩余码 …… 156
§5.3 双环循环码与准循环码 …… 160
§5.4 二进制平方剩余码 …… 165
§5.5 对称码…… …… 167
§5.6 幂次剩余码…… 169
第6章 有限几何码…… 175
§6.1 大数逻辑译码 …… 175
§6.2 射影几何及有关的组合设计 …… 184
§6.3 有限射影几何码 …… 190
§6.4 差集码和极长码 …… 195
§6.5 欧氏几何及有关的组合设计 …… 200
§6.6 有限欧氏几何码 …… 205
§6.7 Reed-Muller 码 …… 211
第7章 复数旋转码…… 215
§7.1 基本定义与定理 …… 215
§7.2 复数旋转码的编码方法 …… 217
§7.3 复数旋转码的译码方法 …… 223
§7.4 组合特性分析 …… 230
§7.5 增加码率的途径 …… 237
§7.6 综合性能评价 …… 243
第8章 自正交码…… 246
§8.1 自正交码的基本概念 …… 246
§8.2 自正交码的组合构造 …… 249
§8.3 自正交准循环码 …… 256
§8.4 差集与自正交准循环码 …… 260

§8.5 自正交卷积码 …… 265
§8.6 光正交码 …… 276
§8.7 大数逻辑可译码的组合构造 …… 280
第9章 不等保护能力码 …… 285
§9.1 不等保护能力码的基本概念 …… 285
§9.2 循环不等保护能力码 …… 289
§9.3 自正交不等保护能力码 …… 291
§9.4 自正交准循环UEP码 …… 295
§9.5 基于复数旋转码的UEP码构造 …… 297
§9.6 大数逻辑可译UEP码 …… 299
§9.7 不等保护能力自正交卷积码 …… 301
第10章 组合编码算法 …… 305
§10.1 素数表的生成 …… 305
§10.2 数据结构及其基本运算 …… 308
§10.3 组合及其等价类的生成 …… 310
§10.4 Gray码与码的重量分布计算 …… 312
§10.5 对偶码的重量分布 …… 318
第11章 在计算机和通信系统中的应用 …… 321
§11.1 计算机系统差错控制的特点 …… 321
§11.2 简单奇偶校验码 …… 322
§11.3 纠一位错的DBBD码 …… 325
§11.4 纠一检二(SEC-DED)码 …… 326
§11.5 拉丁方阵码 …… 328
§11.6 Horiguchi-Morita码 …… 333
§11.7 Rotational码 …… 336
§11.8 在电报系统中的应用 …… 338
§11.9 自适应差错控制系统 …… 342
附录 若干自正交码的码重分布 …… 345
参考文献 …… 347

第 1 章

绪　论

§1.1　差错控制编码历史的回顾

20 世纪的后 50 年，是信息科学技术突飞猛进的大发展年代。以数字式集成电路为标志的微电子技术的诞生、成长并日臻完善，促进了信息传输领域中的现代数字通信技术和信息处理领域中的数字电子计算机技术迅猛发展。正如各类商品都转化为同一的载体——货币，在市场上到处流通和相互交换那样，各种图象、语音、数据和文字信息都以二进序列的统一编码形式，在传输网上流通或者在计算机中进行处理存储。

编码理论(coding theory)包括三方面的内容：

(1) 以保证数字信息传输和处理的可靠性为目的的*差错控制编码*(error-control coding)，又叫*信道编码*(channel coding)；

(2) 以提高数字信息传输、存储处理的有效性为宗旨的*信源编码*(source coding)；

(3) 以增加数字信息传输、存储的安全性为目标的*数据加密编码*(data encryption)。

差错控制编码技术类别繁多，应用面广，在上述三类编码中占有较大的比重。因此，通常使用的编码这一术语，常常是指差错控制编码译码技术。本书所涉及的组合编码原理及应用，也是属于差错控制编码的范畴。

差错控制编码技术是适应数字通信抗噪声干扰的需要而诞生和发展起来的，它始于 1948 年。著名的信息论创始人 Shannon 在当时发表为一篇题为《通信的数学原理》[1]，提出了受干扰信道编码定理。该定理的主要内容如下：

每个受干扰的信道具有确定的信道容量 C。例如，当信道中存在高斯白噪声时，在信道带宽 W，单位频带信号功率 S，单位频带噪声功率 N

下，信道容量可以表示为式(1.1)。

$$C=W\log_2\left(1+\frac{S}{N}\right)b/s \tag{1.1}$$

对于任何小于信道容量 C 的信息传输速率 R，存在一个码长为 n，码率为 R 的分组码，若用最大似然译码，则其译码错误概率由式(1.2)所示。

$$P_e\leqslant Ae^{-nE_b(R)} \tag{1.2}$$

对于码率为 R，约束长度为 n_c 的卷积码，则有下式的结果。

$$P_e\leqslant Be^{-n_cE_c(R)} \tag{1.3}$$

式中 A、B 为常数，$E_b(R)$和 $E_c(R)$称为误差指数。

Shannon 在这个定理中告诉我们，只要传信率不超过信道容量 C，总可以用差错控制编码方法，使信道输出的错误减至任意小。但是，这个定理并没有告诉人们如何去进行满足要求的编码。

Hamming 和 Golay 在 50 年前分别发表了“检错和纠错码”和“评论数字编码”论文，提出了从充分利用监督元的观点来看是最佳的纠正二进制序列中一位错和三位错的完全码 (perfect codes)，开创了与实际应用相结合的编码理论方法的研究。

信息理论的发展和编码方法的成长始终是相互依赖、相互促进的。Hamming 曾经说过：“从逻辑上来说，编码理论导致信息理论，信息理论则为适当的信息编码提供了所能达到的限度”。

从 50 年代以来，编码理论和应用长期是围绕数字通信业务的特点和需要而发展的，也就是以伽罗华域($GF(2^m)$)上的代数编码方法为主体，研究适合于串行通信信道中使用的码率尽可能高、检错纠错能力尽可能强的码型。从结构上来看，可以划分为分组码(block codes)和卷积码(convolution codes) 两大类。在这方面有着内容极为丰富的研究成果，文献[6,7,25,37]是国内外具有代表性的佳作。

随着计算机技术的迅速兴起和发展，存储系统迅速可靠存取数据的问题提到日程上来。以编码译码速度为主要指标的高速并行组合编码理论和方法开始受到人们的重视。IBM 公司的 Hsiao, Patel 等，结合各种类型计算机内存和磁带存储器的需要，提出了多种适于并行高速译码的编码方法。

为了适应计算机运算部分差错控制的需要，用于纠正二进制算术运

算中差错的算术码(arithmetic codes)，也开始受到重视[25 36,29]。

此外，随着信息传输技术和信息处理技术的发展，差错控制编码和调制技术、数据加密技术开始结合起来[35,37]，根据信息位不同而采用不等保护的编码方法也开始引起人们的注意[51,52]。在这些编码方法中，组合设计都起着重要的作用。

可以预计，适应数字通信特点的传统代数编码，和具有高速并行处理特点的组合编码并不互相替代或互相排斥，而是相互补充并各得其所。

§1.2 差错控制编码的基本原理

不论是建立在有限域上的代数编码方法，还是建立在组合设计基础上的组合编码方法，都有一些基本的定义和定理需要遵守，都有一些恰当的表达方法可以使用。

1.2.1 差错控制信道模型

一般来说，数字信息传输系统和计算机信息存储系统，均可表达为图1.1所示的信道模型。

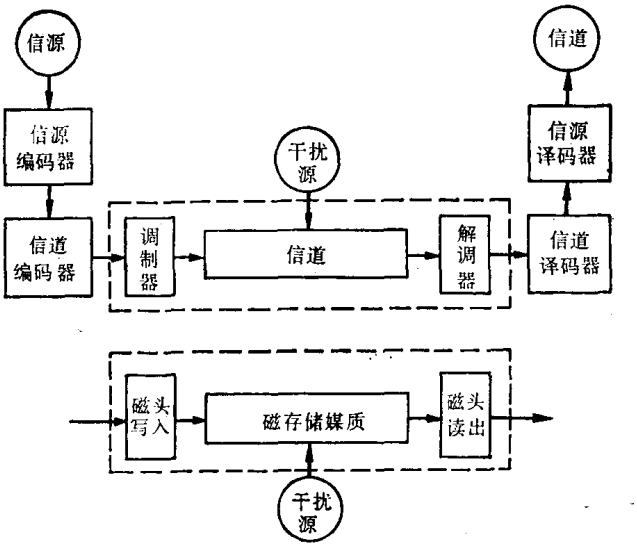

图1.1 数字信息传输系统和数字信息存储系统

为了使信息具有抗噪声干扰能力，可以将信源发出的信息通过信源编码器编成二进数字[0,1]序列，经过信道编码器进行差错控制编码后送入信道。

信道是传送、携带信息信号的媒介。根据使用场合的不同，它们可以是明线、电缆、无线通道、微波通道、卫星通道和光导纤维通道等，也可以是包括磁头写入和读出的磁存储介质。在需要用模拟载波信号进行传送的信道两端，设置有调制器和解调器。

实际信道中存在着各种干扰(有时又叫噪声)。天电干扰、电机或开关电器设备的火花、电气化铁路的畸变性牵引负荷、电子器件的热噪声、来自相邻线路的串音等，是通信信道中常见到的干扰。计算机系统磁存储介质的固有缺陷、表面磨损污染，以及 α 粒子对半导体存储单元的作用等，则是存储媒质中引起差错的干扰来源。

在接收端的解调器将信道中传来的信号波形还原成调制器以前的二进数字序列。由于信道中干扰的影响，还原的数字序列往往和原来输入调制器的数字序列有差别，这就是误码现象。此时信道译码器利用信道编码所提供的冗余度，按一定规则检查或纠正接收的数字序列中的错误，然后经过信源译码器转换为适合收信者接收的消息形式。

差错控制着眼于信道编码器和信道译码器之间（如图 1.1 中虚线方框所示)的噪声干扰误码及其防止问题，这部分包括调制器、信道和解调器，被称为*编码信道*。

1.2.2 码距和码重

设 $\mathbf{X}\in\{0,1\}^n$，$\mathbf{Y}\in\{0,1\}^n$ 为二进制码矢量，则有如下定义。

定义 1.1 $\mathbf{X}$ 和 $\mathbf{Y}$ 的 Hamming 码距定义为式(1.4)。

$$\mathrm{dist}(\mathbf{X},\mathbf{Y}):=|\{i\,|\,1\leqslant i\leqslant n, x_i\neq y_i\}| \tag{1.4}$$

定义 1.2 码矢量(即码字)$\mathbf{X}$ 的重量记为式(1.5)。

$$wt(\mathbf{X}):=\mathrm{dist}(\mathbf{X},\mathbf{0}) \tag{1.5}$$

式中 $\mathbf{0}$ 表示 n 维零矢量。

定义 1.3 一组码字 $\mathbf{Z}=\{\mathbf{Z}_1,\mathbf{Z}_2,\cdots,\mathbf{Z}_N\}$ 的最小码距 d(或 $d\min$)定义为式(1.6)。

$$d=\min\{\mathrm{dist}(\mathbf{Z}_i,\mathbf{Z}_j)\,|\,\mathbf{Z}_i\in\mathbf{Z},\mathbf{Z}_j\in\mathbf{Z},i\neq j\} \tag{1.6}$$

在对称二进信道条件下，由于干扰使符号“1”变成符号“0”的概率，和

使符号"0"变成符号"1"的概率相同，用*误码率* p 来表示。通常 p 远小于1，所以出现多位错的概率比出现较少位错的概率小。例如，在码长为 n 时，两位出错概率远小于一位出错概率，即

$$p^2(1-p)^{n-2} \ll p(1-p)^{n-1}$$

于是，人们有理由按照向邻近码字靠拢的*最大似然译码* (maximum likelihood decoding)原则进行纠错译码。由此可得出如下定理。

定理 1.1 纠错检错能力与码的最小码距 d 间的关系为：

(1) 要检出码字中任意 e 个码元错误，最小码距 d 必须满足式(1.7)

$$d \geqslant e+1 \tag{1.7}$$

(2) 要纠正码字中任意 t 个码元错误，最小码距 d 必须满足式(1.8)

$$d \geqslant 2t+1 \tag{1.8}$$

(3) 要纠正码字中任意 t 个码元错误，并同时发现 e 个错误($e \geqslant t$)，则最小码距 d 必须满足式(1.9)。

$$d \geqslant e+t+1 \tag{1.9}$$

由此可知，纠正错误总比检出同样数目错误难一些，因而纠错编码译码的码结构和设备要比相应的检错码复杂一些。

定义 1.4 码长 n、信息元数 k 和最小码距 d 的线性码用$[n,k,d]$表示。在信息元数 k 和监督元数 r 下，编码效率（简称*码率*）η 为有效信息码元数 k 和总码元数 n 之比，即式(1.10)。

$$\eta = \frac{k}{n} = \frac{k}{k+r} \tag{1.10}$$

对于不满足叠加特性的非线性码，用(n,M,d)表示，其中 M 为码字的数目。

1.2.3 监督矩阵与生成矩阵

首先，我们以熟知的$[n,k,d]=[7,4,3]$Hamming 码的构成为例，来说明监督矩阵 H 和生成矩阵 G 的意义。

设 X_1,X_2,X_3,X_4 为信息元，它们按一致监督方程组为式(1.11)。

$$\begin{cases} X_1+X_2+X_4=X_5 \\ X_1+X_3+X_4=X_6 \\ X_2+X_3+X_4=X_7 \end{cases} \tag{1.11}$$

产生监督元 X_5, X_6, X_7, 式中加号均表示模 2 相加，将式(1.11)写成矩阵的形式如下：

$$\begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \\ X_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

或记为式(1.12)。

$$H\mathbf{X}^T = \mathbf{0}^T \tag{1.12}$$

式中H称为监督矩阵 (check matrix)，$\mathbf{X}^T$为码矢量 $\mathbf{X} = [X_1X_2X_3X_4X_5X_6X_7]$的转置。

一般来说，对码长 n、信息元数 k 和监督元数 r 的码，监督矩阵 H 可以表示为式(1.13)。

$$H = [A \mid I_r] = [A \mid I_{n-k}] \tag{1.13}$$

式中 A 为 $r \times k$ 阶矩阵，I_r 为 $r = n - k$ 阶单位方阵。

当发送码序列 $\mathbf{X}$，在差错序列 $\mathbf{E}$ 影响下形成接收码序列 $\mathbf{Y}$ 时，可以用 H 来检验接收码序列 Y 是否有错，即

$$H\mathbf{Y}^T = H(\mathbf{X} + \mathbf{E})^T = H\mathbf{X}^T + H\mathbf{E}^T = H\mathbf{E}^T$$

记为 $S^T = H\mathbf{E}^T$，或记为式(1.14)。

$$S = \mathbf{E}H^T \tag{1.14}$$

S 被称为校验子或伴随式(syndrome)，如它不为零，表明接收码序列中有错。

如果码矢量 $\mathbf{X} = (X_1X_2\cdots X_kX_{k+1}\cdots X_n)$ 中的前面 k 位表示消息矢量 $\mathbf{U} = (U_1U_2\cdots U_k)$时，由式(1.12)可以写出式(1.15)。

$$X = \mathbf{U}G \tag{1.15}$$

式中如式(1.16)所示的，称为生成矩阵(generator matrix)。

$$G = [I_k \mid -A^T] \tag{1.16}$$

生成矩阵 G 与监督矩阵 H 间满足

$$HG^T = O^T.$$