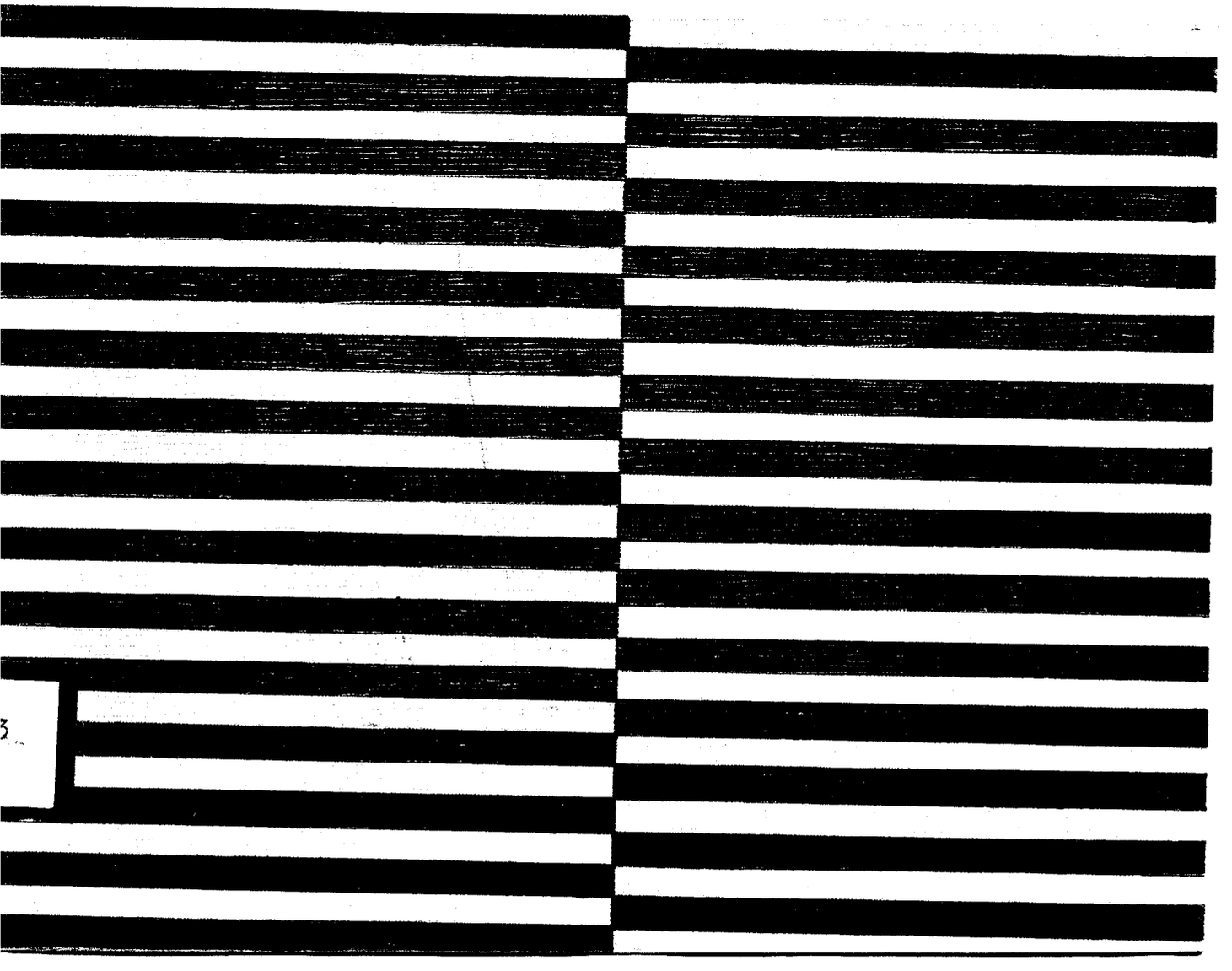
差错控制码的理论与实践

● 〔美〕 R·E·Blahut 著

● 徐秉铮 欧阳景正 冯贵良 译

● 徐秉铮 校

● 华南理工大学出版社



内 容 简 介

本书论述差错控制码的理论与实践，内容包括：代数导论，线性分组码，伽罗华域的算术，循环码，实现循环码的电路，BCH 码，基于谱技术的码，基于谱技术的算法，多维谱技术，快速算法，卷积码，用择多逻辑译码的码和算法，差错控制码的构成和性能，以及噪声信道的有效信号传输。在内容上兼顾到数学的严谨性和工程实用性，而且包括作者多年来应用谱技术研究差错控制码的成果。作者曾以本书在康乃尔大学研究生院，IBM 公司，以及华南工学院（现华南理工大学）进行过教学实践。

本书可供大专院校有关专业的师生以及研究生作为教材，也可供通信与电子系统领域的科技人员参考。

责任编辑　杨昭茂

差错控制码的理论与实践

〔美〕R. E. Blahut　著

徐秉铮　欧阳景正　冯贵良　译

徐秉铮　校

*

华南理工大学出版社出版发行

（广州　五山）

广东省新华书店经销

华南理工大学出版社印刷厂印刷

*

开本 787×1092 1/16　印张26.25　字数588千

1988年12月第1版　1988年12月第1次印刷

印　数 1—1000册

ISBN 7-5623-0035-6/TN·2

定价：5.15元

译 者 的 话

差错控制码的发展大致可以分为三个阶段：第一阶段自1949年至60年代初，这是差错控制码提出、发展，到奠定线性分组码的理论基础的阶段，代表性著作有Peterson的《纠错码》一书。第二阶段自60年代初至60年代末，这是差错控制码发展最活跃的时期，代表性著作有Berlekamp的《代数编码理论》一书。第三阶段自70年代至今，代数编码理论已臻成熟，差错控制码的实际应用显得愈来愈重要，因而本书特别强调编译码技术的实用方面。

本书作者Blahut教授是IBM公司的高级研究员（fellow）兼康乃尔大学的客座教授，长期以来从事差错控制译码器的研制工作，他对发展变换技术在差错控制码中的应用起了重要作用。1981年Blahut教授来华南工学院（现华南理工大学）作短期讲学，采用本书原稿作为教材，参加听讲的有来自全国各地的科技工作者约50人。

1982年IEEE在法国举行的国际信息论会议的Shannon讲座，由I.S.Reed教授主讲，题目是："在编码和有关课题中变换的应用"。Reed在报告中特别提到本书。1985年7月份IEEE信息论汇刊中发表了J.L.Massey教授的书评，认为本书是编码方面杰出的带创造性的一本书。

徐秉铮

1986年9月

目　录

序言 ……（1）

第一章　导论 ……（3）

1.1　离散的通信信道 ……（3）

1.2　差错控制编译码的历史 ……（4）

1.3　应用 ……（5）

1.4　基本概念 ……（6）

1.5　基本码 ……（10）

习题 ……（12）

第二章　代数导论 ……（14）

2.1　二元域与十六元域 ……（14）

2.2　群 ……（16）

2.3　环 ……（20）

2.4　域 ……（21）

2.5　向量空间 ……（23）

2.6　线性代数 ……（27）

习题 ……（33）

附注 ……（35）

第三章　线性分组码 ……（36）

3.1　线性分组码的结构 ……（36）

3.2　线性分组码的矩阵描述 ……（37）

3.3　标准阵列 ……（39）

3.4　汉明码 ……（42）

3.5　完备码和准完备码 ……（44）

3.6　对线性码的简单修改 ……（44）

3.7　Reed－Muller 码 ……（45）

习题 ……（49）

附注 ……（51）

第四章　伽罗华域的算术 ……（52）

4.1　整数环 ……（52）

4.2　基于整数环的有限域 ……（54）

4.3　多项式环 ……（55）

4.4　基于多项式环的有限域 ……（60）

4.5　本原元 ……（63）

4.6　有限域的结构 ……（65）

习题 ……（70）

附注 ……（71）

第五章　循环码 ……（72）

5.1　从扩展域来观察一个码 ……（72）

5.2 循环码的多项式描述……（74）
5.3 极小多项式与共轭……（78）
5.4 循环码的矩阵描述……（83）
5.5 作为循环码的汉明码……（85）
5.6 纠两个错的循环码……（87）
5.7 纠突发差错的循环码……（88）
5.8 二进 Golay 码……（92）
5.9 二次剩余码……（97）
习题 ……（99）
附注 ……（100）
第六章 实现循环码的电路 ……（101）
6.1 有限域算术的逻辑电路……（101）
6.2 数字滤波器……（103）
6.3 用移位寄存器实现的编码器和译码器……（107）
6.4 Meggitt 译码器 ……（109）
6.5 捕错译码……（114）
6.6 缩短循环码……（119）
6.7 Golay 码的 Meggitt 译码器 ……（122）
习题 ……（122）
附注 ……（124）
第七章 BCH码 ……（125）
7.1 码的定义……（125）
7.2 Peterson－Gorenstein－Zierler 译码器……（129）
7.3 Reed－Solomon 码……（135）
7.4 自回归滤波器的综合……（136）
7.5 BCH 码的快速译码 ……（142）
7.6 二元 BCH 码的译码……（149）
7.7 欧几里德算法译码……（151）
7.8 嵌套码……（155）
7.9 Justesen 码 ……（157）
习题 ……（160）
附注 ……（162）
第八章 基于谱技术的码 ……（163）
8.1 有限域中的傅里叶变换……（163）
8.2 共轭约束与幂等……（166）
8.3 循环码的谱描述……（169）
8.4 扩展的 Reed－Solomon 码……（174）
8.5 扩展 BCH 码 ……（176）
8.6 交替码……（180）
8.7 交替码的性能……（183）
8.8 Goppa 码 ……（185）

8.9 Preparata 码 …… (193)
习题 …… (196)
附注 …… (197)
第九章 基于谱技术的算法 …… (198)
9.1 谱技术的译码 …… (198)
9.2 纠正删除和差错 …… (205)
9.3 扩展 Reed－Solomon 码的译码 …… (208)
9.4 扩展 BCH 码的译码 …… (211)
9.5 在时域中的译码 …… (212)
9.6 BCH 界以外的译码 …… (215)
9.7 交替码的译码 …… (218)
9.8 有限域变换的计算 …… (221)
习题 …… (225)
附注 …… (226)
第十章 多维谱技术 …… (227)
10.1 乘积码 …… (227)
10.2 中国剩余定理 …… (229)
10.3 乘积码的译码 …… (232)
10.4 多维谱 …… (237)
10.5 快速 BCH 码 …… (240)
10.6 多维码的译码 …… (242)
10.7 小域中的长码 …… (244)
习题 …… (247)
附注 …… (248)
第十一章 快速算法 …… (249)
11.1 线性卷积和循环卷积 …… (249)
11.2 快速卷积算法 …… (251)
11.3 快速傅里叶变换 …… (257)
11.4 Agarwal－Cooley 卷积算法 …… (262)
11.5 Winograd 快速傅里叶变换 …… (264)
11.6 加速 Berlekamp－Massey 算法 …… (268)
11.7 递归 Berlekamp－Massey 算法 …… (273)
11.8 BCH 码的加速译码 …… (276)
11.9 代用域上的卷积 …… (278)
习题 …… (280)
附注 …… (280)
第十二章 卷积码 …… (281)
12.1 树码和格码 …… (281)
12.2 卷积码的多项式描述 …… (285)
12.3 纠错和距离概念 …… (290)
12.4 卷积码的矩阵描述 …… (292)

12.5 某些简单的卷积码 ……（294）
12.6 校正子译码算法 ……（297）
12.7 纠突发差错卷积码 ……（302）
12.8 Viterbi 译码算法 ……（306）
12.9 网格搜索算法 ……（310）
习题 ……（315）
附注 ……（316）
第十三章 用择多逻辑译码的码和算法 ……（317）
13.1 择多逻辑译码 ……（317）
13.2 择多译码电路 ……（320）
13.3 循环码的仿射置换 ……（323）
13.4 基于置换的循环码 ……（327）
13.5 择多译码的卷积码 ……（330）
13.6 广义 Reed－Muller 码 ……（332）
13.7 欧几里德几何码 ……（337）
13.8 投影几何码 ……（345）
习题 ……（349）
附注 ……（349）
第十四章 差错控制码的构成和性能 ……（350）
14.1 重量分布 ……（350）
14.2 译码错误和译码失败的概率 ……（357）
14.3 卷积码的重量分布 ……（359）
14.4 分组码的最小距离的界 ……（361）
14.5 卷积码的最小距离界 ……（368）
习题 ……（371）
附注 ……（372）
第十五章 噪声信道的有效信号传输 ……（373）
15.1 带通的高斯信道 ……（373）
15.2 比特能量和比特错误比 ……（375）
15.3 分组码的软判决译码 ……（378）
15.4 卷积码的软判决译码 ……（386）
15.5 序贯译码 ……（391）
习题 ……（393）
附注 ……（393）
汉英名词对照索引 ……（395）

序　言

当前，差错控制码这个课题在供从事实际设计的工程师使用的手册中占有重要的一页。今天，人们对这一课题有着浓厚的兴趣，然而在这个课题刚提出来时，人们曾以为，除了在那些特别昂贵的通信系统中需要作这方面的考虑外，这个课题并无多大实际价值。人们现在对控制差错的要求已经比以前迫切得多，而电子电路能力的巨大改进又促使人们对这一课题更加关心。任何一位从事现代通信系统设计或者从事大型数字系统设计的人都已经离不开这方面的实用知识，而且有种种迹象表明，这方面的知识在工程上的重要性还在增加。关于这个课题已经出版过几种好书，但它们主要是侧重数学方面和研究方面，着眼点是探讨如何才能找到最佳的码。这当然是这个课题的发展方向，然而设计者关心的却只是他能为此做些什么。

本书是专为对差错控制码感兴趣的大学生和工程师编写的，他们在实际工作中将会要用到这种差错控制码。当然，这样说并不意味着实践可以不需要理论。一位设计者虽然不一定要掌握对于这一领域的研究者来说必不可少的全部知识，但是，他如果缺乏必要的数学背景知识，对这一课题了解得十分肤浅，那也一定做不好工作。

本书本来是作者为讲授差错控制码课程准备的一份讲稿，曾在康乃尔大学和IBM公司使用过多次。那几次讲课中，当然不可能要求听讲者具备抽象代数方面的知识。因此，我不得不想办法把所需要的代数基本知识讲述得既严格，又够用，而且必须在几周的讲课时间内讲完。正是这个缘故，我在本书中已写进了学习差错控制码所必需的全部代数学知识，而且对于所涉及的每一个问题，不是给出了证明，就是作了有一定说服力的讨论。不掌握这些起码的知识，就不能说一位工程师已经有了扎实的基础。

在我讲授这门课程的那几年中，听课的人原来只有研究生，后来又有了大学生。所以，我不得不注意讲得简单些，并尽可能多用工程上的行话。尽管所使用的数学已降低到不能再低的水平，然而在某些部份，恐怕仍会有读者嫌所用数学程度偏高。

本书以证明一个又一个定理的方式来展开讨论，而不是象多数工程教科书中那样就一个一个的问题来讨论。采用这种方式，读者在必要时可以暂时跳过一些证明，径直去阅读有关结论。同时，为了照顾讲求严密性的读者，整个理论在分拆开来时，每一部分又自成系统。

我把讨论的重点放在有限域傅里叶变换上，因为有工程背景的读者会很快掌握它，而且对工程师来说，它有丰富的直观的内容。傅里叶变换的应用使我们清楚地看到差错控制编译码是数字信号处理的一个分支。虽然用类似的方法来表达同样的概念，我们也可以采用更早的 Mattson－Solomon 多项式，但这要求工科学生在他们已熟悉的术语之外重新学习另一新的术语。

本书经常强调差错控制码课题与数字信号处理课题之间的紧密关系。因为两个课题

有着迥然不同的历史根源：课题之一主要由代数学家所开发，而另一个则主要由工程师所开发。因而大多数的论述往往使二者之间的关系变得模糊不清。其实，两个课题除了用不同的数学系统——伽罗华域或复数域外，所用技术则相类似。两者都以傅里叶变换，FIR 滤波器，循环卷积，以及序列的时域和频域性质的相互关系为基础。

全书的重点放在设计移位寄存器电路来实现编码器和译码器。尽可能使用滤波器理论的工程术语。在设计移位寄存器时尽量把概念阐释清楚。我们往往可以看到，设计时作些修正可以减少元件的数目。诚然，通常编码器和译码器是用软件来实现的，即使这样，一些关于移位寄存器硬件概念对简化程序也有帮助。我认为对一种码或者一个算法的最终检验在于编码器/译码器的价格如何。对于工程师来说，有了优越的最小距离的码如果不知道好的译码算法也将毫无意义。好码需要好的译码器，而好的译码算法却很难找到。最后，对理论家来说，寻找新的码来适应已知的译码器和寻找新的译码器来适应已知的码同样是很有意义的。

在选用符号和术语方面，习惯用法和内在的一致性之间常常发生冲突。在选用符号时已经小心地考虑到习惯用法。但在有些情况，我认为更重要的是要达到数学上的明晰性和一致性。例如，在讨论卷积码时，我宁愿选择强调与分组码类比的符号，尽管这些符号和多数卷积码文献中所用的符号不同。

Richard E. Blahut

致谢

不可能将所有日常交谈的内容和所阅读的材料划分为对本书有重要影响和没有什么影响的两类，如果这样划分的话，那将是一种曲解。最多我只能提及我感到对本书影响大的那些。在编写本书的那些年头 Toby Berger 教授是我的一位朋友和顾问，他经常给我正确的劝告。D. L. Sarwate 教授仔细阅读了绝大部分的书稿，使我避免了许多差错和臃肿的章节。有价值的劝告和批评也来自 C. L. Chen, Elgamel, M. R. Best, N. M. Blachman, T. Hashimoto, K. Kobayashi, M. Shimada, G. Ungerboeck, W. Vanderkulk, S. winograd, 和 S. C. West。在参考文献中列出了直接或间接对本书有较大影响的书籍和学术论文。许多影响较少的学术论文就不一一列举了。

感谢 IBM（国际商用电子计算机公司）公司对本书准备工作的支持，康乃尔大学提供课堂使用本书可以进行教学实践。我在华南理工大学的讲学也使本书益臻成熟。

在本书编写过程，最重要的参加者是我的妻子 Barbara。她的帮助是多方面的，物质上的和心理上的，静静地承受着我的某些挫折，共享着渡过那些欢快的日子。最后，本书奉献给 Edward J. Blahut, Andrew S. Chamer, 和 Carl A. Krachenfels。

第一章 导 论

数字信号处理是一个有着许多分支的工程课题，其中包括差错控制码理论，它是一个有着自己的目标以及自己的数学系统的特殊课题。在这些数学系统中，最有成效的技术是人们所熟悉的信号处理运算——涉及卷积、傅里叶变换、滤波以及移位寄存器的运算。差错控制编译码有自己的历史和魅力，而且触及许多其它课题。

差错控制码的课题所处理的工程问题是针对经过通信信道传输过程所发生的差错，如何保护数字数据。基于丰富的数学理论，这一课题已经发展了许多巧妙的防止差错的技术，并已发展成为一种常用的重要工程课题。

差错控制之所以需要是因为现代大容量数据的传输或存储，对差错非常敏感所致。为了适应这一需要，已经有了好的码和好的编码算法的成熟理论。此外，数字集成电路的快速发展也使这些算法有可能付诸实现。

1.1 离散的通信信道

一个通信系统通过一个信道将数据信源和数据用户连接起来。微波链路、同轴电缆、电话线路，甚至磁带，都是信道的例子。通信系统设计者为信道开发了提供输入和加工输出的器件。习惯上，人们把一个数字通信系统的主要功能按图 1-1 所示的方框图

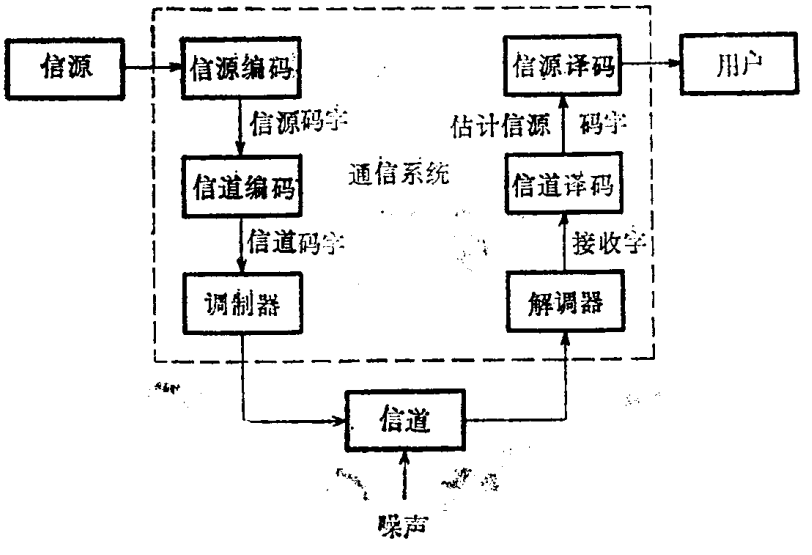


图 1-1 数字通信系统的方框图

划分为若干个部分。数据从信号源进入通信系统后，先被一个信源编码器加工变成更加紧凑的形式。数据在这个中间阶段实际上是一串符号称为信源码字。这以后，又有一个信道编码器对这些数据进行加工，把一串信源码字符号再变成另一串符号，即所谓信道码

字。信道码字已不同于信源码字，它是一串更长的符号，冗余码也更多。信道码字的每一个符号可以用1比特来代表，也可以用一组比特来代表。接着，调制器将信道码字的每一个符号转变为由一个有限模拟符号集合中取出的相对应的模拟符号。然后模拟符号序列通过信道发射出去。由于信道遭受各种类型的噪声和干扰，还有失真，因而信道输出和信道的输入不同。解调器将收到的序列中的每一个信道输出信号转变了一个信道码字符号。每一个经过解调的符号是发射符号的一个最佳估计，但由于信道噪声，解调器产生某些差错。已解调的符号序列称为接收字。由于存在差错，接收字的符号不是都和那些信道码字相吻合。

信道译码器利用信道码字中的冗余码来纠正接收字中的差错，然后产生信源码字的一种估计。如果所有的差错都得到纠正，则估计的信源码字和原来的信源码字相吻合。信源译码器执行信源编码器的逆运算，并将其输出给用户。

本书限于讨论信道编码器与译码器，也就是所谓差错控制编译码。这里不讨论信源编码和信源译码所进行的数据压缩或数据紧化功能，亦不涉及调制器与解调器。今后信道编码器和信道译码器将分别简称为编码器和译码器。

1.2 差错控制编译码的历史

差错控制编译码的历史始于1948年山农（Claude Shannon）发表的一篇著名论文。山农指出，对于任何一个通信信道，总有一个数C（以每秒比特来度量），称为信道容量，它的意义如下。只要一个通信系统所要求的传信率R（表为每秒比特）小于C，就有可能用差错控制码为此信道设计一通信系统使其输出差错概率为任意小。事实上，对于山农信息论的一个重要注释是，建立一个太好的信道纯属浪费；应用一种码往往更为经济。但山农并没有告诉我们如何寻找合适的码；他们的成就在于证明这些码的存在。整个50年代，人们在寻找能给出任意小差错概率的各种码的显式结构方面做了许多工作，但进展不大。60年代的大部分时间，人们较少纠缠于这一雄心勃勃的目标；编译码的研究开始稳定下来，并沿着两条道路进行探索。

第一条道路有着浓厚的代数气息，基本上是分组码。第一个分组码出现于1950年，当时汉明（Hamming）描述一类纠单个差错的分组码。汉明码比山农定理指出的码要弱得多。尽管人们进行了努力探索，但一直到50年代末，尚未找到更好的码类。在这一期间找到了许多码长比较短的码，但一般性理论尚未出现。主要进展在于Bose，Ray-Chaudhuri（1960）和Hocquenhem（1959）找到一大类多个差错纠错码（BCH码），以及Reed与Soloman（1960）为非二进信道找到一类有关的码。这些码至今仍然留在最重要的码类之中。自那以后，这一课题的理论已经大大加强，而且间歇不断地发现新码。

BCH码的发现使得人们开始为研制编码器和译码器去寻找设计硬件或软件的实用方法。第一个好的算法已被Peterson找到。其后，Berlekamp和Massey发现了进行Peterson计算的强有力的算法，而且由于有了新的数字技术，完成这一算法成为实际可行。

编译码研究的第二条道路有着更多的概率气息。早期的研究在于对最佳分组码族估计差错概率，尽管并不知道这些最佳码。这些研究的目的是想从概率论的观点来了解编码和译码，这些尝试导致了序贯译码的概念。序贯译码要求引入一类不定长非分组码，它们可用一棵树来表示，而后用搜索该树的算法来译码。最有用的树码是高度结构码称为卷积码。这些码可用对信息序贯进行卷积运算的线性移位寄存器电路来产生。在50年代后期卷积码已经用序贯译码算法成功地译码。有趣的是，一种简单得多的卷积码的译码算法——Viterbi 算法，一直到1967年才出现。对中等复杂度的卷积码，Viterbi 算法得到广泛的应用，但对更强的卷积码它已无能为力了。

在70年代的十年中，两条研究道路开始结合在一起。代数学者对卷积码理论进行了研究，他们为卷积码理论带来了许多新的见解。在分组码理论方面，这一时期也朝着山农定理所指出的码有所前进，那就是提出了设计同时具有码长很长和性能很好的码族的方案，其一由 Justesen 提出，另一种由 Goppa 提出。但两种方案都有实用上的局限性，有待于进一步发展。当80年代开始时，在新设计的数字通信系统以及数字存储系统中，编码器和译码器开始经常出现。

1.3 应　　用

由于差错控制码的发展主要是受通信问题所激发，差错控制编译码的专有名词都是从通信理论的课题中移植过来。但这些码有着许多其它的应用。例如，在计算机存贮器以及数字磁带和磁盘中用这些码来保护数据，在数字逻辑电路中用这些码来防止电路失常或噪声。这些码也用来作数据压缩，而且编译码理论与统计实验的设计理论也是密切相关的。

在通信问题上码的应用是多方面的。平常二进制码在计算机终端之间，在飞机之间，以及在空间飞船与地面之间进行传输。即使在接收信号功率接近于热噪声功率的情况下，也可以用这些码来达到可靠的通信。当前，由于许多人造信号使得电磁频谱愈来愈拥挤，纠错编码成为更加重要的课题，因为当干扰存在时，纠错码可使通信线路可靠地工作。在军事应用中常需使用纠错码防止敌人的有意干扰。

许多通信系统对发射功率都有限制。例如，在通信中继卫星中功率是非常昂贵的。纠错码是降低所需功率的一种最好办法，因为终端接收到的微弱消息可以借助于纠错码来正确地得到恢复。

在计算机系统内的传输，平常只能允许极低的差错率，因为一个差错就会破坏整个计算机程序。纠错码在这些应用中显得愈来愈重要。通过应用纠错码可在某些计算机存贮器中（例如磁带）将位数组装得更紧凑些。

通信系统的另一种结构，是时分多址系统，其中对众多用户中的每一用户指定某些预定的时隙（间隔）容许该用户传输。将一个长的消息划分为许多信息包，每一信息包在指定的时隙中传输。由于同步或路由的问题，有时信息包会丢失。合适的纠错码可以对这一丢失起防护作用，因为可以从已知的包中导出丢失的包。

在一定系统内的通信也很重要。在现代复杂的数字系统内，其子系统之间可存在巨

大的数据流。数字自动驾驶，数字过程控制系统，数字开关系统，以及数字雷达信号处理都是含有大量数字数据的系统，它们必须由多个相互连接的子系统来分担。这些数据或由专用线或由更高级的时分数据总线系统来传送。在上述的任一情形，为了保证正常的性能，差错控制技术变得愈来愈重要。

纠错码以及编码和译码电路最终将达到它们能掌握大量数据的境地。可以预期将来差错控制技术在所有通信系统中将扮演中心角色。很可能将来唱片，磁带以及电视波形将使用纠错码保护的数字消息。唱片的刮擦或者接收信号的干扰将被编码全部抑制掉，只要设计纠错码的纠错能力比差错更强。

1.4 基本概念

差错控制码这一课题既简单而又困难。所谓简单，是指对于任何受过技术训练的人来说，这一问题很容易解释清楚。所谓困难是指为了发展一个解题——而且只是部分解题，就占据了本书一定的厚度，而且在阐述之前还需要专题论讨近世代数的题目。

假设所有数据都可以用二进（编码的）信息来表示，亦即可表为 0 和 1 的序列。二进信息通过一信道传输，有时会引进差错。编码的目的是将额外的符号加于信息符号，因而在接收处可找到并纠正差错。亦即，数据序列表为某种更长的符号序列具有足够的多余度以保护数据。

大小为 M，码长为 n 的二进制码是长为 n 的 M 个二进字的集合，称为码字。平常对于整数 k，$M=2^k$，此码称为（n，k）二进码。

例如，我们可以造成下列码

$$\mathcal{C}=\begin{Bmatrix}1&0&1&0&1\\1&0&0&1&0\\0&1&1&1&0\\1&1&1&1&1\end{Bmatrix}$$

这是一个很次的（很小的）码，具有 $M=4$，$n=5$，但它满足定义的要求。用下面的（任意的）对应关系，我们可以用这种码来代表 2 位的二进制数。

$$\begin{array}{l}0\ 0\longleftrightarrow 1\ 0\ 1\ 0\ 1\\0\ 1\longleftrightarrow 1\ 0\ 0\ 1\ 0\\1\ 0\longleftrightarrow 0\ 1\ 1\ 1\ 0\\1\ 1\longleftrightarrow 1\ 1\ 1\ 1\ 1\end{array}$$

如果接收到 4 个 5 位码字之一，我们可以断定所对应的 2 位是正确的信息。如果已发生一个差错，我们将收到一个不同的 5 位码字。然后我们试图找出最相象的传输码字作为真正 2 信息位的估计。例如，如果我们收到 0 1 1 0 0，则我们设想传输的码字是 0 1 1 1 0，因而信息码将是 1 0。

上例的码不是好码，因为它不能纠正多种差错的图样。我们希望找到一些码，它的每一码字尽可能和每一个别的码字不同，特别当码长很长时我们希望这样做。

本书的第一个目的是寻找好码。尽管从表面来看好象这一种简单的工作，事实上这

是异常困难的，因而至今许多好码尚未被发现。

对于不熟悉的人来说，可能认为只要对好码的各种要求进行定义，然后让计算机搜索遍所有可能码的集合就解决问题。但对于给定的(n, k)究竟有多少码呢？我们知道，每一码字是一n个二进符号的序列，这种码字总共有2^k个。因而一个码要用$n\cdot 2^k$二个进符号来描述。总共有$2^{n\cdot 2^k}$种方式来选择这些二进符号。这样一来，不同(n, k)码的数目为$2^{n\cdot 2^k}$。当然这些码中有许多是没有价值的（例如当两个码相等时），这就要求搜索时必须核对这些码，或者必须发展某种理论来排除这些无用的码。

例如，取$(n, k)=(40, 20)$，这从今天的标准来看是一种极为普通的码。然而这种码的数目却多于$10^{10,000,000}$，一个难以想象的大数目。由此可见，盲目的搜索步骤是毫无价值的。

一般来说，我们在任意有限字母，譬如说q个符号$\{0,1,2\cdots,q-1\}$的字母上定义分组码。乍看起来，引入二进字母以外的字母，似乎是一种不必要的一般化。但由于效率的原因，今天许多信道是非二进的，因而这些信道的码必须是非二进的。事实上，非二进信道的差错控制码往往是比较好的，这对使用非二进信道的理由是一种加强。将二进源数据表示为q元字母是无意义的事，特别是如果q是2的幂（实践中通常是这样）。

定义1.4.1 在q个符号的字母上，大小为M的分组码，是一个被称为码字的M个长为n的q元序列的集合。

如果$q=2$，则这些符号称为比特（bits）。通常$M=q^k$，k为某一整数，我们将限于讨论这一情况，并将这种码称为(n, k)码。每一k个q元的信息符号序列可伴随以n个q元符号序列构成一个码字。

有两种基本的码类：分组码和树码，如图1-2所示。分组码是用n个符号的码字来代表一组k个信息符号。分组码的码率①R定义为

$$R=\frac{k}{n}$$

树码更复杂些。它以每时间间隔k_0符号的码率取一个无休止的信息符号序列，而且以每时间间隔n_0符号的码率构成一个码字符号的连续序列。我们把树码的讨论推后到第十二章。开始我们限于讨论分组码。

当一个消息包含有许多比特时，从原理上说用较长码长的单一分组码比用由较短的分组码构成的一串码字更好。这是因为随机起伏的性质使得差错的随机图样往往显示出某种差错群聚的形式。即随机图样的某些段的差错比平均数多，而某些段则比平均数少。因而对于相同的速率来说，长码字比短码字对于随机差错要更不敏感得多，当然编码器和译码器也会更复杂些。作为例子，假设用一个可纠正100位差错的（假想的）2000位二进码字来传送1000信息位。试将它和用每分组可纠正10位差错的200位二进码字每次传送100位的方案比较。传送1000位需要10个分组。后一方案同样可以纠正总共100个差

①码率是无量纲的，或者用比特/比特或符号/符号为单位来表示。需要将它和用于信道的另一名词信息率（表为比特/秒）相区别。但另一定义$R=(k/n)\log_e q$，其单位为奈特（nat）/符号，（1奈特$=\log_2 e$比特）也在应用。定义$R=(k/n)\log_2 q$，其单位为比特/符号也很通用。

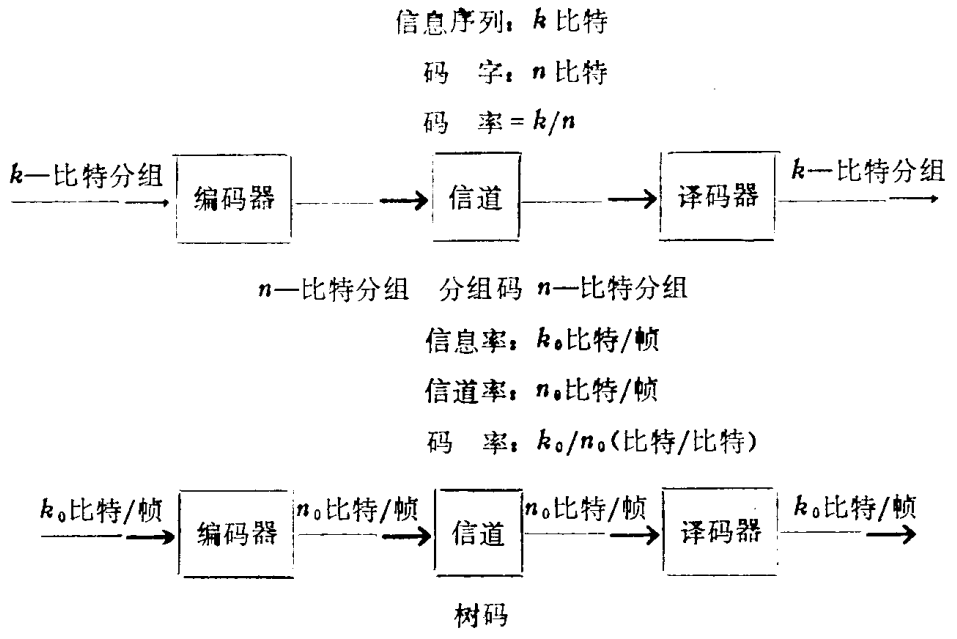


图 1－2　基本的码表

错，但只有当这些差错恰当地分布时——每200位分组有 10 个差错时才是这样。而第一方案不管差错在2000位码字内如何分布，都能纠正 100 个差错。因而第一方案比第二方案要强有力得多。

可以把上述带启发性的论述放在一个强的理论基础上，但这不是我们现在的目的。我们只希望说明这一点，即好码的码长很长，极好码有的码长更长。这些码找起来十分困难，而且一旦找到，可能要求复杂的设备来实现编码和译码运算。

分组码可由三个参量来评价：码长 n，信息长 k，以及最小距离 d^*。最小距离是两个最相似码字之间差别程度的一种量度。下列两定义给出最小距离。

定义1.4.2　两个具有长 n 的 q 元序列 x 与 y 汉明距离 $d(x,y)$ 是它们相异的位的数目。

例如，取 $x=10101$，$y=01100$ 则 $d\ (10101,01100)=3$。另一例子，取 $x=30102$，$y=21103$，则 $d(30102,21103)=3$。

定义1.4.3　令码 $\mathcal{C}=\{c_i, i=0\cdots,M=1\}$。则 $\mathcal{C}$ 的最小距离为具有最小汉明距离的码字对的汉明距离，亦即

$$d^*=\min_{\substack{c_i,c_j\in\mathcal{C}\\ i\neq j}} d(c_i,c_j)$$

具有最小距离 d^* 的一个（n，k）分组码，亦可描述成一个（n，k，d^*）分组码。

本节第一个例子的码 $\mathcal{C}$，

$$d(10100,\ 10010)=3$$
$$d(10100,\ 01110)=4$$
$$d(10100,\ 11111)=2$$
$$d(10010,\ 01110)=3$$
$$d(10010,\ 11111)=3$$

$$d(01110,\ 11111)=2$$

因而这个码的$d^*=2$。

假设我们发送一个码字，信道引起单个差错，则接收到的码字与发送的码字相距1个汉明距离。如果与其它每一码字的距离大于1，则在假定与收到的码字相距最近的码字即为发送码字的前提下，译码器可以适当地纠正这一差错。

更一般地说，如果发生t个差错，而收到的码字与其它每一个码字的距离大于t，且如果假定实际发送的码字是和收到的码字相距最近的码字，则译码器将能恰当地纠正这些差错实现的条件是

$$d^*\geqslant 2t+1$$

即使未能满足上列不等式，有时也可能纠正t个差错的某些差错图样。但是，如结$d^*<2t+1$则不能保证纠正t个差错，因为t个差错的纠正取决于发送了哪个码字，以及码组内t个差错的实际图样。

图1-3表示几何的解释。在所有q维n元的空间内，选择一个n元集合，其元素被指定作为码字。如果d^*为这种码最短距离且t为满足下列条件的最大整数

$$d^*\geqslant 2t+1$$

则围绕每一码字可绘出半径为t的不相交圆球，我们把落在一个圆球内的被接收的码字译码成圆球中心的码字。如果发生t或小于t个差错，则接收到的码字总是在适当的圆球内，因而译码正确无误。

某些接收到的差错大于t的码字将落入其它码字的圆球，因而将被错译。另一些差错大于t的接收码字将落入译码圆球间的间隙内。基于应用的不同要求，这些码字可使下列两种方法之一来处理。

其一是不完全的译码器，它只将落入一个码字的译码圆球内的接收码字译码。其它多于容许差错数目的接收码字，译码器将其看成不可辨识。这种差错图样在不完全的译码器中被称为无法纠正的差错图样。绝大多数平常用的译码器属于不完全译码器。

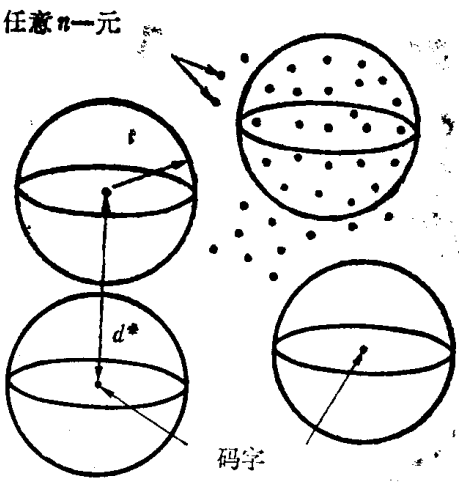


图 1-3 译码圆球

其二是完全译码器，它将每一接收码字译码成最近的码字。用几何术语来说，完全译码器将空隙切开并将切开部分依附于译码圆球，使得每一点都依附于一个最近的圆球。通常总有某些点与几个圆球等距离。这些点可任意指定属于最近圆球之一。当差错大于t时，完全译码器将发生错译，但有时也会找到正确的码字。如果我们认为对消息进行最佳猜测比对消息不作任何猜测为好时就要用完全译码器。

我们也将讨论既产生差错也产生删除的信道。就是说，可以设计一种接收机当接收得模糊不清，或接收机测知发生干扰或瞬时工作不正常时宣告删除一个符号。这种信道输入字母容量为q时输出字母容量为$q+1$，额外增加的符号称为一个删除。例如，从

消息12345的3的一个删除给出12—45。不要把删除和另一称为缺失的概念混淆，缺失将给出1245。

对于上述信道可以使用一种差错控制码。假如该码的最小距离为d^*，如果$d^* \geqslant \rho+1$，则任何具有ρ删除的图样都可以填补上。而且只要满足条件

$$d^* \geqslant 2\nu+1+\rho$$

任何具有ν差错和ρ删除的图样均可译码。作为证明，从所有码字中将接收机指示删除的ρ个分量去掉，这是一个新码。其最小距离不少于$d^*-\rho$；因而只要满足上列不等式，ν个差错可被纠正。这样我们可以将被抹去ρ个分量的缩短了的码字恢复过来。最后，由于$d^* \geqslant \rho+1$，只有一个码字和未删除的分量相符；因而整个码字都可以恢复。

1.5 基本码

某些足够简单的码可以从一开始给予描述。

简单检错码 这是些高速码但差错性能差。给定k位信息，加进一个第$(k+1)$位使得码字中1的总数为偶数。例如，如果$k=4$，

$$0\ 0\ 0\ 0 \longleftrightarrow 0\ 0\ 0\ 0\ 0$$
$$0\ 0\ 0\ 1 \longleftrightarrow 0\ 0\ 0\ 1\ 1$$
$$0\ 0\ 1\ 0 \longleftrightarrow 0\ 0\ 1\ 0\ 1$$
$$0\ 0\ 1\ 1 \longleftrightarrow 0\ 0\ 1\ 1\ 0$$

等等。这是一个$(k+1,\ k)$或$(n,n-1)$码。最小距离为2，因而不能纠错。一个简单的检错码可用来检出（不能纠正）单一差错。

简单重复码 这是一些具有良好差错性能的低速码。给定单一个信息位，重复n次。平常n为奇数。

$$0 \longleftrightarrow 0\ 0\ 0\ 0\ 0$$
$$1 \longleftrightarrow 1\ 1\ 1\ 1\ 1$$

这是一个$(n,1)$码。最小距离为n，假设多数接收位与正确信息位相符则可纠$\frac{1}{2}(n-1)$个差错。

汉明码 这是些能纠正单一差错的码。这里我们用直接描述的方法来介绍这些码。对每一个m来说，有一个$(2^m-1,2^m-1-m)$汉明码。当m很大时，码率接近于1，但总的位数中产生差错的部分很小。（7，4）汉明码可用图1-4 (a)的装置来描述。给定四个信息位(i_1,i_2,i_3,i_4)，令码字的头四个位等于这四个信息位。附加三个一致监督位(p_1,p_2,p_3)定义如下

$$p_1=i_1+i_2+i_3$$
$$p_2=i_2+i_3+i_4$$
$$p_3=i_1+i_2+i_4$$

这里 + 代表模2加$(0+0=0,0+1=1,1+0=1,1+1=0)$。汉明（7，4）码的16个码字示如表1-1。译码器接收一个7位字$v=(i_1',i_2',i_3',i_4',p_1',p_2',p_3')$，对应于发送码字