第三章 信道编码

（计划学时　14）

本章主要内容

- 检错、纠错原理 2学时
- 差错控制理论 2学时
- 线性分组码 2学时
- 循环码 2学时
- 循环码的扩展 2学时
- 卷积码 2学时
- 纠正突发错误的编码 2学时

教学目的与要求

1. 深刻理解信道编码的检、纠错原理。

2. 熟练掌握建立译码规则的原则和计算错误概率的方法。

3. 掌握线性分组码、循环码、卷积码等重要的信道编码的原理与方法。（重点）

4. 了解Shnnon信道编码的内容和意义。

参考文献

1.王新梅：**纠错码—原理与方法**

西安电子科技大学出版社（2001年4月修正版）

2.袁东风：**宽带移动通信中的先进信道编码技术**

北京邮电大学出版社 (2004年3月第一版)

3.吴伟陵：**信息处理与编码**

人民邮电出版社（1999年7月第一版）

4.曹雪虹：**信息论与编码**

北京邮电大学出版社 (2001年8月第一版)

第三章 信道编码

3.1 检错、纠错原理

（第9讲 2007.10.30.）

本节的主要内容

- 检错原理
- 纠错原理
- 汉明重量与最小码距
- 检错纠错能力
- 信道编码的性能指标
- 几种简单的检、纠错码

外语关键词

信道编码：channel coding

检错与纠错：error detection and correction

汉明距离：hamming distance

最小码距：minimum code distance

重复码：repetition code

奇偶校验码：parity-check code

[温旧引新]

- 三大编码：

  信源编码、信道编码和加密编码

- 信源编码：

  压缩代码长度的编码。

- 信道编码：

  为了减少差错而进行的编码。

3.1.1 检错原理

- 误码的必然性：

在有噪声和损失存在的信道中，输入符号与接收符号不能一一对应，传输错误和判断错误的情况总会存在。

- 误码的不可预知性：

误码是随机发生的。接收端并不了解是否发生了误码，更不知晓是哪个码元出现错误。

- 检错-------如何检测有无误码？

办法很多。比如双连编码方式把每个0和1都重复一次，再送入信道。接收端一旦发现有违背“双连”规律的码元，就能判断该码元发生了错误。

- 正误应有别-------许用码字与非法码字

两位一组，有00，01，10，11四种方式。00与11是符合双连规律的，称为许用码字。01与10是违背双连规律的，称为非法码字。从而正与  有了区别。

- 原因----------冗余的作用

两个码元当一个用，添加了一倍的“冗余”，才使码字克服了“非此即彼”的情况。

- 四种天气状况：

晴、	阴、	雨、	风
00	01	10	11
没有冗余	非此即彼	错了也不可知	

晴、	阴、	雨、	风
00000	01101	10110	11011

- 5位二进制码元传输2*bit*的信息，冗余3 *bit*。
- 2^5=32个码字， 许用码字4个，非法码字28个。
- 因为冗余才避免了码字之间非此即彼，错了也不可知的情况，使编码具有了检错能力。

思考题：

信源编码中要压缩掉冗余，信道编码中又要添加进冗余，是不是返工浪费？两种冗余有何不同？

3.1.2 纠错原理

- 发现错误怎么办？

  有三种处理方案：

  （1）重发反馈方式（ARQ）；

  （2）前向纠错方式（FEC）；

  （3）混合纠错方式（FEC）；

- 自动纠错的前提：

  不仅知道有错，还应知道是哪个码元出现错误。

- 双连码只能检错，不能纠错，因为冗余不够。
- 三连重复码当其中任一码元出错时，还有2位没有错，通过比较就能发现是哪一位错了，从而可以进行纠正。
- 万一三连重复码有两位码元出错时，就会判断失误。表明超出了它的纠错能力。
- 三连重复码的纠错能力是1位，五连重复码的纠错能力是2位，添加的冗余越多，纠错能力就越强，然而传输效率就越低。

汉明距离(Hanming distance)：

定义两个许用码字相同码位上不同玛元的个数叫它们的汉明距离。汉明距离反映了两个码字的差别。

- 双连重复码的汉明距离d=(00,11)=2；三连重复码的汉明距离d=(000,111)=3；
- 双连重复码发生1位错的误码是01或10，与许用码字11或00的汉明距离相同，无法判别误码来自谁。
- 当三连重复码有一位码元出错时，比如000变成001、010或100，误码与000的汉明距离为1，而与111的汉明距离为2，据此就可判定误码来自000；

检、纠错原理------添加冗余

- 冗余的添入，增加了码字总数，给误码留下了空间，就给判断正确与错误提供了可能（检错原理）。
- 更多冗余的添入，拉大了许用码字之间的汉明距离，就有可能区别误码与各个许用码字之间汉明距离的不同，从而判断误码可能的来源（纠错原理）。

3.1.3 汉明重量与最小码距

- 汉明重量(Hamming weight)
- 定义码字中码元 1 的个数叫做该码字的汉明重量。

  如：$W[11010]=3$；$W[00000]=0$;
- 汉明重量与汉明距离的关系：

$$d(u, v)=W[u \oplus v]$$

u和v是任意两个码字，$\oplus$代表按位模2加(异或)。u和v码元相同的位模2加为0，不同的位模2加为1，$W[u \oplus v]$就给出了u和v不同位的个数。

最小汉明距离

一种编码有许多个码字，全部两两比较后，找出最小的汉明距离为d_0。

如：C_1=00000；　C_2=01101；　C_3=10110；　C_4=11011；

则：$d(C_1, C_2)=3$；　$d(C_1, C_3)=3$；　$d(C_1, C_4)=4$；

$d(C_2, C_3)=4$；　$d(C_2, C_4)=3$；　$d(C_3, C_4)=3$；

因此最小的汉明距离为$d_0=3$

线性码的最小汉明距离：

$d_0 = W_{min}$　(最小汉明重量）

证明：

- 设u 与v 之间码距最小，于是：

$$d_0 = d(u, v) = W[u \oplus v] \geqslant W_{min}$$

式中$u \oplus v$是另一许用码字，其重量当然不应小于设定的最小重量 W_{min} ；

- 另一方面，设非零码字c具有最小重量，则：

$$W_{min} = W[c] = W[c \oplus 0] = d(c, 0) \geqslant d_0$$

式中d（$c, 0$）是码字c与码字0的距离，它当然不应小于设定的最小汉明距离d_0 ；

- 两式同时考虑， 只能有 $d_0 = W_{min}$

3.1.4 检错、纠错能力

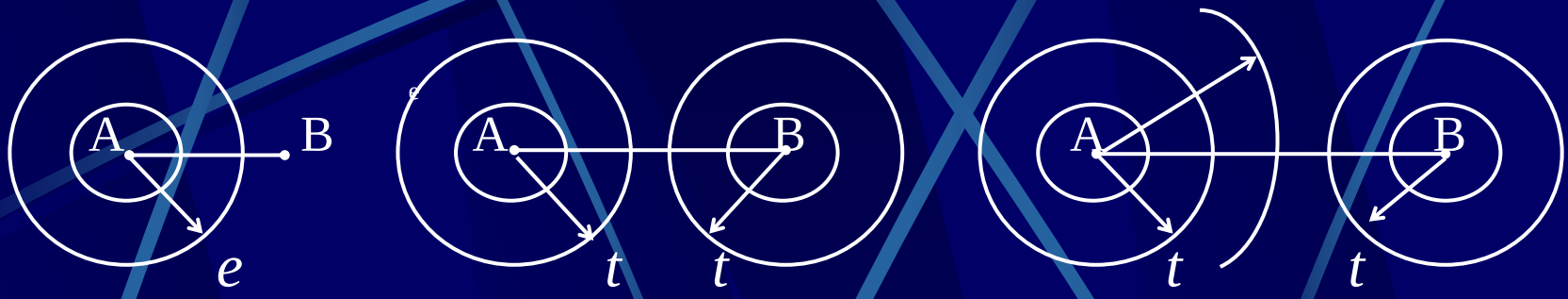
- 检错、纠错能力与最小汉明距离有直接关系
- 设d_0为最小汉明距离，e为可检错位数，t为可纠错位数，则：

(*a*)为了发现e个错误码元，要求最小码距$d_0 \geq e+1$；

(b)为了纠正t个错误码元，要求最小码距$d_0 \geq 2t+1$；

(c) 为发现e个错码元，同时又能纠正其中t ($t<e$)个错误码元，要求最小码距$d_0 \geq e+t+1$；

纠错、检错能力与最小码矩的关系



(a)$d_0 \geqq e+1$　　(b) $d_0 \geqq 2t+1$　　(c) $d_0 \geqq t+e+1$ 且$e>t$

如果只检错不纠错，则最多检错位数是$e=d_0-1$。

如果只纠错不检错，则最多纠错位数是：

t =INT [$(d_0-1)/2$]，INT表示取整数

如果既纠错又检错，则e和t就不能独立，它们应满足关系式：$e+t \leqslant d_0-1$，而且e必须大于t，

[例1]已知汉明距离d_0=7，分析检、纠错能力。

（1）只检不纠：$e=d_0-1=6$；可检到6位错。

（2）只纠不检：$t=(d_0-1)/2=3$；可纠3位错。

（3）又检又纠：e 与t不独立：$e+t\leqslant d_0-1$，且$e>t$

当t=1时，$1<e\leqslant 5$，1位错已纠，只能检第2—5位错。

当t=2时，$2<e\leqslant 4$，2位错已纠，只能检第3—4位错。

当t=3时，e 无解，3位错已纠， 3位以上无能力检

3.1.5 信道编码的性能指标

- 编码效率

  设某种编码的码字长n位，其中信息只有k位，$r = n–k$为冗余位，则该编码的信息率（也叫编码效率）：　$\eta = k / n$

- 漏检率

  把编码检查不出的错误所出现的概率叫做漏检率。

- 差错率

  把编码不能自动纠正的错误所出现的概率叫做差错率。

[例2]讨论双连重复码的编码效率和漏检率。

编码效率　$\eta=k/n=1/2=0.5$

假设对称信道，0与1的错误概率均为p，则：

漏检率=2位同时出错的概率=p^2

[例3]讨论三连重复码的编码效率和差错率。

编码效率　$\eta=k/n=1/3=0.33$

差错率=3位同时出错的概率

+2位出错1位正确的概率=$p^3+3p^2(1-p)$

3.1.6 几种简单的检错、纠错码

（1）n连重复码

- n为奇数时，可纠正 $t=(n-1)/2$位错误。错传概率为p时，

  差错率=n位同时出错的概率+(n−1)位出错1位正确的概率

  +……+(n+1)/2位出错(n−1)/2位正确的概率

  $$=p^n+C_n^1 p^{n-1}(1-p)+\cdots\cdots+C_n^{(n-1)/2} p^{(n+1)/2}(1-p)^{(n-1)/2}$$

- n为偶数时，可纠正(n/2−1)位错误，还能发现第n/2位出错。

  漏检率=n位同时出错的概率+(n−1)位出错1位正确的概率

  +……+(n/2+1)位出错(n/2−1)位正确的概率

  $$=p^n+C_n^1 p^{n-1}(1-p)+\cdots\cdots+C_n^{(n/2-1)} p^{(n/2+1)}(1-p)^{(n/2-1)}$$

- 它的编码效率很低，只有 $\eta=1/n$

（2）奇偶校验码

- k 个信息位后面添加1位“奇偶校验位”，其关系为：

$$\underbrace{a_{n-1} \oplus a_{n-2} \oplus \cdots\cdots \oplus a_1}_{n-1\text{个信息位}} \oplus \underbrace{a_0}_{1\text{个监督位}} = \begin{cases} 0 & (\text{偶校验}) \\ 1 & (\text{奇校验}) \end{cases}$$

- 编码效率

  它的效率很高，达到： $\eta = (n-1)/n = 1-1/n$

- 检、纠错能力

  可查出任何奇数位的错误，却不能发现偶数位的错误，也没有纠错能力。

- 漏检率$= C_n^2 p^2 (1-p)^{n-2} + C_n^4 p^4 (1-p)^{n-4} + \cdots\cdots$

（3）双向监督码

- 为了增加奇偶校验码的检错能力，对列再进行奇偶校验，增加一行“竖直奇偶校验位”，进行双向监督。也称方阵码。

信息位	监督位
1 1 0 1 0	1
0 1 0 0 1	0
1 1 0 0 1	1
1 1 1 0 0	1
1 0 0 0 0	1
0 0 1 1 1	1
0 0 0 0 1	1
列监督	

- 干扰造成的误码，往往发生在连续若干个码元上。

采用上述双向监督，就把行中难以发现的错误，在列中发现。

（4）恒比码

- 从n 位二进制自然码中挑出 1 和 0 的个数之比恒定的一些码字组成许用码集合。
- 五中取三码是从2^5 =32 个二进自然码中挑出含3 个1，2 个 0的码字，共10个（5取2的组合数），用来表达10个阿拉伯数字；
- 七中取四码则是从 2^7 =128 个二进自然码中选出含4个1 ，3 个0 的码字，共35个（7取3的组合数），可用来表达26个英文字母和标点符号。

[例4]讨论恒比码的编码效率和漏检率。

- 五中取三码的编码效率　$\eta=k/n=\log 10/5=66.4\%$
- 七中取四码的编码效率　$\eta=k/n=\log 35/7=73.3\%$
- 它们可以发现多位错误，却不能发现同时出现0误为1且1误为0的错误，也无纠错能力。
- 五中取三码的漏检率为：

$$[C_3^1 p(1-p)^2]\cdot[C_2^1 p(1-p)]+[C_3^2 p^2(1-p)]\cdot[C_2^2 p^2(1-p)^0]$$

$$=6p^2(1-p)^3+3p^4(1-p)$$

- 七中取四码的漏检率请同学们推导。

小结：

❖信道编码的原理：

检错原理-----添加冗余，克服非此即彼，分辨对错。

纠错原理-----添加冗余，拉大码矩，判定错误来源

❖ 汉明距离与检纠错能力：

检错位数$e=d_0-1$， d_0表示汉明距离

纠错位数t =INT [$(d_0-1)/2$]，INT表示取整数

❖ 简单的检纠错码：

n连重复码、奇偶校验码、恒比码

课后复习题

- 思考题：

  讨论七中取四码的编码效率和漏检率。

- 作业题：

  教材第114页习题三第1、4、6、7题；

第三章 信道编码

3.2 差错控制理论

（第10讲 2007.11. 1.）

本节的主要内容

- 译码规则
- 最小平均错误准则
- 最大后验概率准则
- 最大似然准则
- 错误概率的计算
- 信道编码定理

外语关键词

最小平均错误准则:

Minimum mean error rate criterion

最大后验概率:Maximum posterior probability

最大似然准则:

Maximum likelihood decoding criterion

译码规则：Decoding rule

错误概率：Error probability

漏检率：Undetected error rate

差错率：Uncorrected Error rate

[温旧引新]

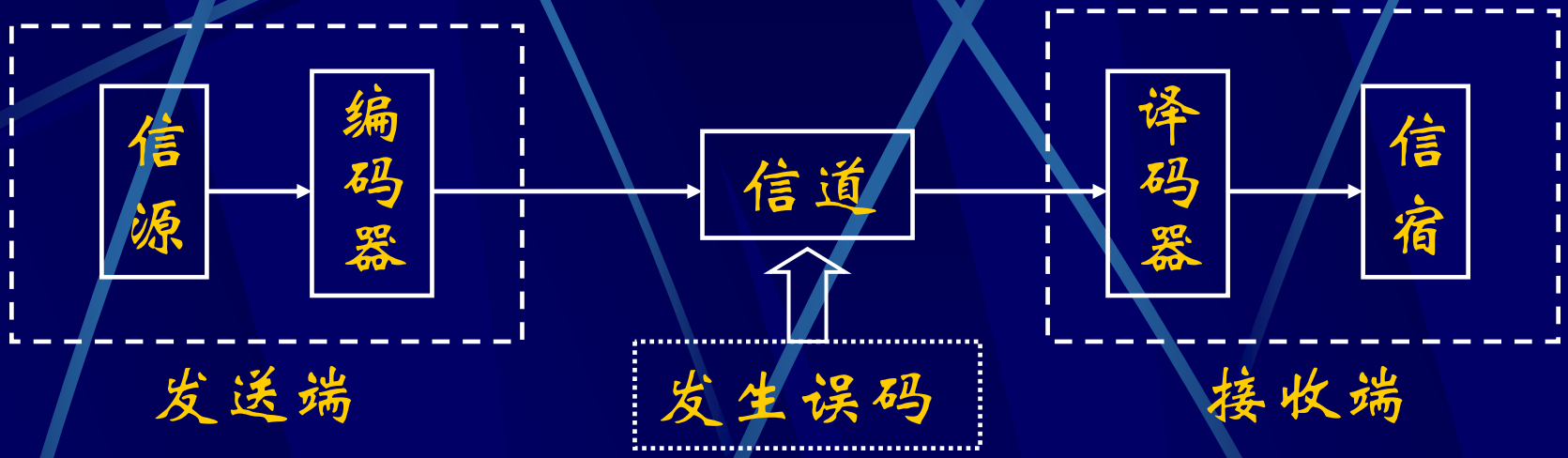
- 检错、纠错原理：冗余的作用。
- 检错、纠错能力：最小汉明距离
- 编码效率：　$\eta = k/n$
- 漏检率：

  实施编码后仍检查不出的错误所出现的概率。

- 差错率：

  实施编码后仍检纠正不了的错误所出现的概率。

3.2.1 译码规则



- 编码在发送端进行；
- 误码在信道中发生；
- 发现与纠正错误必须在接收端进行。

- 译码规则是设计人员预先为译码器设计好的译码指令。
- 每当收到一个符号b_j；系统都应当为它指定一条译码规则：$F(b_j) = a_i$*

  告诉译码器应当把b_j译为哪个 a_i ；
- 比如三连重复码的译码规则是：

  $F(000)=0$； $F(001)=0$； $F(010)=0$； $F(100)=0$；

  $F(111)=1$； $F(011)=1$； $F(110)=1$； $F(101)=1$；

[例1]信源发出符号a_1，a_2，a_3；接收端符号为b_1，b_2，b_3；

对它可以构造多种译码规则。比如：

- 译码规则A: $F(b_1)=a_1^*$；$F(b_2)=a_2^*$；$F(b_3)=a_3^*$
- 译码规则B: $F(b_1)=a_1^*$；$F(b_2)=a_3^*$；$F(b_3)=a_2^*$

同样的信源信道，选用不同的译码方案，将会得到不同的结果。究竟采用哪个译码规则更好呢？一个很自然的考虑就是按该方案译码后，造成的错误概率应最小。

3.2.2 译码错误概率

- 译码规则一旦确定，译码器就按指令办事。
- 设译码规则为$F(b_j) = a_i^*$，当收到一个符号b_j时，按译码规则它就被译为 a_i^*。
- 如果信源发出符号恰为 a_i^*，就翻译对了；如果信源发出的是其它符号，译码器按指令仍然会把它译成为 a_i^*，就翻译错了。
- 根据后验概率计算结果可知，翻译对的概率是 $p(a_i^* \mid b_j)$，翻译错的概率是 $\sum_{i \neq i^*} p(a_i \mid b_j)$。

译码错误概率计算公式

- 对于各种接收符号求平均，翻译对的平均概率是：

$$P = \sum_{j=1}^{s} p(b_j)\, p(a_{i^*} \mid b_j)$$

- 译码错误的平均概率是：

$$P_E = 1 - P = \sum_{j=1}^{s} p(b_j) \sum_{i \neq i^*} p(a_i \mid b_j)$$

- 利用联合概率公式，上两式可写为：

$$P = \sum_{j=1}^{s} p(a_{i^*} b_j) \quad \text{与} \quad P_E = \sum_{j=1}^{s} \sum_{i \neq i^*} p(a_i b_j)$$

[例2] 已知信源和信道的统计性质：

$$\begin{pmatrix} X \\ p(X) \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ 0.3 & 0.4 & 0.3 \end{pmatrix}$$ 和 $$P = \begin{pmatrix} 0.5 & 0.3 & 0.2 \\ 0.2 & 0.3 & 0.5 \\ 0.3 & 0.3 & 0.4 \end{pmatrix}$$

解：因为：$p(a_i b_j) = p(a_i)p(b_j|a_i) = \begin{pmatrix} 0.15 & 0.09 & 0.06 \\ 0.08 & 0.12 & 0.20 \\ 0.09 & 0.09 & 0.12 \end{pmatrix}$

- 按规则A：　$F(b_1)=a_1^*$；对应 $p(a_1^* b_1) = 0.15$

  $F(b_2)=a_2^*$；对应 $p(a_2^* b_2) = 0.12$

  $F(b_3)=a_3^*$；对应 $p(a_3^* b_3) = 0.12$

正确概率 $P = 0.15 + 0.12 + 0.12 = 0.39$

错误概率 $P_E = 1 - P = 0.61$

- 按规则B: $F(b_1)=a_1^*$； 对应 $p(a_1^* b_1) = 0.15$

  $F(b_2)=a_3^*$； 对应 $p(a_3^* b_2) = 0.09$

  $F(b_3)=a_2^*$； 对应 $p(a_2^* b_3) = 0.20$

  正确概率 $P=0.15 + 0.09 + 0.20 = 0.44$

  错误概率$P_E=1-P=0.56$
- 显然，译码规则B优于译码规则A。
- 有没有更好的译码规则呢？
- 最佳的译码规则是什么？怎样确定？

3.2.3 制定译码规则的准则

1．基本准则：

- 最合理的译码规则应当能使译码的平均错误概率最小。这个原则称为平均错误概率最小准则。
- 要使最小平均错误概率最小，译码正确的平均概率就应当最大。其充分条件是对于每一个接收符号b_j时都满足后验概率$p(a_i^* \mid b_j)$确为最大。
- 可见，平均错误概率最小准则等价于后验概率择大准则。

后验概率择大准则

- 当收到一个符号b_j时，系统并不知道发的是什么符号。
- 只能根据信源和信道的统计性质算出收到b_j的条件下，信源发出各个 a_i 的后验概率$p(a_i \mid b_j)$，并进行比较。
- 把后验概率最大的那个a_i *指定为应当翻译的发送符号。此即后验概率择大准则。
- 它要求译码规则：$F(b_j) = a_i^*$

  满足：$p(a_i^* \mid b_j) \geqslant p(a_i \mid b_j)$　　$i=1, 2, \ldots\ldots, m$

- 后验概率比较难算。但由公式：

$$P = \sum_{j=1}^{s} p(b_j)\, p(a_i^* \mid b_j) = \sum_{j=1}^{s} p(a_i^* b_j)$$

$$P_E = \sum_{j=1}^{s} p(b_j) \sum_{i \neq i^*} p(a_i \mid b_j) = \sum_{j=1}^{s} \sum_{i \neq i^*} p(a_i b_j)$$

- 看出，对于每个指定的b_j，后验概率择大与联合概率择大是一致的。
- 因此, 可以通过联合概率来确定译码规则并计算错误概率。

[例3]求[例2]中满足最小错误概率的译码规则。

解：由联合概率矩阵$p(a_ib_j)=\begin{pmatrix}0.15^* & 0.09 & 0.06\\ 0.08 & 0.12^* & 0.20^*\\ 0.09 & 0.09 & 0.12\end{pmatrix}$

每列选出最大者，打上*号，即知：

$$F(b_1)=a_1^*;\quad F(b_2)=a_2^*;\quad F(b_3)=a_2^*$$

不妨称它为译码规则C，它与规则A和规则B都不同。

正确概率 $P=0.15+0.12+0.20=0.47$

错误概率$P_E=1-P=0.53$

显然，译码规则C优于译码规则A与 B 。

2. 等概信源的基本准则：

- 将最大后验概率准则： $p(a_i^*|b_j) \geqslant p(a_i|b_j)$ 。
- 代入贝叶斯公式，即：

$$\frac{p(a_i^*)\,p(b_j|a_i^*)}{p(b_j)} \geqslant \frac{p(a_i)\,p(b_j|a_i)}{p(b_j)}$$

- 信源符号等概率： $p(a_i) = p(a_i^*) = 1/m$ 为常数，
- 则后验概率最大的准则变成了前向概率最大：

$$p(b_j|a_i^*) \geqslant p(b_j|a_i) \quad i=1, 2, \ldots\ldots, m$$

- 在信源符号等概率分布的前提下，可以直接由传输矩阵确定译码规则。

［例4］信源等概，信道与[例2]相同。试确定其译码规则和错误概率。

解：由传输矩阵：

$$P=\begin{pmatrix}0.5^* & 0.3^* & 0.2\\ 0.2 & 0.3 & 0.5^*\\ 0.3 & 0.3 & 0.4\end{pmatrix}$$

每列选最大者打上＊号。第二列元素相同，随便选哪列都行。比如仍选第一行的矩阵元，得到与[例3]相同的译码规则：

$$F(b_1)=a_1^*;\quad F(b_2)=a_1^*;\quad F(b_3)=a_2^*$$

正确概率 $P=\frac{1}{m}\sum_j p(b_j\mid a_i^*)=(0.5+0.3+0.5)/3=0.433$

错误概率P_E=1−P = 0.567

3. 最大似然准则：

- 有时信源并不等概，或不知道它是否等概，为了方便，也往往直接由前向传输矩阵出发，选每列的最大者来制定译码规则，称之为最大似然准则。
- 这样做，显然不能保证所设定的译码会使平均错误为最小。
- 作为一个简单实用的近似准则使用，仍然是可以的。何况很多时候实际信源与等概率信源差别并不大。

［例5］已知信源和信道的统计性质：

$$\begin{pmatrix} X \\ p(X) \end{pmatrix} = \begin{pmatrix} a_1 & a_2 & a_3 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix} \qquad P = \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.2 & 0.3 & 0.5 \\ 0.3 & 0.6 & 0.1 \end{pmatrix}$$

分别按(１)按最小错误准则(２)最大似然准则，制定译码规则并计算错误概率。

解：（1）由 $p(a_i b_j) = p(a_i)p(b_j|a_i) = \begin{pmatrix} 0.10 & 0.075 & 0.075 \\ 0.05 & 0.075 & 0.125^* \\ 0.15^* & 0.30^* & 0.05 \end{pmatrix}$

可见： $F(b_1) = a_3^*$； $F(b_2) = a_3^*$； $F(\mathrm{b}_3) = a_2^*$

$P = 0.15 + 0.3 + 0.125 = 0.575$； $P_E = 1 - P = 0.425$；

(2) 直接在信道传输矩阵每列最大者上打*

$$P=\begin{pmatrix} 0.4^* & 0.3 & 0.3 \\ 0.2 & 0.3 & 0.5^* \\ 0.3 & 0.6^* & 0.1 \end{pmatrix}$$

译码规则：$F(b_1)=a_1^*$；$F(b_2)=a_3^*$；$F(b_3)=a_2^*$

正确概率：$P=\sum_i p(a_i^*)\sum_j p(b_j \mid a_i^*)$

$= 0.4/4 + 0.5/4 + 0.6/2 = 0.525$;

错误概率：$P_E = 1-P = 0.475$;

3.2.4 错误概率与纠错能力

我们已经有了两种评判信道编码效果的方法：检、纠错能力与译码错误概率。

两种评判方法的比较

描述方式	检、纠错能力	差错控制能力
观测点	可检纠错的位数	漏检率与差错率
优点	简单、直观	科学、本质
缺点	不严谨	计算复杂

- 为什么后一种描述更科学？
- 一方面，有些编码很难用检、纠错的位数来描述其功能，比如奇偶校验码能发现奇数个多位的错误，却不能发现偶数个，哪怕两位的错误。
- 另一方面，通信的可靠程度与纠错位数的多少并无简单关系，而与具体编码方案有关。最终可观测的质量指标应当是平均错误概率。
- 因此，平均错误概率是衡量编码优劣的最终判据。“能纠正几位错码”的说法是不严格的，更科学的说法是通过信道编码，把差错率控制到什么水平之下。

差错率的计算：

- 差错率是经过信道编码仍然不能得到纠正的错误所占的比率，可见这就是平均译码错误概率。因此可用计算平均错误概率的方法计算差错率。
- 在ARQ(重发反馈方式)的检错编码中，如果认为凡是发现的错误都要重发，直至正确为止。从这个意义上讲，凡是发现的错误都得到了纠正，没有纠正的就是没有发现的错误。因此差错率=漏检率。
- 漏检率的计算，则要具体分析哪些情况的错误是该种编码发现不了的，它的概率多大。

［例6］讨论三连重复码对降低差错率的作用。

设：二元对称信道传输单符号的错误概率为：

$p=0.01$，（正确概率为 $\bar{p}=1-p=0.99$）

三连重复编码，发出的码字为：

$\alpha_0=000$，$\alpha_1=111$；

各种可能的接收符号为：

$\beta_0=000$，$\beta_1=001$，$\beta_2=010$，$\beta_3=011$，

$\beta_4=100$，$\beta_5=101$，$\beta_6=110$，$\beta_7=111$；

三次扩展信道的转移矩阵为：

$$p(\beta_j|\alpha_i)=\begin{pmatrix} \bar{p}^3 * & \bar{p}^2 p * & \bar{p}^2 p * & \bar{p}p^2 & \bar{p}^2 p * & \bar{p}p^2 & \bar{p}p^2 & p^3 \\ p^3 & \bar{p}p^2 & \bar{p}p^2 & \bar{p}^2 p * & \bar{p}p^2 & \bar{p}^2 p * & \bar{p}^2 p * & \bar{p}^3 * \end{pmatrix}$$

由于 $\bar{p} >> p$，所以矩阵中含 $\bar{p}^3$ 和 $\bar{p}^2 p$ 的矩阵元都较大，按最大似然准则选它们作为 a^*，可确定译码规则为：

$$F(\beta_0) = F(\beta_1) = F(\beta_2) = F(\beta_4) = \alpha_0$$

$$F(\beta_3) = F(\beta_5) = F(\beta_6) = F(\beta_7) = \alpha_1$$

平均错误概率为：

$$P_E = \frac{1}{m}\sum_{i \neq *}\sum_{j} p(b_j \mid a_i^*) = \frac{1}{2}(2p^3 + 6\bar{p}p^2) = p^2(1+2\bar{p}) = 2.98\times10^{-4}$$

与编码前单符号传输相比，差错率由原来的百分之一降低到万分之三。代价是传输效率为原来的三分之一。

3.2.6 信道编码定理

定理： 离散无记忆信道$[X, P(y|x), Y]$，X 是信道编码使用的符号集，Y 是接收符号集，$P(y|x)$为信道传输概率，信道容量为C。当信息传输率$R<C$时，只要编码长度n 足够大，总可以在长度为n的编码符号序列中找到$M(=2^{nR})$个许用码字和相应的译码规则，组成一个编码，使译码的错误概率P_E任意小。

逆定理： 离散无记忆信道$[X, P(y|x), Y]$，$P(y|x)$为信道传输概率，其信道容量为C。当信息传输率$R>C$时，无论码长n多大，也不可能找到一种编码，使译码错误概率任意小。

- 信道编码定理及其逆定理，合起来就是香农第二定理。
- 它指出在有噪声干扰信道中高效率高可靠地传输信息是可能的，可靠指错误概率可任意小，高效指传信率可无限接近信道容量。
- 其办法是使用有足够多冗余进行信道编码。
- 香农第二定理仅指这种编码是存在的，并未给出实现这种编码的途径。

小结：

- 译码规则
- 译码错误概率的定义与计算
- 制定译码规则的准则

  最大后验概率准则-----有最小平的均错误概率

  最大似然准则-----近等概信源的近似准则

- Shannon信道编码定理

课后复习题

- 思考题：

为什么用译码错误概率评判信道编码效果更科学？

- 作业题：

教材第114页习题三第9、13题；

第三章 信道编码

3.3 线性分组码

（第11讲　2007.11.6.）

本节的主要内容

- 线性分组码的基本概念
- 编码方法
- 校验原理
- 纠正多位错的方法
- 纠错能力的讨论

- 外语关键词：

线性分组码：linear block codes

生成矩阵：generator matrix

一致监督矩阵：parity-check matrix

校验原理：parity-check theorem

伴随子向量：syndrome vector

错误格式向量：error pattern vector

纠错能力不等式：error correcting capability inequality

[温旧引新]

- 汉明距离：两码字不同码元的个数
- 最小汉明距离=码字最小重量
- 检纠错位数与最小汉明距离的关系：

$$d_0 \geqslant e + t + 1 \quad (t<e)$$

- 冗余对检纠错的作用。
- 编码效率： $\eta = k/n$ （$r=n-k$为冗余）

3.3.1 基本概念

信道编码从结构上也有块码和流码之分。

- 块码（分组码）：

  将被编信息分组，每组信息分别各自建立与码字之间的对应关系，这种编码方式叫分组码。上节介绍过的奇偶校验码和多连重复码都属于分组码。

- 流码（序列流编码）：

  直接建立输入序列与检、纠错编码序列之间的对应关系。如，以后要讲的卷积码。

今天介绍的线性分组码，首先是分组：

- 信息序列按相同长度分组，k 位信息为一组，符号次序不变，每组后面添加 r 位冗余（称为监督位），合成来构成一个码长为 $n=k+r$ 的码字。

- 我们把码长为n、信息位为k 的分组码记作(n, k)码；如$(7, 3)$ 码 $C = (c_6c_5c_4c_3c_2c_1c_0)$，其中：

  $c_6c_5c_4$为信息位， $c_3c_2c_1c_0$为监督位。

还要求是线性：

- 线性指r个监督元是由k个信息元按一定的线性组合方式构成。比如$(7,3)$码的监督位$c_3c_2c_1c_0$与信息位$c_6c_5c_4$之间满足线性方程：

$$\begin{cases} c_3 = c_6 \oplus c_4 \\ c_2 = c_6 \oplus c_5 \oplus c_4 \\ c_1 = c_6 \oplus c_5 \\ c_0 = c_5 \oplus c_4 \end{cases}$$

信息按三位分组：$k = 3$，共有$2^3 = 8$个不同的组合方式，只要构造出8个许用码字，所有分组的编码就都有了。

由上述方程构成的8个许用码字见右表所示。

码字	信息位	监督位
C_1	000	0000
C_2	001	1101
C_3	010	0111
C_4	011	1010
C_5	100	1110
C_6	101	0011
C_7	110	1001
C_8	111	0100

例如，信息位是011，即$c_6=0, c_5=1, c_4=1$

根据方程，$c_3=c_6 \oplus c_4 =1$，$c_2=c_6 \oplus c_5 \oplus c_4 =0$，

$c_1=c_6 \oplus c_5 =1$，$c_0=c_5 \oplus c_4 =0$，监督位是1010

所以编码是011 1010

$n = 7$，长度为7的二进码共有$2^7 = 128$个,而8个许用码字只是它的一个子集。选取8个为许用码字，可以有很多种选择方案。设计不同的线性方程，就得到不同的编码。

- 思考：选择什么样的方案（线性方程）才是好方案？我们希望的许用码字应是什么样的？
- 答案：是汉明距离最大的那种编码。

3.3.2 编码方法

由信息位生成监督位，再合起来构成一个码字的过程，其实可以一步进行，直接由信息位生成相应码字：

$$\begin{pmatrix} c_6 \\ c_5 \\ c_4 \\ c_3 \\ c_2 \\ c_1 \\ c_0 \end{pmatrix} = \begin{pmatrix} c_6 \\ c_5 \\ c_4 \\ c_6 \oplus c_4 \\ c_6 \oplus c_5 \oplus c_4 \\ c_6 \oplus c_5 \\ c_5 \oplus c_4 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix} \bullet \begin{pmatrix} c_6 \\ c_5 \\ c_4 \end{pmatrix}$$

习惯上，码字写为行矢量：

$$C = (c_6 c_5 c_4 c_3 c_2 c_1 c_0)$$，

信息组亦可表示为行矢量：

$$K = (c_6 c_5 c_4) = (k_2 k_1 k_0)$$；

矩阵方程两边取转置，即得：

$$C = K \cdot G$$

其中：

$$G = \begin{pmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{pmatrix}$$

它具有 $G=[I_k \; Q]$ 的形式。

可以看到生成矩阵的三行正好是三个许用码字：

C_5=(1001110)

C_3=(0100111)

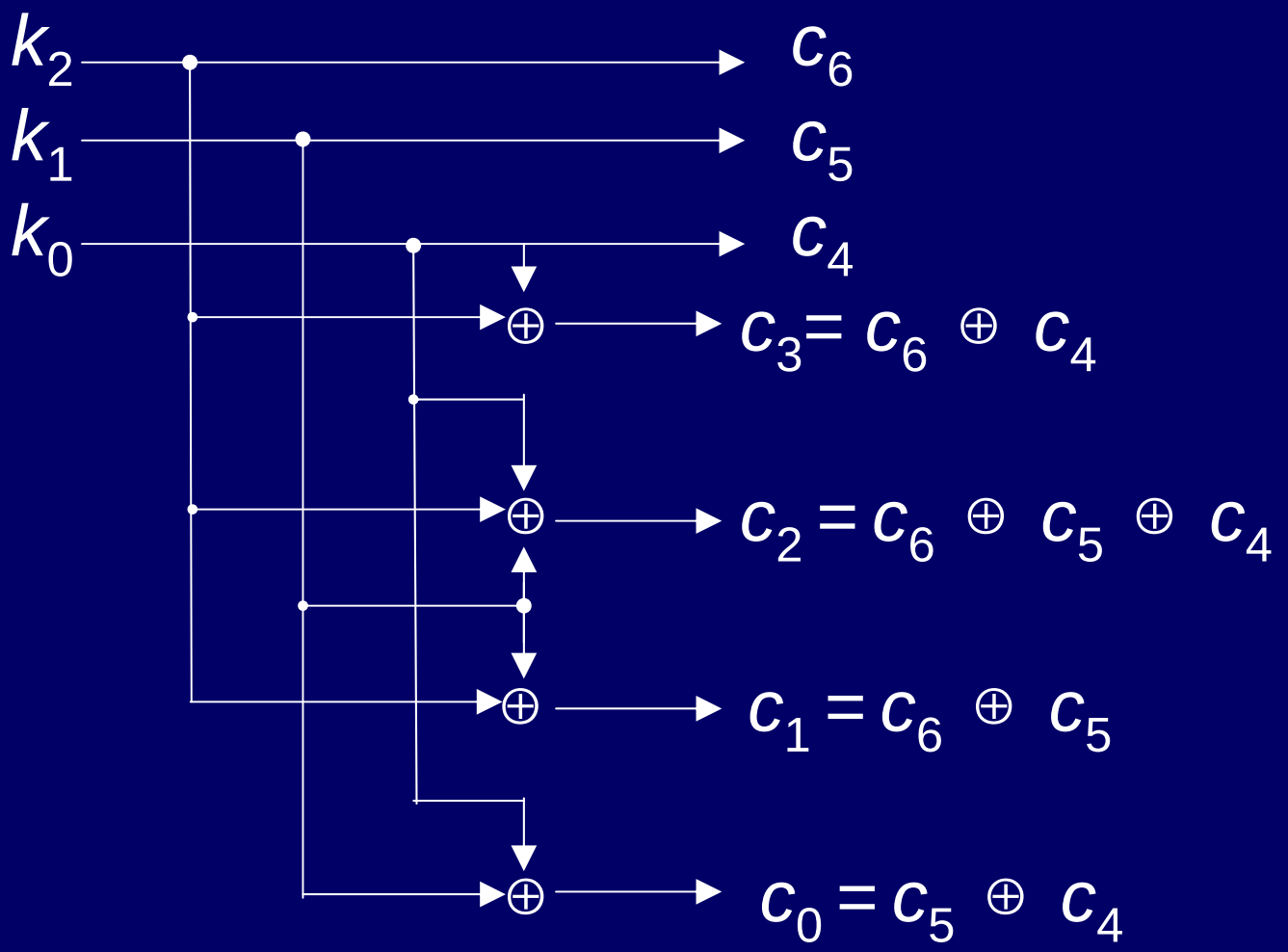
C_2=(0011101)

因此$C = K \cdot G$ 可以写成：

$$C=\begin{pmatrix}k_2 & k_1 & k_0\end{pmatrix}\cdot\begin{pmatrix}C_5\\C_3\\C_2\end{pmatrix}$$

即： $C=k_2C_5+k_1C_3+k_0C_2$

表明任意许用码字都能有这三个“生成码字”的线性组合得到。


k_2
k_1
k_0
c_6
c_5
c_4
$c_3 = c_6 \oplus c_4$
$c_2 = c_6 \oplus c_5 \oplus c_4$
$c_1 = c_6 \oplus c_5$
$c_0 = c_5 \oplus c_4$

生成(7,3)码的逻辑电路

3.3.3 校验原理

- 把输入的信息变成相应许用码字的过程叫编码，编码的核心是生成矩阵。
- 从收到的代码中恢复出原来信息的过程叫译码，译码的关键是一致校验矩阵。

1.一致校验矩阵：

$$\begin{cases} 1\cdot c_6 \oplus 0\cdot c_5 \oplus 1\cdot c_4 \oplus 1\cdot c_3 \oplus 0\cdot c_2 \oplus 0\cdot c_1 \oplus 0\cdot c_0 = 0 \\ 1\cdot c_6 \oplus 1\cdot c_5 \oplus 1\cdot c_4 \oplus 0\cdot c_3 \oplus 1\cdot c_2 \oplus 0\cdot c_1 \oplus 0\cdot c_0 = 0 \\ 1\cdot c_6 \oplus 1\cdot c_5 \oplus 0\cdot c_4 \oplus 0\cdot c_3 \oplus 0\cdot c_2 \oplus 1\cdot c_1 \oplus 0\cdot c_0 = 0 \\ 0\cdot c_6 \oplus 1\cdot c_5 \oplus 1\cdot c_4 \oplus 0\cdot c_3 \oplus 0\cdot c_2 \oplus 0\cdot c_1 \oplus 1\cdot c_0 = 0 \end{cases}$$

写成矩阵形式：$\begin{pmatrix}1&0&1&1&0&0&0\\1&1&1&0&1&0&0\\1&1&0&0&0&1&0\\0&1&1&0&0&0&1\end{pmatrix}\cdot\begin{pmatrix}c_6\\c_5\\c_4\\c_3\\c_2\\c_1\\c_0\end{pmatrix}=0$

令：$H=\begin{pmatrix}1&0&1&1&0&0&0\\1&1&1&0&1&0&0\\1&1&0&0&0&1&0\\0&1&1&0&0&0&1\end{pmatrix}=[P\ I_r]$

H叫**一致校验矩阵**. 它的左半边是r×k矩阵P，它的右半边是r×r单位方阵I，P与生成矩阵中的Q互为转置关系：

$$P=Q^T \quad 或 \quad Q=P^T$$

监督方程可表为：$H\cdot C^T=0$ 或 $C\cdot H^T=0$；

凡是满足此方程的均为正确的许用码字，否则，便是误码。

2.错误格式矢量：

为了表述误码错误所在的位置，我们构造n维错误格式矢量:

$$E=(e_{n-1}e_{n-2}\cdots\cdots e_1e_0);$$

如n=7时，若c_2位有错，则写E=(0000100)；

若c_0 和c_4两位错，则写 E = (0010001)；

设发送入信道的码字为$C = (c_6c_5c_4c_3c_2c_1c_0)$，接收端收到的码字为$R = (r_6\, r_5\, r_4\, r_3\, r_2\, r_1\, r_0)$；

不难知道三者的关系是：

$$E = C \oplus R; \quad R = C \oplus E; \quad C = R \oplus E;$$

3.伴随子向量：

定义伴随子向量： $S = R \cdot H^T$

每收到一个码字R，都可由一致校验矩阵H计算出对应的伴随子向量S。

（1）若无错，R=C，则 $S = R \cdot H^T = C \cdot H^T = 0$

（2）若有错，R≠C，

$\because S = R \cdot H^T = (C \oplus E) \cdot H^T = C \cdot H^T \oplus E \cdot H^T = E \cdot H^T$;

$\therefore S = R \cdot H^T = E \cdot H^T \neq 0$

现在的问题是由S来计算E，知道了E，就知道了错误位置，再由：$C = R \oplus E$ 来纠错。

- 又能分两种情况：只有一位错和多位错。这里先讨论只有一位错的情况。
- 当接收代码R只有一位(第i位)错时，E中只有第 i 位为1，其它均为0，$E \cdot H^T$的乘积，就是H^T矩阵中第 i 行元素构成的行矢量。
- 现在既然$S = E \cdot H^T$，就是把算出的伴随子S与H^T比较，看S与H^T的哪一行相同，就表明错在哪一位。

- 如：$\boldsymbol{R}$ = (1101000)；

  不难求出 $\boldsymbol{S} = \boldsymbol{R} \cdot \boldsymbol{H}^{\mathrm{T}}$ = (0001)≠0，

  容易看到，(0001)是$\boldsymbol{H}$是矩阵的第7列（即 $\boldsymbol{H}^{\mathrm{T}}$矩阵的第7行），表明错在第7位，即知错误格式矢量为：

  $\boldsymbol{E}$ =(0000001)；

  纠错后为：$\boldsymbol{C} = \boldsymbol{R} \oplus \boldsymbol{E}$ = (1101001)；

[例1]已知某线性分组码的生成矩阵：

$$G=\begin{pmatrix}1&0&0&0&1&0&1\\0&1&0&0&1&1&1\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{pmatrix}$$

（思考：它的码长几位？信息几位？）

1. 求它的所有许用码字。
2. 求它的一致监督矩阵。
3. 若收到码字为**(1011010)**，请译码。

(1)由于n=7, k=4, r=3,有2^4=16个许用码字，利用：

生成方程 $C=K\cdot G$容易求出，他们是：

序号	信息	许用码字	序号	信息	许用码字
C_0	0000	（0000 000)	C_8	1000	（1000 101）
C_1	0001	（0001 011)	C_9	1001	（1001 110）
C_2	0010	（0010 110）	C_{10}	1010	（1010 011）
C_3	0011	（0011 101）	C_{11}	1011	（1011 000）
C_4	0100	（0100 111)	C_{12}	1100	（1100 010）
C_5	0101	（0101 100）	C_{13}	1101	（1101 001）
C_6	0110	（0110 001）	C_{14}	1110	（1110 100）
C_7	0111	（0111 010）	C_{15}	1111	（1111 111）

(2)一致校验矩阵：

$$H=\begin{pmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\1&1&0&1&0&0&1\end{pmatrix}\qquad H^T=\begin{pmatrix}1&0&1\\1&1&1\\1&1&0\\0&1&1\\1&0&0\\0&1&0\\0&0&1\end{pmatrix}$$

(3)伴随子向量：

$$S=R\cdot H^T=(1011010)\cdot\begin{pmatrix}1&0&1\\1&1&1\\1&1&0\\0&1&1\\1&0&0\\0&1&0\\0&0&1\end{pmatrix}=(010)$$

$\therefore$ E=(0000010);　C=R $\oplus$ E=(1011000)

4. 译码过程的电路实现

（7，4）码的一致监督矩阵为：

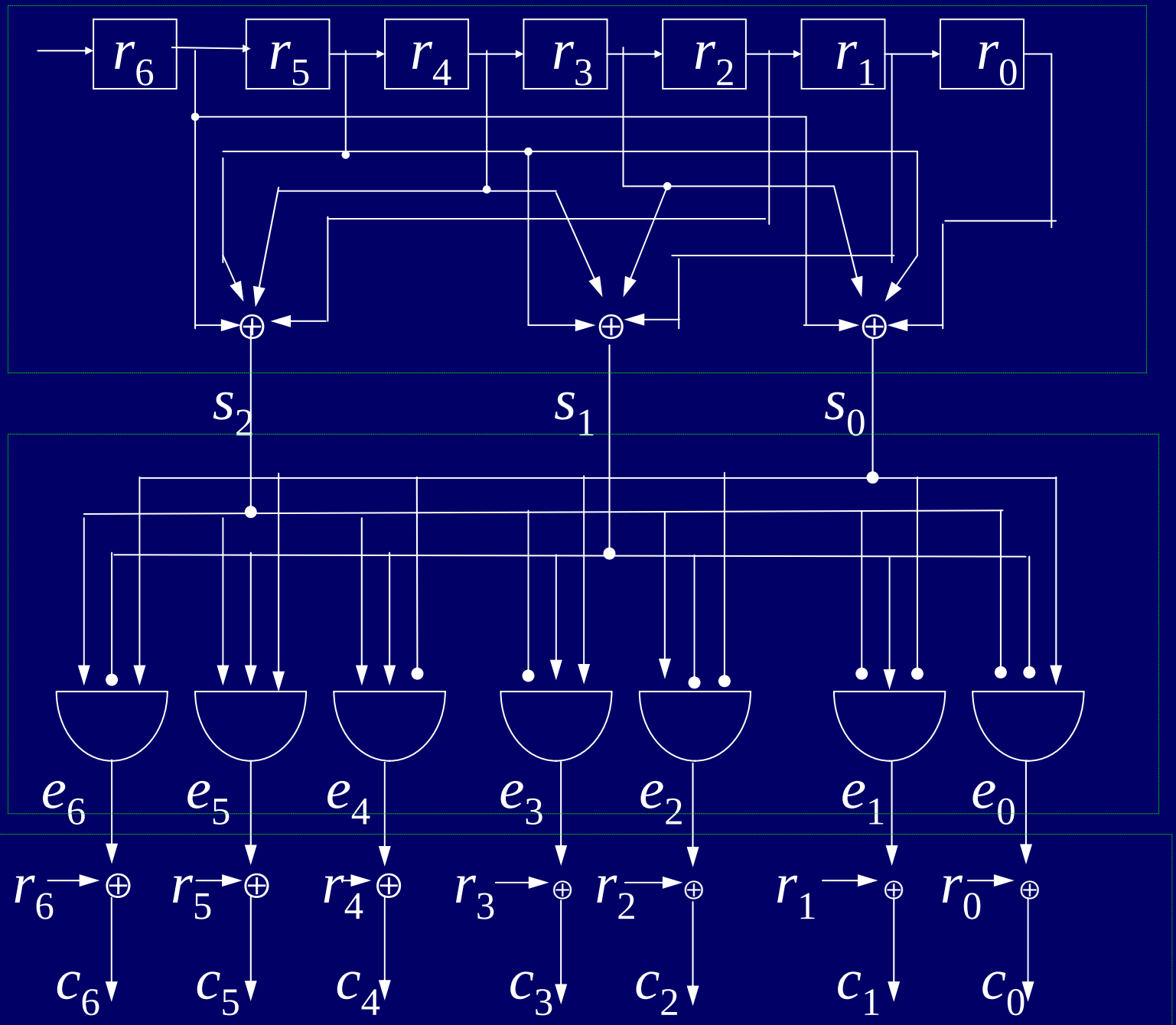
$$H=\begin{pmatrix}1&1&1&0&1&0&0\\0&1&1&1&0&1&0\\1&1&0&1&0&0&1\end{pmatrix}$$

伴随子$S = R\cdot H^{T}$为一行三列的向量$S=(s_2 s_1 s_0)$；

$$s_2=r_6\oplus r_5\oplus r_4\oplus r_2;$$

$$s_1=r_5\oplus r_4\oplus r_3\oplus r_1;$$

$$s_0=r_6\oplus r_5\oplus r_3\oplus r_0;$$



(7,4)码的译码电路原理图

- 上面部分是计算伴随子$S=(s_2s_1s_0)$的逻辑电路；
- 中间七个与门电路构成计算错误格式：
  $E=(e_6e_5e_4e_3e_2e_1e_0)$的电路，
  当S与H的第一列相同时，即$(s_2s_1s_0)=101$时，
  $E=(1000000)$，故有$e_6=s_2\wedge\underline{s}_1\wedge s_0$；
  同理可知：
  $e_5=s_2\wedge s_1\wedge s_0$　$e_4=s_2\wedge s_1\wedge\underline{s}_0$；　$e_3=\underline{s}_2\wedge s_1\wedge s_0$；
  $e_2=s_2\wedge\underline{s}_1\wedge\underline{s}_0$　$e_1=\underline{s}_2\wedge s_1\wedge\underline{s}_0$；　$e_0=\underline{s}_2\wedge\underline{s}_1\wedge s_0$；
- 下面的七个模二加构成纠错电路，E与R逐位模二加，哪位错了便将哪位改回来。

[例2] 码长为4的偶校验码是线性分组码吗？若是，写出生成矩阵和全部许用码字，讨论纠错能力，计算漏检率。

解：设 $C=(c_3, c_2, c_1, c_0)$；

则 c_3, c_2, c_1,为信息位， c_0为监督位。

监督关系为： $c_0=c_3 \oplus c_2 \oplus c_1$

可见它是（4,3）线性分组码。

$n=4, k=3, r=1$

8个许用码字和生成矩阵为：$C = K \cdot G =$

$$G=\begin{pmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\end{pmatrix}$$

$$\begin{cases}000 & 0\\001 & 1\\010 & 1\\011 & 0\\100 & 1\\101 & 0\\110 & 0\\111 & 1\end{cases}$$

一致校验矩阵：

$H=[P\ I_r]$=（1111）；　$H^T=\begin{pmatrix}1\\1\\1\\1\end{pmatrix}$

伴随子向量：　$S=R\cdot H^T$一定是1行1列的，即0或1；

S=0表示无错，　S=1表示有错。

虽然可以检查到错误，但是不能纠正错误，因为H^T的4行都是1，各行都与 S 一样，无法判定错在哪一位。

实际上，　S 只有一种状态，无法区别不同位置发生的错误。

对于多位错的情况，奇偶校验码不能发现同时发生偶数位的错误。如同时2位错及如同时4位错的情况。

设:信道中单个符号的错误概率为p

则: 同时2位错的概率是:

$$P1=C42p2(1-p)\ 2 = 6p2(1-p)$$

同时4位错的概率是:

$$P2= C44p4(1-p)\ 0 = p4$$

因此 (4,3)码的漏检率为：$P1+P2=6p2(1-p)+p4$

3.3.4 纠正多位错的方法

- 接收码字$\boldsymbol{R}$发生一位错时，因为错误格式$\boldsymbol{E}$与$\boldsymbol{R}$是一一对应的，$\boldsymbol{R}$与S也是一一对应的，观察S是H^{T}的那一行就能确定$\boldsymbol{E}$，译码不存在多义性。
- 但当$\boldsymbol{R}$发生两位或多位错时，E 中不止含有一个1，由方程$S=E\cdot H^{\mathrm{T}}$知，$\boldsymbol{S}$应等于H^{T}中若干行的模二加。这样一来，不同的错误格式$\boldsymbol{E}$可能算出相同的伴随子S。比如，对于

$$H=\begin{pmatrix}1&0&1&1&0&0&0\\1&1&1&0&1&0&0\\1&1&0&0&0&1&0\\0&1&1&0&0&0&1\end{pmatrix}$$

$E_1=(1100000)$、$E_2=(0001001)$和 $E_3=(0110011)$都能得到$S=(1001)$；根据接收码字$\boldsymbol{R}$算出的$\boldsymbol{S}=(1001)$来译码时，很难唯一地判定错误格式究竟是哪一个E。

鉴于此，实用中多是采用查表法来译码。

预先对每一个许用码字C_i写出它在纠错能力之内的各种可能的接收码字$R_j(C_i)$，按无错、一位错、二位错、……、t位错的次序将相应的R排序，在C_i下面，编制出一个R-C对照表。

	错误格式	正、误码对照	
无错	（000）	（000）	（111）
一位错	（001）	（001）	（110）
	（010）	（010）	（101）
	（100）	（100）	（011）
二位错	（011）	（011）	（100）
	（110）	（110）	（001）
	（101）	（101）	（010）

[例3] 证明五连重复码是(5,1)线性分组码。并写出生成矩阵、一致校验矩阵，错误格式向量和伴随子向量。

解：五连重复码只有2个许用码字：00000 和 11111

视首位为信息，后4位为监督，即k=1, r =4, n=5；

监督位c_3 ,c_2 , c_1与信息位c_4之间有线性关系：

$$c_3 = c_4\,;\quad c_2 = c_4\,;\quad c_1 = c_4\,;\quad c_0 = c_4$$

连同$c_4 = c_4$ 五个方程写成矩阵形式：

$$\begin{pmatrix} c_4 \\ c_3 \\ c_2 \\ c_1 \\ c_0 \end{pmatrix} = \begin{pmatrix} c_4 \\ c_4 \\ c_4 \\ c_4 \\ c_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \bullet \begin{pmatrix} c_4 \end{pmatrix}$$

- 转置得到生成矩阵：G=(1 1111)，

  由G写出一致监督矩阵是：

$$H=\begin{pmatrix}1&1&0&0&0\\1&0&1&0&0\\1&0&0&1&0\\1&0&0&0&1\end{pmatrix}\qquad H^T=\begin{pmatrix}1&1&1&1\\1&0&0&0\\0&1&0&0\\0&0&1&0\\0&0&0&1\end{pmatrix}$$

- 发生1位错的格式及相应的伴随子是C_5^1=5个：

  E_1=(10000);　S_1=(1111);

  E_2=(01000);　S_2=(1000);

  E_3=(00100);　S_3=(0100);

  E_4=(00010);　S_4=(0010);

  E_5=(00001);　S_5=(0001);

- 发生2位错的格式及相应的伴随子是$C_5^2=10$个：

$E_6=(11000)$; $S_6=(0111)$;

$E_7=(10100)$; $S_7=(1011)$;

$E_8=(10010)$; $S_8=(1101)$;

$E_9=(10001)$; $S_9=(1110)$;

$E_{10}=(01100)$; $S_{10}=(1100)$;

$E_{11}=(01010)$; $S_{11}=(1010)$;

$E_{12}=(01001)$; $S_{12}=(1001)$;

$E_{13}=(00110)$; $S_{13}=(0110)$;

$E_{14}=(00101)$; $S_{14}=(0101)$;

$E_{15}=(00011)$; $S_{15}=(0011)$;

- (5,1)码是能纠正两位错误的完备码。纠错能力以内的误码共30个，它们是互不相同的，加上2个正确码字，正好是全部码字的总数：$30+2=32=2^5$ 个。
- 将它们列成表，就能从接收到的码字直接查表进行译码。

	错误格式E	码字C_1	码字C_2
无错C	(00000）	(00000)	(11111)
一位错	(10000）	(10000)	(01111)
一位错	(01000）	(01000)	(10111)
一位错	(00100)	(00100)	(11011)
一位错	(00010）	(00010)	(11101)
一位错	(00001)	(00001)	(11110)
两位错	(11000)	(11000)	(00111)
两位错	(10100)	(10100)	(01011)
两位错	(10010)	(10010)	(01101)
两位错	(10001)	(10001)	(01110)
两位错	(01100)	(01100)	(10011)
两位错	(01010)	(01010)	(10101)
两位错	(01001)	(01001)	(10110)
两位错	(00110)	(00110)	(11001)
两位错	(00101)	(00101)	(11010)
两位错	(00011)	(00011)	(11100)

- 首先，即使在一般情况下，也不难证明，在t位误码的范围内列表，所有列出的误码是互不相同的，因而错误格式E和误码R是一一对应的。

这是因为纠错能力为 t，最小码距则为$2t+1$，用含有t 个1的格式向量E与所有许用码字C模二加，得到的所有误码R之间，必定至少还有一位不同，于是 t 位误码范围内列出的对照表中，所有的误码都是各不相同的。因此由接收到的误码查找相应的许用码字是一一对应的。

- 其次，把多于t位的错误所造成的误码也追加在表的后面，是无益的。因为它所造成的误码，在不大于t位的误码表格中已经出现过，按照译码规则，它会按照较少错位的情况来译码。多于t位的误码是无法纠正的，它将被错误译码。

比如，**(5,1)**码的纠错能力是**2**位，**C=(00000)**错**2**位，与**C=(11111)**错**3**位，都能得到**R=(01010)**。

按照译码规则它被译为**(00000)**。 所以错3位的情况不被列入表中。

- 若收到一个码字是**R=(01010)**

  先计算伴随子**$S=R\cdot H^T=(1010)$**

  表中它唯一对应着错误格式**$E_{11}=(01010)$**

  于是译码为：**C= E ⊕ R=(00000)**

- 如果发生**3**位错**E=(10101)**，会使**C=(11111)**也变为**R=(01010)**；

  但它已超出纠错能力，译码时仍会译为：

  C=(00000)；造成错译。

3.3.5 纠错能力的讨论

1.最小码距与监督位数的关系：

$$d_{min} \leqslant r+1$$

证明：设 C_{min}为重量最小的非零码字，它的重量为d_{min}

因为C_{min}中的非零码元是d_{min}个，$C_{min} \cdot H^T$相乘实际上就是H^T中对应于C_{min}非零码元的那几行的逐位模二加，而这样的行数是d_{min}个。监督方程要求$C_{min} \cdot H^T=0$，意味着H^T中d_{min}行相加的结果等于零，因此H^T中线性无关的行最多不超过$(d_{min}-1)$个。

H^T是n行r列的矩阵，因为$n > r$，H^T的秩至多为r，可见，$(d_{min}-1) \leqslant r$，即$d_{min} \leqslant r+1$

2. 纠错位数不等式

- H^T是n行r 列的矩阵，所以$S = R \cdot H^T$是一行r列的向量，每位只能取0或1，它共有2^r个不同的取值，扣除了S=0的一个无错状态，用它来区分不同的错误格式，最多不超过$2^r - 1$个。

- 另一方面，码长为n 的码字，发生一位错码的方式有C_N^1种，同时发生二位错的方式有C_N^2 种……，发生t位错的方式有 C_N^t种。

- 为了能用伴随子S区分 t位以内的所有错误格式，

应当有： $2^r - 1 \geqslant C_n^1 + C_n^2 + \cdots\cdots + C_n^t$

或： $2^r \geqslant C_n^0 + C_n^1 + C_n^2 + \cdots\cdots + C_n^t$

- 下面以$n=7$的线性分组码为例进行讨论：

$k=6$时：　$r=1$，$2^r=2>C_7^0=1$，　　$t=0$

$k=5$时：　$r=2$，$2^r=4>C_7^0=1$，　　$t=0$

$k=4$时：　$r=3$，$2^r=8=C_7^0+C_7^1=1+7=8$，　$t=1$

$k=3$时：　$r=4$，$2^r=16>C_7^0+C_7^1=1+7=8$，　$t=1$

$k=2$时：　$r=5$，$2^r=32>C_7^0+C_7^1+C_7^2$

$=1+7+21=29$，　　$t=2$

$k=1$时：　$r=6$，$2^r=64=C_7^0+C_7^1+C_7^2$

$+C_7^3=1+7+21+35=64$，　　$t=3$

3. 完备码：

纠错不等式中，满足取等号的线性分组码称为完备码，如 **(7, 4)**码和**(7,1)**码都是完备码。在相同的纠错能力下，完备码的冗余最少，效率最高；

4. 汉明（Hamming）码：

能纠正一位错的完备码叫做汉明码，它满足：

$$2^r = 1 + C_n^1 = 1 + n;$$

或：　$n = 2^r - 1;$

汉明码举例：

- $r = 2$时：$n = 2^r-1=3$，　构成（3, 1）汉明码；
- $r = 3$时：$n = 2^r-1=7$，　构成（7, 4）汉明码；
- $r = 4$时：$n = 2^r-1=15$，构成（15, 11）汉明码；

[例4] 试构造（3, 1）汉明码。

因为$n = 3$，$k=1$, 信息只有1位（0或1）；许用码字只有两个。考虑到必存在一个全零码（000）；所以为了码字间距尽量大，另一个码字应是（111），此即三连重复码。它可纠正一位错。

[例5] 二元对称信道单个码元的错传概率为 $p=0.01$，试讨论（7, 4）汉明码对减小差错率的作用。

（1）编码前：每个（7, 4）码字中含4位信息，4位都正确的概率为$(1-p)^4$;差错率为：

$$1-(1-p)^4=1-0.99^4=0.04$$

（2）编码后：(7, 4）码能纠正1位错。

7位码元中1位错6位对的概率是：$7p(1-p)^6$;

7位全对的概率是$(1-p)^7$,

所以差错率为：$1-(1-p)^7-7p(1-p)^6$

$$=1-0.99^7-7\times0.01\times0.99^6=0.002$$

小结：

※线性分组码(n,k)的编码方法：

k行n列的生成矩阵----- $G=(I_k P)$

生成方程-----$C=K \cdot G$

※线性分组码的译码方法：

n行r列的一致监督矩阵------$H^T=\begin{pmatrix} P \\ I_r \end{pmatrix}$

1行r列的随子向量------$S=RH^T$是否等于零

错误格式向量非0列位置由S在H^T中的行号决定

◆纠错位数不等式 $2^r \geq \sum_{i=0}^{t} C_n^i$

- 思考题：

  1.(3,2)奇偶校验码是不是线性分组码？如果是，它的生成矩阵和一致监督矩阵是什么？

  2.构造线性分组码生成矩阵所追求的目标是什么?

- 作业题：

  P113页 2，3，5 题。