第六章 有噪信道编码

- 6.1 信道编码的相关概念
- 6.2 有噪信道编码定理
- 6.3 纠错编码
- 6.4 线性分组码

一、信道编码的相关概念

1. 信道的问题
2. 信道编码
3. 错误概率和译码规则
4. 错误概率与编码方法
5. 线性码和码的距离

- 所谓信道编码，就是按一定的规则给信息序列增加一些多余的码元，使不具有规律性的信息序列变为具有某种规律性的信道码字序列X，也就是说码字序列X的码元之间是相关的。在接收端，信道译码器利用这种相关性(也就是已知的编码规则)来译码，检验接收到的码字序列Y中是否有错，并且纠正其中的差错。

- 根据相关性来检测和纠正传输过程中产生的差错就是信道编码的基本思想。

在有噪信道中，传输信息发生错误的**错误概率**与什么有关?

1、信道的统计特性

2、**编码方法**和**译码规则**

❖ 我们已经知道，错误概率与信道的统计特性有关。信道的统计特性可由信道的传递矩阵来表示，由信道矩阵可以求出错误概率。例如在图6.2的二元对称信道中，单个符号的错误传递概率是p，单个符号的正确传递概率为 $\bar{p}=1-p$ ，

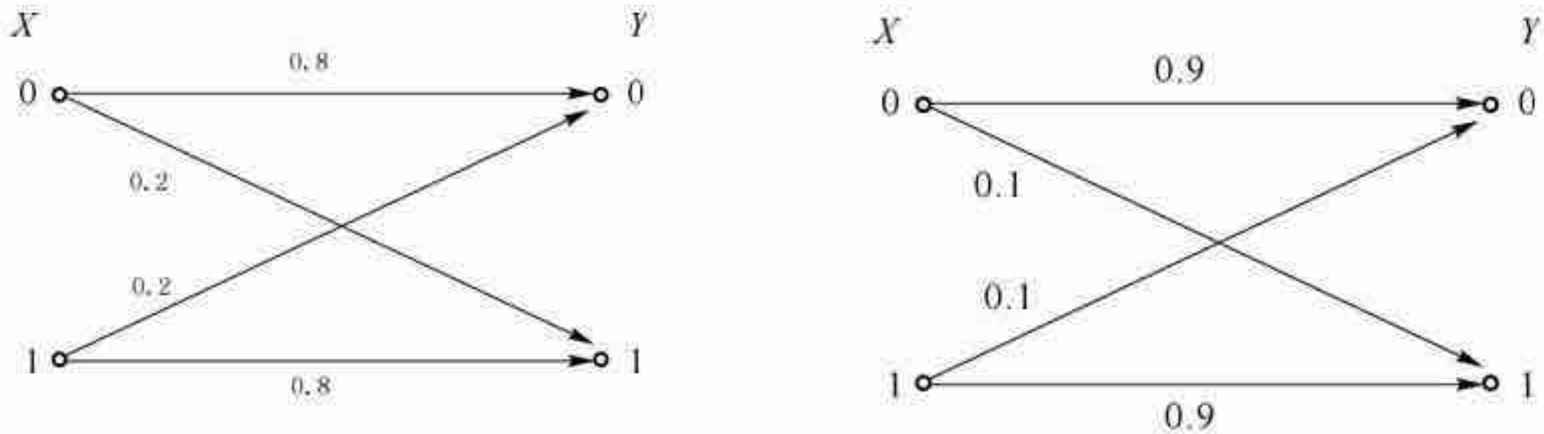
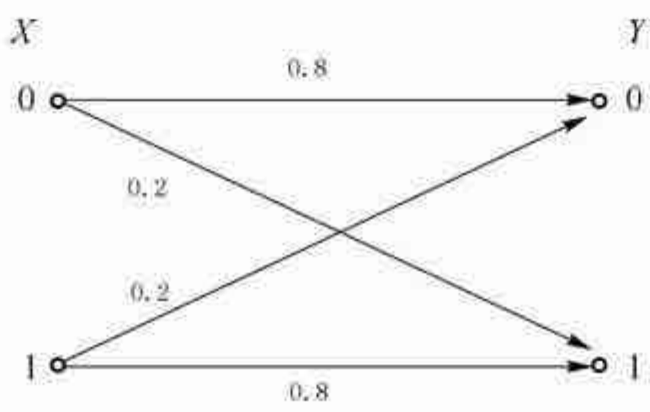


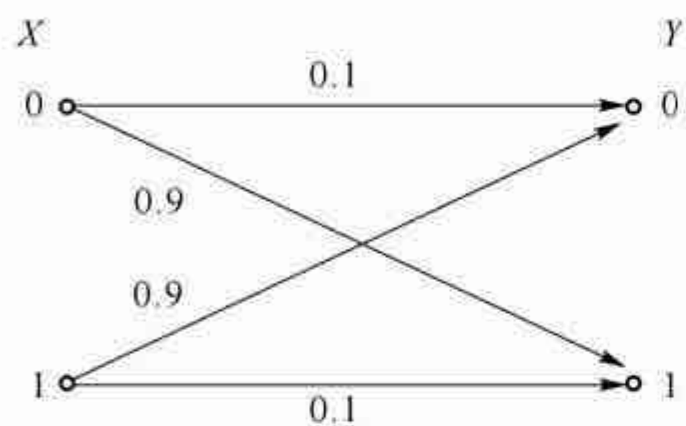
图6.2 二元对称信道

因此错误概率与信道的统计特征有关。

译码过程和译码规则

- 但是通信的过程并不是信息传输到信道输出端就结束了，
- 还要经过译码过程才能到达信宿，所以译码过程和译码规则对系统的错误概率影响很大。





译码规则的定义

定义 6.1 设信道的输入符号集 $X=\{x_i, i=1,2,\cdots,r\}$，输出符号集 $Y=\{y_j, j=1,2,\cdots,s\}$，若对每一个输出符号 y_j，都有一个确定的函数 $F(y_j)$，使 y_j 对应于唯一的一个输入符号 x_i，则称这样的函数为译码规则，记为

$$F(y_j)=x_i \qquad i=1,2,\cdots,r;j=1,2,\cdots,s \tag{6.1}$$

对于有 r 个输入，s 个输出的信道而言，输出 y_j 可以对应 r 个输入中的任何一个，所以译码规则共有 r^s 种.

【例 6.2】

设有一信道，信道矩阵为

$$P=\begin{bmatrix}0.5 & 0.3 & 0.2\\ 0.2 & 0.3 & 0.5\\ 0.3 & 0.3 & 0.4\end{bmatrix}$$

根据此信道矩阵，可以设计一个译码规则如下：

$$A:\begin{cases}F(y_1)=x_1\\ F(y_2)=x_2\\ F(y_3)=x_3\end{cases}$$

也可以设计另一个译码规则

哪一种好？为什么？

$$B:\begin{cases}F(y_1)=x_1\\ F(y_2)=x_3\\ F(y_3)=x_2\end{cases}$$

由于 $r=3$, $s=3$，总共可以设计出 $r^s=27$ 种译码规则，应该怎样选择译码规则呢？一个很自然的准则就是使平均错误概率最小.

错误概率的定义

- 在确定译码规则 $F(y_j)=x_i$ 后，若信道输出端接收到符号 y_j，则一定译成 x_i。如果发送端发送的确实就是 x_i，就是正确译码；反之，若发送端发送的不是 x_i 就认为是错误译码。

- 于是收到符号 y_j 条件下，译码的条件正确概率为：

$$p[F(y_j)|y_j] = p(x_i|y_j)$$

- 而条件错误概率：

$$p(e|y_j) = 1 - p(x_i|y_j) = 1 - p[F(y_j)|y_j]$$

- 译码后的平均错误概率 P_E 是条件错误概率 $p(e|y_j)$ 对Y空间取平均值，即

$$P_E = E[p(e|y_j)] = \sum_{j=1}^{s} p(y_i)p(e|y_j)$$

- 它表示经过译码后平均接收到一个符号所产生的错误大小。

译码规则的选择

如何设计译码规则F（y_j）=x_i使P_E最小呢？

$$P_E = E[p(e|y_j)] = \sum_{j=1}^{s} p(y_j)p(e|y_j)$$

可以选择译码规则使每一项为最小，则所得 P_E 为最小．因为 $p(y_j)$与译码规则无关，所以只要设计译码规则 $F(y_j)=x_i$ 使条件错误概率$p(e|y_j)$最小，也就是要选择$p[F(y_j)|y_j]$最大．这就是最大后验概率准则．

最大后验概率译码规则

- **定义6.2** 选择译码函数 $F(y_j)=x^*$，使之满足条件

$$p(x^*|y_j)\ge p(x_i|y_j) \qquad \forall i, x^*\in X$$

- 则称为**最大后验概率译码规则**，又称为**最小错误概率准则**、**最优译码**、**最佳译码**、**理想观测者规则**。

因为一般是已知信道的前向概率 $p(y_j|x_i)$和输入符号的先验概率 $p(x_i)$,

所以根据贝叶斯定律,式(6.5)又可写成

$$\boxed{\frac{p(y_j|x^*)p(x^*)}{p(y_j)} \geqslant \frac{p(y_j|x_i)p(x_i)}{p(y_j)}}$$

这样,最大后验概率译码规则就可表示为:选择译码函数 $F(y_j)=x^*, x^* \in X$, $y_j \in Y$,使其满足 $p(y_j|x^*)p(x^*) \geqslant p(y_j|x_i)p(x_i), x_i \in X$.

极大似然译码准则下的错误概率

当输入符号的先验概率 $p(x_i)$ 相等时，上式又可写成 $p(y_j|x^*) \geq p(y_j|x_i)$，因此又定义了一个极大似然译码规则.

如果输入为等概分布，即 $p(x_i)=\frac{1}{r}$，则

$$P_E=\frac{1}{r}\sum_{Y,X-x^*} p(y_j|x_i) \tag{6.11}$$

式(6.11)表明，在输入等概的情况下，译码错误概率可用信道矩阵中的元素 $p(y_j|x_i)$ 的求和来表示，求和是除去每列中对应于 $F(y_j)=x^*$ 的那一项后，矩阵中其余元素之和.

极大似然译码的特点

当输入符号等概时，这两个译码规是等价的，根据极大似然译码准则可以直接从信道矩阵的传递概率中去选定译码函数：当收到 y_j 后，译成信道矩阵 $\boldsymbol{P}$ 第 j 列中最大的转移概率所对应的 x_i.

极大似然译码准则本身不依赖于先验概率 $p(x_i)$，当先验概率为等概率分布时，它与最大后验概率译码准则是等价的，可以使平均错误概率 P_E 达到最小. 如果先验概率不相等或不知道时，仍可以采用这个准则，但不一定能使 P_E 最小.

设有一信道，信道矩阵为

$$P=\begin{bmatrix}0.5 & 0.3 & 0.2\\ 0.2 & 0.3 & 0.5\\ 0.3 & 0.3 & 0.4\end{bmatrix}$$

- 分别求

  输入等概分布

  输入分布为 $p(x_1)=1/4,,p(x_2)=1/4,p(x_3)=1/2,$

时最大后验概率译码规则和极大似然译码规则得出来的译码结果。

【例 6.2】

设有一信道，信道矩阵为　**两种译码规则**

$$P=\begin{bmatrix}0.5 & 0.3 & 0.2\\0.2 & 0.3 & 0.5\\0.3 & 0.3 & 0.4\end{bmatrix}\qquad A:\begin{cases}F(y_1)=x_1\\F(y_2)=x_2\\F(y_3)=x_3\end{cases}\qquad B:\begin{cases}F(y_1)=x_1\\F(y_2)=x_3\\F(y_3)=x_2\end{cases}$$

解　当输入为等概分布时，译码规则 B 就是极大似然译码规则，两种译码规则所对应的平均错误概率分别为

$$P_E(A)=\frac{1}{3}\sum_{Y,X-x^*}p(y_j\mid x_i)$$

$$=\frac{1}{3}\times[(0.2+0.3)+(0.3+0.3)+(0.2+0.5)]=0.6$$

$$P_E(B)=\frac{1}{3}\sum_{Y,X-x^*}p(y_j\mid x_i)$$

$$=\frac{1}{3}\times[(0.2+0.3)+(0.3+0.3)+(0.2+0.4)]=0.5667$$

所以在输入为等概分布时，极大似然译码规则可使信道平均错误概率最小。

计算最小错误概率准则下的错误概率

当输入为不等概分布时，假设某个输入概率分布 $p(x_1)=1/4, , p(x_2)=1/4, p(x_3)=1/2$，则因为极大似然译码规则与输入概率分布无关，所以与上例相同，

$$P_E(B)=\frac{1}{3}\sum_{Y,X-x^*} p(y_j|x_i)=\frac{1}{3}\times[(0.2+0.3)+(0.3+0.3)+(0.2+0.4)]=0.5667$$

如何计算最小错误概率准则下的错误概率？

由其联合概率矩阵

$$\boldsymbol{P}(XY)=\begin{bmatrix} 0.125 & 0.075 & 0.050 \\ 0.050 & 0.075 & 0.125 \\ 0.150 & 0.150 & 0.200 \end{bmatrix}$$

可得译码函数

$$C:\begin{cases} F(y_1)=x_3 \\ F(y_2)=x_3 \\ F(y_3)=x_3 \end{cases}$$

所以此时错误概率为 **公式(6.8)**

$$P_E(C)=[(0.125+0.075+0.050)+(0.050+0.075+0.125)]=0.5$$

所以输入不是等概分布时最大似然译码准则的平均错误概率不是最小.

4、错误概率与编码方法

在有噪信道中，传输信息发生错误的错误概率与什么有关?

1、信道的统计特性

2、编码方法和译码规则

错误概率与编码方法的关系

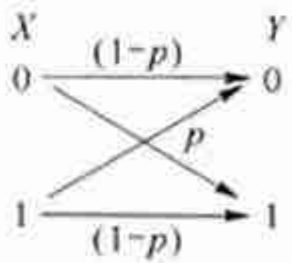


图 6.4　二元对称信道

设有二元对称信道如图 6.4 所示. 相应的信道矩阵为

$$P=\begin{bmatrix}0.99 & 0.01\\ 0.01 & 0.99\end{bmatrix}$$

选择最佳译码规则为

$$\begin{cases}F(y_1)=x_1\\ F(y_2)=x_2\end{cases}$$

总的平均错误概率在输入分布为等概的条件下为

$$P_E=\frac{1}{r}\sum_{Y,X-x^*} p(y|x)=\frac{1}{2}\times(0.01+0.01)=10^{-2}$$

对于一般数字通信系统，这个错误概率是非常大的，一般数字通信要求错误概率在 $10^{-6}\sim10^{-9}$ 的范围内，有的甚至要求更低的错误概率.

错误概率与编码方法的关系

上述例题中已经采用最佳译码方法了，那采用什么编码方法可以降低错误概率了？

实际经验告诉我们：只要在发送端把消息重复发几遍，就可以使接收端接收消息时错误减小，从而提高通信的可靠性。

例如，发信源符号"0"时，重复发送 3 个"0"，发"1"时，重复发送 3 个"1"，这可以看成离散无记忆信道的 3 次扩展信道. 在信道输入端有两个码字"000"和"111"，在输出端由于信道干扰，各个码元都可能发生错误，则有 8 个可能的输出序列. 输入是 3 次扩展信道的 8 个可能出现的二元序列中选两个作为消息，而输出端这 8 个可能的二元序列都是接收序列，如表 6.1 所示.

所用到的编码方法

表 6.1　简单重复编码举例

输入状态	没有使用的码字	用作消息的码字	输出端接收序列	输出状态
x_1		000	000	y_1
	001		001	y_2
	010		010	y_3
	011		011	y_4
	100		100	y_5
	101		101	y_6
	110		110	y_7
x_2		111	111	y_8

如何译码？错误概率为多少？

这时信道矩阵为

$$\begin{array}{cc} & \begin{array}{cccccccc} 000 & 001 & 010 & 011 & 100 & 101 & 110 & 111 \end{array} \\ \begin{array}{c} 000 \\ 111 \end{array} & \mathbf{P} = \begin{bmatrix} \bar{p}^3 & \bar{p}^2 p & \bar{p}^2 p & \bar{p} p^2 & \bar{p}^2 p & \bar{p} p^2 & \bar{p} p^2 & p^3 \\ p^3 & \bar{p} p^2 & \bar{p} p^2 & \bar{p}^2 p & \bar{p} p^2 & \bar{p}^2 p & \bar{p}^2 p & \bar{p}^3 \end{bmatrix} \end{array}$$

假设输入符号为等概分布，采用极大似然译码规则，即取信道矩阵中每列数值最大的元素所对应的译码函数为

$$\begin{cases} F(y_1)=x_1 \\ F(y_2)=x_1 \\ F(y_3)=x_1 \\ F(y_4)=x_2 \\ F(y_5)=x_1 \\ F(y_6)=x_2 \\ F(y_7)=x_2 \\ F(y_8)=x_2 \end{cases}$$

$$\boldsymbol{P}=\begin{bmatrix} 0.99 & 0.01 \\ 0.01 & 0.99 \end{bmatrix}$$

译码后的平均错误概率为

$$\boxed{P_E}=\sum_{Y,X-x^*} p(y|x)p(x)$$

$$=\frac{1}{M}\sum_{Y,X-x^*} p(y|x)$$

$$=\frac{1}{2}(p^3+\bar{p}p^2+\bar{p}p^2+\bar{p}p^2+\bar{p}p^2+\bar{p}p^2+\bar{p}p^2+p^3)$$

$$=p^3+3\bar{p}p^2$$

$$\approx 3\times 10^{-4}$$

未采用该编码之前的错误概率？

$$P_E=\frac{1}{r}\sum_{Y,X-x^*} p(y|x)=\frac{1}{2}\times(0.01+0.01)=10^{-2}$$

刚才采用的是：极大似然译码规则

我们也可以直观地采用择多译码，即根据输出端接收序列是“0”多还是“1”多来译码。如果有两个以上是“0”则译码器就判决为“0”，如果有二个以上是“1”则判决为“1”。根据择多译码规则，同样可得到

$$
\begin{aligned}
P_E &= P_r\{\text{错 3 个码元的概率} + \text{错 2 个码元的概率}\} \\
&= C_3^3 p^3 + C_3^2 \bar{p} p^2 \\
&= p^3 + 3\bar{p}p^2 \\
&\approx 3\times 10^{-4}
\end{aligned}
$$

择多译码与极大似然译码的平均错误概率是一样的。若接收码字中有一位码元发生错误，译码器还能正确译出所发送的码字，所以说这种简单重复编码能纠正一位码元的错误，使得错误概率降低。

显然，如果进一步增大重复次数n，则会继续降低平均错误概率。可算得

$n=5 \quad P_E=10^{-5}$

$n=7 \quad P_E=4\times10^{-7}$

$n=9 \quad P_E=10^{-8}$

$n=11 \quad P_E=5\times10^{-10}$

重复次数多大最合适？

可见，当 n 很大时，使 P_E 很小是可能的. 但这时带来了一个新问题，当 n 很大时，信息传输会降低很多. 我们把编码后的信息传输率表示为

$$\boxed{R=\frac{\log_2 M}{n}} \tag{6.20}$$

含义为：M 个信源符号（一般已接近等概分布），平均每个信源符号所携带的信息量为 $\log_2 M$ bit，用 n 个码元的信道编码来传输，平均每个码符号所携带的信息量即为信息传输率 R.

平均错误概率和信息传输率

当 $M=2$ 时可依次求得简单重复编码的信息传输率和信息传输速率：

$n=1$	$R=1$	$\vdots$	$\vdots$
$n=3$	$R=1/3$	$n=11$	$R=1/11$
$n=5$	$R=1/5$		

由此可见，利用简单重复编码减小平均错误概率 P_E 是以降低信息传输率 R 为代价的，那么有没有可能找到一种编码方法，使平均错误概率 P_E 充分小而信息传输率又不至于太低呢？

降低平均错误概率对信息传输率的影响

首先看一下简单重复编码为什么使信息传输率降低？在未重复以前，输入端有 2 个消息，假设为等概率分布则每个消息携带的信息量是 $\log_2 M=1$ bit. 简单重复编码（$n=3$）后，可以把信道看成是无记忆信道的三次扩展信道，这时输入端有 8 个二元序列可以作为消息，但是我们只选择了两个二元序列作为消息 $M=2$，每个消息携带的平均信息量仍为 1 bit，而传送一个消息需要 3 个二元码符号，所以 R 就降低到 1/3.

如果在扩展信道的输入端把 8 个可能作为消息的二元序列都用上作为消息，则 $M=8$，每个消息携带的平均信息量就是 $\log_2 M=\log_2 8=3$ bit，而传递一个消息所需的码符号仍为 3 个二元码，这样 R 就提高到 1 比特/码符号.

译码时接收端 8 个接收序列译成与它对应的发送序列，只要接收序列中有一个码元发生错误就会变成其他的码字序列，使译码造成错误. 只有接收序列中每个码元都不发生错误才能正确传输，所以得到正确传输概率为 $1-P_E=\bar{p}^3$. 于是错误概率为

$$P_E=1-\bar{p}^3\approx 3\times 10^{-2} \qquad (p=0.01)$$

这时 P_E 反而比单符号信道传输的 P_E 大 3 倍.

简单重复编码实际上是增加了信息的冗余度

若在 3 次无记忆扩展信道中，取 $M=4$，用如下 4 个符号序列作为消息：

$$000 \qquad 011 \qquad 101 \qquad 110$$

按照最大似然译码规则，可计算出错误概率为 $P_E \approx 2\times 10^{-2}$

信息传输率为 $R=\frac{\log_2 4}{3}=\frac{2}{3}$

与 $M=8$ 的情况相比，错误概率降低了，而信息传输率也降低了.

因此看到这样一个现象：在一个二元信道的 n 次无记忆扩展信道中，输入端有 2^n 个符号序列可以作为消息. 如果选出其中的 M 个作为消息传递，则：M 大一些，P_E 也大些，R 也大；M 取小一些，P_E 就降低，而 R 也要降低.

因此 仅仅依靠改变消息符号的个数不能解决PE和R的矛盾

消息个数一样，编码方法不同时，错误概率不同

再进一步看，从 2^n 个符号序列中取 $M=4$ 个作为消息可以有 C_8^4 共 70 种选择方法，选取的方法也就是编码方法不同，错误概率不同，现在来比较下面两种取法：

- $M=4$　第 1 种选法　000，011，101，110
- $M=4$　第 2 种选法　000，001，010，100

各自的平均错误概率如何计算？先猜一下，谁大谁小

可求得第 1 种选法的错误概率为 $\boxed{P_E}=2\times10^{-2}$　$\boxed{R}=\frac{2}{3}$

第 2 种选法的错误概率为 $\boxed{P_E}=2.28\times10^{-2}$　$\boxed{R}=\frac{2}{3}$

比较的结论：

两者 R 相同．第 2 种码的平均错误概率大．

对于第 1 种码来说，当接收到的发送的 4 个符号序列中任一码元发生错误，就可判断消息在传输中发生了错误，但无法判断由哪个消息发生错误而来；

对第 2 种码，当发送消息“000”时，传输中任一码元发生错误就变成了其他 3 个可能发送的消息，根本无法判断传输时有无发生错误．

为什么码1有这样的性质，码2没有？

可见错误概率与编码方法有很大关系．

5、(5,2)线性码和码的距离

考察这样一个例子：信道输入端所选的消息数不变，任取 $M=4$，而增加码字的长度，取 $n=5$. 这时信道为二元对称信道的五次扩展信道，这个信道输入端有 $2^5=32$ 个不同的二元序列. 我们选取其中 4 个作为发送消息. 这时信息传输率为

$$\boxed{R=\frac{\log_2 4}{5}=\frac{2}{5}}$$

设输入序列 $\boldsymbol{x}_i=x_{i_1}x_{i_2}x_{i_3}x_{i_4}x_{i_5}$，$x_{i_k}\in\{0,1\}$，$i=1,2,3,4,5$，其中 x_{i_k} 为 $\boldsymbol{x}_i$ 序列中第 k 个分量，若 $\boldsymbol{x}_i$ 中各分量满足方程

$$\begin{cases} x_{i_1}=x_{i_1} \\ x_{i_2}=x_{i_2} \\ x_{i_3}=x_{i_1}\oplus x_{i_2} \\ x_{i_4}=x_{i_1} \\ x_{i_5}=x_{i_1}\oplus x_{i_2} \end{cases}$$

或者写成矩阵形式

$$\boldsymbol{x}_i=\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ \boxed{1} & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}\begin{pmatrix} x_{i_1} \\ x_{i_2} \\ x_{i_3} \\ x_{i_4} \\ x_{i_5} \end{pmatrix}$$

其中，$\oplus$ 为模二和运算，也叫异或.

由上述编码方法得到一种(5,2)线性码. 如果译码采用极大似然译码规则,它的译码规则如表 6.2 所示.

选用此码,接收端译码能纠正码字中所有发生一位码元的错误,也能纠正其中两个两位码元的错误,所以可计算得

正确译码概率 $\overline{P_E}=\bar{p}^5+5\bar{p}^4p+2\bar{p}^3p^2$

错误译码概率 $P_E=1-\overline{P_E}\approx 7.8\times 10^{-4}(p=0.01)$

与之前M=4, n=3时相比

$$\boxed{P_E}=2.28\times 10^{-2}\quad \boxed{R}=\frac{2}{3}$$

信息传输率略低，错误概率降低很多

与之前M=2, n=3简单重复码相比

$$P_E\approx 3\times 10^{-4}$$

错误概率差不多，信息传输率大了

表 6.2 (5,2)线性码译码规则

接收码字	译码输出	接收码字	译码输出
00000	00000	10000	00000
00001	00000	10001	00000
00010	00000	10010	11010
00011	00000	10011	10111
00100	00000	10100	10111
00101	01101	10101	10111
00110	10111	10110	10111
00111	10111	10111	10111
01000	00000	11000	11010
01001	01101	11001	11010
01010	11010	11010	11010
01011	11010	11011	11010
01100	01101	11100	01101
01101	01101	11101	01101
01110	01101	11110	11010
01111	01101	11111	10111

编码方法

因此采用增大n，并且适当增大M和采用恰当的编码方法，既能使P_E降低，又能使信息传输率R不会减少太多。

编码方法的讨论

当码字长度n和消息数M确定时，采用什么样的编码方法能最大限度的降低错误概率？

回顾前面n=3，M=4，两种编码方法的比较

再进一步看，从 2^n 个符号序列中取 $M=4$ 个作为消息可以有 C_8^4 共 70 种选择方法，选取的方法也就是编码方法不同，错误概率不同，现在来比较下面两种取法：

- $M=4$ 第 1 种选法 000，011，101，110
- $M=4$ 第 2 种选法 000，001，010，100

码的距离

定义 6.4 长度为 n 的两个符号序列(码字)$\boldsymbol{x}_i$ 与 $\boldsymbol{y}_j$ 之间的距离是指序列 $\boldsymbol{x}_i$ 和 $\boldsymbol{y}_j$ 对应位置上码元不同的位置的个数,通常又称为汉明距离.

对于二元码序列,令 $\boldsymbol{x}_i = x_{i_1}x_{i_2}\cdots x_{i_n}$, $x_{i_k}\in\{0,1\}$, $\boldsymbol{y}_j = y_{j_1}y_{j_2}\cdots y_{j_n}$, $y_{j_k}\in\{0,1\}$

则 $\boldsymbol{x}_i$ 和 $\boldsymbol{y}_j$ 的汉明距离为 $D(\boldsymbol{x}_i,\boldsymbol{y}_j) = \sum_{k=1}^{N} x_{i_k}\oplus y_{j_k}$ (6.22)

二元序列 $\boldsymbol{x}_i = 101111$, $\boldsymbol{y}_j = 111100$ 则得 $D(\boldsymbol{x}_i,\boldsymbol{y}_j) = 3$

四元序列 $\boldsymbol{x}_i = 1320120$ $\boldsymbol{y}_j = 1220310$ 则得 $D(\boldsymbol{x}_i,\boldsymbol{y}_j) = 3$

码的距离

这样定义的码字距离满足距离公理，即汉明距离满足以下性质：

(1) 非负性　$D(x_i, y_j) \geqslant 0$，当且仅当 $x_i = y_j$ 时等号成立；

(2) 对称性　$D(x_i, y_j) = D(y_j, x_i)$；

(3) 三角不等式　$D(x_i, z_k) + D(z_k, y_j) \geqslant D(x_i, y_j)$.

码的距离（续）

定义 6.5 码 C 中，任意两个码字的汉明距离的最小值称为该码 C 的最小距离.

$$d_{\min}=\min D(w_i,w_j)\quad w_i\neq w_j;w_i,w_j\in C \quad (6.23)$$

码的最小距离 $d_{\min}$ 与译码错误概率密切相关. 我们用距离概念来考察以下 5 个码，如表 6.3 所示.

表 6.3　码的最小距离与平均译码错误概率的关系

	码 1	码 2	码 3	码 4	码 5
码字	000	000	000	00000	000
	111	011	001	01101	001
		101	010	10111	010
		110	100	11010	011
					100
					101
					110
消息数 M	2	4	4	4	8
信息传输率 R	1/3	2/3	2/3	2/5	1
码的最小距离 $d_{\min}$	3	2	1	3	1
错误概率 P_E(极大似然译码)	3×10^{-4}	2×10^{-2}	2.28×10^{-2}	7.8×10^{-4}	3×10^{-2}

码的距离与错误概率

显然，$d_{\min}$越大，P_E越小．码的最小距离$d_{\min}$越大，受干扰后，越不容易把一个码字错成另一个码字，因而错误概率小；$d_{\min}$越小，受干扰后越容易把一个码字变成另一个码字，因而错误概率大．这就说明：在编码选择码字时，要使码字之间的距离尽可能地大．

第二节 有噪信道编码定理

1 有噪信道编码定理

2 有噪信道编码逆定理

3 错误概率的上界

1、有噪信道编码定理

- 定理6.2　设有一个离散无记忆平稳信道，其信道容量为C。当信息传输率$R<C$，只要码长n足够长，则总存在一种编码，可以使平均译码错误概率任意小。
- 这个定理称为有噪信道编码定理，又称为香农第二定理。

- 通过一个有噪信道可以实现几乎无失真传输，这与人们的直观想象似乎是大相径庭的，而定理的证明也是非常巧妙的。香农不去构造理想的好码，而是用随机编码的方法得到所有可能生成的码的集合，然后在码集合中随机选择一个码作为输入码序列，最后计算这样随机选择的一个码在码集合上的平均性能。在译码时，利用了联合典型序列的概念，即将接收序列译成与其联合典型的码字．这种译码方法不是最优译码，但便于理论分析。

香农在证明这一结论时采用了出人意料的方法

•通常想法：先要构造一个理想的好码，然后计算这一码用于传输时的误码率。但这两点都很难实现:

✓ •构造具有理想性能的好码是一个极其复杂的问题，在当时根本无望解决。

✓ •在N很大时计算这一理想好码在理想译码器或最大似然译码器下的误码率也是及其困难的。

•香农在证明过程中巧妙的回避了这两个难题，而是按照随机编码和联合典型译码方法进行论证

证明的基本思路是：

(1) 允许平均错误概率 P_E 任意小而非零；

(2) 在 n 次无记忆扩展信道中讨论，且 n 足够大，这样可以使用大数定理；

(3) 随机编码，在随机编码的基础上计算整个码集的码的平均错误概率，由此证明至少有一种好码存在，因为是随机编码，所以求错误译码概率时与特定的码字无关.

所谓随机编码，是指在 n 长的输入序列中，随机选择 M 个作为输入码字序列组成一个码 C_k，$C_k=\{x_1,x_2,\cdots,x_M\}$，$M$ 为信源消息数.

- 香农第二定理指出，若R<C，则存在某种编码可以使传输错误概率任意小；反之，若R>C，则可以使传输错误概率任意小的编码不存在。

- 它从理论上证明平均错误译码概率Pe趋于零、信道信息传输速率R无限接近于信道容量C的抗干扰信道编码是存在的。
- 但从实用观点来看，理论的证明尚不能令人满意。因为在证明的过程中是完全“随机地”去选择一个码。这个码是完全无规律的，因此，就无法具体构造这个码，也就无法实现和应用。但人们在理论指导下，赋予码以各种形式的代数结构，出现了代数编码、卷积码等

2、有噪信道编码逆定理

定理 6.3 设有一个离散无记忆平稳信道，其信道容量为 C. 对于任意 $\varepsilon>0$，若选用码字总数 $M=2^{n(C+\varepsilon)}$，则无论 n 取多大，也找不到一种编码，使译码错误概率 P_E 任意小.

3、错误概率的上界

错误概率与什么有关？ 1、信道的统计特性 2、编码方法和译码规则

在实际应用中常见这样的问题，若信道的统计特性已知，给定错误概率的上界，如何确定编码方法和译码规则？

离散无记忆信道中 P_E 趋于零的速度与 n 成指数关系，即当 $R<C$ 时，平均错误概率

$$P_E \leqslant \exp[-nE_r(R)] \quad (6.60)$$

式中，$E_r(R)$ 称为随机编码指数，又称为可靠性函数或加拉格(Gallager)函数．一般可靠性函数 $E_r(R)$ 与信息传输率 R 的关系曲线如图 6.5 所示，它是一条下凸函数曲线．在 $R<C$ 范围内 $E_r>0$，所以随 n 增大 P_E 以指数趋于零．实际编码的码长 n 不需选得很大．

可靠性函数 $E_r(R)$ 在信道编码中有极其重要的意义

可以帮助我们选择信息传输率 R 和编码长度 n.

C 是可达的最大的可靠信息传输率.

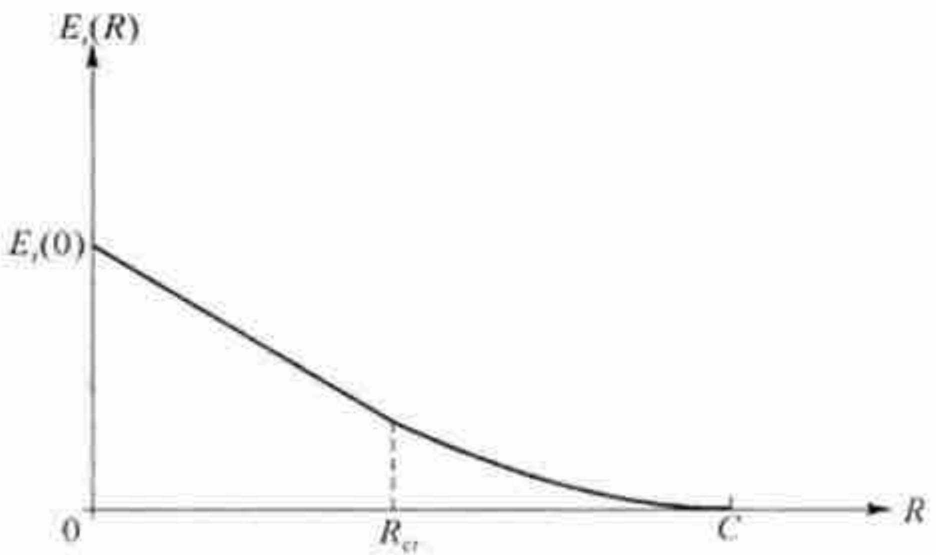


图 6.5　$E_r(R)$ 曲线图

- 从编码的角度来看，平均误码率与信道编码的码字长度N有关，同时也与信道上所传消息的信息传输速率R有关

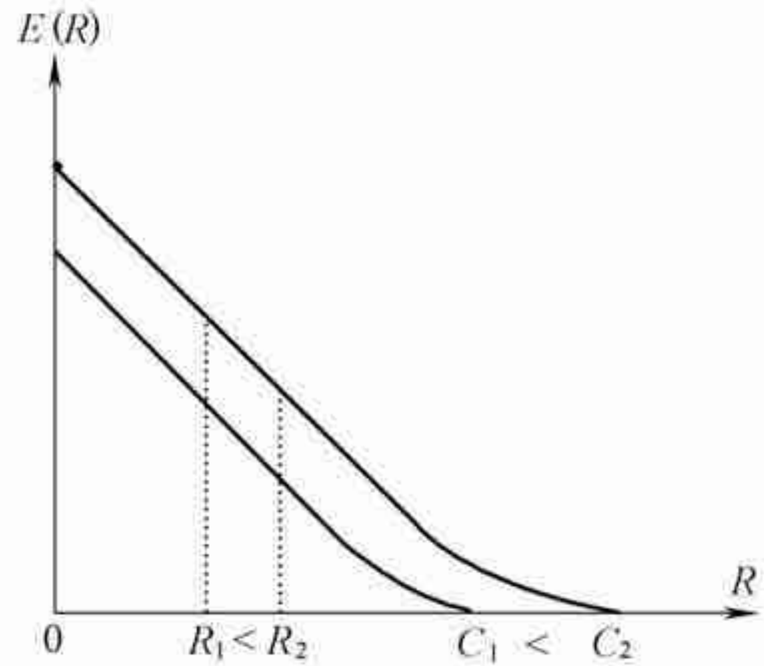
- 由上式可知，要减少误码率Pe就应当增大码长N或增大可靠性函数Er(R)。而对于同样的信息传输速率，信道容量C大者其可靠性函数肯定越大；对于同样的信道容量C,R小者其可靠性函数Er(R)则大。因此，想增大Er(R)就要加大信道容量C或减少信息传输速率R。

鉴于以上分析，可采取以下措施来减少误码率：

①增大信道容量**C**;

②减小信息传输速率**R**;

③增加码长**N**。



增大E(R)的途径

差错控制的途径

❖ **从工程的角度，可通过以下方法提高纠错能力、减少误码率**

A. 利用冗余度
B. 噪声均化(随机化)

(1) 利用冗余度

冗余度:在信息流中插入冗余比特,这些冗余比特与信息比特之间存着特定的相关性。

冗余的资源:时间,频带,功率,设备复杂度

(2) 噪声均化

噪声均化:让差错随机化，以便更符合编码定理的条件从而得到符合编码定理的结果。

基本思想:设法将危害较大的、较为集中的噪声干扰分摊开来，使不可恢复的信息损伤最小。

小结

从香农第一定理和香农第二定理可以看出,要做到有效和可靠地传输信息,可以将编码分成信源编码和信道编码两部分.首先,通过信源编码,用尽可能少的信道符号来表达信源,也就是对信源数据用最有效的表达方式表达,尽可能减少编码后数据的冗余度.然后对信源编码后的数据设计信道编码,也就是适当增加一些冗余度以纠正和克服信道中干扰引起的错误.这两部分是分别独立考虑的.

❖ 从香农第一定理和香农第二定理可以看出，要做到有效和可靠地传输信息，可以将编码分成信源编码和信道编码两部分。这种分两部分编码的方法在实际通信系统中有着重要的意义。在实际数字通信系统中，信道常常是共用的数字信道(二元信道)，输入端只是二元序列，所以信道编码只需针对信道特性进行，纠正信道中带来的错误，这样，可以大大降低通信系统设计的复杂度。可以证明，这种分两步编码处理的方法与一步编码处理方法是一样有效的。

香农第二定理也只是一个存在定理，它说明错误概率趋于零的好码是存在的，但是没有说明如何构造这个好码。尽管如此，香农第二定理仍然具有重要的理论意义和实践指导作用，它可以指导各种通信系统的设计，有助于评价各种通信系统及编码效率。

第三节 纠错编码

1 纠错编码

2 纠错码的分类

3 纠错码的基本概念

1、纠错编码

自香农给出信道编码定理以来，引起了人们对信道编码的极大兴趣，但是香农只是证明了满足这种特性（$R<C$ 时 $P_E \to 0$）的码的存在，还不能按其证明方法得到这种好码. 证明过程是采用随机编码的方法，由于随机编码所得的码集很大，通过搜索得到好码的方法在实际上很难实现，而且即使找到其中的好码，这种码的码字也是毫无结构的，这意味着译码时只能用查表的方法，而在 N 很大时译码表所需的存储量也是很难被接受的，因此真正实用的信道编码还需通过各种数学工具来构造，使码具有很好的结构性以便于译码. 抽象代数（也称近世代数）就是编码理论的最重要的数学工具，它包括群论、环论、域论、格论、线性代数等许多分支.

- 信道编码的目的是提高信号传输的可靠性，广义的信道编码还包括为特定信道设计的传输信号，如NRZ(不归零)码、HDB3码、伪随机序列码都属于信道编码。纠错编码是提高传输可靠性的最主要措施之一。

2、 纠错码分类

- 由于信道中干扰和噪声的存在，使得经信道传输后的接收码字与原来的发送码字存在着差异，也就是差错。一般信道中噪声干扰越大，码字产生错误的概率也越大。葍
- 信道中的干扰和噪声一般分为两种形式：一是随机噪声，它主要来源于设备的热噪声和散弹噪声以及传播媒介的热噪声，是通信系统中的主要噪声；二是脉冲干扰和信道衰落，它的特点是突发出现，主要来源于雷电、通电开关、负荷突变或设备故障等。

- 根据信道中干扰和噪声的形式，可将信道分为3类：随机信道、突发信道和混合信道。

- 随机噪声产生的错误是独立随机出现的，称为随机错误。它的特点是各码元是否发生错误是随机的，且相互独立，因而不会出现成片的错误．只产生随机错误的信道称为随机信道。

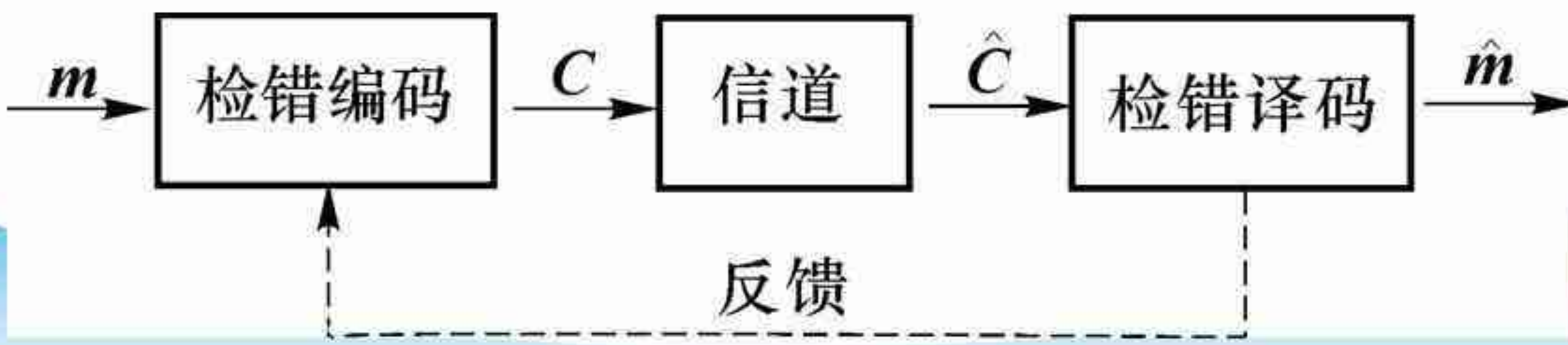
- 脉冲干扰和信道衰落产生的错误是成串出现的，错误间具有相关性，这类错误称为突发错误。产生突发错误的信道称为突发信道。

- 有些实际信道既有随机错误又有突发错误，称为混合信道。

- 根据不同的信道类型设计的信道编码分为纠随机错误码、纠突发错误码和纠混合错误码。在通信系统中，纠检错的工作方式有反馈重传纠错、前向纠错和混合纠错等。

按照纠检错工作方式分类

- 1. 反馈重传纠错
- **图6.6即反馈重传纠错图示**
- **发送端发出的是能够发现错误的检错码，接收端对接收到的码字进行译码，发现有错时，通过反馈系统向发送端请求重传已发送的全部或部分码字，直到接收端认为没有错误为止。**



- 2. 前向纠错
- 图6.7即前向纠错图示.
-

图6.7 前向纠错

- 前向纠错也称为自动纠错。发送端发出的是具有纠错能力的纠错码，接收端根据编码规则进行解码。当误码个数在码的纠错能力范围之内时，译码器可以自动纠正错误。

3. 混合纠错萹

对发送端进行适当的编码，当错误不严重时，在码的纠错能力之内，采用自动纠错，当产生的差错超出码的纠错能力时，则通过反馈系统向发送端要求重发，这种同时具有反馈重传纠错和自动纠错工作方式的纠错称为混合纠错。

- 检错码和纠错码在不加区别时统称为纠错码。

按照分组方式得出的纠错码分类

- 纠错码的分类如下：
- 葍
- (1)根据信息码元与校验码元之间是否存在线性关系可分为线性码和非线性码。
- (2) 根据不同的分组及映射方式，纠错码可以分成分组码和树码。

分组码和树码

① 分组码

把信息序列以每 k 个码元分组，然后把每组 k 个信息元按一定规律产生 r 个多余的校验元，输出码序列每组长为 $n = k + r$. 每一码字的 r 个校验元只与本组的 k 个信息元有关，与别的组的信息位无关，记为分组码 (n, k).

本章主要介绍线性分组码，线性分组码中有一类特别重要的是循环码，它具有完整的代数结构，编译码都比较简单和易于实现.

② 树码

信息序列以每 k_0(通常较小)个码元分段，编码器输出该段的校验元 $r = n - k_0$ 不仅与本段的 k_0 个信息元有关，而且还与其前面若干段的信息元有关，称为**树码或链码**. 树码中最重要的一类是**卷积码**，它的校验元与信息之间是线性关系.

- 纠错码按结构分类如下：

- 纠错码
  - 分组码
    - 线性分组码
      - 循环码
      - 非循环码
    - 非线性分组码
  - 树码
    - 线性—卷积码
    - 非线性码

- 目前的通信系统大多采用二进制的数字系统，所以以下提到的纠错码都是指二元码。

3、纠错码的基本概念

信源编码把信源符号用二元序列来表示，这个二元序列称为信息序列，信源编码主要解决的是通信的有效性问题，即用尽可能少的码符号来表示信源符号或信源符号序列. 信道编码要解决的问题是通信的可靠性问题. 信息序列送到信道进行传输之前还需要经过信道编码变成具有纠检错能力的码序列. 通过在信息序列中插入冗余码元(称为校验元或监督元)，使新序列的码元之间具有相关特性，然后再送入信道进行传输. 在接收端，信道译码器根据这个相关特性对接收序列进行译码，在纠错能力范围内可以对差错进行自动纠正，恢复原发送码序列.

- 对信道编码的一般要求是：
- （1）纠错检错能力强，可发现和纠正多个错误；
- （2）信息传输率高，信息传输率也称为码率：$R=\dfrac{\log M}{n}$，M为信息序列数，所以码长n应尽可能短；
- （3）编码规律简单，实现设备简单且费用合理；
- （4）与信道的差错统计特性相匹配。

对信道编码的一般要求

信道编码就是综合考虑以上因素的情况下选择和设计合理的编译码实现方案.

每个码字 $\boldsymbol{C}=C_1C_2\cdots C_n$ 中 k 位称为信息位,其余 $n-k$ 位为校验位或监督位.信息序列的个数为 $M=2^k$,而长为 n 的二元码一共有 2^n 个,选出其中的 M 个作为码字,称为许用序列,而其他序列为禁用序列.

【例 6.5】

下面给出一种编码如表 6.4 所示.

表 6.4 (7,3)码

消息序列	码字	消息序列	码字
000	0000000	100	1001110
001	0011101	101	1010011
010	0100111	110	1101001
011	0111010	111	1110100

上例中信息位为 $k=3$,码长为 $n=7$,监督位为 $r=4$,用 $\eta=k/n$ 表示码字中信息位所占的比重,称编码效率,它表明了信道的利用率. η 越大,编码效率越高,它是衡量码性能的一个重要参数.上例中 $\eta=\dfrac{3}{7}\approx 43\%$,而

$$\boxed{R}=\frac{\log_2 M}{n}=\frac{\log_2 2^k}{n}=\frac{k}{n}=\boxed{\eta}.$$

有关概念和定义

如果 n 长码字的每一位与原始信息序列 $\boldsymbol{d}=d_1d_2\cdots d_k$ 的 k 个信息位之间的函数关系 $c_i=f_i(d_1,d_2,\cdots,d_k)$, $i=1,2,\cdots,n$ 是线性关系,则称该分组码为**线性分组码**,否则称为**非线性分组码**.

若 (n,k) 分组码字中 k 个信息位与原始信息序列的 k 个信息位相同,且位于 n 长码字的前(或后)k 位,而校验位位于其后(或前)$n-k$ 位,则该分组码为**系统码**,否则为**非系统码**.

在 6.1.2 节我们介绍了汉明距离和码的最小距离.码的最小汉明距离是衡量码的纠检错能力的重要参数,码的最小距离越大,其纠检错能力越强.

对于某一码字,其非零元素的个数称为该码字的**汉明重量**.

对于二元码,其码字的重量是码字中 1 的个数.若码字为 $\boldsymbol{c}_i=c_{i_1}c_{i_2}\cdots c_{i_n}$,则其重量可以表示为 $W(\boldsymbol{c}_i)=\sum_{k=1}^{n}c_{i_k}$.例如,码字 $\boldsymbol{c}_1=1010101$,其重量为 4.