信息论基础

李　莹
liying2009@ecust.edu.cn

第六章：有噪信道编码

一、信道编码的相关概念

二、有噪信道编码定理

三、纠错编码

1. 有噪信道编码定理

定理6.2 （香农第二定理）

设有一离散无记忆平稳信道，其信道容量为C，只要待传送的信息传输率$R<C$，当码长 n 足够大时，则至少存在一种编码，使译码错误概率任意小。

说明：

1) 信息传输速率

2) 香农第二定理仅指出了满足这种要求的信道编码的存在性，没有给出具体的编码方法。

定理6.3 （有噪信道编码逆定理）

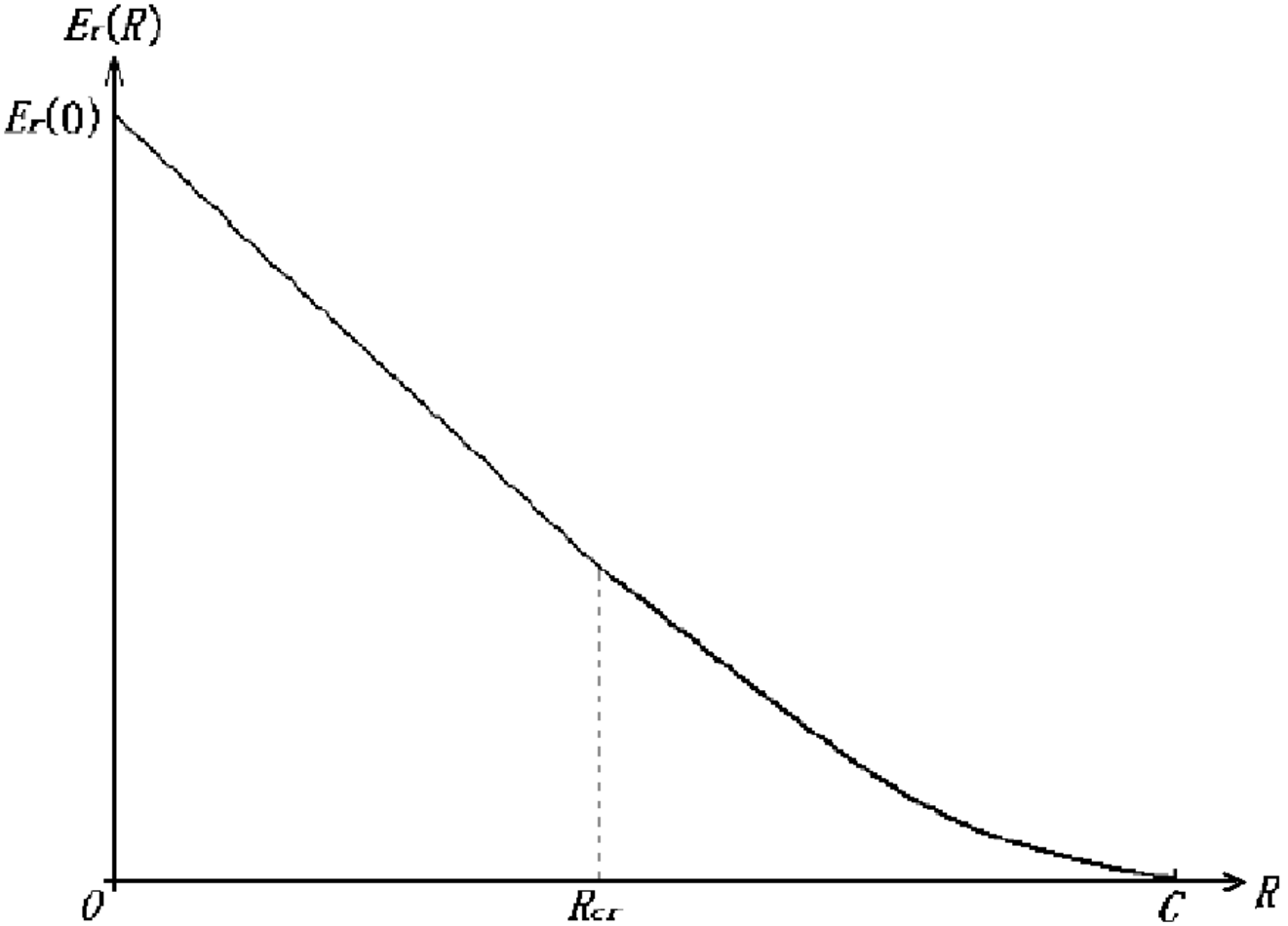
设有一离散无记忆平稳信道，其信道容量为C，如果信息传输率$R>C$，即 $M > 2^{nC}$，则无论码长 n 取多大，也不可能使译码错误概率任意小。

结论：

信道容量是在信道中可靠传输信息的最大信息传输率。

2. 错误概率的上界

$$P_E \le \exp[-nE_r(R)]$$



纠错编码

1 纠错码的分类

2 纠错码的基本概念

3 线性分组码

4 汉明码

5 循环码

概述

- 香农第二定理证明，当 $R < C$ 时 $P_E \to 0$ 的码存在。
- 证明过程采用的是随机编码的方法：
  - 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；
  - 即使找到了好码，这种码的码字也没有规律，不便于译码。
- 真正实用的信道编码方法还需要通过各种数学工具来构造，使码具有好的结构性以便于译码。

概述

- 近世代数是信道编码理论用到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。
- 广义信道编码包括：调制、成形滤波、扩频、上下变频等。
- 纠错编码是提高传输可靠性的最主要措施之一。

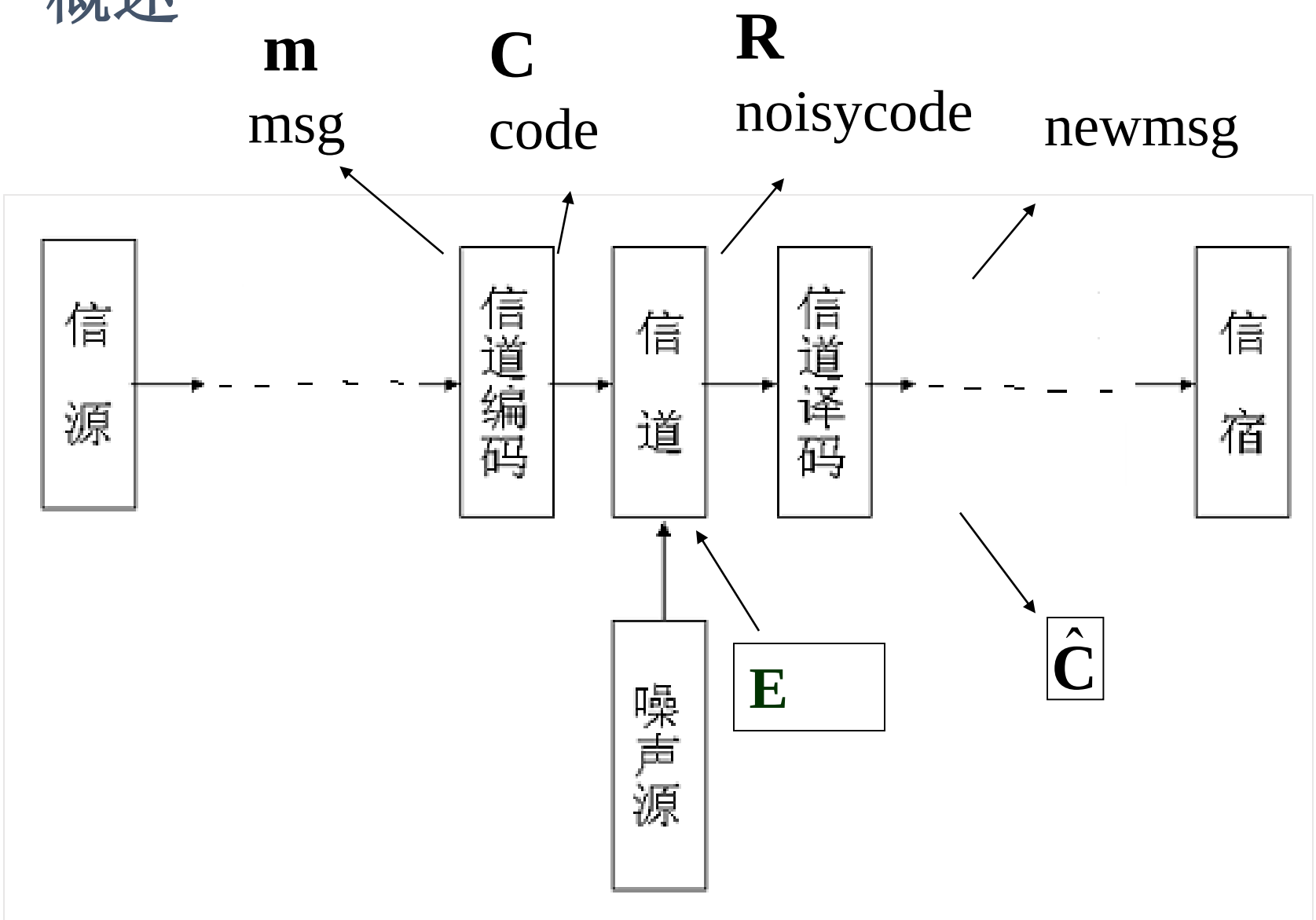
概述

- 纠错编码的基本思路：

根据一定的规律在待发送的信息码元中人为的加入一些冗余码元，这些冗余码元与信息码元之间以某种确定的规则相互关联（约束）。

在接收端按照既定的规则检验信息码元与监督码元之间的关系。如果传输过程出错，则信息码元与监督码元之间的关系将受到破坏，从而可以发现错误乃至纠正错误。

概述



干扰一般分为两种形式：

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，它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。

信道可分为三类：

1. 只产生随机错误的信道称为随机信道。比如卫星信道、同轴电缆、光缆信道以及大多数微波中继信道。

2. 产生突发错误的信道称为突发信道。实际的短波信道、移动通信信道、由于擦伤造成成串差错的光盘和磁盘，均为这一类信道。

3. 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。

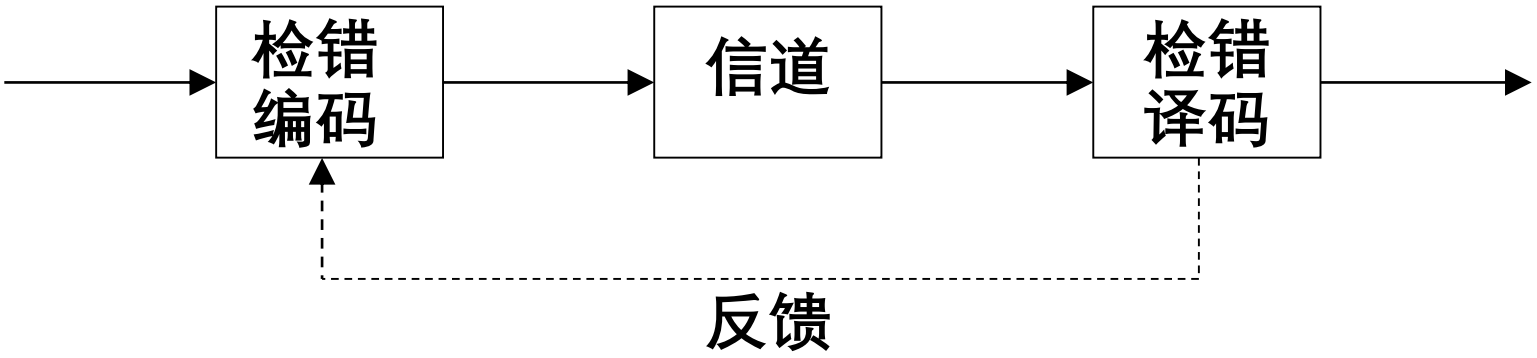
在通信系统中，纠检错的工作方式有：

(1) 反馈重传(ARQ)

(2) 前向纠错(FEC)

(3) 混合纠错

(1) 反馈重传(ARQ)



发送端经编码后发出能够发现错误的码，接收端收到后经检验，如果发现传输中有错误，则通过反馈系统把这一判断结果反馈回发端，然后发送端把前面发出的信息重新传送一次，直到接收端认为正确地收到信息为止。

(2) 前向纠错(FEC)

发送端发出的是具有纠错能力的纠错码，接收端根据译码规则进行译码。当误码个数在码的纠错能力范围内时，译码器可以自动纠正错误。

特点：

1）前向纠错方式不需要反馈信道，特别适合于只能提供单向信道的场合。

2）由于能自动纠错，不要求检错重发，因而延时小，实时性好。

3）随着纠错能力的增强，译码设备也变得复杂。

(3) 混合纠错

对发送端进行适当的编码。当错误不严重，在码的纠错能力范围之内时，采用自动纠错；当产生的差错超出码的纠错能力范围时，通过反馈系统要求发端重发。

1 纠错码的分类

(1) 按功能分：

- 检错码：仅能检测误码
- 纠错码：可纠正误码
- 纠删码：兼纠错和检错能力

} 纠错码

(2) 按信息码元与监督码元之间的检验关系分：

- 线性码：满足线性关系
- 非线性码：不存在线性关系

(3) 按信息码元与监督码元之间的约束方式不同分：

- 分组码：本码组的监督码元仅和本码组的信息元相关。
- 树码：本码组的监督码元不仅和本码组的信息元相关，而且与前面码组的信息码元有关。如果是线性关系则称为卷积码。
- 系统码：信息码元与监督码元在分组内有确定位置，编码后的信息码元保持不变；

(4) 按信息码元在编码后是否保持原形式不变：

- 非系统码：信息位打乱，与编码前不同。

(5)按纠正差错的类型可分为：

- 纠随机错误码
- 纠突发错误码
- 纠随机和突发错误码

纠错码按结构分类如下：

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

2 纠错码的基本概念

- 分组码的表示方法：（二元分组码）
  - 信息码组由 k 个信息码元（信息位）组成，共有 2^k 个不同的信息码组；
  - 附加 $r = n - k$ 个校验码元（校验位或监督位），每个校验码元是该信息码组的某些信息码元模2和；
  - 编码器输出长度为 n 的码字；
  - 码字的数目共有 2^k ；
  - 这 2^k 个码字的集合称为 (n,k) 分组码；

- 对 二进制(n, k)线性分组码，合法码字数为2^k，可用编码空间的序列数为2^n个。许用序列，禁用序列
- 任一种2^k信息集合到二进制序列集合(2^n)的映射都是一种(n, k)码，因此总共可能的编码方案有 $C_{2^n}^{2^k}$ 种。

- **信息传输率（码率）**
- **编码效率**

$$R = \frac{\log M}{n} = \frac{\log 2^k}{n} = \frac{k}{n}$$

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

- 发现或构造好码是信道编码研究的主要问题。
- 线性分组码是最具实用价值的一类码，比如汉明码、循环码、BCH码、RS码等。

对信道编码的一般要求是：

①纠错检错能力强；

②信息传输率高；

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。

汉明距离 $\mathbf{c}_i = c_{i_1} c_{i_2} \cdots c_{i_n}$ $\mathbf{c}_j = c_{j_1} c_{j_2} \cdots c_{j_n}$

$$d(\mathbf{c}_i, \mathbf{c}_j) = \sum_{k=1}^{n} c_{i_k} \oplus c_{j_k}$$

汉明距离满足距离公理

(1) 非负性 $d(\mathbf{c}_i, \mathbf{c}_j) \geq 0$

(2) 对称性 $d(\mathbf{c}_i, \mathbf{c}_j) = d(\mathbf{c}_j, \mathbf{c}_i)$

(3) 三角不等式 $d(\mathbf{c}_i, \mathbf{c}_j) \leq d(\mathbf{c}_i, \mathbf{c}_l) + d(\mathbf{c}_l, \mathbf{c}_j)$

码C 的最小距离

$$d_{\min} = \min\left\{d(\mathbf{c}_i, \mathbf{c}_j) : \mathbf{c}_i, \mathbf{c}_j \in \mathbf{C}, i \neq j\right\}$$

汉明重量

$$W(\mathbf{c}_i) = \sum_{k=1}^{n} c_{i_k}$$

$$d(\mathbf{c}_i, \mathbf{c}_j) = \sum_{k=1}^{n} c_{i_k} \oplus c_{j_k}$$

$$= W(\mathbf{c}_i \oplus \mathbf{c}_j) = W(\mathbf{c}_l) = d(\mathbf{c}_l, \mathbf{0})$$

线性分组码的最小距离等于非零码字的最小重量。

码1	码2	码3	码4	码5	码6
000 111	000 001	000 011 101 110	000 001 100 010	00000 01101 10111 11010	000 001 010 011 100 101 110 111

3 线性分组码

3.1 校验矩阵与生成矩阵

(1) 校验矩阵

$$\mathbf{C} = C_1 C_2 C_3 C_4 C_5 C_6 C_7$$

$$\begin{cases} C_4 = C_1 + C_3 \\ C_5 = C_1 + C_2 + C_3 \\ C_6 = C_1 + C_2 \\ C_7 = C_2 + C_3 \end{cases}$$

$$\begin{cases} C_1 + C_3 + C_4 = 0 \\ C_1 + C_2 + C_3 + C_5 = 0 \\ C_1 + C_2 + C_6 = 0 \\ C_2 + C_3 + C_7 = 0 \end{cases}$$

$$\begin{cases} C_1 + C_3 + C_4 = 0 \\ C_1 + C_2 + C_3 + C_5 = 0 \\ C_1 + C_2 + C_6 = 0 \\ C_2 + C_3 + C_7 = 0 \end{cases}$$

$$\left[\begin{array}{ccc:cccc} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right] \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \\ C_5 \\ C_6 \\ C_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$\mathbf{H} = [\mathbf{Q} \quad \mathbf{I}]$$

$$\mathbf{H}\mathbf{C}^T = \mathbf{0}^T$$

$$\mathbf{C}\mathbf{H}^T = \mathbf{0}$$

- $\boldsymbol{H}$ 被称为校验矩阵。
- 对(n,k)线性分组码，校验矩阵为$(n-k)\times n$ 维矩阵。
- 对于系统码，校验矩阵可以表示为

$$\mathbf{H}=\begin{bmatrix}\mathbf{Q} & \mathbf{I}\end{bmatrix}$$

其中$\boldsymbol{Q}$ 为$(n-k)\times k$维矩阵， 为$(n-k)\times(n-k)$维单位矩阵。

(2) 生成矩阵

由校验方程，得到

$$\begin{cases} C_4 = C_1 + C_3 \\ C_5 = C_1 + C_2 + C_3 \\ C_6 = C_1 + C_2 \\ C_7 = C_2 + C_3 \end{cases}$$

$$\longrightarrow$$

$$\begin{cases} C_1 = C_1 \\ C_2 = C_2 \\ C_3 = C_3 \\ C_4 = C_1 + C_3 \\ C_5 = C_1 + C_2 + C_3 \\ C_6 = C_1 + C_2 \\ C_7 = C_2 + C_3 \end{cases}$$

$$\begin{cases} C_1 = C_1 \\ C_2 = C_2 \\ C_3 = C_3 \\ C_4 = C_1 + C_3 \\ C_5 = C_1 + C_2 + C_3 \\ C_6 = C_1 + C_2 \\ C_7 = C_2 + C_3 \end{cases} \qquad \mathbf{C} = [C_1 \quad C_2 \quad C_3] \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 1 \end{bmatrix}$$

$$\mathbf{G} = [\mathbf{I} \quad \mathbf{P}]$$

令 $\mathbf{m} = [c_1, c_2, c_3]$

$$\Longrightarrow \quad \mathbf{C = mG}$$

- $\mathbf{G}$ 被称为生成矩阵。
- 对(n,k)线性分组码，生成矩阵为 $k \times n$ 维矩阵。
- 对于系统码，生成矩阵可以表示为

$$\mathbf{G} = \begin{bmatrix} \mathbf{I} & \mathbf{P} \end{bmatrix}$$

其中 $\mathbf{P}$ 为$k \times (n-k)$维矩阵，$\mathbf{I}$ 为 $k \times k$ 维单位矩阵。

- 把生成矩阵的每一行用一个行向量 $\boldsymbol{G}_i, i=1,2,\cdots,k$ 来表示，则生成矩阵可以表示为

$$\mathbf{G}=\begin{bmatrix}\mathbf{G}_1\\ \mathbf{G}_2\\ \vdots\\ \mathbf{G}_k\end{bmatrix}$$

- 令 $\mathbf{m}=[m_1 \quad m_2 \quad \cdots \quad m_k]$，则

$$\mathbf{C}=\mathbf{mG}=\sum_{i=1}^{k} m_i\mathbf{G}_i$$

(3) 校验矩阵和生成矩阵的关系

- 由于生成矩阵$\boldsymbol{G}$的每一行都是一个码字，所以$\boldsymbol{G}$ 的每行都满足 $\boldsymbol{H}\boldsymbol{G}_i^{\mathrm{T}} = \boldsymbol{0}^{\mathrm{T}}$，则有

$$\boldsymbol{H}\boldsymbol{G}^{\mathrm{T}} = \boldsymbol{0}$$

- 对于标准形式的校验矩阵和监督矩阵，有

$$\boldsymbol{H}\boldsymbol{G}^{\mathrm{T}} = \begin{bmatrix}\boldsymbol{Q} & \boldsymbol{I}\end{bmatrix}\begin{bmatrix}\boldsymbol{I} & \boldsymbol{P}\end{bmatrix}^{\mathrm{T}} = \boldsymbol{Q} + \boldsymbol{P}^{\mathrm{T}} = \boldsymbol{0}$$

$$\Longrightarrow \boldsymbol{Q} = \boldsymbol{P}^{\mathrm{T}}$$

- 线性分组码的封闭性：线性分组码中任意两个码字之和仍然是该码的码字。

证明：设C_1和C_2分别是码C 中的两个码字，因此有

$$\left.\begin{aligned} HC_1^{\mathrm{T}} &= 0^{\mathrm{T}} \\ HC_2^{\mathrm{T}} &= 0^{\mathrm{T}} \end{aligned}\right\} \Rightarrow H(C_1 + C_2)^{\mathrm{T}} = HC_1^{\mathrm{T}} + HC_2^{\mathrm{T}} = 0^{\mathrm{T}}$$

即$C_1 + C_2$满足监督方程，所以是码C中的一个码字。

例1：3重复码是一个（3,1）线性分组码。其生成矩阵为

$$G=\begin{bmatrix}1 & 1 & 1\end{bmatrix}$$

$$\mathbf{C}=C_1C_2C_3=[m_1]\begin{bmatrix}1 & 1 & 1\end{bmatrix}=[m_1 \quad m_1 \quad m_1]$$

例2：（4,3）偶校验码是一个（4,3）线性分组码，其生成矩阵为

$$\mathbf{G}=\begin{bmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\end{bmatrix}$$

$$\mathbf{C}=C_1C_2C_3C_4=[m_1\quad m_2\quad m_3]\begin{bmatrix}1&0&0&1\\0&1&0&1\\0&0&1&1\end{bmatrix}$$

$$=[m_1\quad m_2\quad m_3\quad m_1+m_2+m_3]$$

例3： 已知生成矩阵为

$$\mathbf{G}=\begin{bmatrix}1&0&0&1&1&1&0\\0&1&0&0&1&1&1\\0&0&1&1&1&0&1\end{bmatrix}$$

求生成的线性分组码及由H 生成的线性分组码。

$$\mathbf{C}=\mathbf{mG}$$

m	C
000	0000000
001	0011101
010	0100111
011	0111010
100	1001110
101	1010011
110	1101001
111	1110100

$$\mathbf{G}=\begin{bmatrix}1&0&0&1&1&1&0\\0&1&0&0&1&1&1\\0&0&1&1&1&0&1\end{bmatrix}$$

$$\mathbf{G}=\begin{bmatrix}\mathbf{I}&\mathbf{P}\end{bmatrix}$$

$$\mathbf{P}=\begin{bmatrix}1&1&1&0\\0&1&1&1\\1&1&0&1\end{bmatrix}$$

$$\mathbf{P}^{T}=\mathbf{Q}$$

$$\mathbf{H}=\begin{bmatrix}\mathbf{Q}&\mathbf{I}\end{bmatrix}=\begin{bmatrix}1&0&1&1&0&0&0\\1&1&1&0&1&0&0\\1&1&0&0&0&1&0\\0&1&1&0&0&0&1\end{bmatrix}$$

$$\mathbf{C} = \mathbf{mH}$$

$$\mathbf{H} = [\mathbf{Q} \quad \mathbf{I}] = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

m	C
0000	0000000
0001	0110001
0010	1100010
0011	1010011
0100	1110100
0101	1000101
0110	0010110
0111	0100111
1000	1011000
1001	1101001
1010	0111010
1011	0001011
1100	0101100
1101	0011101
1110	1001110
1111	1111111

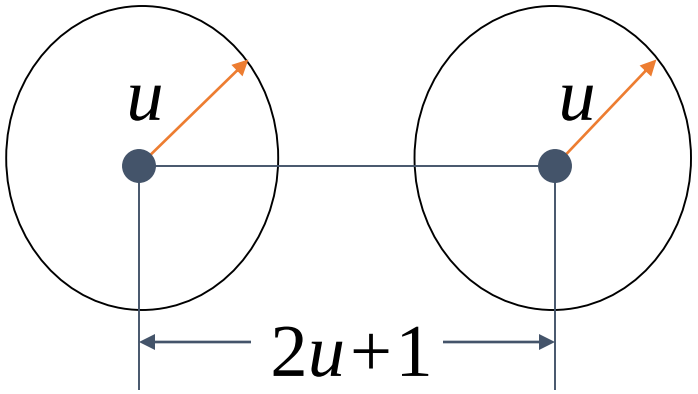
3.2 线性分组码的纠、检错能力

对于一个二进制对称信道，当输入为2^k个等可能的n长码字，则最大后验概率准则等效于最小汉明距离译码准则。

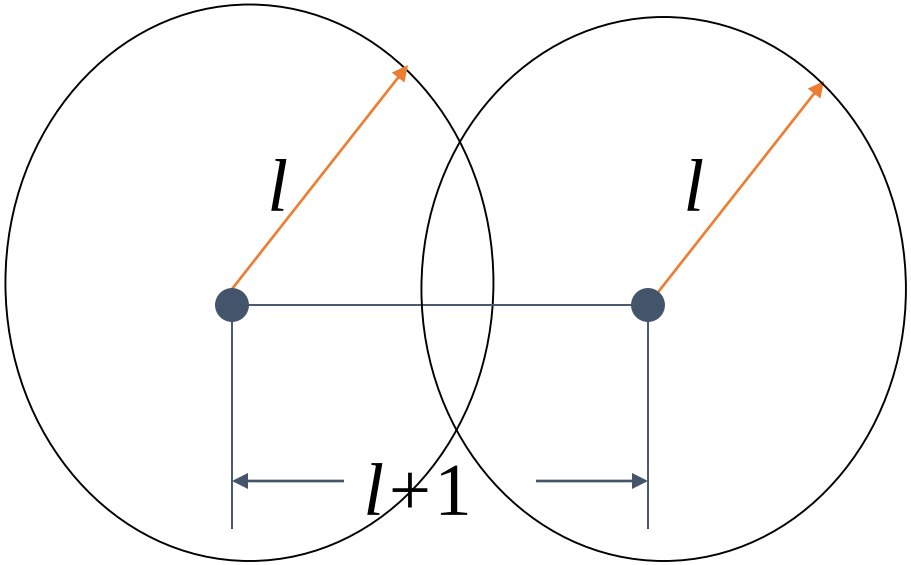
$$p(\mathbf{y}_j \mid \mathbf{x}_i) = \prod_{k=1}^{n} p(y_{j_k} \mid x_{i_k}) = p^D \overline{p}^{n-D}$$

关于码的最小距离与纠、检错能力的关系有以下结论：对于（n，k）线性分组码，设 $d_{\min}$ 为码的最小距离则

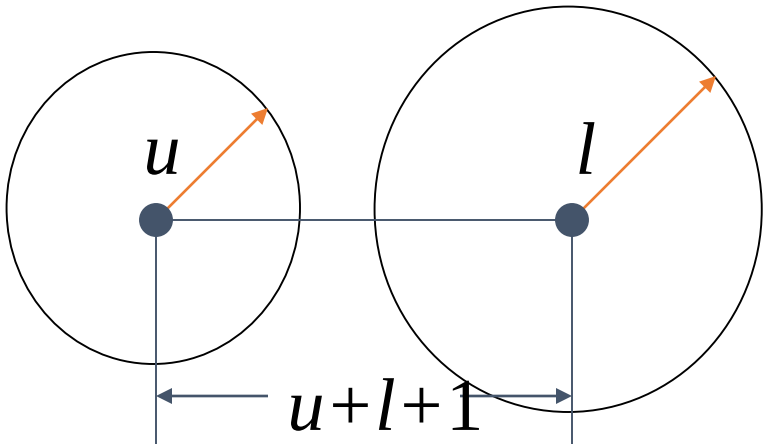
（1）这组码有纠正 u 个错误的充要条件是 $d_{\min}=2u+1$



（2）具有检测l个错误的充要条件是 $d_{\min} = l + 1$



（3）具有纠正 u 个错误，同时可以发现 l 个错误的充分必要条件为 $d_{\min} = u + l + 1$



码的纠错能力u与码字的长度n和消息数M

满足以下关系：

$$M\sum_{i=0}^{u}C_n^i \le 2^n$$

3.3 校验矩阵与码的最小距离的关系

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

对于（n，k）线性分组码：

校验矩阵H中的任意t列线性无关而t+1列线性相关，则码的最小距离(码字的最小重量)为t+1。

反过来说，若码的最小距离(码字的最小重量)为t+1则H 的任意t列线性无关而t+1列线性相关。

m	C
000	0000000
001	0011101
010	0100111
011	0111010
100	1001110
101	1010011
110	1101001
111	1110100

3.4 线性分组码的伴随式

$\mathbf{CH}^T = \mathbf{0}$　　R=C+E　　E=[$e_1\ e_2 \ldots e_n$]

$$\mathbf{S} = \mathbf{RH}^T = (\mathbf{E}+\mathbf{C})\mathbf{H}^T = \mathbf{EH}^T + \mathbf{CH}^T = \mathbf{EH}^T$$

1) $S = 0$，说明R 是一个码字;

2) $S \neq 0$，说明R 不是码字，传输过程产生了误码。

$S^T = HR^T$　　$S^T = HE^T$

例：某（5,2）系统线性码的生成矩阵是

$$\mathbf{G}=\begin{bmatrix}1 & 0 & 1 & 1 & 1\\ 0 & 1 & 1 & 0 & 1\end{bmatrix}$$

设收码是 $\mathbf{R}=(10101)$，问它是否是码字。

令 $E=\begin{bmatrix} e_1 & e_2 & \cdots & e_n \end{bmatrix}$ $\mathbf{H}=\begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \cdots & \mathbf{H}_n \end{bmatrix}$

（其中 表示 的列向量）

则

$$\mathbf{S}^T=\mathbf{H}\mathbf{E}^T=\begin{bmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \cdots & \mathbf{H}_n \end{bmatrix}\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}=\sum_{i=1}^{n} e_i\mathbf{H}_i$$

结论：

1) 当传输过程没有错误时，即 $E=\begin{bmatrix} 0 & 0 & \cdots & 0 \end{bmatrix}$，$S^T=0$

2)当发生一位错误时，S^T是校验矩阵的某一列。

3)当发生多个错误时，S^T为校验矩阵对应列的模2和。

例： 设(7,3)线性分组码的校验矩阵为

$$\boldsymbol{H}=\begin{bmatrix}1&0&1&1&0&0&0\\1&1&1&0&1&0&0\\1&1&0&0&0&1&0\\0&1&1&0&0&0&1\end{bmatrix}$$

$$d_{\min}=3+1$$

（1）接收码字R=(1010011)，

$$S^{T} = HR^{T} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$\Rightarrow$ 传输过程中没有误码，$\hat{C} = R$

（2）接收码字$\boldsymbol{R}$=(1110011)，

$$\boldsymbol{S}^{\mathrm{T}} = \boldsymbol{H}\boldsymbol{R}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$\boldsymbol{S}^{\mathrm{T}} = \boldsymbol{H}_2$ ，第2位出错， $\boldsymbol{E} = (0100000)$

$$\hat{C} = R + E = (1010011)$$

（3）接收码字$\boldsymbol{R}$=(0011011)，

$$\boldsymbol{S}^{\mathrm{T}} = \boldsymbol{H}\boldsymbol{R}^{\mathrm{T}} = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}$$

$\boldsymbol{S}^{\mathrm{T}}$与$\boldsymbol{H}$中的任一列都不相同，

$$\boldsymbol{S}^{\mathrm{T}} = \boldsymbol{H}_1 + \boldsymbol{H}_4 = \boldsymbol{H}_2 + \boldsymbol{H}_7 = \boldsymbol{H}_5 + \boldsymbol{H}_6$$

不能确定到底是哪两位出错，不能正确译码。

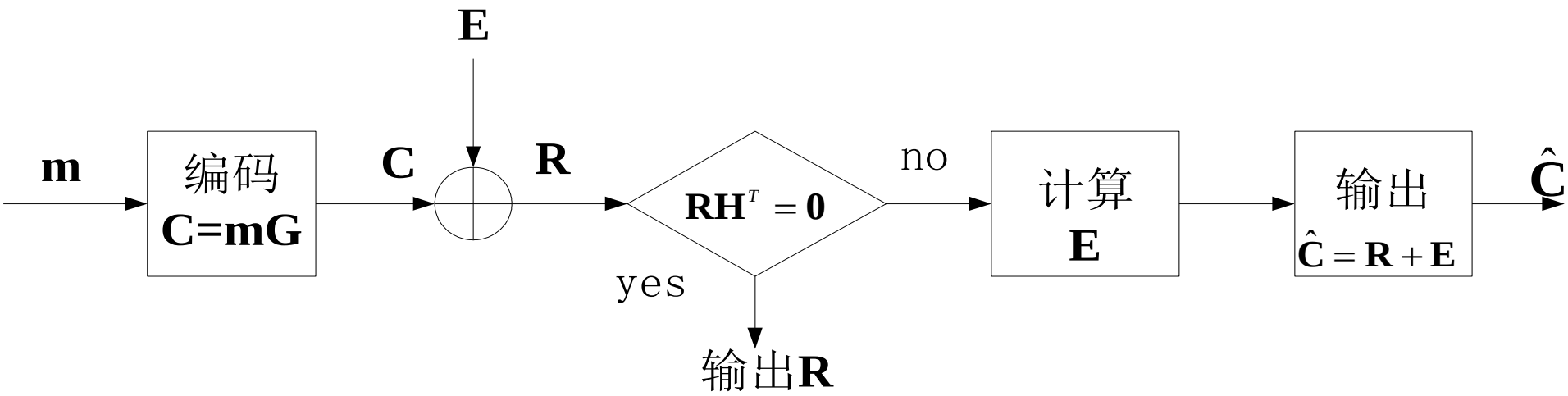
线性分组码的伴随式译码

$$\mathbf{S}^T = \mathbf{H}\mathbf{E}^T = [\mathbf{H}_1 \quad \mathbf{H}_2 \quad \cdots \quad \mathbf{H}_n]\begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix}$$

$$= e_1\mathbf{H}_1 + e_2\mathbf{H}_2 + \cdots + e_n\mathbf{H}_n$$

$$= \sum_{i=1}^{n} e_i\mathbf{H}_i$$

$$\mathbf{S} = \mathbf{E}\mathbf{H}^T \rightarrow \mathbf{E} \rightarrow \hat{\mathbf{C}} = \mathbf{R} + \mathbf{E}$$


E
m
编码
C=mG
C
R
$\mathbf{RH}^T=\mathbf{0}$
no
计算
E
输出
$\hat{\mathbf{C}}=\mathbf{R}+\mathbf{E}$
$\hat{\mathbf{C}}$
yes
输出R

4 汉明码

若（n，k）线性分组码能够纠正 u 个错误，则其校验位的数目必须满足

$$2^{n-k} \geq \sum_{i=0}^{u} C_n^i$$

$$M \sum_{i=0}^{u} C_n^i \leq 2^n$$

上式等号成立则称为完备码

如果是能纠正一位错误的完备码则 $2^r = n+1$

完备码具有下述特性：

（1）以每个发送码字为球心，以u为半径画一个球，那么每一个接收码字都落在其中一个球中，因此接收码字与发送码字的距离至多为u；

（2）所有差错数小于等于u的接收码字都能得到纠正；

（3）差错数大于等于$u+1$的接收码字，因为落在另一个球内被纠正为其他的发送码字。

完备码并不多见，我们知道的有$u=1$的汉明码、$u=3$的高莱码，以及$(n,1)$中n为奇数的重复码等。

000
000
001
010
100
111
110
101
011

$d_{\min} = 3$

完备码

$$\frac{2^n}{2^k} = 2^{n-k} = \sum_{i=0}^{u} C_n^i$$

00000
00001
00010
00100
01000
10000
01101
01100
01111
01001
00101
11101
00011
10011
01110
11100
10111
11010

00000
10111
.
.
.
01101
11010
.
.
.

非完备码

- 汉明码是一种能够纠正单个错误的完备码。
  - 汉明码最小码距 $d_{\min} = 3$
  - 设监督码共有r 位，对于汉明码必然有 $n = 2^r - 1$。
  - 通常汉明码可以表示成 $(2^r - 1, 2^r - 1 - r)$ 。

在同样的纠错能力下，汉明码的码率是最高的

$$R = \frac{2^r - 1 - r}{2^r - 1} = 1 - \frac{r}{2^r - 1}$$

- 汉明码监督矩阵构成的两种方式：
  - 按 r 位的二进制数的自然顺序从左到右排列（不包括全0列）。当发生可纠的单个错误时，伴随式为 $\boldsymbol{H}$ 阵中对应的列，译码比较方便。
  - 构成 H 阵的标准形式，$\mathbf{H} = [\mathbf{Q} \quad \mathbf{I}]$ 。非标准形式的监督矩阵可以通过列置换变成标准形式的监督矩阵，纠错能力保持不变。

例：构造一个r=3的二元（7,4）汉明码

解：r=3的汉明码，$n = 2^r - 1 = 7$

$$k = n - r = 4$$

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \xrightarrow{\text{列置换}} \left[\begin{array}{cccc:ccc} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 \end{array}\right] = \begin{bmatrix} \mathbf{Q} & \mathbf{I} \end{bmatrix}$$

$$\mathbf{Q}=\begin{bmatrix}1&1&1&0\\0&1&1&1\\1&1&0&1\end{bmatrix}$$

$$\mathbf{P}=\mathbf{Q}^{\mathrm{T}}=\begin{bmatrix}1&0&1\\1&1&1\\1&1&0\\0&1&1\end{bmatrix}$$

$$\mathbf{G}=\begin{bmatrix}\mathbf{I} & \mathbf{P}\end{bmatrix}=\left[\begin{array}{cccc|ccc}1&0&0&0&1&0&1\\0&1&0&0&1&1&1\\0&0&1&0&1&1&0\\0&0&0&1&0&1&1\end{array}\right]$$

信息比特	码字（循环1）	信息比特	码字（循环2）	信息比特	码字
0001	0001011	0011	0011101	0000	0000000
0010	0010110	0100	0100111		
0101	0101100	0111	0111010		
0110	0110001	1001	1001110	1111	1111111
1000	1000101	1010	1010011		
1011	1011000	1101	1101001		
1100	1100010	1110	1110100		

- 如果给汉明码添加一位奇偶校验位，可得到扩展汉明码：
  - 信息位保持不变，监督位增加一位。
  - 最小码距$d_{\min}=4$，$d_{\min}=4=1+2+1$ 可纠正一位错误，同时发现两位错误。
- 扩展汉明码的监督方程：

$$\mathbf{H}'=\begin{bmatrix} & & & 0\\ & \mathbf{H} & & 0\\ & & & \vdots\\ & & & 0\\ 1 & 1 & \cdots & 1 \end{bmatrix}=\left[\begin{array}{ccccccc:c} 0&1&1&1&1&0&0&0\\ 1&0&1&1&0&1&0&0\\ 1&1&0&1&0&0&1&0\\ \hline 1&1&1&1&1&1&1&1 \end{array}\right]$$

5 循环码

信息比特	码字（循环1）	信息比特	码字（循环2）	信息比特	码字
0001	0001011	0011	0011101	0000	0000000
0010	0010110	0100	0100111		
0101	0101100	0111	0111010		
0110	0110001	1001	1001110	1111	1111111
1000	1000101	1010	1010011		
1011	1011000	1101	1101001		
1100	1100010	1110	1110100		

- 循环码是线性分组码的一个重要子集。
- 循环码有严密的代数学理论基础，检错和纠错能力较强，而且编码和解码设备都不太复杂。
- 循环码除了具有线性分组码的一般性质外，还具有循环性：循环码中任一码字经过循环移位后，所得到的码字仍然是该码的码字。

- 设循环码的码字为$C=[c_{n-1} \quad c_{n-2} \quad \cdots \quad c_0]$，用码多项式表示为

$$C(x)=c_{n-1}x^{n-1}+c_{n-2}x^{n-2}+\ldots+c_1x+c_0$$

- 码字（**1100101**）可以表示为：

$$C(x)=1\bullet x^6+1\bullet x^5+0\bullet x^4+0\bullet x^3+1\bullet x^2+0\bullet x^1+1$$

$$=x^6+x^5+x^2+1$$

1)对于二进制码，码多项式的每个系数不是**0**就是**1**。

2) 仅是码元位置的标记。我们并不关心x的取值。

$$C(x) = c_{n-1}x^{n-1} + \cdots + c_2x^2 + c_1x + c_0$$

$$C^{(1)}(x) = xC(x) = c_{n-2}x^{n-1} + \cdots + c_1x^2 + c_0x + c_{n-1}$$

$$C^{(2)}(x) = x^2C(x) = c_{n-3}x^{n-1} + \cdots + c_0x^2 + c_{n-1}x + c_{n-2}$$

$$\cdots$$

$$C^{(n-1)}(x) = x^{n-1}C(x) = c_0x^{n-1} + \cdots + c_3x^2 + c_2x + c_1$$

循环码的循环特性可以用码多项式来证明：

●在整数运算中，有模*n*运算。若一个整数m可以表示为:

$$\frac{m}{n} = q + \frac{p}{n}, \qquad p < n$$

则

$$m \equiv p \qquad (\text{模} n)$$

在模n运算下，一整数m等于其被n除所得的余数。

- 在码多项式运算中也有类似的按模运算法则。

若一任意多项式 $M(x)$被一个n次多项式 $N(x)$ 除，得到商式 $Q(x)$和一个次数小于 n 的余式 $P(x)$ ，也就是：

$$\frac{M(x)}{N(x)} = Q(x) + \frac{P(x)}{N(x)}$$

则可以写为：$M(x) \equiv P(x) \quad (\text{模} N(x))$

- 在循环码中，若$C(x)$是一个长为n的许用码字，则$x^i C(x)$在模 $x^n + 1$运算下，也是一个许用码字。

例：某循环码的一个码字为1100101，则

$$C(x) = x^6 + x^5 + x^2 + 1$$

若将此码左移一位，得：

$$xC(x) = x^7 + x^6 + x^3 + x$$

$$x^7 + x^6 + x^3 + x \mod x^7 + 1 = x^6 + x^3 + x + 1$$

对应的码字为1001011

(即将码字1100101

循环左移一位）。

$$\begin{array}{r} 1 \\ x^7+1 \overline{)\, x^7 + x^6 + x^3 + x} \\ x^7 \qquad\qquad +1 \\ \hline x^6 + x^3 + x + 1 \end{array}$$

信息比特	码字（循环1）	信息比特	码字（循环2）	信息比特	码字
0001	0001011	0011	0011101	0000	0000000
0010	0010110	0100	0100111		
0101	0101100	0111	0111010		
0110	0110001	1001	1001110	1111	1111111
1000	1000101	1010	1010011		
1011	1011000	1101	1101001		
1100	1100010	1110	1110100		

- 在(n,k)循环码中，除了全0码字以外，其他码字的连0个数最多只有k-1个。

- 从码中取出一个前面k-1位都是0的码字，定义这个码字的码多项式为生成多项式 $g(x)$。

- 生成多项式的次数为 $n-k$，且常数项不为0。

- 为了保证构成的生成矩阵**G**的各行线性不相关，通常用 $g(x)$ 这样来构造生成矩阵。

$$\boldsymbol{G}(x)=\begin{bmatrix} x^{k-1}g(x) \\ x^{k-2}g(x) \\ \vdots \\ xg(x) \\ g(x) \end{bmatrix} \qquad g(x)=x^{n-k}+\cdots+g_2x^2+g_1x+1$$

$$\mathbf{G}=\begin{bmatrix}
1 & g_{n-k-1} & g_{n-k-2} & \cdots & g_1 & 1 & 0 & 0 & \cdots & 0 \\
0 & 1 & g_{n-k-1} & g_{n-k-2} & \cdots & g_1 & 1 & 0 & \cdots & 0 \\
0 & 0 & 1 & g_{n-k-1} & g_{n-k-2} & \cdots & g_1 & 1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & 1 & g_{n-k-1} & g_{n-k-2} & \vdots & \cdots & 1
\end{bmatrix}$$

k-1 个 0

- 设信息码元为$(m_{k-1}m_{k-2}\cdots m_0)$时，由生成矩阵得到相应码字的码多项式：

$$C(x)=[m_{k-1}m_{k-2}\cdots m_0]\mathbf{G}(x)=[m_{k-1}m_{k-2}\cdots m_0]\begin{bmatrix} x^{k-1}g(x)\\ x^{k-2}g(x)\\ \vdots\\ xg(x)\\ g(x)\end{bmatrix}$$

$$=m_{k-1}x^{k-1}g(x)+m_{k-2}x^{k-2}g(x)+\cdots+m_0g(x)$$

$$=(m_{k-1}x^{k-1}+m_{k-2}x^{k-2}+\cdots+m_0)g(x)$$

所有码多项式必定为$g(x)$的倍式。

$$C(x)=(m_{k-1}m_{k-2}\cdots m_1m_0)\mathbf{G}(x)=\sum_{i=0}^{k-1}m_i x^i g(x)=m(x)g(x)$$

$$x^n+1=h(x)g(x)$$

$$=(h_k x^k+\cdots+h_1x+h_0)(x^{n-k}+\cdots+g_2x^2+g_1x+1)$$

$$h(x)=h_k x^k+\cdots+h_1x+h_0=\sum_{i=0}^{k}h_i x^i$$ —致校验多项式

$$C(x)=\sum_{i=0}^{n-1}c_i x^i \qquad E(x)=\sum_{i=0}^{n-1}e_i x^i \qquad R(x)=\sum_{i=0}^{n-1}r_i x^i$$

$$R(x)=C(x)+E(x) \qquad S(x)=\frac{R(x)}{g(x)}=\frac{C(x)+E(x)}{g(x)}=\frac{E(x)}{g(x)}$$

$$S(x)=\frac{E(x)}{g(x)}=E(x)[\mathrm{mod}\, g(x)]$$

循环码的伴随多项式 $S(x)$ 就是用接收码多项式除以生成多项式 $g(x)$ 所得的余式。

- 译码可分为三步：

1)由接收到的码多项式 $R(x)$计算伴随多项式 $S(x)$；

2)由伴随多项式$S(x)$ 确定错误图样$E(x)$；

3)将错误图样 $E(x)$ 与$R(x)$ 相加，纠正错误。