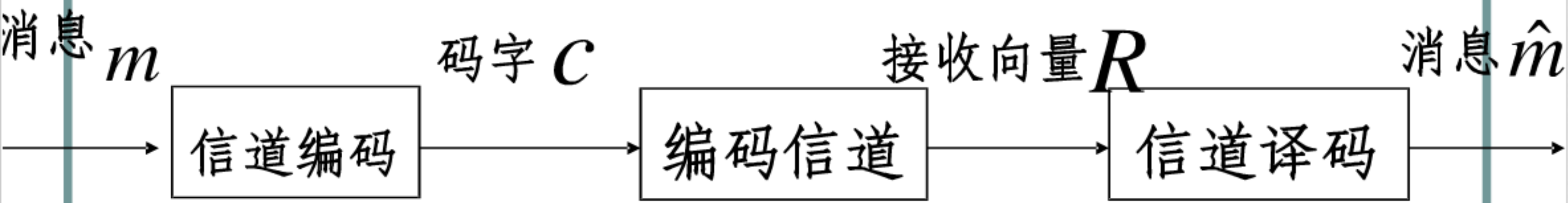
信道编码

- In the seminal paper ‘A mathematical theory of communication’ published in 1948, Shannon show the theorem about the capacity of the channel. This marked the birth of coding theory.
- It has found wide spread applications:
  communication systems, compact disc player, cryptography, storage technology, etc.

- 信道编码主要解决信息在信道上的正确传输为目标的编码。
- 纠错编码

  用于检测与纠正信号传输过程中因噪声干扰

  导致的差错。



编码信道模型

$$c=\left(c_0,c_1,\cdots,c_{n-1}\right),\qquad c_i\in\{0,1\}$$

$$R=\left(r_0,r_1,\cdots,r_{n-1}\right),\qquad r_i\in\{0,1\}$$

- For example, consider the source encoding of four fruits:

apple	banana	cherry	grape
00	01	10	11
000	011	101	110

(there is only one error introduced, detect one error)

00000	01111	10110	11001

(there is one error introduced, correct one error)

Goal of channel coding:

- fast encoding of messages;
- easy transmission of encoded messages;
- fast decoding of received messages;
- maximum transfer of information per unit time;
- maximal detection or correction capability.

Definition. Let $A = \{a_1, a_2, \cdots, a_q\}$ be a set of size q, which we refer to a code alphabet and whose elements are called code symbols.

1. A q-ary word of length n over A is a sequence $\mathbf{w} = w_1 w_2 \cdots w_n$ with each $w_i \in A$ for all i. Equivalently, $\mathbf{w}$ may also be regraded as the vector $(w_1, \cdots, w_n)$.
2. A q-ary block code of length n over A is a nonempty set C of q-ary words having the same length n.
3. An element of C is called a codeword in C.

4. The number of codewords in C, denoted by $|C|$, is called the size of C.

5. The rate of a code C of length n is defined to be $log_q|C|/n$.

6. A code of length n and size M is called an (n, M)-code.

几个概念

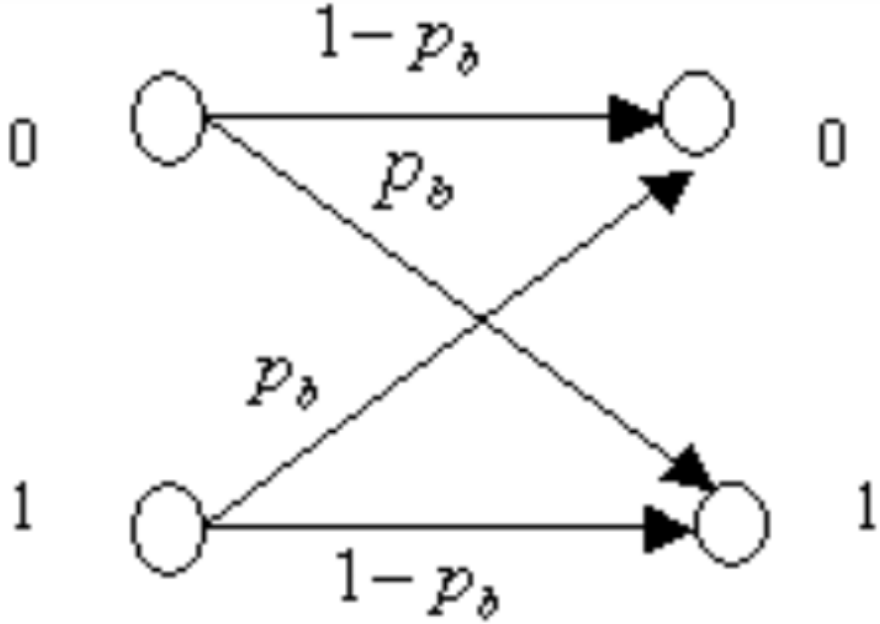
- 当码字 C和接收向量R均由二元序列表示，称编码信道为二进制信道。
- 如果对于任意的n 都有

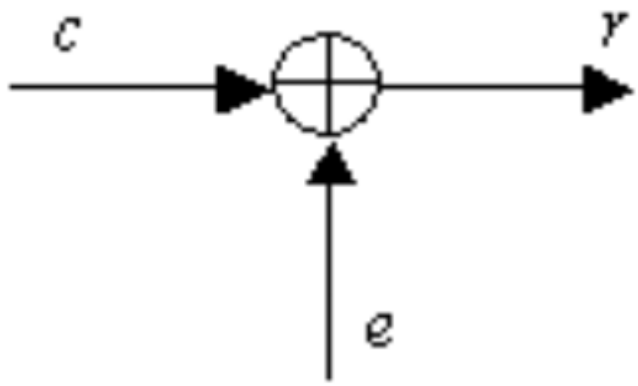
$$P(R|C)=\prod P(r_i|c_i)$$

则称此二进制信道为无记忆二进制信道.

- 如果 $P(0|1)=P(1|0)=p_b$

则称此信道为无记忆二进制对称信道 BSC.





BSC转移概率 **BSC编码信道**

BSC输入输出关系等效为

$$\begin{cases} r = c + e \qquad \bmod 2 \\ p(e=1) = p_b,\, p(e=0) = 1 - p_b \end{cases}$$

差错图案

差错图案：随机序列 (e_i) 或 $e=[e_0,e_1,\cdots;,e_{n-1}]$

第 i 位上的一个随机错误：$e_i=1$

例如:C=[10000],e=[01000],R=?

R=(11000)

码元的组成及关系

- 消息序列m总以k个码元为一组传输，称k个码元的码组为信息码组。
- 信道编码器按一定规则对每个信息码组附加一些多余的码元，构成长为n个码元的码组c（信道编码）。
- 附加的r=n-k个码元称为监督码元

检错和纠错能力

- 常用汉明距离来描述检纠差错的数目，对于两个n长的向量 $u=(u_1,\dots,u_n),v=(v_1,\dots,v_n)$,它们的汉明距离为

$$d(u,v)=\sum_{i=1,u_i\neq v_i}^{n} 1.$$

- 码C的最小汉明距离 $d_{\min}$：任意两码字之间的汉明距离的最小值

$$d_{\min}=\min_{c\neq c'} d(c,c').$$

汉明距离的性质

定理　对任意的 $x,y,z \in \{0,1\}^n$,汉明距离具有如下性质：

(1) (非负性) $d(x,y) \geq 0;$

(2) (自反性) $d(x,y)=0$ 当且仅当 x=y.

(3) (对称性) $d(x,y)=d(y,x).$

(4) (三角不等式) $d(x,y) \leq d(x,z)+d(z,y).$

- Definition. A code C of length n, size M and distance d is referred to as an (n,M,d)-code. The numbers n, M and d are called the parameters of the code.
- $M=|C|$.

Definition. Let l be a positive integer. A code C is l-error-detecting if , whenever a codeword incurs at least one but at most l errors, the resulting word is not a codeword.

Definition. Let t be a positive integer. A code C is t-error-correcting if minimum distance decoding is able to correct t or fewer errors.

大数逻辑译码 (Maximum likelihood decoding)

Definition. The maximum likelihood decoding (MLD) rule will conclude that c_x is the most likely codeword transmitted if c_x maximizes the forward channel probability, i.e.,

$$P(x \text{ received} | c_x \text{ sent}) = \max_{c \in C} P(x \text{ received} | c \text{ sent}).$$

There are two kinds of MLD:

- Complete maximum likelihood decoding(CMLD). If a word **x** is received, find the most likely codeword transmitted. If there are more than one such codewords, select one of them arbitrarily.
- Incomplete maximum likelihood decoding(CMLD). If a word **x** is received, find the most likely codeword transmitted. If there are more than one such codewords, request a retransmission.

Minimum distance decoding (最小距离译码)

Definition. If a word $\mathbf{x}$ is received, the minimum distance decoding rule will decode $\mathbf{x}$ to $\mathbf{c}_{\mathbf{x}}$ is minimal among all the codewords in C, i.e.,

$$d(\mathbf{x}, \mathbf{c}_{\mathbf{x}}) = \min_{\mathbf{c} \in C} d(\mathbf{x}, \mathbf{c}).$$

Similarly, complete and incomplete minimum distance decoding rule.

检错和纠错能力

定理 对一个最小距离为 d_{min} 纠错码，如下三个结论仅有其中任意一个结论成立，

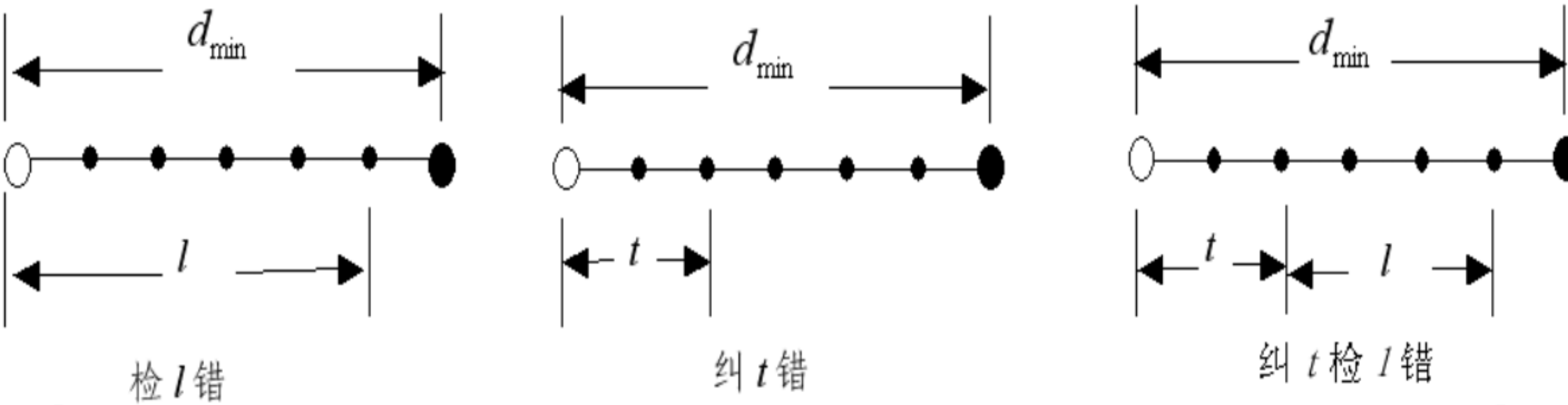
（1） 可以检测出任意小于等于 $l = d_{min} - 1$ 个差错；

（2） 可以纠正任意小于等于 $t = \left[\dfrac{d_{min} - 1}{2} \right]$ 个差错；

（3） 可以检测出任意小于等于 l 同时纠正小于等于 t 个差错，其中 l 和 t 满足

$$\begin{cases} l + t \le d_{min} - 1 \\ t < l \end{cases}$$

检错和纠错能力



最小码距与检纠错能力

Proof. (1). Suppose $d(C) \geq l+1$. If $\mathbf{c} \in C$ and $\mathbf{x}$ are such that $1 \leq d(\mathbf{c}, \mathbf{x}) \leq l < d(C)$, then $\mathbf{x} \in C$; hence, C is l-error-detecting.

(2) Suppose that $d(C) \geq 2t+1$. Let $\mathbf{c}$ be the codeword sent and let $\mathbf{x}$ be the word received. If t or fewer errors occur in the transmission, then $d(\mathbf{x}, \mathbf{c}) \leq t$. Hence, for any codeword $\mathbf{c}' \in C$, $\mathbf{c} \neq \mathbf{c}'$, we have

$$\begin{aligned} d(\mathbf{x}, \mathbf{c}') &\geq d(\mathbf{c}, \mathbf{c}') - d(\mathbf{x}, \mathbf{c}) \\ &\geq 2t+1-t \\ &= t+1 \\ &> d(\mathbf{x}, \mathbf{c}). \end{aligned}$$

- (3) holds from (1) and (2).

定义　设 n 是一个正整数，A={0,1}, 定义一个长为n 的二元奇偶校验码 C是 A^n中 包含偶数个1字构成的集合。

例　设A={0,1}, n=4, 则有

A^n = { 0000, 0001,0010, 0011, 0100, 0101, 0110, 0111,1000,1001,1010,1011,1100,1101,1110,1111 }

C={0000, 0011, 0101, 0110, 1001, 1010, 1100, 1111}

这个奇偶校验码能够检测1个错误而不能纠正一个错误。Why？

- 定义： 设 n 是一个正整数，A={0,1}，则 A^n 中有 2^n个元素。定义一个长为n的二元重复码C是只包含两个元素的集合，即只包含全0和全1的字符串。
- 例　设A={0,1}, n=5.

C={00000,11111}.

最小距离为5，可以纠2个错。

码的最小距离越大， 它的检错和纠错能力也就相应的越大。因此在纠错码中总是要求码具有较大的最小距离。

Let $C = \{00000, 00111, 11111\}$ be a binary code.
Then $d(C) = 2$ since

$$\begin{aligned} d(00000, 00111) &= 3 \\ d(00000, 11111) &= 5 \\ d(00111, 11111) &= 2. \end{aligned}$$

Hence, C is a binary $(5, 3, 2)$-code.