信息论与编码知识点梳理

- 信息的特征是不确定性。
- **自信息量**：度量某一事件、信源某一具体符号的不确定性。 $I(x_i) = -\log p(x_i)$
- **信源熵**：信源的平均不确定度。 $H(X) = E[I(X)] = -\sum_i p(x_i)\log p(x_i)$
- **条件熵H(Y | X)**：已知X 后，关于Y 的平均不确定度。 $H(Y|X) = -\sum_{ij} p(x_i, y_j)\log p(y_j|x_i)$
- **联合熵H(X , Y)** ：表示 X 和 Y 同时发生的不确定度。$H(X,Y) = -\sum_{ij} p(x_i, y_j)\log p(x_i, y_j)$

$$\boxed{H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)}$$

- **互信息量**：已知某一条件Y，使得对X的不确定度减少了。衡量条件Y提供了多少关于X的信息量

$$I\left(x_i;y_j\right)=\log\frac{p(x_i\mid y_j)}{p(x_i)}=\log\frac{\text{后验概率}}{\text{先验概率}}\quad(i=1,2,\ldots,n,\ j=1,\ldots,m)$$

- **平均互信息量**：平均意义上的互信息量

$$I(X;Y)=\sum_{i,j}p(x_i,y_j)\log\frac{p(x_i/y_j)}{p(x_i)}$$

$$I(X;Y)=H(X)-H(X/Y)$$

信息量与信源熵主要公式

信源

信源的单个符号

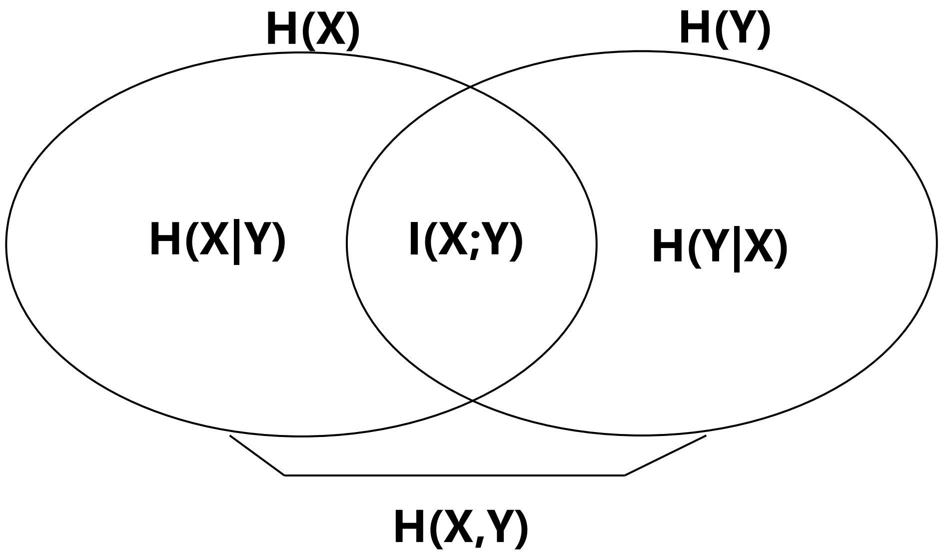
- 自信息量 $I(x_i) = -\log p(x_i)$
- 条件自信息量$I(x_i|y_j) = -\log p(x_i|y_j)$
- 互信息量 $I(x_i;y_j) = \log\frac{p(x_i|y_j)}{p(x_i)}$

离散信源

- 信源熵 $H(X) = E[I(X)] = -\sum_i p(x_i)\log p(x_i)$
- 条件熵 $H(Y|X) = -\sum_{ij} p(x_i, y_j)\log p(y_j|x_i)$
- 联合熵 $H(X,Y) = -\sum_{ij} p(x_i, y_j)\log p(x_i, y_j)$
- 平均互信息$I(X;Y) = \sum_{i,j} p(x_i, y_j)\log\frac{p(x_i/y_j)}{p(x_i)}$

各概念的关系

- $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$
- $I(X;Y) = H(X) - H(X|Y)$
- $I(Y;X) = H(Y) - H(Y|X) = I(X;Y)$
- $I(X;Y) = H(X) + H(Y) - H(X,Y)$



通信模型

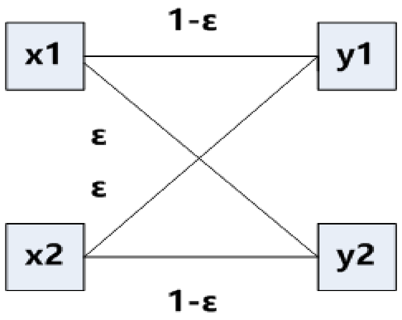
信源　　信道　　信宿

从信息量传输的角度看：发出的信息量H(X) → **信道中损失的信息量H(X|Y)** → **接收端获得的信息量I(X;Y)**

- H(X|Y)：疑义度，表示由于信道上存在干扰和噪声而损失掉的平均信息量。
- H(Y|X)：噪声熵。
- 全损信道：干扰很大，难以从Y中提取X的有效信息，信源发出的所有信息都损失在信道中。
  - I（X;Y）=0
  - 例如：加密编码
- 无损信道：没有干扰，接收端能完全收到信源发出 的信息。
  - I（X;Y）=H（X）

第3章 信道容量

- **信道容量**：信道所能传送的最大信息量。 $C=\max\limits_{p(x_i)} I(X;Y)$
- **单位时间的信道容量**：单位时间内信道所能传送的最大信息量。$C_t=\frac{1}{t}\max\limits_{p(x_i)} I(X;Y)$
- **对称DMC信道**
  - 如果转移概率矩阵P的每一行都是第一行的置换(即包含同样元素)，称该矩阵是输入对称的。
  - 如果转移概率矩阵P的每一列都是第一列的置换(即包含同样元素)，称该矩阵是输出对称的。
  - 输入输出都对称的离散无记忆信道称为对称DMC信道。
  - **信道容量**： $C=\log m-H(Y/x_i)=\log m+\sum_{j=1}^{m} p_{ij}\log p_{ij}$
- **BSC信道（二进制对称DMC信道）**
  - **信道容量**： $C=\log 2-H(\varepsilon,1-\varepsilon)=1-H(\varepsilon)$



- **香农公式**：AWGN信道的信道容量。$C = W\log(1+SNR) = W\log(1+\frac{P_S}{N_0 W})$
  - W：信道频带宽度，简称带宽，单位Hz。
  - SNR：signal to noise ratio，信噪比，是信号功率（单位为W）与噪声功率（单位为W）的比值。
  - P_s：信号发射功率。
  - N_0：高斯白噪声的单边功率谱密度。

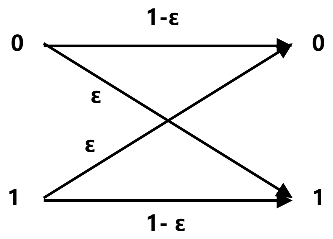
- **提高信道容量的方式**
  - 提升信道带宽
  - 提升信噪比
    - 提升发射功率
    - 降低信道噪声

- **香农限**：当带宽不受限制时，传送1比特信息，信噪比最低只需-1.6dB。

信道主要公式

信道

- BSC信道



$$P=\begin{bmatrix}1-\varepsilon & \varepsilon \\ \varepsilon & 1-\varepsilon\end{bmatrix}$$

$$C=\log 2-H(\varepsilon,1-\varepsilon)=1-H(\varepsilon)$$

- DMC信道 → 对称DMC信道

$$C=\log m-H(Y/x_i)=\log m+\sum_{j=1}^{m}p_{ij}\log p_{ij}$$

- AWGN信道 → 香农公式

$$C=W\log(1+\frac{P_S}{N_0W})$$

- 数学模型：转移矩阵
- 信道容量 $C=\max\limits_{p(x_i)} I(X;Y)$
- 单位时间的信道容量 $C_t=\frac{1}{t}\max\limits_{p(x_i)} I(X;Y)$

- 提高信道容量的方法：提高带宽、提高发射功率、降低信道噪声。
- 香农限：当带宽不受限制时，传送1比特信息，信噪比最低只需-1.6dB

第5章 信源编码

信源编码 —— 无失真信源编码 —— 哈夫曼编码

- 编码方法

- 目的：提高通信系统的有效性
- 码树

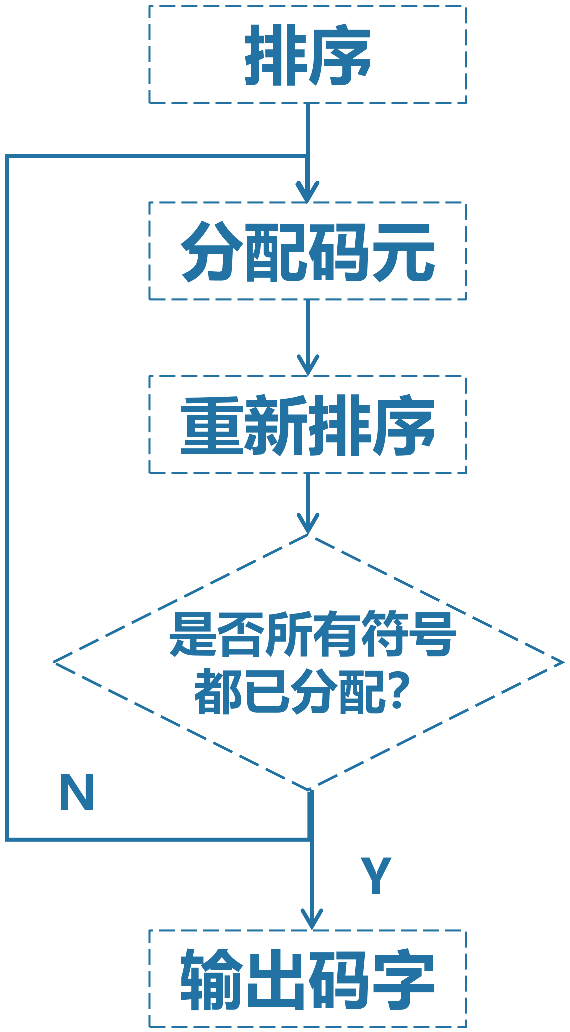
- 定长编码
- 无失真条件 $R \geq H(\mathbf{X})$
- 输出信息率 $R = \frac{K_L}{L}\log m$
- 编码效率 $\eta = \frac{H(\mathbf{X})}{R}$

- 变长编码
- 无失真条件 $H(\mathbf{X}) \leq \overline{R} < H(\mathbf{X}) + \frac{\log m}{L}$
- 平均输出信息率 $\overline{R} = \frac{\overline{K_L}}{L}\log m$
- 编码效率 $\eta = \frac{H(\mathbf{X})}{\overline{R}}$
- 码字平均长度 $\overline{K_L} = \sum_i p(\mathbf{x}_i)K_i$
- 平均码长 $\overline{K} = \overline{K_L} / L$

- L：输入编码器的信息位长度
- m：进制数
- K_L：编码后的码字长度
  - 定长编码中，K_L是定值，
  - 变长编码中，$\overline{K_L}$ 是码字平均长度
- η：编码前后的信息量比值
- 平均码长$\overline{K}$：每一个信息位用几位编码来表示
  - $L = 1$、二进制的情况下，$\overline{K_L} = \overline{K}$

哈夫曼编码

➢哈夫曼编码



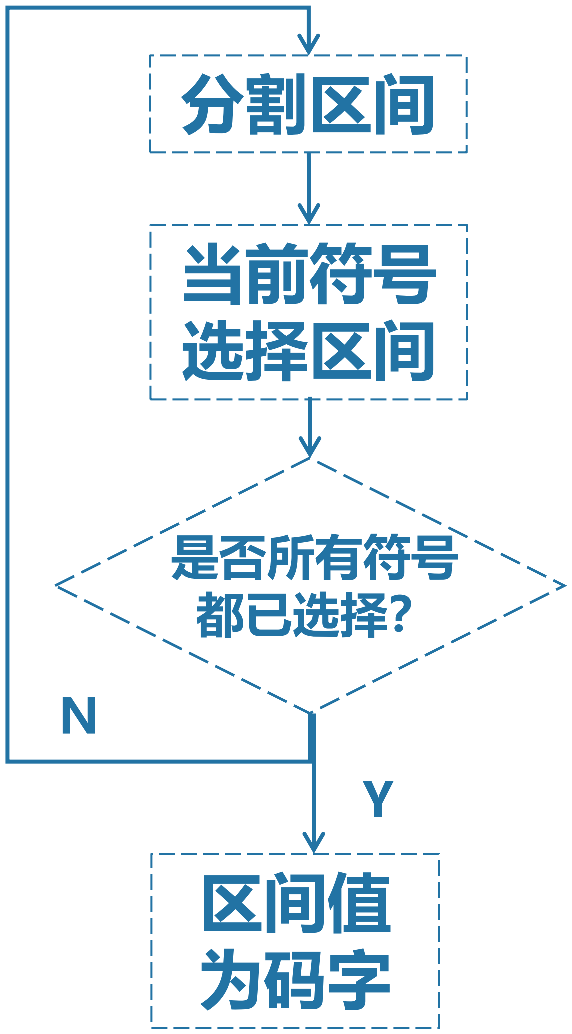
- 将信源符号按概率递减的次序排列
- 取概率最小的两个符号，分配“0”和“1”码元
- 将这两个符号的概率相加，视为一个新的符号
- 与未分配的符号一起，重新排序
- 重复以上步骤，直到所有符号都分配完毕
- 从最后一级返回得到各个符号对应的码字

信源符号xi	概率p(xi)	编码过程					码字
x1	0.20	0.20	0.26	0.35	0.39	0.61 —0	10
x2	0.19	0.19	0.20	0.26	0.35 —0	0.39 —1	11
x3	0.18	0.18	0.19	0.20 —0	0.26 —1		000
x4	0.17	0.17	0.18 —0	0.19 —1			001
x5	0.15	0.15 —0	0.17 —1				010
x6	0.10 —0	0.11 —1					0110
x7	0.01 —1						0111

➢ 注意：

- 写码字的方法：从最后一级往回走，走回信源符号，然后按**从后往前**的顺序，写下路径上经过的码元。
- 组合概率排序靠前
- 同级分支的0/1可以互换，**答题时统一要求：概率较大的分配0，较小的分配1**

➢算术编码的编码步骤

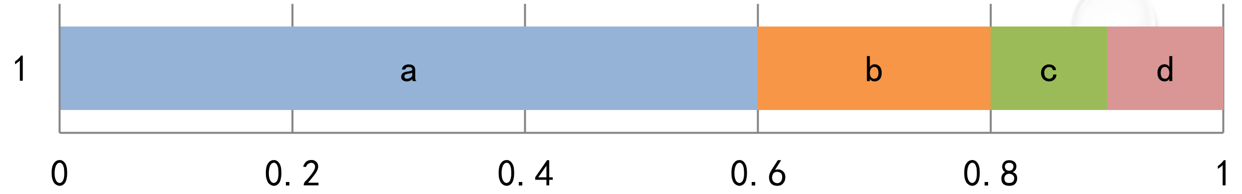


初始区间T=[0, 1)
根据信源符号出现的概率把T分成若干个子区间
区间长度取决于出现概率
为当前符号选择区间
将此区间作为继续分割的基础

重复以上步骤，直到所有符号都分配完毕

当前的区间值为码字

➢编码例子：信源符号：a、c、数据结束符



- □第1个符号是a
  - 落在区间[0, 0. 6）之间
- □第2个符号是c
  - 将区间[0, 0. 6)分成子区间：
  - a 的区间是 [0， 0. 36)——[0， 0. 6) 的 60%
  - b 的区间是 [0. 36， 0. 48)——[0， 0. 6) 的 20%
  - c 的区间是 [0. 48， 0. 54)——[0， 0. 6) 的 10%
  - 数据结束符 的区间是 [0. 54， 0. 6)——[0， 0. 6) 的 10%
- □第3个符号是数据结束符
  - a 的区间是 [0. 48， 0. 516)
  - b 的区间是 [0. 516， 0. 528)
  - c 的区间是 [0. 528， 0. 534)
  - 数据结束符 的区间是 [0. 534， 0. 540)
- □十进制的情况下，编码结果可以在[0. 534， 0. 540)之间任意选择一个数，如0. 538
- □转成整数538，编码结束

第6章 信道编码

- **信源编码目的：**提高数字通信系统的有效性。
- **信道编码目的：**提高数字通信系统的可靠性。
- **差错控制系统**
  - **前向纠错**：发端信息经过纠错编码后实行传送，而接收端通过独立的纠错译码，自动纠正传递过程的差错。
  - **自动反馈重传**：接收端若发现接收码字有错，则通过反向信道通知发送端重发该码字，如此反复，直到接收端认为接收正确为止。
  - **混合纠错**：前向纠错和反馈重传的结合，兼有检错和纠错两种能力。
- **有扰离散信道编码定理**
  - 信道编码器的码率R≤信道容量C时，一定存在一种信道编码方式可以实现无差错传输。
- **汉明距离：**两个（相同长度）码字对应位不同的数量。
  - 对两个码字进行异或运算，并统计结果为1的个数。
- **汉明重量：**码字对应于同样长度的零字符串的汉明距离，也就是码字中非零元素的个数。

第6章 知识点小结

信道编码

- 信道编码的目的：提高可靠性
- 差错图样E
- 差错控制系统FEC、ARQ、HEC
- 有扰离散信道编码定理$P_e < e^{-NE(R)}$
  - 减小Pe的三种方法：
    - ✓ 增大信道容量C
    - ✓ 减小传信率R
    - ✓ 增大码长N
- 最大后验概率译码、最大似然译码
  - BSC信道下，简化为最小汉明距离译码

线性分组码

- 生成矩阵**G**
- 校验矩阵**H**
- 系统形式的**G**和**H**
- 伴随式**S**
- 最小距离d_{min}
- 纠检错能力

汉明码

- 定义、约束条件
- 译码方法

循环码

CRC校验码

- 码多项式
- 生成多项式
- 生成矩阵
- 系统循环码的构造方法
- 判断收码是否有错

卷积码

- 状态图
- 网格图
- 维特比译码

线性分组码

- **(n，k) 线性分组码**：
- k=信息位长度，n=码长，n-k=校验位长度
- 两个矩阵
  - 生成矩阵G **MG=C**
  - 校验矩阵H **$CH^T=0$ $RH^T=EH^T=S$**
  - 系统码 $G=[I_k \mid P]$ $H=[P^T \mid I_{n-k}]$

- M：消息
- G：生成矩阵
- C：许用码字
- H：校验矩阵
- R：接收到的码字
- E：差错图案
- S：伴随式

- 对一个给定的码集，最小距离 d_{min}=非零码字的最小汉明重量
- 对（n，k）线性分组码所能组成的码集，$d_{min} \le n-k+1$
- 检错能力 $e=d_{min}-1$
- 纠错能力 $t=\left\lfloor \frac{d_{min}-1}{2} \right\rfloor$

汉明码

➢**汉明码**：纠1个错误的线性分组码，d_{min} = 3的完备码。

- 汉明码的 n 和 k 满足：

$$\begin{cases} m = n - k \quad (m \geq 3) & \text{——校验位长度} \\ n = 2^m - 1 & \text{——码字长度} \\ k = 2^m - 1 - m & \text{——信息位长度} \end{cases}$$

➢汉明码纠错解题思路

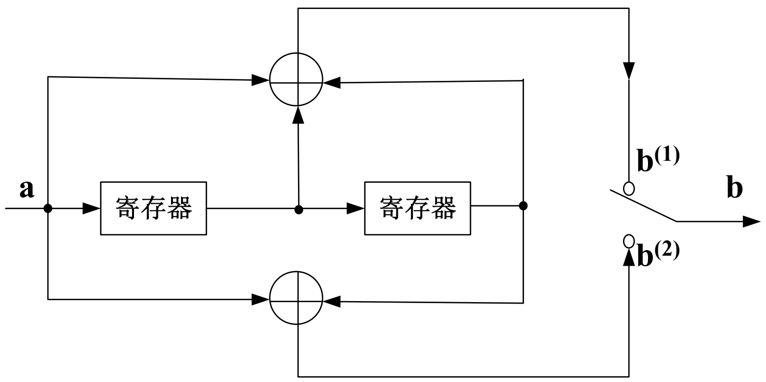
- （1）计算S=RH^T
- （2）汉明码只能纠一位错误，观察S等于H矩阵的哪一列，说明收码的哪一位符号出错
- （3）修改R的出错位，得到C

循环码

- 生成多项式
  - 对(x^n+1)作因式分解，找出其$(n-k)$次因式
- 系统码字编码方法
  - 步骤1：将信息多项式$m(x)$预乘 即左移$(n-k)$位
  - 步骤2：将 $m(x)x^{n-k}$ 除以 $g(x)$ 得余式 $r(x)$
  - 步骤3：$C(x)=x^{n-k}m(x)+r(x)$
- 循环码的生成矩阵（参考《通信原理》P347 ）
  - 由生成多项式推出生成矩阵
  - $G(x)=\begin{bmatrix} x^{k-1}g(x) \\ x^{k-2}g(x) \\ \vdots \\ xg(x) \\ g(x) \end{bmatrix}$ →生成矩阵**G**

卷积码

- **卷积码**
- 卷积码表示为（n，k，m），n：码字长度，k：信息长度，m：移位寄存器个数
- 约束长度=m+1



二进制(2,1,2)卷积码编码器

- **状态图**

写表达式

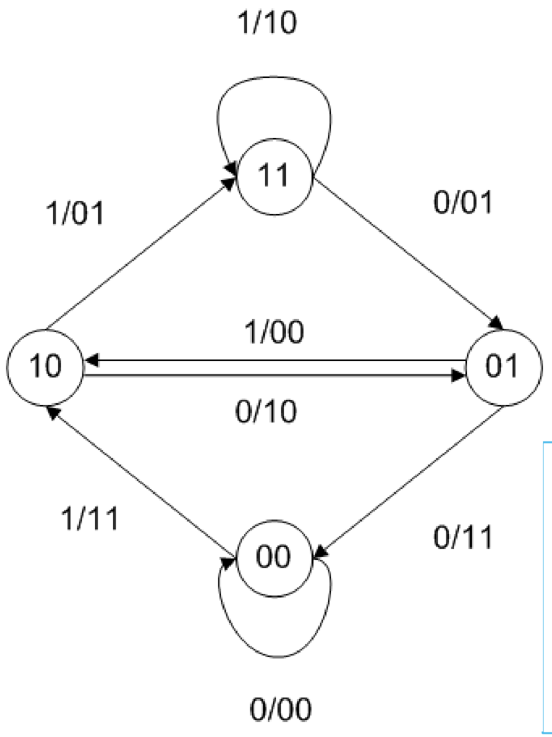
- 根据编码器电路图，写出输出和输入之间的逻辑运算表达式

列表

- 列表：寄存器当前时刻状态、输入、输出、寄存器下一时刻状态

画图

- 按寄存器当前时刻状态和下一时刻状态画出转移图



a/bb：输入信息a时，输出码字为bb
框内的xx：表示寄存器状态
箭头：寄存器状态转移

➢卷积码的网格图

➢画图步骤

排列状态

- 纵向依次排列各个状态，S0排第一行

连接状态

- 从状态S0开始，根据状态图，连接状态转移关系
- 在连接线上标出相应的输出

- 注意：输入信息序列、编码输出序列和网格图中的一条路径是唯一对应的
- 虚线表示输入0，实线表示输入1

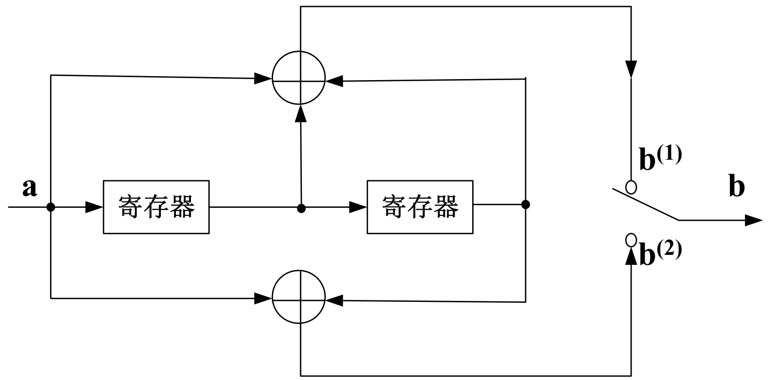
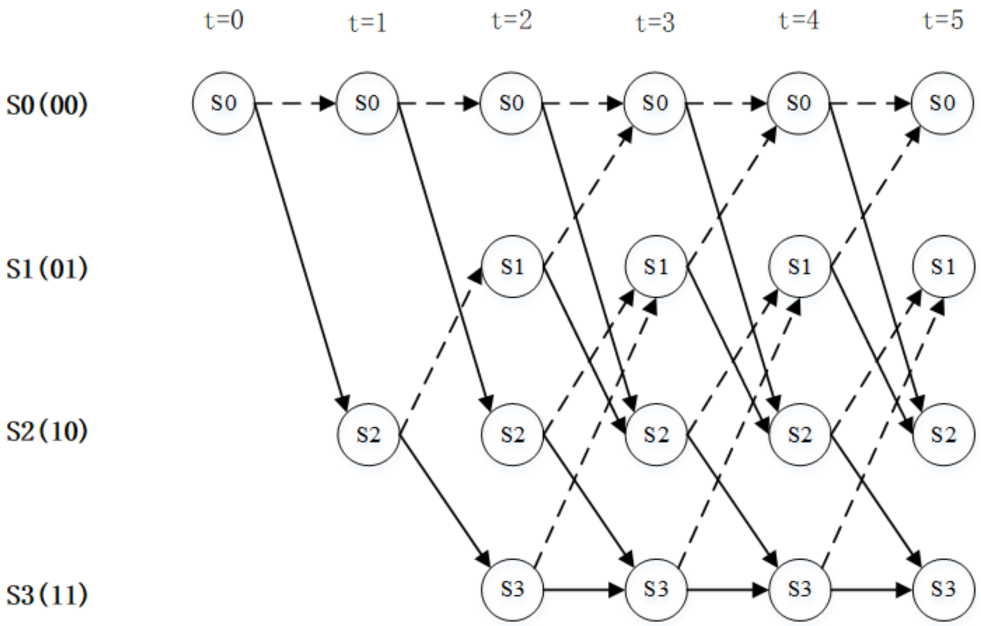


图4. 21 二进制(2,1,2)卷积码编码器



- 网格图简化
  - 以4状态的网格图为例，$t = 5$开始是重复的
  - $t = 0 \sim t = 2$ 不完整

熟练之后可以省去前面的步骤，
直接画出简化的网格图