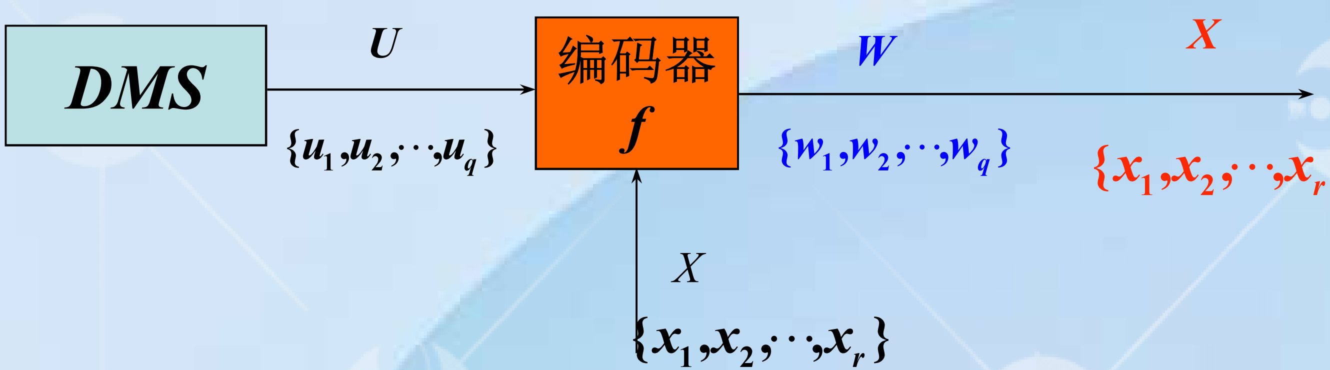
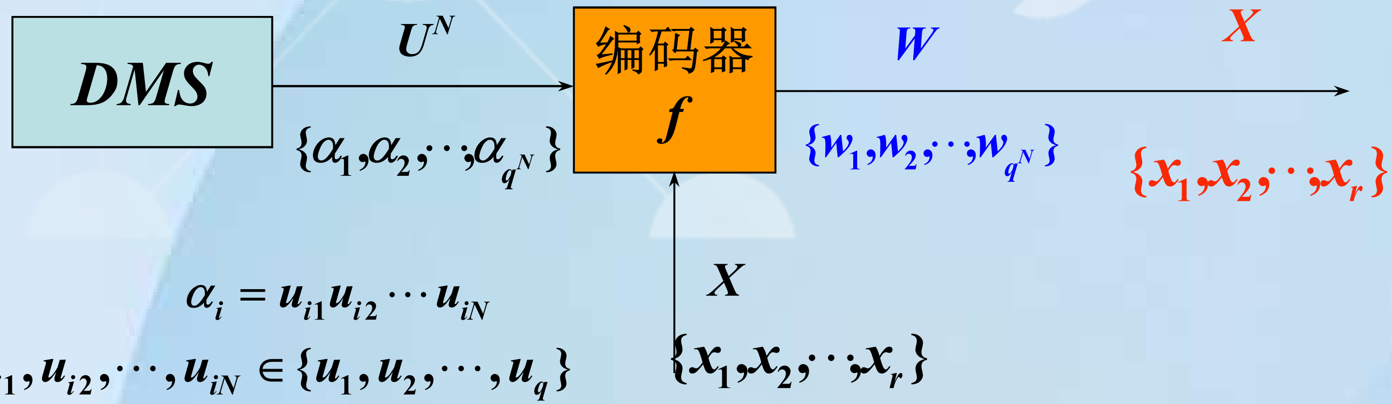
4.3 定长编码定理和定长编码方法

1、对信源输出的符号序列进行编码

对信源U的
单个符号
进行编码



对信源U的
N长符号串
进行编码

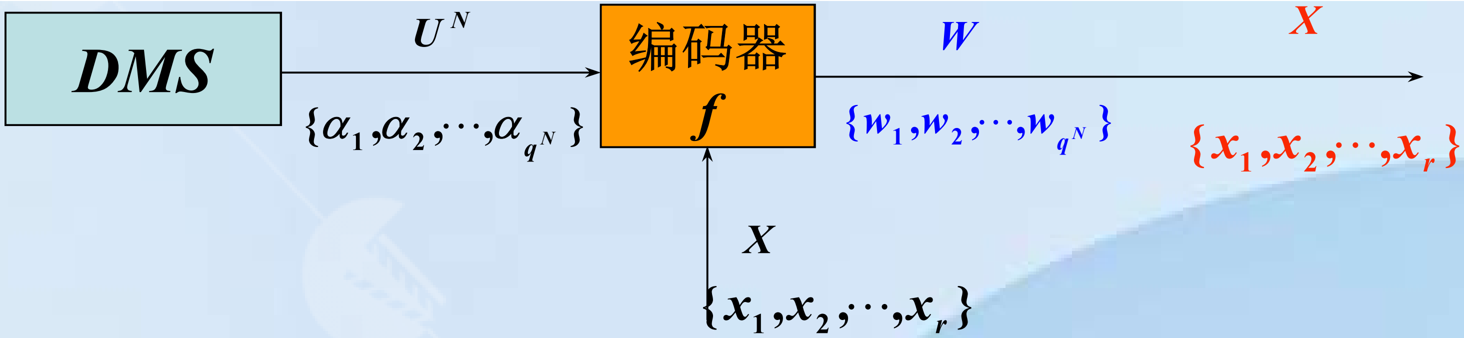


对扩展信源
U^N的单个符
号进行编码

$$\alpha_i = u_{i1}u_{i2}\cdots u_{iN}$$

$$u_{i1},u_{i2},\cdots,u_{iN} \in \{u_1,u_2,\cdots,u_q\}$$

2、定长编码定理



r 进制定长编码，码长为l_N，可用的码字数目：r^{l_N}

唯一可译 $\Longrightarrow r^{l_N} \ge q^N \Longrightarrow \dfrac{l_N}{N} \ge \dfrac{\log q}{\log r} = \dfrac{H_{\max}(U)}{\log r} = H_{r\max}(U)$

信息传输率 $R = H(X) = \dfrac{H(U)}{l_N / N}$ bit/码元

编码效率 $\eta_c = \dfrac{H(X)}{H_{\max}(X)} = \dfrac{H(U)}{\dfrac{l_N}{N}\log r}$

定长无失真编码定理：用r元符号表对离散无记忆信源U的N长符号序列进行定长编码，N长符号序列对应的码长为l_N，若对于任意小的正数ε，有不等式:

$$\frac{l_N}{N} \geq \frac{H(U)+\varepsilon}{\log r} = H_r(U)+\varepsilon'$$

就几乎能做到无失真编码，且随着序列长度N的增大，译码差错率趋于0。反过来，若

$$\frac{l_N}{N} \leq \frac{H(U)-2\varepsilon}{\log r} = H_r(U)-2\varepsilon'$$

就不可能做到无失真编码，且随着N的增大，译码差错率趋于1。

3、定长编码方法（例）

$$\begin{bmatrix} U \\ P_U \end{bmatrix} = \begin{bmatrix} u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 \\ \frac{1}{2} & \frac{1}{2^2} & \frac{1}{2^3} & \frac{1}{2^4} & \frac{1}{2^5} & \frac{1}{2^6} & \frac{1}{2^6} \end{bmatrix}$$

对U的单个符号进行2进制定长编码。

码元集：$X=\{0, 1\}$

解：$l = \frac{l_N}{N} \geq \frac{\log q}{\log r} = \frac{\log 7}{\log 2} \approx 2.8$ 码元/符号

取$l = 3$

$$\begin{matrix} U: & u_1 & u_2 & u_3 & u_4 & u_5 & u_6 & u_7 \\ W: & 001 & 010 & 011 & 100 & 101 & 110 & 111 \end{matrix}$$

$$H(U) = -\sum_{i=1}^{7} P(u_i)\log P(u_i) = \frac{63}{32} \text{ bit/符号}$$

$\bar{l} = l = 3$ 码元/符号 $\quad \eta_c = \frac{H(U)}{\bar{l}\log r} = \frac{\frac{63}{32}}{3\times\log 2} = 65.625\%$

提高编码效率的方法：对符号串进行编码；引入一定的失真。