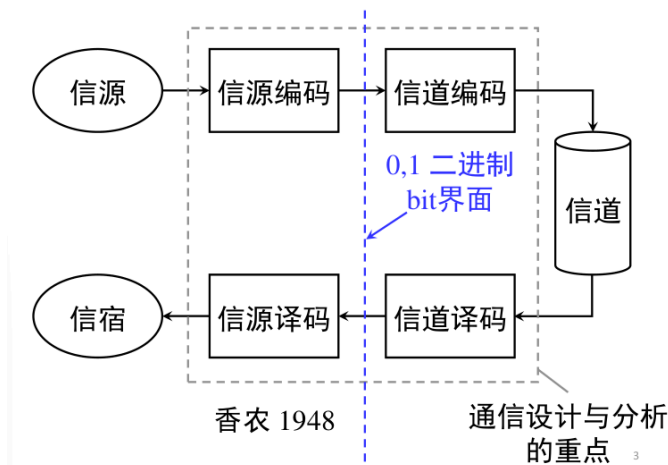


§1 信息论

信息论的基本模型:



§1.1 信源

信源的概念

(*) Def. 信源

信息的产生者称为**信源 (source of information)**, 产生一个**随机过程**发出信息。

离散无记忆信源

离散信源 (discrete source) 是时间、取值上都离散的信源, 可以表示为

$$X[k] \in \{x_1, x_2, \dots, x_N\}$$

为简化讨论, 可以假设 $X[k]$ 是**独立同分布的随机过程**, 即:

(*) Def. 离散无记忆信源

持续产生**独立同分布**的符号 $X \in \{x_1, x_2, \dots, x_N\}$ 的信源, 称为**离散无记忆信源 (discrete memoryless source, DMS)**, 记为

$$X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_N \\ p_1 & p_2 & \dots & p_N \end{pmatrix}, \quad p_i = \Pr\{X = x_i\}$$

信源编码的基本要求

将信源产生的符号 x 映射为 0, 1 比特串 $f(x)$ 的过程称为**信源编码 (source coding)**, 产生的比特串称为**码字 (codeword)**, 其长度 l 称为**码长 (code length)**。

- 若不同 x_i 映射出的码长 l_i 相同, 称为**定长码 (fixed-length code)**;
- 若不同 x_i 映射出的码长 l_i 不同, 称为**变长码 (variable-length code)**。

定长码的可解码条件

对**定长码**, 要求

$$f(x_i) \neq f(x_j), \quad \forall i \neq j$$

因此固定码长 l 应满足

$$N \leq 2^l \implies l \geq \lceil \log N \rceil$$

此处 \log 均为以 2 为底的对数。

变长码的可解码条件

对**变长码**, 要求任意码字不能是另一个码字的前缀, 因此又称为**前缀码 (prefix code)**。

引入**平均码长 (average code length)**

$$\bar{l} = \sum_{i=1}^N p_i l_i$$

我们希望在可解码的条件下, 使 \bar{l} 尽可能小。

可以证明, 信源 $X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_N \\ p_1 & p_2 & \dots & p_N \end{pmatrix}$ 的最小平均码长为

$$\bar{l}_{\min} = - \sum_{i=1}^N p_i \log p_i$$

信源的熵

离散信源的熵

(*) Def. 离散信源的熵

离散无记忆信源 $X \sim \begin{pmatrix} x_1 & x_2 & \dots & x_N \\ p_1 & p_2 & \dots & p_N \end{pmatrix}$ 的**熵 (entropy)** 定义为

$$H(X) = - \sum_{i=1}^N p_i \log p_i$$

$H(X)$ 可以写为

$$H(X) = \mathbb{E}_X[-\log p(X)]$$

因而 $-\log \Pr\{X = x_i\}$ 刻画事件 $\{X = x_i\}$ 所包含的信息量, **概率越小, 信息量越大**。

熵的性质

考虑 $H(X) = \bar{l}_{\min}$ 的物理含义, 应有 $0 \leq H(X) \leq \log N$ 。

- 当且仅当

$$\exists i \in \{1, 2, \dots, N\}, \quad p_i = 1$$

时, $H(X) = 0$, 此时信源没有不确定性, 不包含信息;

- 当且仅当

$$p_i = \frac{1}{N}, \quad i = 1, 2, \dots, N$$

时, $H(X) = \log N$, 此时信源不确定性最大, 包含信息量最大。

对离散信源, **均匀分布的不确定度最大**。

联合熵、条件熵

(*) Def. 联合熵

考虑两个离散无记忆信源 X 和 Y , 事件 $\{X = x_i, Y = y_j\}$ 的概率为 $p_{ij} = \Pr\{X = x_i, Y = y_j\}$, 则 X 和 Y 的**联合熵 (joint entropy)** 定义为

$$H(X, Y) = \mathbb{E}_{XY}[-\log p(X, Y)] = - \sum_{i=1}^N \sum_{j=1}^M p_{ij} \log p_{ij}$$

在没有歧义的情况下, 联合熵 $H(X, Y)$ 可记为 $H(XY)$ 。

联合熵刻画了将 X 和 Y 一起编码所需的最小平均码长, 即综合考虑 X 和 Y 的信息量。显然, $H(X, Y) \geq H(X)$ 。

(*) Def. 条件熵

考虑 X 和 Y 的联合分布 $p_{ij} = \Pr\{X = x_i, Y = y_j\}$, 在 Y 已知的条件下, 事件 $\{X = x_i\}$ 发生的概率为

$$p_{i|j} = \Pr\{X = x_i \mid Y = y_j\} = \frac{\Pr\{X = x_i, Y = y_j\}}{\Pr\{Y = y_j\}} = \frac{p_{ij}}{p_j}$$

则以 Y 为条件的 X 的**条件熵 (conditional entropy)** 定义为

$$H(X \mid Y) = \mathbb{E}_{XY}[-\log p(X \mid Y)] = - \sum_{i=1}^N \sum_{j=1}^M p_{ij} \log p_{i|j}$$

条件熵 $H(X \mid Y)$ 刻画了在已知 Y 的条件下, X 所包含的信息量, 即在观测 Y 后 X 残存的不确定度。显然, $0 \leq H(X \mid Y) \leq H(X)$ 。

信源的熵 $H(X)$ 刻画了信源 X 所包含的信息量，其值即对 X 进行无失真编码时所需的最小平均码长。

- 考虑对一个离散无记忆信源 X 编码传输，当信道速率（每传输一个信源符号所传输的平均比特数） $R \geq H(X)$ 时，可以实现无失真传输，即信源译码环节可无失真恢复 X ；
- 考虑对两个离散无记忆信源 X 和 Y 做联合信源编码传输，当信道速率 $R \geq H(X, Y)$ 时，在信源译码环节可无失真恢复 (X, Y) ；
- 考虑对离散无记忆信源 X 编码传输，编译码器可以共同观测另一个信源，当信道速率 $R \geq H(X | Y)$ 时，在信源译码环节可无失真恢复 X 。

► 熵的链式法则

🔗 Thm. 熵的链式法则

对于任意两个离散随机变量 X 和 Y ，有

$$H(X, Y) = H(Y) + H(X | Y) = H(X) + H(Y | X)$$

进一步地，

- 若 X 与 Y 独立（记为 $X \perp Y$ ），则 $p_{ij} = p_i p_j$ ， $p_{i|j} = \frac{p_i p_j}{p_j} = p_i$ ，于是

$$\begin{aligned} H(X | Y) &= - \sum_{i=1}^N \sum_{j=1}^M p_{ij} \log p_{i|j} = - \sum_{i=1}^N \sum_{j=1}^M p_i p_j \log p_i \\ &= - \sum_{i=1}^N p_i \log p_i = H(X) \end{aligned}$$

$$H(X, Y) = H(X | Y) + H(Y) = H(X) + H(Y)$$

即**独立随机变量的联合熵等于各自熵之和**。观测 Y 不会减少 X 的不确定性。

- 若 X 是 Y 的**确定性映射**（记为 $X = f(Y)$ ），则 $p_{i|j}$ 要么为 0，要么为 1，因此 $H(X | Y) = 0$ ，于是

$$H(X, Y) = H(Y)$$

即**确定性映射不会增加不确定性**。观测 Y 会完全消除 X 的不确定性。

进而，一般地有

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

► 互信息

$H(X | Y)$ 表征通过观测 Y 后 X 残存的不确定度，这种观测所消除不确定度的程度，即称为**互信息 (mutual information)**。

显然，有

$$\begin{aligned} H(X) - H(X | Y) &= H(X) - (H(X, Y) - H(Y)) \\ &= H(X) + H(Y) - H(X, Y) \\ &= H(Y) - H(Y | X) \end{aligned}$$

即，通过观测 Y 消除的 X 的不确定度，等于通过观测 X 消除的 Y 的不确定度。 X 与 Y 的**互信息是对称的**。

🔗 Def. 互信息

离散随机变量 X 和 Y 的**互信息 (mutual information)** 定义为

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

一般地，

$$I(X; Y) = H(X) + H(Y) - H(X, Y) = \sum_{i=1}^N \sum_{j=1}^M p_{ij} \log \frac{p_{ij}}{p_i p_j}$$

考虑两个离散无记忆信源 X 和 Y ，

- 若 $X \perp Y$ ，则

$$I(X; Y) = H(X) - H(X | Y) = H(X) - H(X) = 0$$

即**独立信源间没有互信息**；另可证明，没有互信息的信源是独立的。

- 若 $X = f(Y)$ ，则

$$I(X; Y) = H(X) - H(X | Y) = H(X) - 0 = H(X)$$

即**确定性映射的信源间互信息是其全部不确定度**。

- $0 \leq H(X | Y) \leq H(X)$ ，因此 $0 \leq I(X; Y) \leq \min \{H(X), H(Y)\}$ ，**信源之间不存在欺骗**。

■ 连续信源的微分熵

对于连续随机变量 X ，其概率分布由**概率密度函数 (probability density function, PDF)** $p_X(x)$ 描述，满足

$$\int_{-\infty}^{+\infty} p_X(x) dx = 1, \quad p_X(x) \geq 0$$

🔗 Def. 微分熵

微分熵 (differential entropy) 定义为

$$b(X) = \mathbb{E}_X[-\log p_X(X)] = - \int_{-\infty}^{+\infty} p_X(x) \log p_X(x) dx$$

微分熵 $b(X)$ **刻画了信源 X 的相对不确定度**，其值可以为负数。

🔗 Ex. 均匀分布的微分熵

对 $X \sim \mathcal{U}(a, b)$ ，有 $p_X(x) = \frac{1}{b-a}$ ，因此

$$\begin{aligned} b(X) &= - \int_a^b p_X(x) \log p_X(x) dx \\ &= - \int_a^b \frac{1}{b-a} \log \frac{1}{b-a} dx = \boxed{\log(b-a)} \end{aligned}$$

这是给定有限区间长度约束下的最大微分熵。

🔗 Ex. Gauss 分布的微分熵

对 $X \sim \mathcal{N}(\mu, \sigma^2)$ ，有 $p_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ ，因此

$$\begin{aligned} b(X) &= - \int_{-\infty}^{+\infty} p_X(x) \log p_X(x) dx \\ &= - \int_{-\infty}^{+\infty} p_X(x) \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \right) dx \\ &= - \int_{-\infty}^{+\infty} p_X(x) \left(-\frac{(x-\mu)^2}{2\sigma^2} \log e - \log \sqrt{2\pi\sigma^2} \right) dx \\ &= \frac{\log e}{2\sigma^2} \int_{-\infty}^{+\infty} (x-\mu)^2 p_X(x) dx + \log \sqrt{2\pi\sigma^2} \int_{-\infty}^{+\infty} p_X(x) dx \\ &= \frac{\log e}{2\sigma^2} \mathbb{E}[(X-\mu)^2] + \log \sqrt{2\pi\sigma^2} \\ &= \frac{\log e}{2\sigma^2} (\mathbb{E}^2[(X-\mu)] + \text{Var}[(X-\mu)]) + \log \sqrt{2\pi\sigma^2} \\ &= \frac{\log e}{2\sigma^2} \cdot \sigma^2 + \log \sqrt{2\pi\sigma^2} = \boxed{\log \sqrt{2\pi e \sigma^2}} \end{aligned}$$

这是给定双边无限区间约束下的最大微分熵，也是给定有限方差约束下的最大微分熵。

🔗 Ex. 指数分布的微分熵

对 $X \sim \text{Exponential}(\lambda)$ ，有 $p_X(x) = \lambda e^{-\lambda x}$ ， $x \geq 0$ ，因此

$$\begin{aligned} b(X) &= - \int_0^{+\infty} p_X(x) \log p_X(x) dx \\ &= - \int_0^{+\infty} \lambda e^{-\lambda x} \log (\lambda e^{-\lambda x}) dx \\ &= - \int_0^{+\infty} \lambda e^{-\lambda x} (\log \lambda - \lambda x \log e) dx \\ &= -\lambda \log \lambda \int_0^{+\infty} e^{-\lambda x} dx + \lambda^2 \log e \int_0^{+\infty} x e^{-\lambda x} dx \\ &= -\lambda \log \lambda \cdot \frac{1}{\lambda} + \lambda^2 \log e \cdot \frac{1}{\lambda^2} = \boxed{1 - \log \lambda} \end{aligned}$$

这是给定单边无限区间约束下的最大微分熵。

► 联合微分熵、条件微分熵

(*) Def. 联合微分熵

考虑两个连续随机变量 X 和 Y ，其联合概率密度函数为 $p_{X,Y}(x,y)$ ，则 X 和 Y 的**联合微分熵 (joint differential entropy)** 定义为

$$\begin{aligned} b(X, Y) &= \mathbb{E}_{X,Y}[-\log p_{X,Y}(X, Y)] \\ &= -\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{X,Y}(x, y) \log p_{X,Y}(x, y) \, dx \, dy \end{aligned}$$

在没有歧义的情况下，联合微分熵 $b(X, Y)$ 可记为 $b(XY)$ 。

(*) Def. 条件微分熵

考虑 X 和 Y 的联合概率密度函数 $p_{X,Y}(x,y)$ ，条件随机变量 $X | Y$ 的概率密度函数为

$$p_{X|Y}(x, y) = \frac{p_{X,Y}(x, y)}{p_Y(y)}$$

则以 Y 为条件的 X 的**条件微分熵 (conditional differential entropy)** 定义为

$$\begin{aligned} b(X | Y) &= \mathbb{E}_{X,Y}[-\log p_{X|Y}(X, Y)] \\ &= -\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{X,Y}(x, y) \log p_{X|Y}(x, y) \, dx \, dy \end{aligned}$$

类似地，微分熵也有**链式法则**

$$b(X, Y) = b(Y) + b(X | Y) = b(X) + b(Y | X)$$

► 微分熵的互信息

(*) Def. 微分熵的互信息

连续随机变量 X 和 Y 的**互信息 (mutual information)** 定义为

$$I(X; Y) = b(X) - b(X | Y) = b(Y) - b(Y | X)$$

仍有 $b(X) \geq b(X | Y)$ ，因此 $I(X; Y) \geq 0$ 。

一般地，

$$\begin{aligned} I(X; Y) &= b(X) - b(X | Y) = b(X) + b(Y) - b(X, Y) \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p_{X,Y}(x, y) \log \frac{p_{X,Y}(x, y)}{p_X(x)p_Y(y)} \, dx \, dy \end{aligned}$$

§1.2 信道

离散幅值信道

■ 离散无记忆信道模型

离散无记忆信道 (Discrete Memoryless Channel, DMC) 模型假设信道在每次传输时的行为独立且不依赖于之前的传输结果。DMC 由一个条件概率来描述，即**给定信源输入符号 $X = x_i$ 时信宿输出符号 $Y = y_j$ 的概率分布 $p_{j|i}$** 。

当给定 X 的分布 p_i 时，可以得到联合分布

$$p_{ij} = \Pr\{X = x_i, Y = y_j\} = \Pr\{Y = y_j | X = x_i\} \Pr\{X = x_i\} = p_i p_{j|i}$$

信宿输出 Y 的边缘分布为

$$p_j = \Pr\{Y = y_j\} = \sum_i p_{ij} = \sum_i p_i p_{j|i}$$

这样，**互信息 $I(X; Y)$ 就可以给定**为

$$\begin{aligned} I(X; Y) &= H(X) + H(Y) - H(X, Y) \\ &= -\sum_i p_i \log p_i - \sum_j p_j \log p_j + \sum_i \sum_j p_{ij} \log p_{ij} \\ &= -\sum_i \sum_j p_{ij} \log p_i - \sum_i \sum_j p_{ij} \log p_j + \sum_i \sum_j p_{ij} \log p_{ij} \\ &= \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{p_i p_j} = \sum_i \sum_j p_{ij} \log \frac{p_{j|i}}{p_j} \\ &= \sum_i \sum_j p_i p_{j|i} \log \frac{p_{j|i}}{\sum_i p_i p_{j|i}} \end{aligned}$$

若允许通过映射**调整输入分布 p_i** ，则可以在此意义下**最大化互信息 $I(X; Y)$** ，这一最大值称为**信道容量 (channel capacity)**，记为

$$C = \max_{p_i} I(X; Y) = \max_{p_i} \sum_i \sum_j p_i p_{j|i} \log \frac{p_{j|i}}{\sum_i p_i p_{j|i}}$$

这一优化问题还有两个约束条件，即概率归一化条件 $\sum_i p_i = 1$ 和非负性条件

$p_i \geq 0$ 。这是一个**非凸**的带约束优化问题，有一定的求解难度。一般 DMC 的容量需要通过数值方法求解，如 Blahut-Arimoto 算法。

■ 对称二进制信道 (BSC)

我们希望得到容量的解析闭式解，因此需要在**数字通信**背景下简化 DMC 模型。

(*) Def. 对称二进制信道

对称二进制信道 (Binary Symmetric Channel, BSC) 是这样的 DMC：

- 输入和输出符号均为二进制（0 和 1），
- 每个比特在传输过程中都有一个固定的**差错概率 (cross-over probability)** ϵ ，即输出比特有 ϵ 的概率变为输入比特的反码。

BSC 可以等效为

$$Y = X \oplus Z, \quad Z \sim \text{Bernoulli}(\epsilon) = \begin{pmatrix} 0 & 1 \\ 1 - \epsilon & \epsilon \end{pmatrix}$$

其中 Z 是与输入 X 独立的噪声变量， \oplus 表示**按位异或**运算。

使用这一等效，BSC 的容量为

$$\begin{aligned} C &= \max_{p_i} I(X; Y) = \max_{p_i} I(X; X \oplus Z) \\ &= \max_{p_i} (H(X \oplus Z) - H(X \oplus Z | X)) \\ &= \max_{p_i} (H(X \oplus Z) - H(Z)) \\ &= \max_{p_i} H(X \oplus Z) + \underbrace{\epsilon \log \epsilon + (1 - \epsilon) \log (1 - \epsilon)}_{\text{const}} \end{aligned}$$

因此，最大化互信息等价于**最大化 $H(X \oplus Z)$** 。由于 $X \oplus Z \in \{0, 1\}$ ，其熵的最大值为 1，**比特均匀分布**时取得，即

$$\begin{aligned} C &= 1 + \epsilon \log \epsilon + (1 - \epsilon) \log (1 - \epsilon) \quad \text{iff} \quad Y \sim \text{Bernoulli}(0.5) \\ &\iff X \sim \text{Bernoulli}(0.5) \end{aligned}$$

■ 更多离散信道模型

► 擦除信道

二进制擦除信道 (Binary Erasure Channel, BEC) 是另一种常见的离散信道模型。与 BSC 不同，BEC 在传输过程中不会将比特翻转为反码，而是有一定概率 ϵ **将比特「擦除」**，即**输出一个特殊符号 e** ，表示该比特丢失。

在 BEC 中，已知输出 Y 为 0 或 1 时，可以确定输入 X 的值，**没有任何不确定性**；而当输出 Y 为 e 时，输入 X **等概地**可能是 0 或 1，存在不确定性 $H(X | Y = e) = 1$ 。这样，BEC 的容量为

$$\begin{aligned} C &= \max_{p_i} I(X; Y) = \max_{p_i} (H(X) - H(X | Y)) \\ &= \max_{p_i} (H(X) - \epsilon H(X | Y = e)) \\ &= \max_{p_i} (H(X) - \epsilon \cdot 1) = 1 - \epsilon \end{aligned}$$

当输入比特均匀分布时取得，即 $X \sim \text{Bernoulli}(0.5)$ 。

推广到**多进制擦除信道 (M-ary Erasure Channel, MEC)**，输入输出符号集均为 M 个符号 $\{x_1, x_2, \dots, x_M\}$ ，每个符号在传输过程中有概率 ϵ **被擦除**为特殊符号 e ，不变的概率为 $1 - \epsilon$ 。MEC 的容量为

$$C = \log M - \epsilon \log M = (1 - \epsilon) \log M$$

当输入符号均匀分布时取得，此时 $p_i \equiv \frac{1}{M}$ 。

► 多进制对称信道

将 BSC 推广到多进制符号集，可以得到**多进制对称信道 (M-ary Symmetric Channel, MSC)**。MSC 的输入输出符号集均为 M 个符号 $\{x_1, x_2, \dots, x_M\}$ ，每个符号在传输过程中有概率 ϵ **被均匀地**翻转为其他 $M - 1$ 个符号中的任意一个，不变的概率为 $1 - (M - 1)\epsilon$ 。

对 MSC，有

$$\begin{aligned}
I(X; Y) &= H(Y) - H(Y|X) \\
&= -\sum_{j=1}^M p_j \log p_j + \sum_{i=1}^M p_i \sum_{j=1}^M p_{j|i} \log p_{j|i} \\
&= -\sum_{j=1}^M p_j \log p_j + \sum_{i=1}^M p_i \left((1 - (M-1)\epsilon) \log(1 - (M-1)\epsilon) \right. \\
&\quad \left. + (M-1)\epsilon \log \epsilon \right) \\
&= -\sum_{j=1}^M p_j \log p_j + \underbrace{(1 - (M-1)\epsilon) \log(1 - (M-1)\epsilon)}_{\text{const}} \\
&\quad + \underbrace{(M-1)\epsilon \log \epsilon}_{\text{const}} \\
&\leq \log M + \text{const}
\end{aligned}$$

因此, MSC 的容量为

$$C = \log M + (1 - (M-1)\epsilon) \log(1 - (M-1)\epsilon) + (M-1)\epsilon \log \epsilon$$

此时, 输入符号均匀分布, $p_i \equiv \frac{1}{M}$ 。

连续幅值信道

类似于离散幅值信道, **连续幅值信道 (continuous amplitude channel)** 也可以用条件概率密度函数 $p_{Y|X}(y, x)$ 来描述, 所传递的信息量仍然用互信息 $I(X; Y)$ 来衡量。

若允许通过映射 **调整输入分布** $p_X(x)$, 则可以在此意义下 **最大化互信息** $I(X; Y)$, 这一最大值仍然称为 **信道容量 (channel capacity)**, 记为

$$C = \max_{p_X(x)} I(X; Y)$$

Gauss 信道

Gauss 信道 (Gaussian channel) 是一种特殊的连续幅值信道, 假设信道输出 Y 是输入 X 与 Gauss 白噪声 Z 的和, 即

$$Y = X + Z, \quad \text{where } Z \sim \mathcal{N}(0, \sigma^2)$$

其中 X 有功率约束 $\mathbb{E}[X^2] = E_s$ 。

Gauss 信道的容量为

$$\begin{aligned}
C &= \max_{p_X(x)} I(X; Y) = \max_{p_X(x)} I(X; X + Z) = \max_{p_X(x)} (b(X + Z) - b(X + Z | X)) \\
&= \max_{p_X(x)} (b(X + Z) - b(Z)) = \max_{p_X(x)} b(X + Z) - \log \sqrt{2\pi e \sigma^2}
\end{aligned}$$

因此, 最大化互信息等价于 **最大化 $b(X + Z)$** 。由于 X, Z 独立,

$$\mathbb{E}[(X + Z)^2] = \mathbb{E}[X^2] + \mathbb{E}[Z^2] + 2\underbrace{\mathbb{E}[X]\mathbb{E}[Z]}_0 = E_s + \sigma^2$$

当且仅当 $X \sim \mathcal{N}(0, E_s)$ 时, $X + Z \sim \mathcal{N}(0, E_s + \sigma^2)$, $b(X + Z)$ 最大化为

$$b(X + Z) = \log \sqrt{2\pi e(E_s + \sigma^2)}$$

从而

$$C = \log \sqrt{2\pi e(E_s + \sigma^2)} - \log \sqrt{2\pi e \sigma^2} = \boxed{\log \sqrt{1 + \frac{E_s}{\sigma^2}}}$$

这里 C 无量纲, **单位为 bit/次**。

带宽受限的 Gauss 信道

在实际通信系统中, 信道通常是 **带宽受限 (bandwidth-limited)** 的, 即信号只能在有限的频率范围内传输。设信道的带宽为 W , 则根据 **Nyquist 采样定理**, 单位时间最多可以传输 $2W$ 个独立符号。

考虑带宽受限为 W 的 Gauss 信道, 由于这一信道中的噪声是与信号 **时域相加** 的、功率谱在 $[-W, W]$ 均匀的 Gauss 噪声, 因此称为 **加性白 Gauss 噪声 (Additive White Gaussian Noise, AWGN) 信道**。设其中噪声 $N(t)$ 的方差为 σ^2 , 则

$$R_N(t, s) = \begin{cases} \mathbb{E}[N^2(t)] = \sigma^2, & t = s \\ 0, & t \neq s \end{cases} = \sigma^2 \delta(t - s)$$

因此 $N(t)$ 是 **宽平稳** 的, $R_N(\tau) = \sigma^2 \delta(\tau)$, 其功率谱密度为

$$S_N(f) = \int_{-\infty}^{+\infty} R_N(\tau) e^{-j2\pi f\tau} d\tau = \sigma^2, \quad |f| \leq W$$

将正负频部分合并, 得到单边功率谱密度为

$$S_N^{(\text{single})}(f) = 2\sigma^2 =: n_0, \quad 0 \leq f \leq W$$

同时, 记通信功率 P 为单位时间内传输的平均能量, 则每个符号的平均能量为 $E_s = \frac{P}{2W}$ 。则 AWGN 信道的容量为

$$C = 2W \log \sqrt{1 + \frac{P}{2W\sigma^2}} = W \log \left(1 + \frac{P}{Wn_0} \right)$$

Shannon 公式

带宽受限为 $|f| \leq W$ 、噪声功率谱密度为 $S_N(f) = \frac{n_0}{2}$ 的 **AWGN 信道** 在功率 P 下的容量为

$$C = W \log \left(1 + \frac{P}{Wn_0} \right)$$

其中 C 的**单位为 bit/s**。

信号功率与噪声功率的比值 $\frac{P}{Wn_0}$ 称为 **信噪比 (signal-to-noise ratio, SNR)**, 可记为 SNR。因此, Shannon 公式也可以写为

$$C = W \log(1 + \text{SNR})$$

- 当 $\text{SNR} \rightarrow 0$ 时, 即在 **低 SNR 区**,

$$C \rightarrow W \log e \cdot \text{SNR} = \frac{P}{n_0} \log e \approx 1.44 \frac{P}{n_0}$$

容量 C 与带宽 W 无关, 而与功率 P 呈线性关系;

- 当 $\text{SNR} \rightarrow +\infty$ 时, 即在 **高 SNR 区**, $C \rightarrow W \log \text{SNR}$, 引入信噪比的分贝值 $\text{SNR}_{\text{dB}} = 10 \log_{10} \text{SNR} = \frac{10}{\log 10} \log \text{SNR}$, 则

$$C \rightarrow W \cdot \frac{\log 10}{10} \text{SNR}_{\text{dB}} \approx 0.3322W \cdot \text{SNR}_{\text{dB}}$$

Remark 题型: Shannon 公式的能量极限

典型题型 给定 AWGN 信道的带宽 W 、噪声功率谱密度 $S_N(f) = \frac{n_0}{2}$,

求在单位时间内传输 1 bit 所需的最小能量 $E_{b, \min}$ 。

以容量 C 传输时, **每比特耗时 $\frac{1}{C}$** , 因此每比特能量为

$$E_b = \frac{P}{C} = \frac{P}{W \log \left(1 + \frac{P}{Wn_0} \right)}$$

对 E_b 关于 P 求导, 由 $\ln(1+x) \geq \frac{x}{1+x}$ 可知

$$\frac{\partial E_b}{\partial P} = \frac{W \log \left(1 + \frac{P}{Wn_0} \right) - P W \frac{\frac{1}{Wn_0}}{\left(1 + \frac{P}{Wn_0} \right) \ln 2}}{\left(W \log \left(1 + \frac{P}{Wn_0} \right) \right)^2} \geq 0$$

因此, E_b 随 P **单调递增**, 其最小值在 $P \rightarrow 0$ 时取得, 即

$$E_{b, \min} = \lim_{P \rightarrow 0} E_b = \lim_{P \rightarrow 0} \frac{P}{W \log \left(1 + \frac{P}{Wn_0} \right)} = \boxed{\frac{n_0 \ln 2}{1}}$$

此时传输时长无限大。

这一结果与 Shannon 公式中的 **低 SNR 极限** 一致, 且与 W 无关。

§1.3 加性白 Gauss 噪声 (AWGN)

在实际通信系统中, 信道通常受到各种噪声的影响, 其中最常见和重要的一种噪声类型是 **加性白 Gauss 噪声 (Additive White Gaussian Noise, AWGN)**。AWGN 噪声具有以下几个特点:

- 加性 (additive)**, 即噪声直接叠加在信号上, 接收信号为发送信号与噪声的和。
- 白 (white)**, 即噪声在整个频谱上均匀分布, 具有恒定的功率谱密度 $S_Z(f) = \frac{n_0}{2}$ 。
- Gaussian**, 即噪声服从 Gauss 分布, 具有均值为零的正态分布特性。

■ 离散时间 AWGN 信源

一个离散时间 AWGN 信源可以表示为

$$Z_i \sim \mathcal{N}(0, \sigma^2), \quad i = 1, 2, 3, \dots$$

其中 $\sigma^2 = \frac{n_0}{2}$ 。该信源的每个样本 Z_i **独立同分布**，服从均值为 0、方差为 σ^2 的 Gauss 分布，因此可只考虑单个样本 Z 的性质。

对 $Z \sim \mathcal{N}(\mu, \sigma^2)$ ，有 $p_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right)$ ，因此其**微分熵**为

$$\begin{aligned} h(Z) &= -\int_{-\infty}^{+\infty} p_Z(z) \log p_Z(z) \, \mathrm{d}z \\ &= -\int_{-\infty}^{+\infty} p_Z(z) \log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) \right) \, \mathrm{d}z \\ &= -\int_{-\infty}^{+\infty} p_Z(z) \left(-\frac{(z-\mu)^2}{2\sigma^2} \log e - \log \sqrt{2\pi\sigma^2} \right) \, \mathrm{d}z \\ &= \frac{\log e}{2\sigma^2} \int_{-\infty}^{+\infty} (z-\mu)^2 p_Z(z) \, \mathrm{d}z + \log \sqrt{2\pi\sigma^2} \int_{-\infty}^{+\infty} p_Z(z) \, \mathrm{d}z \\ &= \frac{\log e}{2\sigma^2} \mathbb{E}[(Z-\mu)^2] + \log \sqrt{2\pi\sigma^2} \\ &= \frac{\log e}{2\sigma^2} (\mathbb{E}^2[(Z-\mu)] + \text{Var}[(Z-\mu)]) + \log \sqrt{2\pi\sigma^2} \\ &= \frac{\log e}{2\sigma^2} \cdot \sigma^2 + \log \sqrt{2\pi\sigma^2} = \boxed{\log \sqrt{2\pi e \sigma^2}} \end{aligned}$$

■ 离散时间 AWGN 信道

离散时间 AWGN 信道即 **AWGN 电平信道**，其噪声源是离散时间 AWGN 信源 $z \sim \mathcal{N}(0, \sigma^2)$ 。信道输入 x 和输出 y 之间有条件概率关系

$$p(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right)$$

因此，在给定 x 下输出 y 向单侧偏移 A 以上的概率为

$$\begin{aligned} \Pr\{y < x - A | x\} &= \int_{-\infty}^{x-A} p(y|x) \, \mathrm{d}y = \int_{-\infty}^{-A} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}} \, \mathrm{d}z \\ &= \int_A^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}} \, \mathrm{d}z = \int_{A/\sigma}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d}z \\ &= Q\left(\frac{A}{\sigma}\right) \\ \Pr\{y > x + A | x\} &= \int_{x+A}^{+\infty} p(y|x) \, \mathrm{d}y = \int_A^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}} \, \mathrm{d}z \\ &= Q\left(\frac{A}{\sigma}\right) \end{aligned}$$

这里 $Q(x)$ 是 **Q 函数 (Q-function)**，定义为

$$Q(x) = \int_x^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d}z$$

在 CASIO fx-991CN X 计算器中，可以通过「统计」页面「单变量统计计算」中的「正态分布」>「R()」选项计算 $Q(x)$ 。

■ 连续时间 AWGN

■ 连续时间 AWGN 信源

一个连续时间 AWGN 信源 $z(t)$ 满足

- 对任意时刻 t ， $z(t)$ 服从均值为 0 的 Gauss 分布；
- 对任意时刻序列 t_1, t_2, \dots, t_n ，随机变量组 $\{z(t_1), z(t_2), \dots, z(t_n)\}$ 服从均值为 $\vec{0}$ 的 n 元 Gauss 分布；
- 功率谱密度函数 $S_z(f) \equiv \frac{n_0}{2}$ ，其中 n_0 为常数。

即， $z(t)$ 是一个 **Gauss 过程**，其自相关函数为

$$R_z(\tau) = \mathbb{E}[z(t)z(t+\tau)] = \int_{-\infty}^{+\infty} S_z(f) e^{j2\pi f\tau} \, \mathrm{d}f = \int_{-\infty}^{+\infty} \frac{n_0}{2} e^{j2\pi f\tau} \, \mathrm{d}f = \frac{n_0}{2} \delta(\tau)$$

任意时刻其方差都为 $\mathbb{E}[z^2(t)] = R_z(0) = +\infty$ 。

■ 连续时间 AWGN 信道

连续时间 AWGN 信道即 **AWGN 波形信道**，其噪声源是连续时间 AWGN 信源 $z(t)$ 。信道输入 $x(t)$ 和输出 $y(t)$ 之间满足

$$y(t) = x(t) + z(t)$$

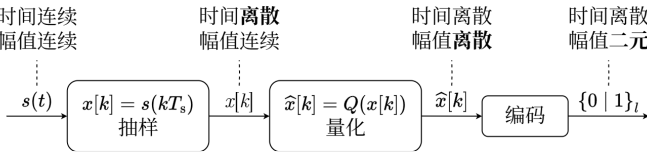
其化归为**标准等价电平信道**的过程中，将对 $z(t)$ 做线性处理 $z = \int_{-\infty}^{\infty} z(t)g(t) \, \mathrm{d}t$ ，所得 z 的方差为

$$\begin{aligned} \mathbb{E}[z^2] &= \mathbb{E}\left[\left(\int_{-\infty}^{+\infty} z(t)g(t) \, \mathrm{d}t\right)^2\right] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(t)g(s) \mathbb{E}[z(t)z(s)] \, \mathrm{d}t \, \mathrm{d}s \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(t)g(s) R_z(t-s) \, \mathrm{d}t \, \mathrm{d}s \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(t)g(s) \cdot \frac{n_0}{2} \delta(t-s) \, \mathrm{d}t \, \mathrm{d}s \\ &= \int_{-\infty}^{+\infty} g^2(t) \cdot \frac{n_0}{2} \, \mathrm{d}t = \frac{n_0}{2} \int_{-\infty}^{+\infty} g^2(t) \, \mathrm{d}t \end{aligned}$$

即 $z \sim \mathcal{N}\left(0, \frac{n_0}{2} \int_{-\infty}^{\infty} g^2(t) \, \mathrm{d}t\right)$ 。当 $g(t)$ 能量归一化时， $z \sim \mathcal{N}\left(0, \frac{n_0}{2}\right)$ 。

§2 模拟信源的数字编码

模拟信源是**时间连续、幅值连续**的信源，其输出 $s(t)$ 是一个连续时间的**随机过程**，将其数字化的过程如下：



在信道另一侧，接收端对接收到的二值序列进行**译码 (digital decoding)**，恢复出数字信号，然后通过**电平重建 (level reconstruction)** 和**内插 (interpolation)**，即恢复出模拟信号 $s(t)$ 。

■ 抽样与内插恢复

抽样 (sampling) 是将连续时间信号 $s(t)$ 在时间上离散化的过程，得到离散时间信号 $x[k] = s(kT_s)$ ，其中 T_s 是**抽样周期 (sampling period)**， $f_s = \frac{1}{T_s}$ 是**抽样频率 (sampling frequency)**。

▮ *Thm.* Nyquist 抽样定理

设 $s(t)$ 是**带宽受限**的信号，其频谱 $S(f) = \mathcal{F}\{s(t)\}$ 满足

$$S(f) = 0, \quad |f| > W$$

则当抽样频率 $f_s \geq 2W$ 时，可以从抽样所得信号 $x[k] = s(kT_s)$ **无失真**地恢复出 $s(t)$ 。这个最低抽样频率 $f_s = 2W$ 称为 **Nyquist 频率 (Nyquist frequency)**。

从序列 $x[k]$ 恢复出 $s(t)$ ，可以先以 $x[k]$ 调制冲激序列得到

$$\tilde{s}(t) = \sum_k x[k] \delta(t - kT_s), \quad \text{其 Fourier 变换为}$$

$$\tilde{S}(f) = \mathcal{F}\{\tilde{s}(t)\} = \sum_k x[k] \mathcal{F}\{\delta(t - kT_s)\} = \sum_k s(kT_s) e^{-j2\pi f k T_s}$$

另一方面，

$$\begin{aligned} \tilde{S}(f) &= \mathcal{F}\left\{s(t) \sum_k \delta(t - kT_s)\right\} = \frac{1}{2\pi} S(f) * \mathcal{F}\left\{\sum_k \delta(t - kT_s)\right\} \\ &= \frac{1}{T_s} S(f) * \sum_k \delta(f - kf_s) \end{aligned}$$

因而需使用增益为 T_s 的**低通滤波器 (low-pass filter, LPF)** 截去 $\tilde{S}(f)$ 中 $|f| > W$ 的部分，才能恢复出 $S(f)$ ，输出即为

$$\hat{S}(f) = \frac{1}{f_s} \sum_k s(kT_s) e^{-j2\pi f k T_s} \times \mathbb{1}_{|f| \leq W}$$

其中 $\mathbb{1}_P$ 为事件 P 的指示函数。由此，时域恢复的信源为

$$\begin{aligned} \hat{s}(t) &= \mathcal{F}^{-1}\{\hat{S}(f)\} = \int_{-\infty}^{+\infty} \hat{S}(f) e^{j2\pi f t} \, \mathrm{d}f \\ &= \int_{-W}^W \frac{1}{f_s} \sum_k s(kT_s) e^{-j2\pi f k T_s} e^{j2\pi f t} \, \mathrm{d}f \\ &= \frac{1}{f_s} \sum_k s(kT_s) \int_{-W}^W e^{j2\pi f(t - kT_s)} \, \mathrm{d}f = \sum_k s(kT_s) \frac{\sin(2\pi W(t - kT_s))}{\pi f_s(t - kT_s)} \end{aligned}$$

令抽样频率 $f_s = 2W$ 取 Nyquist 频率，则

$$\hat{s}(t) = \sum_k s \left(\frac{k}{2WT} \right) \text{sinc}(2Wt - k)$$

📌 *Note* 抽样前的预滤波

实际采集到的真实信号一定是全频带的信号，为避免混叠 (aliasing) 干扰，通常先通过预滤波器 (pre-filter) 截去高频成分，保留期望频带内的信号，然后再进行抽样。

人类的语音信号主要能量集中在 300 ~ 3400 Hz，通常预滤波器的带宽为 3.4 kHz，抽样频率为 8 kHz。

量化与电平重建

量化 (quantization) 是将幅值连续的信号 $x[k]$ 离散化为数字信号 $\hat{x}[k]$ 的过程。设 n -bit 量化器 Q 有 $L = 2^n$ 个量化级别，则第 i 个量化区间上

$$Q(x) = y_i, \quad x \in (x_i, x_{i+1}], \quad i = 1, 2, \dots, L$$

其中，

- x_i 是**分层电平**， y_i 是**表示电平**，直接用于重建时即**重建电平**；
- $I_i = (x_i, x_{i+1}]$ 称为第 i 个**量化区间 (quantization interval)**，其长度 $\Delta_i = x_{i+1} - x_i$ 称为第 i 个**量化间隔 (quantization step)**。

当对每个 $i = 1, 2, \dots, L$ 有 $\Delta_i \triangleq \Delta$ 时，可有**均匀量化 (uniform quantization)**，否则只能为**非均匀量化 (non-uniform quantization)**。

量化器 Q 对输入 x 的**量化误差 (quantization error)** 定义为

$$e(x) = x - Q(x)$$

因为 x 可看作**随机变量 X** ， $e(x)$ 也是随机变量。良好的量化器应使得 $\mathbb{E}[e(X)] = 0$ ，此时其方差为

$$\begin{aligned} \sigma^2 &= \mathbb{E}[e^2(X)] = \int_{-\infty}^{+\infty} (x - Q(x))^2 p_X(x) dx \\ &= \sum_{i=1}^L \int_{x_i}^{x_{i+1}} (x - y_i)^2 p_X(x) dx \end{aligned}$$

🔗 *Def.* 量化的均方误差

对量化器 Q ，其量化误差的方差

$$\sigma^2 = \mathbb{E}[e^2(X)] = \sum_{i=1}^L \int_{x_i}^{x_{i+1}} (x - y_i)^2 p_X(x) dx$$

称为**量化的均方误差 (mean square error, MSE)**。

若信源 x 超出量化器的表示范围 $[x_1, x_{L+1}]$ ，则会发生**过载失真 (saturation distortion)**，此时量化误差相当于将过载部分全部映射为 y_1 或 y_L ，量化误差可拆分为

$$\sigma^2 = \sigma_q^2 + \sigma_o^2, \quad \begin{cases} \sigma_q^2 = \sum_{i=1}^L \int_{x_i}^{x_{i+1}} (x - y_i)^2 p_X(x) dx, \\ \sigma_o^2 = \int_{-\infty}^{x_1} (x - y_1)^2 p_X(x) dx + \int_{x_{L+1}}^{\infty} (x - y_L)^2 p_X(x) dx \end{cases}$$

一般而言，功率较大的信源对量化噪声的容忍程度也比较高，因而我们可以引入量化器 Q 的**信噪比 (signal-to-noise ratio, SNR)**，定义为

$$\text{SNR}_q = \frac{\mathbb{E}[X^2]}{\mathbb{E}[e^2(X)]} = \frac{\int_{-\infty}^{\infty} x^2 p_X(x) dx}{\sum_{i=1}^L \int_{x_i}^{x_{i+1}} (x - y_i)^2 p_X(x) dx}$$

均匀量化

对于均匀量化器，量化间隔 $\Delta_i \triangleq \Delta$ ，且取**重建电平为区间中点**

$$y_i = \frac{x_i + x_{i+1}}{2}, \quad i = 1, 2, \dots, L$$

均匀量化一定限制在有限范围 $[x_{\min}, x_{\max}]$ 内，量化区间长度为 $\Delta = \frac{x_{\max} - x_{\min}}{L}$ 。我们常令 $x_{\min} = -V$ ， $x_{\max} = V$ ，则 $\Delta = \frac{2V}{L}$ 。

若假设**量化输入 X 均匀分布于 $[-V, V]$** ，则 n -bit 均匀量化的**量化噪声**为

$$\sigma^2 = \frac{L}{2V} \int_{-V}^{V} x^2 dx = \frac{\Delta^2}{12} = \frac{(2V)^2}{12L^2} = \frac{V^2}{3 \cdot 4^n}$$

因此量化器的**信噪比**为

$$\text{SNR}_q = \frac{\mathbb{E}[X^2]}{\sigma^2} = \frac{\int_{-V}^V x^2 \cdot \frac{1}{2V} dx}{\frac{V^2}{3 \cdot 4^n}} = \frac{V^2/3}{V^2/(3 \cdot 4^n)} = 4^n \approx 6.02n \text{ dB}$$

即每增加 1 bit 的量化精度，信噪比提高约 6.02 dB。

对一般的量化输入 X ，当 Δ 足够小时，可近似认为均匀量化器**每个量化区间内的 X 服从均匀分布**，此时 n -bit 均匀量化的量化噪声近似为

$$\begin{aligned} \sigma^2 &= \sum_{i=1}^L \int_{x_i}^{x_{i+1}} (x - y_i)^2 p_X(x) dx \approx \sum_{i=1}^L \int_{x_i}^{x_{i+1}} (x - y_i)^2 dx \int_{x_i}^{x_{i+1}} p_X(x) dx \\ &= \frac{\Delta^2}{12} \sum_{i=1}^L \int_{x_i}^{x_{i+1}} p_X(x) dx = \frac{V^2}{3 \cdot 4^n} \int_{-V}^V p_X(x) dx \end{aligned}$$

噪声功率与信源在量化范围内的分布关联不大，但**信噪比**将依赖于 $\frac{\mathbb{E}[X^2]}{V^2}$ ，信源分布越不均匀，信噪比越低。**均匀量化对均匀分布最优**。

非均匀量化

非均匀量化器的设计

非均匀量化器的设计目标是**最小化量化误差的均方值**，即求

$$\arg \min_{\{x_i\}, \{y_i\}} \sigma^2 = \arg \min_{\{x_i\}, \{y_i\}} \sum_{i=1}^L \int_{x_i}^{x_{i+1}} (x - y_i)^2 p_X(x) dx$$

- 对 x_i 求偏导，得

$$\frac{\partial \sigma^2}{\partial x_i} = (x_i - y_i)^2 p_X(x_i) - (x_i - y_{i-1})^2 p_X(x_i) = 0 \implies x_i = \frac{y_i + y_{i-1}}{2}$$

由此导出**最近邻居准则**：每个分层电平 x_i 应取为相邻两个重建电平 y_{i-1} 和 y_i 的中点，这与 x 的分布无关。

- 对 y_i 求偏导，得

$$\frac{\partial \sigma^2}{\partial y_i} = -2 \int_{x_i}^{x_{i+1}} (x - y_i) p_X(x) dx = 0 \implies y_i = \frac{\int_{x_i}^{x_{i+1}} x p_X(x) dx}{\int_{x_i}^{x_{i+1}} p_X(x) dx}$$

由此导出**重心准则**：每个重建电平 y_i 应取为对应量化区间内 x 的条件期望，这与 x 的分布有关。

🔗 *Ex.* 非均匀量化器设计示例

设信源 X 服从概率密度函数 $p_X(x) = \begin{cases} 1+x, & -1 \leq x \leq 0, \\ 1-x, & 0 \leq x \leq 1 \end{cases}$ ，一个 2-bit 量化器若将分层电平设置在 -1 、 $-\frac{1}{2}$ 、 0 、 $\frac{1}{2}$ 、 1 处，则根据重心准则，对应的重建电平应为

$$\begin{aligned} y_1 &= \frac{\int_{-1}^{-1/2} x(1+x) dx}{\int_{-1}^{-1/2} (1+x) dx} = -\frac{2}{3}, & y_4 &= -y_1 = \frac{2}{3}, \\ y_2 &= \frac{\int_{-1/2}^0 x(1+x) dx}{\int_{-1/2}^0 (1+x) dx} = -\frac{2}{9}, & y_3 &= -y_2 = \frac{2}{9} \end{aligned}$$

显然，这组重建电平不满足最近邻居准则，因此需要重新调整分层电平，迭代进行直到收敛。

Lloyd-Max 算法综合考虑最近邻居准则和重心准则，通过迭代优化非均匀量化器的分层电平 $\{x_i\}$ 和重建电平 $\{y_i\}$ 。

GIVEN —

初始重建电平 $y_1^{(0)} < y_2^{(0)} < \dots < y_L^{(0)}$

REPEAT —

1. 更新分层电平：

$$x_i^{(k)} = \frac{y_i^{(k)} + y_{i+1}^{(k)}}{2}, \quad i = 1, 2, 3, \dots, L-1$$

2. 更新重建电平：

$$y_i^{(k+1)} = \frac{\int_{x_{i-1}^{(k)}}^{x_i^{(k)}} x p_X(x) dx}{\int_{x_{i-1}^{(k)}}^{x_i^{(k)}} p_X(x) dx}, \quad i = 1, 2, 3, \dots, L$$

UNTIL — 量化噪声 $\sigma_{(k+1)}^2$ 小于预设阈值

OUTPUT —

最终分层电平 $\{x_i^*\}$ 和重建电平 $\{y_i^*\}$

在此部分计算中，常见到如下积分：

$$\int x(kx+b) dx = \frac{kx^3}{3} + \frac{bx^2}{2} + C,$$

$$\int (x-y_0)^2(kx+b) dx = \frac{b(x-y)^3}{3} + k \left(\frac{x^4}{4} - \frac{2yx^3}{3} + \frac{y^2x^2}{2} \right) + C$$

压扩量化

对于幅值分布高度集中的信源，如语音信号，其幅值通常服从 **Laplace 分布** 或 **Gauss 分布**，均匀量化器难以取得较好的量化效果。注意到**非线性映射**可以改变随机变量的分布形状，因此可以先对信源进行**压扩 (companding)** 处理，再进行均匀量化，从而达到非均匀量化的效果。

常用的压扩函数有：

- **μ 律压扩**函数 $g(x) = x_{\max} \operatorname{sgn}(x) \frac{\log \left(1 + \mu \frac{|x|}{x_{\max}} \right)}{\log(1 + \mu)}$ ，其中 $\mu > 0$ 是压扩参数；
- **A 律压扩**函数 $g(x) = \begin{cases} A \operatorname{sgn}(x) \frac{|x|}{1 + \ln A}, & 0 \leq |x| \leq \frac{x_{\max}}{A}, \\ x_{\max} \operatorname{sgn}(x) \frac{1 + \ln \left(A \frac{|x|}{x_{\max}} \right)}{1 + \ln A}, & \frac{x_{\max}}{A} < |x| \leq x_{\max} \end{cases}$ ，

其中 $A > 1$ 是压扩参数。

高分辨率量化

当量化级数 L 很大时，量化间隔 Δ_i 很小，可以近似认为在量化区间

$I_i = (x_i, x_{i+1}]$ 上，**信源概率密度函数** $p_X(x)$ 近似为常数 $p_X(x_i^*)$ ，其中 $x_i^* \in I_i$ 。此时，量化噪声可近似为

$$\sigma^2 \approx \sum_{i=1}^L p_X(x_i^*) \int_{x_i}^{x_{i+1}} (x - y_i)^2 dx = \sum_{i=1}^L p_X(x_i^*) \frac{\Delta_i^3}{12}$$

由 Lagrange 中值定理，可取 $p_X(x_i^*) = \frac{1}{\Delta_i} \int_{x_i}^{x_{i+1}} p_X(x) dx = \frac{P_i}{\Delta_i}$ ，从而

$$\sigma^2 \approx \frac{1}{12} \sum_{i=1}^L P_i \Delta_i^2 \stackrel{A_i \equiv \Delta_i}{=} \frac{A^2}{12}$$

对其做无损压缩，**最小码长**为

$$\begin{aligned} H(Q(X)) &= - \sum_{i=1}^L (p_X(x_i^*) \Delta_i) \log(p_X(x_i^*) \Delta_i) \\ &= - \sum_{i=1}^L p_X(x_i^*) \Delta_i \log p_X(x_i^*) - \sum_{i=1}^L p_X(x_i^*) \Delta_i \log \Delta_i \\ &= - \int_{-\infty}^{+\infty} p_X(x) \log p_X(x) dx - \log A \\ &= b(X) - \log A \approx b(X) - \log \sigma - 1.8 \end{aligned}$$

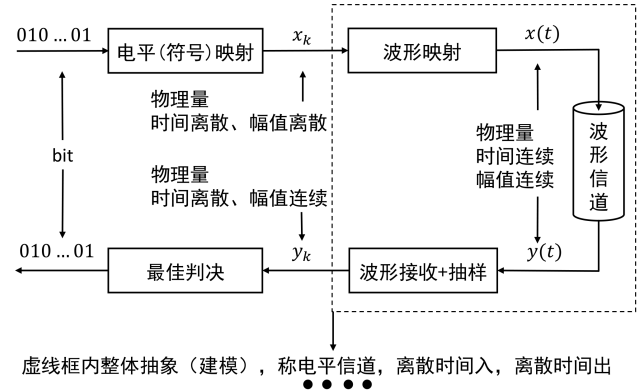
其中 $b(X) = - \int_{-\infty}^{+\infty} p_X(x) \log p_X(x) dx$ 是信源 X 的**微分熵**。

§3 数字调制

我们希望在实际物理信道中有效、可靠地传输 0/1 比特。为此，需要将离散的数字信号映射为适合物理信道传输的连续幅值信号，这一过程称为**数字调制 (digital modulation)**。

数字调制要解决的问题有：

- 0/1 比特的**逻辑量**需要映射为适合物理信道传输的**物理量**，如电压、电流、电磁波等；
- 映射后的信号需要适应物理信道的**带宽**、**噪声**限制，避免信号失真和误码；
- 映射方案需要考虑**频谱效率**、**功率效率**等性能指标，提升通信系统的整体性能。



§3.1 电平信道

电平信道的建模

电平信道是**数字调制**过程中**电平映射**后至**最佳判决**前的环节，输入 x_k 时间离散、幅值离散，输出 $y(t)$ 时间离散、幅值连续。

假定信道无记忆，电平信道可以建模为**条件概率分布** $p(y | x)$ ，表示在输入电平 x 时输出电平 y 的概率密度函数 (PDF)。

(*) Def. 加性白 Gauss 噪声电平信道

加性白 Gauss 噪声 (Additive White Gaussian Noise, AWGN) 电平信道是**离散时间 AWGN 信道**，其输入输出关系为

$$y = x + z, \quad \text{其中 } z \sim \mathcal{N}(0, \sigma^2)$$

即输出电平 y 是输入电平 x 与高斯白噪声 z 的和。对应的条件概率密度函数为

$$p(y | x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(y-x)^2}{2\sigma^2} \right)$$

二进制电平传输

本课程中，一律假设信源编码的输出比特 0/1 等概分布，即遵循**最优信源编码**的先验假设。若约定传输「0」时映射为电平 $-A$ ，传输「1」时映射为电平 $+A$ ，则输入电平 x 的概率分布为

$$p(A) = p(-A) = \frac{1}{2}$$

这里简记 $p(A) = \Pr\{x = A\}$ ， $p(-A) = \Pr\{x = -A\}$ 。

这样，对于给定的电平信道 $p(y | x)$ ，可以计算输出 y 的分布为

$$p(y) = p(y | A)p(A) + p(y | -A)p(-A)$$

如对上述**加性白 Gauss 噪声电平信道**，有

$$\begin{aligned} p(y) &= \frac{1}{2} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(y-A)^2}{2\sigma^2} \right) + \frac{1}{2} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left(-\frac{(y+A)^2}{2\sigma^2} \right) \\ &= \frac{1}{2\sqrt{2\pi\sigma^2}} \left[\exp \left(-\frac{(y-A)^2}{2\sigma^2} \right) + \exp \left(-\frac{(y+A)^2}{2\sigma^2} \right) \right] \end{aligned}$$

■ 最佳判决

由于一般的信道（如上述加性白 Gauss 噪声电平信道）上噪声是随机的，电平信道的输出也是随机的，而且

$$\forall y \in \mathbb{R}, \quad p(y|A) > 0, \quad p(y|-A) > 0$$

因此，需要一种准则在电平信道输出 y 处进行**最佳判决 (optimal decision)**，即根据观测到的 y 判定发送的是「0」还是「1」，以最小化**误码率 (bit error rate, BER)**。

► 最大后验概率判决准则

直观上，对于观测到的输出电平 y ，如果给出 y 的条件为 $+A$ 的概率更大，则判定发送的是「1」；如果给出 y 的条件为 $-A$ 的概率更大，则判定发送的是「0」。这一直观准则可以形式化为**最大后验概率 (maximum a posteriori, MAP) 判决准则**：

$$\hat{x} = \arg \max_{x \in \{-A, +A\}} \Pr\{x|y\}$$

► 最大似然判决准则

由 **Bayes 公式**，有 $\Pr\{x=A|y\} = \frac{p(y|A)p(A)}{p(y)}$ ，
 $\Pr\{x=-A|y\} = \frac{p(y|-A)p(-A)}{p(y)}$ ，而有**等概率条件** $p(A) = p(-A)$ ，因此 MAP 判决准则等价于

$$p(y|A) \underset{\hat{x}=-A}{\overset{\hat{x}=A}{\geq}} p(y|-A)$$

即**最大似然 (maximum likelihood, ML) 判决准则**。

► 最小 Euclid 距离判决准则

对上述加性白 Gauss 噪声电平信道，据 ML 判决准则有：

- $y = 0$ 时， $p(y|A) > p(y|-A)$ ，判决 $\hat{x} = A$ ；
- $y = 0$ 时， $p(y|A) < p(y|-A)$ ，判决 $\hat{x} = -A$ 。

因此，可以得到**最佳判决阈值 (optimal decision threshold)** 为 0，即

$$\hat{x} = \begin{cases} +A, & y \geq 0 \\ -A, & y < 0 \end{cases}$$

即对于 AWGN 电平信道，ML 判决准则则成为**最小 (Euclid) 距离 (minimum Euclidean distance, MED) 判决准则**，分别称 $y < 0$ 和 $y > 0$ 为判决到 $\hat{x} = -A$ 和 $\hat{x} = +A$ 的**判决域 (decision region)**。

■ 误符号率

设实际发送的电平为 x ，最佳判决的输出为 \hat{x} ，则**误符号率 (symbol error rate, SER)** 定义为

$$P_s = \Pr\{\hat{x} \neq x\}$$

对于上述加性白 Gauss 噪声电平信道，误符号率为

$$\begin{aligned} P_s &= \Pr\{\hat{x} \neq x\} \\ &= p(A) \Pr\{y < 0 | x = A\} + p(-A) \Pr\{y > 0 | x = -A\} \\ &= \frac{1}{2} \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-A)^2}{2\sigma^2}} dy + \frac{1}{2} \int_0^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y+A)^2}{2\sigma^2}} dy \\ &= \frac{1}{2} \int_A^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy + \frac{1}{2} \int_{-A}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy \\ &= \int_A^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{y^2}{2\sigma^2}} dy = \int_{A/\sigma}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \end{aligned}$$

积分 $\int_u^{+\infty} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt$ 没有初等函数表达式，通常用 **Q 函数** $Q(u)$ 表示，即上面的误符号率为

$$P_s = Q\left(\frac{A}{\sigma}\right)$$

为讨论方便，使用信息承载符号的**二阶矩** $\sigma_x^2 = \mathbb{E}[x^2]$ 表示信号功率，则对于二进制电平信道，有 $\sigma_x^2 = A^2$ 。定义**信噪比 (signal-to-noise ratio, SNR)** 为

$$\text{SNR} = \frac{\sigma_x^2}{\sigma^2} = \frac{A^2}{\sigma^2}$$

则**误符号率**为

$$P_s = Q\left(\frac{\sigma_x}{\sigma}\right) = Q\left(\sqrt{\text{SNR}}\right)$$

► 多进制实电平传输

在二进制电平传输中，一个电平一次传输仅承载 1 个 bit 的信息。为了提高频谱效率，可以扩大电平值 x 的取值集合，采用**多进制实电平传输**。

■ 多进制电平信道的构造

设电平信道每个符号承载 n 个 bit 信息，则电平信道的输入 x 需要取 $M = 2^n$ 个不同的实数值，即 **M 进制传输**。由于一个符号承载 $n = \log_2 M$ 个 bit 信息，因此考虑对 x 的二阶矩**相对承载 bit 数归一化**，定义

$$\bar{\sigma}_x^2 = \frac{\sigma_x^2}{\log_2 M} = \frac{\mathbb{E}[x^2]}{\log_2 M}$$

本课程中，不考虑非 2 的整数次幂的 M 。

为简化分析，通常采用**等间距电平**，即电平集合为

$$\mathcal{A} = \{-(M-1)A, -(M-3)A, \dots, -3A, -A, A, 3A, \dots, (M-3)A, (M-1)A\}$$

\mathcal{A} 中相邻两个电平的距离均为 $2A$ ，且电平关于 0 对称分布。一般取 M 为 2 的整数次幂，因此为偶数，但以下分析的结果也适用于奇数 M 。

≡ 8.9. 多进制电平的编码

对每个电平编码时，通常采用 **Gray 编码**，使得相邻电平仅有 1 bit 不同，以降低误码率。

例如， $M = 4$ 时，电平集合为 $\mathcal{A} = \{-3A, -A, A, 3A\}$ ，可以采用如下 Gray 编码：

电平 x	$-3A$	$-A$	A	$3A$
Gray 码	00	01	11	10

$M = 8$ 时，电平集合为 $\mathcal{A} = \{-7A, -5A, -3A, -A, A, 3A, 5A, 7A\}$ ，可以采用如下 Gray 编码：

电平 x	$-7A$	$-5A$	$-3A$	$-A$	A	$3A$	$5A$	$7A$
Gray 码	000	001	011	010	110	111	101	100

■ 最佳判决

对于多进制电平信道，最佳判决仍然从**最大后验概率 (MAP) 判决准则**出发，认为

$$\begin{aligned} \hat{x} &= \arg \max_{x \in \mathcal{A}} p(x|y) && (\text{MAP 准则}) \\ &= \arg \max_{x \in \mathcal{A}} \frac{p(y|x)p(x)}{p(y)} = \arg \max_{x \in \mathcal{A}} p(y|x)p(x) \\ &\stackrel{\text{等概假设}}{=} \arg \max_{x \in \mathcal{A}} p(y|x) && (\text{ML 准则}) \\ &\stackrel{\text{AWGN}}{=} \arg \max_{x \in \mathcal{A}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-x)^2}{2\sigma^2}\right) \\ &= \arg \min_{x \in \mathcal{A}} (y-x)^2 && (\text{MED 准则}) \end{aligned}$$

即在输入遵循**最优信源编码的等概**先验、信道为 AWGN 电平信道的条件下，MAP 准则可逐步化归到 ML 和 MED 准则。根据 MED 准则，最佳判决即为选择与输出电平 y 距离最近的输入电平 $x \in \mathcal{A}$ ，因此**判决门限**在

$$\begin{aligned} &\left\{ \frac{(x_i + x_{i+1})}{2} \mid x_i, x_{i+1} \in \mathcal{A}, i = 1, 2, \dots, M-1 \right\} \\ &= \{-(M-2)A, -(M-4)A, \dots, -2A, \\ &\quad 0, 2A, \dots, (M-4)A, (M-2)A\} \end{aligned}$$

■ 差错分析

► 误符号率

对 M 进制传输，电平与判决门限之间的距离仍为 A ，但不同于**二进制情形**的是，**中间 $(M-2)$ 个输入电平有两个方向可能判错**，因此误符号率为

$$\begin{aligned} P_s &= \sum_{x \in \mathcal{A}} p(x) \Pr\{\hat{x} \neq x | x\} \\ &= 2 \times \frac{1}{M} \times Q\left(\frac{A}{\sigma}\right) + (M-2) \times \frac{1}{M} \times 2Q\left(\frac{A}{\sigma}\right) \\ &= \frac{2(M-1)}{M} Q\left(\frac{A}{\sigma}\right) \end{aligned}$$

与**二进制情形**类似，定义**信噪比 (SNR)** 为

$$\text{SNR} = \frac{\sigma_x^2}{\sigma^2} = \frac{\mathbb{E}[x^2]}{\sigma^2} = \frac{1}{\sigma^2} \frac{(M^2-1)A^2}{3}$$

即 $\frac{A}{\sigma} = \sqrt{\frac{3}{M^2-1}\text{SNR}}$ ，则误符号率为

$$P_s = \frac{2(M-1)}{M}Q\left(\sqrt{\frac{3}{M^2-1}\text{SNR}}\right)$$

► 误比特率

当 M 很大时，近似取

$$P_s \approx 2Q\left(\sqrt{\frac{3}{M^2-1}\text{SNR}}\right)$$

误符号率 P_s 并不能直接反映通信系统的性能，因为每个符号承载 $n = \log_2 M$ 个 bit 信息，需要计算**误 bit 率 (bit error rate, BER)** P_b 。在 Gray 编码下，**高信噪比**时可认为只有**相邻电平**之间的判决错误，即每个符号错误只错 1 bit。因此近似有

$$P_b \approx \frac{P_s}{\log_2 M} \approx \frac{2}{\log_2 M}Q\left(\sqrt{\frac{3}{M^2-1}\text{SNR}}\right)$$

逐 bit 意义下，信噪比 $\text{SNR} = \frac{\sigma_x^2}{\sigma^2}$ 替换为**归一化信噪比 (normalized SNR)** $\frac{\sigma_x^2}{\sigma^2}$ ，即

$$P_b \approx \frac{2}{\log_2 M}Q\left(\sqrt{\frac{3 \log_2 M}{M^2-1} \frac{\sigma_x^2}{\sigma^2}}\right)$$

一般地，保留原始误符号率表达式，有

$$\begin{aligned} P_b \approx \frac{P_s}{\log_2 M} &= \frac{2(M-1)}{M \log_2 M}Q\left(\sqrt{\frac{3}{M^2-1}\text{SNR}}\right) \\ &= \frac{2(M-1)}{M \log_2 M}Q\left(\sqrt{\frac{3 \log_2 M}{M^2-1} \frac{\sigma_x^2}{\sigma^2}}\right) \end{aligned}$$

Note 含直流偏置的多进制电平传输

若存在直流偏置 D ，即电平集合为

$$\mathcal{A} = \{D - (M-1)A, D - (M-3)A, \dots, D - 3A, D - A, D + A, D + 3A, \dots, D + (M-3)A, D + (M-1)A\}$$

则符号能量 $\mathbb{E}[x^2] = \frac{M^2-1}{3}A^2 + D^2$ 。

差错概率并不建立在符号能量上，而是建立在符号间距 $2A$ 上，因此引入

$$\zeta = \frac{D}{A\sqrt{\frac{M^2-1}{3}}}, \text{ 则 } \mathbb{E}[x^2] = \frac{M^2-1}{3}A^2(1+\zeta^2), \text{ 误符号率为}$$

$$P_s = \frac{2(M-1)}{M}Q\left(\frac{A}{\sigma}\right) = \frac{2(M-1)}{M}Q\left(\sqrt{\frac{3}{M^2-1} \frac{1}{1+\zeta^2} \frac{\mathbb{E}[x^2]}{\sigma^2}}\right)$$

复电平传输

复电平传输是将信息映射为**复数域**的电平信号进行传输，以提高频谱效率和抗干扰能力。复电平信道的输入 $x = x_I + \mathbb{j}x_Q$ 和输出 $y = y_I + \mathbb{j}y_Q$ 均为复数。

Note 复电平实部、虚部分量的下标含义

在通信领域，对复数信号值通常使用下标 I、Q 而不是 x 、 y 或 r 、 i ，其中 I 表示同相分量 (in-phase component)，Q 表示正交分量 (quadrature component)。

考虑 AWGN 复电平信道，输入输出关系为

$$y = x + z, \quad \text{其中 } z = z_I + \mathbb{j}z_Q \sim \mathcal{CN}(0, 2\sigma^2)$$

即输出电平 y 是输入电平 x 与**复 Gauss 白噪声** z 的和， z 满足

$$p(z_I, z_Q) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{z_I^2 + z_Q^2}{2\sigma^2}\right)$$

类似于实 AWGN 电平信道，复 AWGN 电平信道的 MAP 判决仍然等价到**最小 Euclid 距离 (MED) 判决准则**，即选择与输出电平 y 距离最近的输入电平 $x \in \mathcal{A}$ 。

■ QAM 输入的分析

考虑复电平信道的输入 x 取自

$$\mathcal{A} = \{x_I + \mathbb{j}x_Q \mid x_I, x_Q \in \{-(L-1)A, -(L-3)A, \dots, -A, A, \dots, (L-3)A, (L-1)A\}\}$$

实轴、虚轴均取 L 个等间距电平，共有 $M = L^2$ 个复电平。由于这种信源电平分布来自于**正交振幅调制 (quadrature amplitude modulation, QAM)**，因此称为 **M -QAM 星座图**。

由于复 Gauss 白噪声的实部和虚部相互独立，分别服从均值为 0、方差为 σ^2 的 Gauss 分布，因此复 AWGN 电平信道可视为两路独立的**实 AWGN 电平信道**的组合，恰对应于实轴和虚轴的 L 进制电平传输。

两路信道相互独立，出现差错的情况是**任意一路实 AWGN 电平信道出现差错**。因此，信道的误符号率为

$$\begin{aligned} P_s &= 1 - \Pr\{\hat{x}_I = x_I\} \times \Pr\{\hat{x}_Q = x_Q\} \\ &= 1 - (1 - P_{s,I})(1 - P_{s,Q}) = P_{s,I} + P_{s,Q} - P_{s,I}P_{s,Q} \end{aligned}$$

其中， $P_{s,I}$ 和 $P_{s,Q}$ 分别为实轴和虚轴 AWGN 电平信道的误符号率，均为

$$P_{s,I} = P_{s,Q} = \frac{2(\sqrt{M}-1)}{\sqrt{M}}Q\left(\frac{A}{\sigma}\right)$$

直接忽略掉二次项 $P_{s,I}P_{s,Q}$ ，得到近似误符号率为

$$P_s \approx \frac{4(\sqrt{M}-1)}{\sqrt{M}}Q\left(\frac{A}{\sigma}\right)$$

输入电平的二阶矩为

$$\sigma_x^2 = \mathbb{E}[|x|^2] = 2 \times \frac{(\sqrt{M})^2 - 1}{3}A^2 = \frac{2(M-1)}{3}A^2$$

仍定义**信噪比 (SNR)**为 $\frac{\sigma_x^2}{\sigma^2}$ ，则有 $\frac{A}{\sigma} = \sqrt{\frac{3}{2(M-1)}\text{SNR}}$ ，误符号率为

$$P_s \approx \frac{4(\sqrt{M}-1)}{\sqrt{M}}Q\left(\sqrt{\frac{3}{2(M-1)}\text{SNR}}\right)$$

类似地，取**Gray 映射**，近似误比特率为

$$P_b \approx \frac{4}{\log_2 M} \left(1 - \frac{1}{\sqrt{M}}\right)Q\left(\sqrt{\frac{3 \log_2 M}{2(M-1)} \frac{\sigma_x^2}{\sigma^2}}\right)$$

当 M 较大时， $P_b \approx \frac{4}{\log_2 M}Q\left(\sqrt{\frac{3 \log_2 M}{2(M-1)} \frac{\sigma_x^2}{\sigma^2}}\right)$ 。

■ PSK 输入的分析

另一种常见的复电平传输需求来自**相移键控 (phase shift keying, PSK)**，其输入电平取自

$$\mathcal{A} = \{Ae^{\mathbb{j}2\pi k/M} \mid k = 0, 1, \dots, M-1\}$$

均匀分布在以原点为中心、半径为 A 的圆周上，幅角差为 $\theta = \frac{2\pi}{M}$ 。由于 PSK 星座图的输入电平**模长相同**，因此输入电平的二阶矩为 $\sigma_x^2 = A^2$ ，信噪比为 $\text{SNR} = \frac{A^2}{\sigma^2}$ 。

对于 PSK 星座图，MED 判决给出的**判决门限**为相邻复电平之间的角平分线，判决域张角为 θ 。由**对称性**，其差错分析只用分析任一个电平的差错概率，我们选 $x = A$ 进行分析。设判决输出为 \hat{x} ，则误符号率为

$$\begin{aligned} P_s &= \Pr\{\hat{x} \neq A \mid x = A\} \\ &= \Pr\left\{\arg(A + z) > \frac{\theta}{2} = \frac{\pi}{M} \mid x = A\right\} \end{aligned}$$

求解比较困难。考虑 M 较大、 $\frac{A}{\sigma}$ 较大的情形，可近似认为 $x = A$ 点附近的判决域边界为两条**平行直线**，此时**只有 z_Q 引起差错**，因此有近似误符号率为

$$\begin{aligned} P_s &\approx 2 \times \Pr\left\{z_Q > A \sin \frac{\pi}{M}\right\} = 2Q\left(\frac{A \sin \frac{\pi}{M}}{\sigma}\right) \\ &= 2Q\left(\sin \frac{\pi}{M} \sqrt{\text{SNR}}\right) \end{aligned}$$

类似地，取**Gray 映射**，近似误比特率为

$$P_b \approx \frac{2}{\log_2 M} Q \left(\sin \frac{\pi}{M} \sqrt{\log_2 M \cdot \frac{\sigma_x^2}{\sigma^2}} \right) \\ \approx \frac{2}{\log_2 M} Q \left(\frac{\pi}{M} \sqrt{\log_2 M \cdot \frac{\sigma_x^2}{\sigma^2}} \right)$$

§3.2 波形信道

波形传输

电平是一个物理量，但不能脱离物理实体存在。在通信系统中，**电平**通常通过**波形**来传输，因此需要将**电平信道**延伸到**波形信道**。

■ 波形映射

将离散的电平映射到连续的波形上的过程称为**波形映射**。常见的波形映射方式有：

- **脉冲振幅调制 (PAM)**：将不同的电平映射为不同幅度的脉冲信号。
- **频移键控 (FSK)**：将不同的电平映射为不同频率的正弦波。
- **相移键控 (PSK)**：将不同的电平映射为不同相位的正弦波。
- **正交振幅调制 (QAM)**：结合幅度和相位的变化来表示不同的电平。
- **正交频分复用 (OFDM)**：将数据分散到多个正交的子载波上进行传输。

本课程专注于**线性调制 (linear modulation)**，即通过线性调幅一个**成形脉冲 (shaping pulse)** $p(t)$ 来实现波形映射。传输电平为 x 的符号时，对应的波形为

$$x(t) = xp(t)$$

其中， $p(t)$ 是**归一化**的成形脉冲，满足

$$\int_{-\infty}^{+\infty} p^2(t) dt = 1$$

这样，传输一个符号的能耗即为

$$E_s = \mathbb{E} \left[\int_{-\infty}^{+\infty} x^2(t) dt \right] = \mathbb{E}[x^2] \int_{-\infty}^{+\infty} p^2(t) dt = \mathbb{E}[x^2] = \sigma_x^2$$

■ 波形信道模型

本课程重点关注**加性白 Gauss 噪声 (AWGN)** 对数字通信的影响，因此引入如下波形信道模型：

(*) *Def.* 加性白 Gauss 噪声波形信道

加性白 Gauss 噪声波形信道是**连续时间 AWGN 信道**，信道输入为波形 $x(t)$ ，输出为波形 $y(t)$ ，满足

$$y(t) = x(t) + z(t)$$

其中， $z(t)$ 是白 Gauss 噪声 (white Gaussian noise)，满足

1. 对任意时刻 t ， $z(t)$ 服从均值为 0 的 Gauss 分布；
2. 对任意时刻序列 t_1, t_2, \dots, t_n ，随机变量组 $\{z(t_1), z(t_2), \dots, z(t_n)\}$ 服从均值为 $\mathbf{0}$ 的 n 元 Gauss 分布；
3. 功率谱密度函数 $S_z(f) \equiv \frac{n_0}{2}$ ，其中 n_0 为常数。

二元波形的最佳接收

■ 简单二元波形的最佳接收

考虑最简单的二元波形传输系统，传输的波形形式为

$$x(t) = \begin{cases} V, & t \in [0, T], \quad \text{传输「1」}, \\ -V, & t \in [0, T], \quad \text{传输「0」} \end{cases}$$

即取成形脉冲为矩形脉冲

$$p(t) = \frac{1}{\sqrt{T}}, \quad t \in [0, T]$$

► 直接抽样的接收

尝试**直接抽样**接收波形 $y(t)$ ，注意到

$$\mathbb{E}[x^2(t_i)] = R_z(t_i - t_i) = \frac{n_0}{2} \delta(0) = +\infty$$

因此，直接抽样接收的噪声功率为无穷大，显然不可行。

► 区段积分的接收

观察 $p(t)$ 与 $z(t)$ 的波形特征，考虑通过**区段积分**的方式处理波形 $y(t)$ ，即计算

$$y = \int_0^T y(t) dt = \underbrace{\int_0^T x(t) dt}_x + \underbrace{\int_0^T z(t) dt}_{z \sim \mathcal{N}\left(0, \frac{n_0 T}{2}\right)}$$

得到一个**等效电平信道**，其输入为**原来的符号电平** $x \in \{VT, -VT\}$ ，输出为等效输出电平 y ，满足

$$y = x + z$$

其中， z 服从均值为 0、方差为 $\frac{n_0 T}{2}$ 的 Gauss 分布。这个电平信道的**信噪比**为

$$\text{SNR} = \frac{\sigma_x^2}{\sigma^2} = \frac{V^2 T^2}{n_0 T/2} = \frac{V^2 T}{n_0/2}$$

■ 标准等价电平信道

对于上面**简单二元波形的最佳接收**问题，我们希望能获得一个**最易于处理的等效电平信道**，于是考虑将噪声的方差固定为 $\frac{n_0}{2}$ ，这要求处理时

$$y = \frac{1}{\sqrt{T}} \int_0^T y(t) dt = \int_0^T y(t) \cdot \frac{1}{\sqrt{T}} dt = \int_0^T y(t) \cdot p(t) dt$$

此时等效电平输入为 $x \in \{V\sqrt{T}, -V\sqrt{T}\}$ 。

事实上，对于任意**能量归一化**成形脉冲 $p(t)$ ，都可以通过如下方式获得一个**标准等价电平信道**：

$$y = \int_{-\infty}^{+\infty} y(t) \cdot p(t) dt = \int_{-\infty}^{+\infty} [x(t) + z(t)] \cdot p(t) dt = x + z$$

其中， $x = \int_{-\infty}^{+\infty} x(t) \cdot p(t) dt$ ， $z = \int_{-\infty}^{+\infty} z(t) \cdot p(t) dt \sim \mathcal{N}\left(0, \frac{n_0}{2}\right)$ 。

由于 $p(t)$ 是归一化的，因此**等效电平信道的输入 x 与最初波形映射时的符号电平 x** 相同，即信噪比为

$$\text{SNR} = \frac{\sigma_x^2}{\sigma^2} = \frac{E_s}{n_0/2}$$

考虑使用任意能量归一化的 $g(t)$ 做类似的线性处理

$$\hat{y} = \int_{-\infty}^{+\infty} y(t) \cdot g(t) dt = \int_{-\infty}^{+\infty} x(t) \cdot g(t) dt + \int_{-\infty}^{+\infty} z(t) \cdot g(t) dt = \hat{x} + \hat{z}$$

其中，仍有 $\hat{z} \sim \mathcal{N}\left(0, \frac{n_0}{2}\right)$ 。由**Cauchy-Schwarz 不等式**，这一处理的信噪比为

$$\widehat{\text{SNR}} = \frac{\sigma_{\hat{x}}^2}{\sigma_{\hat{z}}^2} = \frac{\mathbb{E} \left[\left(\int_{-\infty}^{+\infty} x(t) \cdot g(t) dt \right)^2 \right]}{n_0/2} = \frac{\mathbb{E}[x^2] \left(\int_{-\infty}^{\infty} p(t) g(t) dt \right)^2}{n_0/2} \\ = \frac{E_s}{n_0/2} \cdot \left(\int_{-\infty}^{+\infty} p(t) g(t) dt \right)^2 \\ \leq \frac{E_s}{n_0/2} \cdot \left(\int_{-\infty}^{+\infty} p^2(t) dt \right) \left(\int_{-\infty}^{+\infty} g^2(t) dt \right) = \frac{E_s}{n_0/2}$$

因此，使用成形脉冲 $p(t)$ 进行线性处理可以获得最大的信噪比。

► *Thm.* 二进制波形传输的最佳接收

对于一般的二进制传输波形 $x(t) = xp(t)$ （其中 $x \in \{-\sqrt{E_s}, \sqrt{E_s}\}$ ，成形脉冲 $p(t)$ 满足能量归一化条件），在 AWGN 波形信道 $y(t) = x(t) + z(t)$ 上传输，其最佳接收处理为

$$y = \int_{-\infty}^{+\infty} y(t) \cdot p(t) dt = x + z$$

其中 $z \sim \mathcal{N}\left(0, \frac{n_0}{2}\right)$ ，该等效电平信道的信噪比为 $\text{SNR} = \frac{E_s}{n_0/2}$ 。

对于任给的 $g(t)$ ，线性处理 $\hat{y} = \int_{-\infty}^{+\infty} x(t) g(t) dt$ 的信噪比不超过 $\frac{E_s}{n_0/2}$ 。

多进制波形的最佳接收

■ 多进制标准等价电平信道

10 / 21 考虑一般的 M 进制波形传输系统，传输的波形形式为

$$x(t) = xp(t)$$

其中, $x \in \mathcal{A}$ 来自 M 进制 电平信道, $p(t)$ 为能量归一化的成形脉冲。

Thm. 多进制波形传输的最佳接收

对于一般的 M 进制传输波形 $x(t) = xp(t)$ (其中 $x \in \mathcal{A}$, 成形脉冲 $p(t)$ 满足能量归一化条件), 在 AWGN 波形信道 $y(t) = x(t) + z(t)$ 上传输, 其最佳接收处理为

$$y = \int_{-\infty}^{+\infty} y(t) \cdot p(t) dt = x + z$$

其中 $z \sim \mathcal{N}(0, \frac{n_0}{2})$, 该等效电平信道的信噪比为 $\text{SNR} = \frac{E_s}{n_0/2}$ 。

■ 成形脉冲未归一化的情况

对于未归一化的成形脉冲 $p(t)$, 不妨设 $p(t) = \sqrt{E_p} \hat{p}(t)$ 且 $\hat{p}(t)$ 满足能量归一化条件, 考虑使用相同的线性处理方式, 则

$$\begin{aligned} y &= \int_{-\infty}^{+\infty} y(t) \cdot p(t) dt = \int_{-\infty}^{+\infty} x(t) \cdot p(t) dt + \int_{-\infty}^{+\infty} z(t) \cdot p(t) dt \\ &= x \int_{-\infty}^{+\infty} p^2(t) dt + \int_{-\infty}^{+\infty} z(t) \cdot p(t) dt \\ &= E_p x + z \end{aligned}$$

其中等价噪声电平

$$z = \int_{-\infty}^{+\infty} z(t) \cdot p(t) dt \sim \mathcal{N}\left(0, \frac{n_0}{2} \int_{-\infty}^{+\infty} p^2(t) dt\right) = \mathcal{N}\left(0, \frac{n_0}{2} E_p\right)$$

即该等效电平信道的信噪比为

$$\text{SNR} = \frac{E_p^2 \sigma_x^2}{(n_0/2) E_p} = \frac{E_p E_s}{n_0/2}$$

最佳接收的实现

■ 线性调制波形接收的实现

对于以归一化成形脉冲 $p(t)$ 线性调制的波形传输系统, 已知最佳接收处理为

$$y = \int_{-\infty}^{+\infty} y(t) \cdot p(t) dt$$

除开使用积分器实现该处理外, 还可以使用匹配滤波器 (matched filter) 实现为:



其中, 滤波器的冲激响应为 $p(-t)$, 输出即为

$$y(t) = y(t) * p(-t) = \int_{-\infty}^{+\infty} y(\tau) p(-(t-\tau)) d\tau = \int_{-\infty}^{+\infty} y(\tau) p(\tau-t) d\tau$$

因此, 在时刻 $t = 0$ 处抽样, 即可得到最佳接收结果

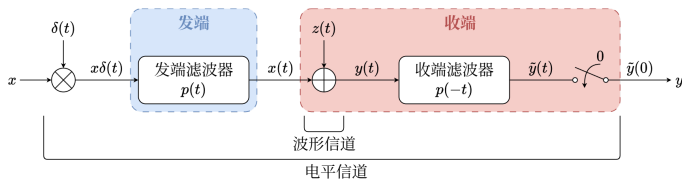
$$y = y(0) = \int_{-\infty}^{+\infty} y(\tau) p(\tau) d\tau$$

■ 波形传输的收发联合模型

波形信道的发送端为

$$x(t) = xp(t) = x\delta(t) * p(t)$$

同样具有滤波器的实现形式。因此, 可以把收发两端的滤波器形式放在一个视野中, 得到如下收发联合模型:

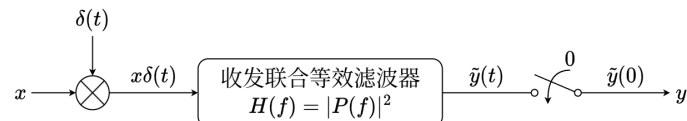


信号部分的整个传输过程同样是 LTI 系统, 整体的冲激响应为

$$b(t) = p(t) * p(-t)$$

从频域来看, 传输系统的频率响应为

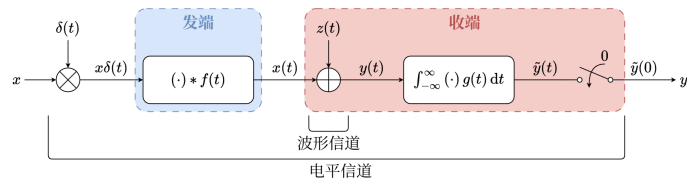
$$H(f) = \mathcal{F}\{p(t) * p(-t)\} = P(f) \cdot P^*(f) = |P(f)|^2$$



■ 成形脉冲与匹配脉冲的放缩

考虑使用一般的成形脉冲 $f(t)$ 和匹配脉冲 $g(t)$, 接收处理为

$$\begin{aligned} y &= \int_{-\infty}^{+\infty} y(t) g(t) dt = \int_{-\infty}^{+\infty} x(t) g(t) dt + \int_{-\infty}^{+\infty} z(t) g(t) dt \\ &= x \int_{-\infty}^{+\infty} f(t) g(t) dt + \int_{-\infty}^{+\infty} z(t) \cdot g(t) dt \end{aligned}$$



考虑最佳接收的条件, 要求成形脉冲与匹配脉冲满足

$$f(t) = \beta_1 p(t), \quad g(t) = \beta_2 p(t)$$

其中 $p(t)$ 满足能量归一化条件, $\beta_1, \beta_2 > 0$ 。此时, 等效电平信道为

$$y = \beta_1 \beta_2 x + z, \quad z \sim \mathcal{N}\left(0, \frac{n_0}{2} \beta_2^2\right)$$

其信噪比为

$$\text{SNR} = \frac{\beta_1^2 \beta_2^2 \sigma_x^2}{\frac{n_0}{2} \beta_2^2} = \frac{\beta_1^2 \sigma_x^2}{n_0/2}$$

因此, 成形脉冲的放缩系数 β_1 决定了信噪比的高低, 而匹配脉冲的放缩系数 β_2 则不会影响信噪比。

若只给定等效系统的 $b(t)$ 或 $H(f)$, 有

$$\int_{-\infty}^{+\infty} H(f) df = b(0) = \int_{-\infty}^{+\infty} f(\tau) g(\tau) d\tau = \beta_1 \beta_2$$

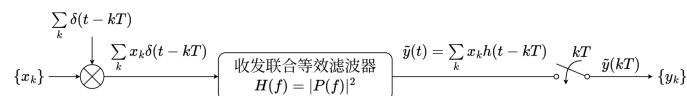
当可确定 $f(t) = g(t)$ 时, 即知 $\beta_1 = \sqrt{b(0)} = \sqrt{\int_{-\infty}^{+\infty} H(f) df}$, 可由此确定信噪比; 否则, 只能确定波形能量与电平能量的比值。

§3.3 载波传输

■ 无失真传输的 Nyquist 准则

我们会在一个波形信道上前后传输多个符号, 因此希望前后符号之间互不干扰。考虑在收发联合模型中传输一系列符号 $\{x_k\}$, 符号调制的冲激串

$$\sum_{k=-\infty}^{\infty} x_k \delta(t - kT) \text{ 间隔为 } T。$$



要满足无失真传输 (distortion-free transmission), 需要收发联合滤波器的冲激响应 $b(t)$ 满足

$$b(kT) = \begin{cases} 1, & k = 0, \\ 0, & k = \pm 1, \pm 2, \pm 3, \dots \end{cases}$$

即采样时刻 $t = kT$ 处, 只有当前符号对输出有贡献, 其他符号均无贡献。

对上述收发联合系统的约束仅在离散点 $t = kT$ 处，因此系统可以在其他时刻有任意响应形式，从而可以设计出满足该约束的多种成形脉冲 $p(t)$ 。

为讨论其频域特征，我们对 $b(t)$ 加窗，只关注约束部分的性质。由于约束部分为离散点，需使用宽度趋近于 0 而增益趋于 $+\infty$ 的窗函数，即使用 Dirac 梳状函数 (Dirac comb) $W_T(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT)$ 加窗，得到

$$b_T(t) = b(t)W_T(t) = \sum_{k=-\infty}^{\infty} b(kT)\delta(t - kT)$$

使用 Dirac 梳状函数 $W_T(t) = \sum_{k=-\infty}^{\infty} \delta(t - kT)$ 加窗，得到

$$b_T(t) = b(t)W_T(t) = \sum_{k=-\infty}^{\infty} b(kT)\delta(t - kT) = \delta(t)$$

其频域特征为

$$\begin{aligned}\mathcal{F}\{b(t)W_T(t)\} &= H(f) * \mathcal{F}\{W_T(t)\} = H(f) * \frac{1}{T}W_{1/T}(f) \\ &= \frac{1}{T} \sum_{n=-\infty}^{\infty} H\left(f - \frac{n}{T}\right) = 1\end{aligned}$$

🔗 *Thm.* Nyquist 准则

要实现**无失真传输**，收发联合滤波器的频率响应 $H(f)$ 需满足

$$\sum_{n=-\infty}^{\infty} H(f - nR_s) = T$$

其中， $R_s = \frac{1}{T_s}$ 为**符号速率**， T_s 为符号发送的间隔时间；右侧一般取 $T = T_s$ ，但也可以取其他值以调整系统增益。

为确保因果性，我们对符号延时 T 后抽样，即 Dirac 梳状函数加窗结果为

$$b_T(t) = b(t)W_T(t) = \sum_{k=-\infty}^{\infty} b(kT)\delta(t - kT) = \delta(t - T)$$

其频域特征为

$$\mathcal{F}\{b(t)W_T(t)\} = H(f) * \frac{1}{T}W_{1/T}(f) = \frac{1}{T} \sum_{n=-\infty}^{\infty} H\left(f - \frac{n}{T}\right) = \mathrm{e}^{-\mathrm{j}2\pi fT}$$

因此，Nyquist 准则改写为

$$\sum_{n=-\infty}^{\infty} H(f - nR_s) = T \cdot \mathrm{e}^{-\mathrm{j}2\pi fT}$$

■ 带限波形信道的 Nyquist 准则

Nyquist 准则表明，为实现无失真传输，收发联合滤波器的频率响应 $H(f)$ 需满足其在频域上**以符号速率 R_s 为间隔周期重复**时，叠加结果为**常数 T** 。考虑 $H(f)$ 带限于 $|f| \leq W$ ，则

- 若 $R_s > 2W$ ，则各周期不重叠，无法满足 Nyquist 准则；
- 若 $R_s = 2W$ ，则各周期在边界处相接，称为**临界速率 (critical rate)**，此时 $H(f)$ 须为理想低通滤波器；
- 若 $R_s < 2W$ ，则各周期之间有重叠的**过渡带**，存在多种满足 Nyquist 准则的 $H(f)$ 形式。

即，Nyquist 准则要求带限波形信道的**符号速率 R_s 不得超过信道带宽 W 的两倍**，否则无法实现无失真传输。

(*) *Def.* 频谱效率

频谱效率 (spectral efficiency) 描述了在单位带宽内支持传输的信息量，定义为

$$\eta = \frac{R_b}{W} = \frac{R_s}{W} \cdot \log_2 M$$

其中， R_b 为比特率 (bit rate)， R_s 为符号速率 (symbol rate)， W 为信道带宽， $M = |\mathcal{A}|$ 为每个符号所承载的电平集合大小。

对于带限波形信道，Nyquist 准则对频谱效率的限制为

$$\eta = \frac{R_s}{W} \cdot \log_2 M < 2 \log_2 M$$

■ 过渡带的残留对称条件

考虑 $H(f)$ 在过渡带内的形式，设符号速率为 R_s ，则 Nyquist 准则要求

$$H(f) + H(f - R_s) = T, \quad f \in [0, R_s]$$

若 $H(f)$ 为偶函数（即成形脉冲 $p(t)$ 为实值信号），则在过渡带内还满足

$$H(f) + H(R_s - f) = T, \quad f \in [0, R_s]$$

即 $H(f)$ 在过渡带内**关于 $f = \frac{R_s}{2}$ 对称**，称为**残留对称条件 (residual symmetry condition)**。

▶ 数字基带传输

数字基带传输 (digital baseband transmission) 指的是**不经过频率变换**，直接在低频段（基带）上传输数字信号的方式。数字基带传输系统通常使用**升余弦滤波器**作为成形脉冲 $p(t)$ ，以实现无失真传输。

■ 升余弦滤波系统

用余弦函数的半个周期作为 $H(f)$ 的过渡带，即得到**升余弦滤波器 (raised-cosine filter)**，其频率响应为

$$H(f) = \begin{cases} T, & |f| \leq \frac{1-\alpha}{2}R_s, \\ \frac{T}{2} \left(1 + \cos\left(\frac{\pi T_s}{\alpha} \left(|f| - \frac{1-\alpha}{2T_s}\right)\right)\right), & \frac{1-\alpha}{2}R_s < |f| \leq \frac{1+\alpha}{2}R_s, \\ 0, & |f| > \frac{1+\alpha}{2}R_s \end{cases}$$

其中， $\alpha \in [0, 1]$ 为**滚降系数 (roll-off factor)**，决定过渡带的宽度；升余弦滤波器的带宽为 $W = \frac{1+\alpha}{2}R_s = \frac{1+\alpha}{2T_s}$ ，频谱效率为

$$\eta = \frac{R_s}{W} \cdot \log_2 M = \frac{2 \log_2 M}{1 + \alpha}$$

特别地，

- 当 $\alpha = 0$ 时，升余弦滤波器退化为**理想低通滤波器**，带宽为 $W = \frac{R_s}{2}$ ；
- 当 $\alpha = 1$ 时，升余弦滤波器的过渡带宽度最大，带宽为 $W = R_s$ 。

🔗 *Remark* 题型：升余弦基带信道

典型题型 给定 R_b 和 W （即给定 η ），求满足无失真传输的 M 和对应的 α 。

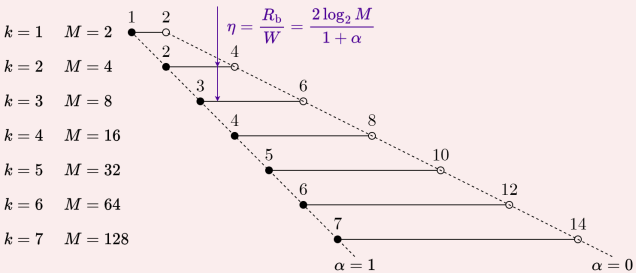
这个问题的约束条件为

$$\frac{R_b}{W} = \eta = \frac{2 \log_2 M}{1 + \alpha}$$

对此欠定问题，一般补充两个条件：

- M 取 **2 的整数次幂**，即 $M = 2^k, k \in \mathbb{N}^*$ ；
- α 不取退化为理想低通滤波器的情况，即 $0 < \alpha \leq 1$ 。

此时，可通过枚举 k 的方式求解该问题。



另外，可以推导出

- 升余弦滤波器的时域**冲激响应**为

$$b(t) = \frac{\sin(\pi R_s t)}{\pi R_s t} \cdot \frac{\cos(\pi \alpha R_s t)}{1 - (2\alpha R_s t)^2} = \text{sinc}(R_s t) \cdot \frac{\cos(\pi \alpha R_s t)}{1 - (2\alpha R_s t)^2}$$

- 升余弦滤波器对应成形脉冲 $p(t)$ 的**能量**为

$$\|p(t)\|^2 = \int_{-\infty}^{+\infty} |P(f)|^2 \mathrm{d}f = \int_{-\infty}^{+\infty} H(f) \mathrm{d}f = 1$$

- 升余弦滤波系统的**功率**为

$$P = E_s R_s = \frac{2WE_s}{1 + \alpha}$$

■ 基带信号的功率谱

通信信号 $x(t)$ 是一个**随机过程 (stochastic process)**，我们希望求解其**功率谱密度 (power spectral density)**

$$S_X(\omega) = \lim_{T \rightarrow +\infty} \frac{1}{2T} \mathbb{E} \left[\left| \int_{-T}^T x(t) e^{-j2\pi f t} dt \right|^2 \right]$$

给定符号序列 $\{x_k\}_{-\infty}^{\infty}$ 时， $x(t)$ 的取值 $\sum_k x_k p(t - kT)$ 成为一个确定的**样本轨道 (sample path)**，对其可以用确定性信号的频域分析方法。对这一样本轨道，期望函数为 $\mathbb{E}[x(\tau)] = \mathbb{E}[x_k] \sum_k p(t - kT)$ ，自相关函数为

$$\begin{aligned} \tilde{R}_x(t + \tau, t) &= \mathbb{E} \left[\sum_k x_k p(t + \tau - kT) \cdot \sum_m x_m p(t - mT) \right] \\ &= \sum_k \sum_m \mathbb{E}[x_k x_m] p(t + \tau - kT) p(t - mT) \\ &= \sum_k \sum_m R_x[k - m] p(t + \tau - kT) p(t - mT) \end{aligned}$$

均为周期为 T 的周期函数，因此 $x(t)$ 不是**宽平稳随机过程**。不过，可以通过**对时间 t 平均**的方式，得到一个与时间无关的自相关函数

$$\overline{R_x(\tau)} = \frac{1}{T} \int_0^T \tilde{R}_x(t + \tau, t) dt$$

于是，据**Wiener-Khinchin 关系**，可以定义通信信号 $x(t)$ 的**功率谱密度 (power spectral density)** 为

$$S_x(f) = \mathcal{F} \left\{ \overline{R_x(\tau)} \right\}$$

考虑**线性调制 (linear modulation)** 信号 $x(t) = \sum_k x_k p(t - kT)$ ，其中 x_k 平稳，

相关为 $R[n]$ 。 $x(t)$ 可视为冲激串 $x_\delta(t) = \sum_k x_k \delta(t - kT)$ 经过成形脉冲 $p(t)$ 滤波

后的结果，由**功率谱密度的性质**可知， $x(t)$ 的功率谱密度为

$$S_X(f) = S_{X_\delta}(f) |P(f)|^2。$$

对冲激串 $x_\delta(t)$ ，有

$$\begin{aligned} \tilde{R}_{x_\delta}(t + \tau, t) &= \mathbb{E} \left[\sum_k x_k \delta(t + \tau - kT) \cdot \sum_m x_m \delta(t - mT) \right] \\ &= \sum_k \sum_m \mathbb{E}[x_k x_m] \delta(t + \tau - kT) \delta(t - mT) \\ &= \sum_n \sum_m R[n] \delta(t + \tau - (m + n)T) \delta(t - mT) \\ \overline{R_{x_\delta}(\tau)} &= \frac{1}{T} \int_0^T \tilde{R}_{x_\delta}(t + \tau, t) dt \\ &= \frac{1}{T} \sum_n R[n] \int_0^T \sum_m \delta(t + \tau - (m + n)T) \delta(t - mT) dt \\ &= \frac{1}{T} \sum_n R[n] \delta(\tau - nT) \\ S_{X_\delta}(f) &= \mathcal{F} \left\{ \overline{R_{x_\delta}(\tau)} \right\} = \frac{1}{T} \sum_n R[n] e^{-j2\pi f n T} \end{aligned}$$

从随机过程的角度，到此就可以了。但从通信角度，希望进一步细化 $R[n]$ 的形式。考虑符号序列 $\{x_k\}$ **独立同分布 (i.i.d.)**，则

$$R[n] = \begin{cases} \mathbb{E}[x_k^2] = \text{Var}(X) + m_X^2, & n = 0, \\ \mathbb{E}[x_k] \mathbb{E}[x_{k+n}] = m_X^2, & n \neq 0 \end{cases}$$

于是

$$\begin{aligned} S_{X_\delta}(f) &= \frac{1}{T} \left(\text{Var}(X) + m_X^2 + m_X^2 \sum_{n \neq 0} e^{-j2\pi f n T} \right) \\ &= \frac{\text{Var}(X)}{T} + \frac{m_X^2}{T} \sum_{n=-\infty}^{\infty} e^{-j2\pi f n T} \\ &= \frac{\text{Var}(X)}{T} + \frac{m_X^2}{T^2} \sum_{n=-\infty}^{\infty} \delta\left(f - \frac{n}{T}\right) \end{aligned}$$

■ 线性调制信号的功率谱

对线性调制信号 $x(t) = \sum_k x_k p(t - kT)$ ，其功率谱为

$$S_X(f) = \underbrace{\frac{\text{Var}(X)}{T} |P(f)|^2}_{\text{连续谱}} + \underbrace{\frac{m_X^2}{T^2} \sum_{n=-\infty}^{\infty} \left| P\left(\frac{n}{T}\right) \right|^2 \delta\left(f - \frac{n}{T}\right)}_{\text{线谱}}$$

其中， $m_X = \mathbb{E}[x_k]$ ， $\text{Var}(X) = \mathbb{E}[(x_k - m_X)^2]$ 分别为符号的均值和方差， $P(f) = \mathcal{F}\{p(t)\}$ 为成形脉冲的频谱。

当符号电平 x_k **均值为 0** 时，功率谱仅包含连续谱部分。

■ 矩形包络载波传输

在实际通信系统中，数字信号往往需要通过**载波传输 (carrier transmission)** 的方式，将数字信号**调制到一个高频载波信号上**进行传输，以适应无线信道的频率特性和传输需求。

■ 矩形包络二进制载波传输

考虑使用**矩形包络载波信号**作为传输波形，即用 $\pm V \cos(2\pi f_c t)$ 作为成形脉冲表示电平 $\pm A$ ，其中 f_c 为载波频率。考虑到每个符号占用时间 T 、能量 E_s ，则有

$$x(t) = \begin{cases} \sqrt{\frac{2E_s}{T}} \cos(2\pi f_c t) \times \mathbb{1}_{\{kT \leq t < (k+1)T\}}, & x = \sqrt{E_s}, \\ -\sqrt{\frac{2E_s}{T}} \cos(2\pi f_c t) \times \mathbb{1}_{\{kT \leq t < (k+1)T\}}, & x = -\sqrt{E_s} \end{cases}$$

其中，要求 $f_c T$ 为**正整数**，以保证不同符号间正交。

这种包络形式下，BPSK 的各符号可以切开独立处理，只需专注讨论一个符号的波形

$$x(t) = x \cdot p(t) \sqrt{2} \cos(2\pi f_c t)$$

其中， x 等概率分布于 $\pm \sqrt{E_s}$ ， $p(t) = \sqrt{\frac{1}{T}} \times \mathbb{1}_{\{0 \leq t < T\}}$ 是**能量归一化**的矩形脉冲。可记**频移后的成形脉冲**为 $p_1(t) = p(t) \sqrt{2} \cos(2\pi f_c t)$ ，则

$$\int_{-\infty}^{+\infty} p_1^2(t) dt = \int_0^T \frac{2}{T} \cos^2(2\pi f_c t) dt = \frac{2}{T} \int_0^T \frac{1 + \cos(4\pi f_c t)}{2} dt = 1$$

即 $p_1(t)$ 仍是能量归一化的，每个符号所使用的波形能量仍为 $E_s = \sigma_x^2$ ，比特能量 $E_b = E_s$ ，**信号功率**为 $P = \frac{E_s}{T} = E_s R_s = E_b R_b$ 。

由**二进制波形的最佳接收**可知，矩形包络载波传输的**最佳接收**方式为

$$y = \int_0^T y(t) p_1(t) dt = x + \underbrace{\int_0^T z(t) \cdot \sqrt{\frac{2}{T}} \cos(2\pi f_c t) dt}_{z \sim \mathcal{N}\left(0, \frac{n_0}{2}\right)}$$

其中 x 仍为原始等概率分布的符号电平。这一接收的**误符号率**由**二进制等概电平信道的误符号率**给出，为

$$P_s = Q\left(\sqrt{\text{SNR}}\right) = Q\left(\sqrt{\frac{E_s}{n_0/2}}\right)$$

■ 矩形包络多进制单路载波传输

增大每个符号所承载的信息量，可以使用 M 进制符号电平 $x \in \mathcal{A}$ 进行传输。此时，传输波形为

$$x(t) = x \cdot p(t) \sqrt{2} \cos(2\pi f_c t), \quad x \in \mathcal{A}$$

一般取

$$\mathcal{A} = \{\pm A, \pm 3A, \pm 5A, \dots, \pm (M-1)A\}, \quad A = \sqrt{\frac{3E_s}{M^2 - 1}}$$

同理，最佳接收方式为

$$y = \int_0^T y(t) p_1(t) dt = x + \underbrace{\int_0^T z(t) \cdot \sqrt{\frac{2}{T}} \cos(2\pi f_c t) dt}_{z \sim \mathcal{N}\left(0, \frac{n_0}{2}\right)}$$

其中 x 服从均匀分布于 \mathcal{A} 。此时，**误符号率**由**多进制等概电平信道的误符号率**给出，为

$$P_s = \frac{2M-2}{M} Q \left(\sqrt{\frac{3}{M^2-1}} \frac{E_s}{n_0/2} \right)$$

误 bit 率由多进制等概电平信道的误 bit 率给出，为

$$P_b = \frac{2M-2}{M \log_2 M} Q \left(\sqrt{\frac{3 \log_2 M}{M^2-1}} \frac{E_b}{n_0/2} \right)$$

■ 矩形包络 I、Q 路载波传输

前面均只使用余弦载波移动信号频谱到高频段，称为**单路载波传输 (single-carrier transmission)**。为了更高效地利用带宽，可以使用**正交载波 (orthogonal carriers)**同时传输两个独立的信号，称为**I、Q 路载波传输 (I/Q carrier transmission)**。

利用余弦和正弦载波的正交性

$$\begin{aligned} \int_0^T \cos(2\pi f_c t) \cdot \sin(2\pi f_c t) dt &= 0, \\ \int_0^T \cos^2(2\pi f_c t) dt &= \int_0^T \sin^2(2\pi f_c t) dt = \frac{T}{2} \end{aligned}$$

可以同时传输两个独立的符号 x_I 和 x_Q ，分别调制在余弦和正弦载波上，得到

$$x(t) = x_I \cdot \underbrace{p(t)\sqrt{2} \cos(2\pi f_c t)}_{p_I(t)} + x_Q \cdot \underbrace{p(t)\sqrt{2} \sin(2\pi f_c t)}_{p_Q(t)} = x_I(t) + x_Q(t)$$

对 $x_I(t)$ 的最佳接收为

$$\begin{aligned} y_I &= \int_0^T y(t) p_I(t) dt = \int_0^T (x_I(t) + x_Q(t) + z(t)) \cdot \sqrt{\frac{2}{T}} \cos(2\pi f_c t) dt \\ &= x_I \cdot \underbrace{\int_0^T \sqrt{\frac{2}{T}} \cos^2(2\pi f_c t) dt}_1 + x_Q \cdot \underbrace{\int_0^T \sqrt{\frac{2}{T}} \sin(2\pi f_c t) \cdot \cos(2\pi f_c t) dt}_0 \\ &\quad + \int_0^T z(t) \cdot \sqrt{\frac{2}{T}} \cos(2\pi f_c t) dt \\ &= x_I + \underbrace{\int_0^T z(t) \cdot \sqrt{\frac{2}{T}} \cos(2\pi f_c t) dt}_{z_I \sim \mathcal{N}(0, \frac{n_0}{2})} \end{aligned}$$

同理，对 $x_Q(t)$ 的最佳接收为

$$y_Q = \int_0^T y(t) p_Q(t) dt = x_Q + \underbrace{\int_0^T z(t) \cdot \sqrt{\frac{2}{T}} \sin(2\pi f_c t) dt}_{z_Q \sim \mathcal{N}(0, \frac{n_0}{2})}$$

对一路的最佳接收自然地消除了另一路的干扰，I、Q 两路信号可以独立解调。

不妨设复电平 $x = x_I + jx_Q$ ，复噪声 $z = z_I + jz_Q \sim C\mathcal{N}(0, n_0)$ ，则最佳接收可写为标准电平信道 $y = x + z$ 的形式。

► 矩形包络 M-QAM 载波传输

取 M-QAM 星座图的电平集合

$$\begin{aligned} \mathcal{A} &= \{x_I + jx_Q \mid x_I, x_Q \in \{\pm A, \dots, \pm(L-3)A, \pm(L-1)A\}\}, \\ A &= \sqrt{\frac{3E_s}{2(L^2-1)}} \end{aligned}$$

实轴、虚轴均取 L 个等间距电平，共有 $M = L^2$ 个复电平。

由上述分析，矩形包络 M-QAM 载波传输等效到 [QAM 输入的复电平信道](#)，其**误符号率**为

$$P_s = \frac{4(\sqrt{M}-1)}{\sqrt{M}} Q \left(\sqrt{\frac{3}{M-1}} \frac{E_s}{n_0} \right)$$

误 bit 率为

$$P_b = \frac{4(\sqrt{M}-1)}{\sqrt{M} \log_2 M} Q \left(\sqrt{\frac{3 \log_2 M}{M-1}} \frac{E_b}{n_0} \right) \stackrel{M \text{ 较大}}{\approx} \frac{4}{\log_2 M} Q \left(\sqrt{\frac{3 \log_2 M}{M}} \frac{E_b}{n_0} \right)$$

► 矩形包络 M-PSK 载波传输

取 M-PSK 星座图的电平集合

$$\mathcal{A} = \left\{ \sqrt{E_s} \cdot e^{j2\pi m/M} \mid m = 0, 1, \dots, M-1 \right\}$$

由上述分析，矩形包络 M-PSK 载波传输等效到 [PSK 输入的复电平信道](#)，其**误符号率**为

$$P_s = 2Q \left(\sqrt{\frac{2E_s}{n_0}} \sin \frac{\pi}{M} \right)$$

误 bit 率为

$$P_b = \frac{2}{\log_2 M} Q \left(\sqrt{\frac{2E_b \log_2 M}{n_0}} \sin \frac{\pi}{M} \right) \stackrel{M \text{ 较大}}{\approx} \frac{2}{\log_2 M} Q \left(\frac{\pi}{M} \sqrt{\frac{2E_b \log_2 M}{n_0}} \right)$$

► 带限载波传输

[矩形包络载波传输](#)实现简单，但发送的信号**不带限**，不具备实际传输条件。为了适应带限信道的传输需求，需要对载波传输系统进行带限设计，只使用 $|f - f_c| \leq W$ 的频率范围进行传输。

■ 单路带限载波传输

► 系统模型

取定带限的成形脉冲 $p(t)$ ，其频谱 $P(f)$ 带限于 $|f| \leq W$ ，则单路带限载波传输的发送信号为 $x_{BB}(t) = \sum_k x_k p(t - kT)$ ，调制到载波上得到

$$x(t) = x_{BB}(t) \cdot \sqrt{2} \cos(2\pi f_c t)$$

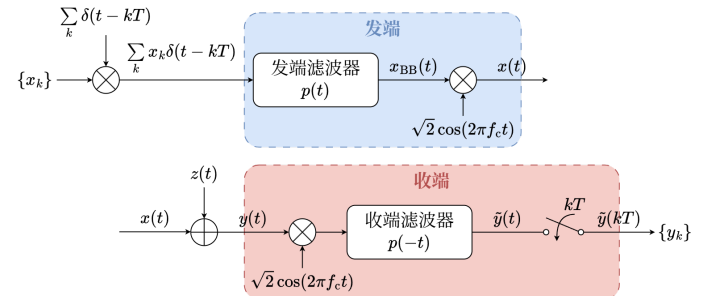
接收端**使用载波解调**，经**低通滤波**后得到基带信号

$$y_{BB}(t) = (y(t) \cdot \sqrt{2} \cos(2\pi f_c t)) * b_{LPF_W}(t)$$

再对 $y_{BB}(t)$ 进行**最佳接收**，得到

$$\begin{aligned} y(kT) &= \int_{-\infty}^{+\infty} y_{BB}(\tau) p(\tau - kT) d\tau \\ &= (y(t) \cdot \sqrt{2} \cos(2\pi f_c t)) * b_{LPF_W}(t) * p(-t) \Big|_{t=kT} \\ &= (y(t) \cdot \sqrt{2} \cos(2\pi f_c t)) * (b_{LPF_W}(t) * p(-t)) \Big|_{t=kT} \\ &= (y(t) \cdot \sqrt{2} \cos(2\pi f_c t)) * p(-t) \Big|_{t=kT} \end{aligned}$$

其中最后一步是因为 $\mathcal{F}\{p(-t)\} = P^*(f)$ 带限于 $|f| \leq W$ ，经过带宽为 W 的低通滤波器后不变。这样，当**载频 $f_c > W$** 时，可以含噪声地恢复出基带信号 $x_{BB}(t)$ ，从而单路带限载波传输等效到**基带传输**，其成形脉冲为 $p(t)$ ，符号电平为 x_k 。



► 传输性能

设 $p(t)$ 为**能量归一化**且**带限于 $|f| \leq W$** 的成形脉冲，载频 $f_c > W$ ，符号电平 x_k 独立同分布，均值为 0，方差为 σ_x^2 ，则单路带限载波传输等效到基带传输系统，**信号功率**为

$$P = E_s R_s = \frac{\sigma_x^2}{T} = E_b R_b$$

最佳接收过程展开为

$$\begin{aligned} y_k &= \tilde{y}(kT) = \int_{-\infty}^{+\infty} y(t) \cdot \sqrt{2} \cos(2\pi f_c t) \cdot p(t - kT) dt \\ &= \int_{-\infty}^{+\infty} (x(t) + z(t)) p(t - kT) \cdot \sqrt{2} \cos(2\pi f_c t) dt \\ &= \int_{-\infty}^{+\infty} \left(\sum_i x_i p(t - iT) \right) p(t - kT) \cdot 2 \cos^2(2\pi f_c t) dt \quad (\text{信号}) \\ &\quad + \int_{-\infty}^{+\infty} z(t) p(t - kT) \cdot \sqrt{2} \cos(2\pi f_c t) dt \quad (\text{噪声}) \end{aligned}$$

- 噪声项中， $p^2(t - kT)$ 与 $\cos(4\pi f_c t)$ 频域无重叠，故**时域亦正交**，因此噪声项服从分布 $z_k \sim \mathcal{N}(0, \frac{n_0}{2})$ 。
- 信号项中， $p(t - ik) p(t - kT)$ 与 $\cos(4\pi f_c t)$ 亦正交，因此

$$\begin{aligned}
(\text{信号}) &= \sum_i x_i \int_{-\infty}^{+\infty} p(t-iT) p(t-kT) dt \\
&= \sum_i x_i \int_{-\infty}^{+\infty} p(t) p(t-(k-i)T) dt \\
&= \sum_i x_i b((k-i)T) = \sum_i x_i \delta[k-i] = x_k
\end{aligned}$$

其中 $b(t) = p(t) * p(-t)$ 应满足采样点无失真准则，以保证无码间干扰。

这样，单路带限载波传输的最佳接收可写为

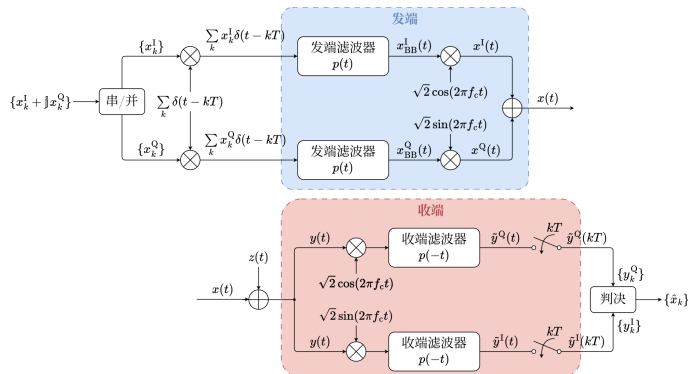
$$y_k = x_k + z_k, \quad z_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{n_0}{2}\right)$$

其中 x_k 仍为原始符号电平。

二进制和多进制单路带限载波传输的误符号率和误 bit 率，与对应的基带传输系统相同，分别由二进制和多进制等电平信道的分析给出。

■ I、Q 路带限载波传输

直接类比矩形包络 I、Q 路载波传输，一般的 I、Q 路带限载波传输系统模型如下：



同样有功率为

$$P = E_s R_s = \frac{\sigma_x^2}{T} = E_b R_b$$

展开其最佳接收过程，对 Q 路有

$$\begin{aligned}
y_k^Q &= \int_{-\infty}^{+\infty} y(t) \cdot \sqrt{2} \sin(2\pi f_c t) \cdot p(t-kT) dt \\
&= \int_{-\infty}^{+\infty} (x^I(t) + x^Q(t) + z(t)) p(t-kT) \cdot \sqrt{2} \sin(2\pi f_c t) dt \\
&= \int_{-\infty}^{+\infty} \left(\sum_i x_i^I p(t-iT) \right) p(t-kT) \cdot 2 \sin(2\pi f_c t) \cos(2\pi f_c t) dt \quad (\text{I 路干扰}) \\
&\quad + \int_{-\infty}^{+\infty} \left(\sum_i x_i^Q p(t-iT) \right) p(t-kT) \cdot 2 \sin^2(2\pi f_c t) dt \quad (\text{Q 路信号}) \\
&\quad + \int_{-\infty}^{+\infty} z(t) p(t-kT) \cdot \sqrt{2} \sin(2\pi f_c t) dt \quad (\text{噪声})
\end{aligned}$$

- 噪声项服从分布 $z_k^Q \sim \mathcal{N}\left(0, \frac{n_0}{2}\right)$ 。
- 干扰项中， $p(t-ikT)p(t-kT)$ 与 $\sin(4\pi f_c t)$ 频域无重叠，故时域亦正交，因此干扰项为 0。
- 信号项中，与单路带限载波传输类似，在 $b(t) = p(t) * p(-t)$ 满足采样点无失真准则的前提下，得其值为原始符号电平 x_k^Q 。

即其最佳接收过程可等效为复电平信道，QAM、PSK 等调制方式下的误符号率和误 bit 率分别由QAM 输入和 PSK 输入的复电平信道的分析给出。

Note I、Q 路带限载波传输的噪声

在 I、Q 路带限载波传输系统中，信道中的 AWGN $z(t)$ 仍是 $S_z(f) \equiv \frac{n_0}{2}$ 的实 Gauss 过程，但经过最佳接收得到的噪声电平 $z_k = z_k^I + jz_k^Q$ 是复 Gauss 随机变量，且 $z_k^I, z_k^Q \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \frac{n_0}{2}\right)$ ，通常记 $z_k \sim \mathcal{CN}(0, n_0)$ 。

若载波下变频过程中使用的解调波形不能量归一化，则噪声方差会相应地放大或缩小。例如，若解调波形为 $\beta \cos(2\pi f_c t)$ 、 $\beta \sin(2\pi f_c t)$ ，则各路噪声将放大为 $\frac{\beta^2}{2}$ 倍，即 $z_k \sim \mathcal{CN}\left(0, \frac{\beta^2}{2} n_0\right)$ 。此变化不影响信噪比、误符号率等性能指标的计算。

■ 带限载波传输的频谱效率

在带限于 $||f| - f_c| \leq W$ 的频率范围内传输信号时，占用带宽为 $B = 2W$ 。定义频谱效率 (spectral efficiency) 为每单位带宽传输的比特率，即

$$\eta = \frac{R_b}{B} = \frac{R_b}{2W} = \frac{R_s}{2W} \cdot \log_2 M$$

若用升余弦滤波系统作为成形脉冲，则有 $W = \frac{1+\alpha}{2T}$ ，其中 α 为滚降系数 (roll-off factor)，则频谱效率可写为

$$\eta = \frac{\log_2 M}{1+\alpha}$$

Remark 题型：升余弦带通信道

典型题型 类似于基带的升余弦滤波系统，给定 R_b 和带通范围 $[f_{\min}, f_{\max}]$ ，求满足无失真传输的 M 和对应的 α 。

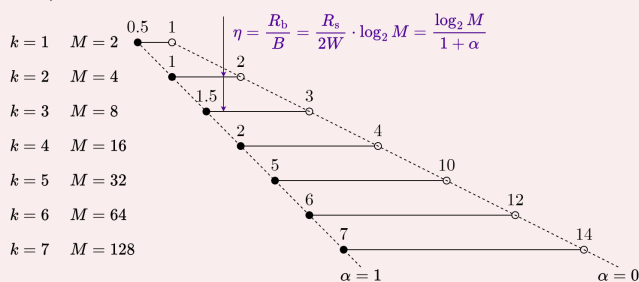
一般地，只在正频轴上讨论，居中取载频 $f_c = \frac{f_{\min} + f_{\max}}{2}$ ，两侧分别有基带带宽 $W = \frac{f_{\max} - f_{\min}}{2}$ ，占用带宽 $B = 2W$ 。此时，数字载波传输系统的频谱效率为

$$\eta = \frac{R_b}{B} = \frac{R_b}{f_{\max} - f_{\min}} = \frac{R_s}{2W} \cdot \log_2 M = \frac{\log_2 M}{1+\alpha}$$

同样，对此欠定问题，一般补充两个条件：

- M 取 2 的整数次幂，即 $M = 2^k, k \in \mathbb{N}^*$ ；
- α 不取退化情况，即 $0 < \alpha \leq 1$ 。

此时，可通过枚举 k 的方式求解该问题。



■ 带限载波传输信号的功率谱

对信号做频谱搬移时，功率谱也会相应地搬移。考虑 $x_{\text{BB}}(t) = \sum_k x_k p(t-kT)$ 频谱搬移到载波频率 f_c ，即传输信号为

$$x(t) = x_{\text{BB}}(t) \cdot \sqrt{2} \cos(2\pi f_c t)$$

则基带信号的功率谱也会相应地搬移，得到带限载波传输信号的功率谱

$$\begin{aligned}
S_X(f) &= \frac{1}{2} (S_{X_{\text{BB}}}(f-f_c) + S_{X_{\text{BB}}}(f+f_c)) \\
&= \frac{\text{Var}(X)}{2T} (|P(f-f_c)|^2 + |P(f+f_c)|^2) \\
&\quad + \frac{m_X^2}{2T^2} \sum_{n=-\infty}^{\infty} \left| P\left(\frac{n}{T}\right) \right|^2 \left(\delta\left(f-f_c-\frac{n}{T}\right) + \delta\left(f+f_c-\frac{n}{T}\right) \right)
\end{aligned}$$

由于 $\sqrt{2} \cos(2\pi f_c t)$ 能量归一化，因此频谱搬移能量守恒。

§4 差错控制

■ 差错控制概述

■ 差错控制的位置

差错控制的位置可以分为两类：

- 在物理层、链路层实现的差错控制，主要用于检测和纠正点到点传输过程中产生的随机差错。
- 在传输层实现的差错控制，主要用于检测和纠正端到端传输过程中产生的差错。

下面主要讨论传输层的差错控制，即默认中间一系列节点不可靠，在发端和收端之间进行差错控制。

■ 典型的差错控制方案

► 反馈确认

在收端 D 收到数据后，将收到的 bit 原样返回给发端 S ，发端 S 将收到的 bit 与发送的 bit 进行比较，如果一致则认为数据正确接收，发送下一个 bit 组；否则认为数据出错，重新发送该 bit 组。

这种方案的问题在于：

1. 反馈确认需要额外的带宽资源；
2. 反馈确认增加了传输延迟，时间效率低；
3. 反馈信道本身也可能出错，导致误判。

► 检错重发 (Automatic Repeat reQuest, ARQ)

在发端 S 发送数据时，附加一些冗余信息（如**校验和**、**CRC 校验码**等），收端 D 在收到数据后进行检错，如果发现数据有误，则请求发端重发该数据。典型的 ARQ 协议有**停一等 ARQ**、**连续 ARQ**等。

► 前向纠错 (Forward Error Correction, FEC)

在发端 S 发送数据时，附加足够的冗余信息，使得收端 D 即使在部分数据出错的情况下，也能通过冗余信息恢复出正确的数据，而无需请求重发。

■ 码本的检错、纠错能力

信息 bit 串经过编码映射为传输**码字 (codeword)**，码字的集合连同其编码的映射关系构成**码本 (codebook)**。本课程中，将信息 bit 串和码字均视为 GF_2 上的**行向量**，以 \vec{m} 、 \vec{c} 标记。此后的向量记号亦均指行向量。

例 \mathcal{E}_3 (1, 3) 冗余码本的检错、纠错能力

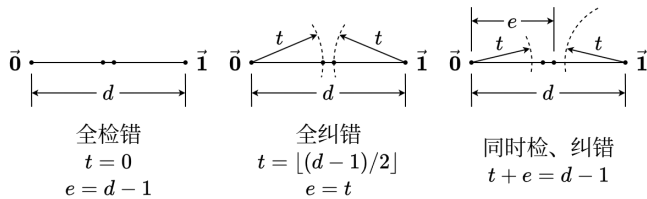
考虑如下码本：

$$1 \rightarrow 111 = \vec{1}_3, \quad 0 \rightarrow 000 = \vec{0}_3$$

- 以 ARQ 方式处理，110、101、011、100、001、010 均可被检测为出错，但无法纠正，即检 2 位错而不纠错；
- 以 FEC 方式处理，110、101、011 认为是 1，100、001、010 认为是 0，即纠 1 位错。

对同一个串的处理不同，其检错、纠错能力也不同，检错与纠错不完全兼容。

两个 bit 串之间对应位相异 bit 的个数体现了其「相似」的程度，称为**Hamming 距离 (Hamming distance)**，记为 d 。上述检错、纠错的位数，实际也是指 Hamming 距离。



在一般码本中，考虑任意两个码字之间 Hamming 距离的**最小值** $d_{\min} = \min_{\vec{c}_1 \neq \vec{c}_2} d(\vec{c}_1, \vec{c}_2)$ ，即**最小 Hamming 距离**或**码距 (code distance)**。这个距离上，码本的检错、纠错能力与上图类似，即有

Thm. 码本的纠错位数、检错位数与码距的关系

给定码距 d_{\min} 的码本，其纠错位数 t 和检错位数 e 满足

$$t + e \leq d_{\min} - 1$$

进而，

- 纠错位数 t 满足 $2t + 1 \leq d_{\min}$ ，
- 检错位数 e 满足 $e + 1 \leq d_{\min}$ 。

► 检错码的构造与应用

■ 奇偶校验

► 简单奇偶校验

最简单的**检错**方式是**奇偶校验 (parity check)**，即在信息 bit 串后附加一个**校验位 (parity bit)**，使得整个码字中 1 的个数为偶数。

设给定 k bits 长信息 $\vec{m} = (m_1 \ m_2 \ \cdots \ m_k)$ ，则编码后的码字为

$$c_i = \begin{cases} m_i, & i = 1, 2, \dots, k, \\ \sum_{j=1}^k m_j, & i = k+1 \end{cases} \quad \text{i.e.} \quad \vec{c} = \vec{m} \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 \\ 0 & 1 & \cdots & 0 & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 1 \end{pmatrix}$$

即 $\vec{c} = \vec{m} \mathbf{G}$ ，**生成矩阵 (generator matrix)** $\mathbf{G} = (\mathbf{I}_k \ \vec{1}_k)$ 。解码时，应有

$$\sum_{i=1}^{k+1} c_i = \vec{c} \cdot \vec{1}_{k+1} = \vec{c} \mathbf{1}_{k+1}^T = 0$$

即**校验矩阵 (parity-check matrix)** $\mathbf{H} = \vec{1}_{k+1}$ ，当接收码字 \vec{y} 满足 $\vec{y} \mathbf{H}^T \neq 0$ 时，认为码字出错。

该码字的**码距** $d_{\min} = 2$ ，因此必有 $t = 0$ ，只能**检 1 位错**，**不能纠错**。同时，还有发生差错但无法检出的情况，**漏检概率**为

$$p_m = \sum_{i=1}^{\lfloor (k+1)/2 \rfloor} \binom{k+1}{2i} P_b^{2i} (1 - P_b)^{k+1-2i}$$

► 循环冗余校验

循环冗余校验 (cyclic redundancy check, CRC) 是奇偶校验的推广，通过**多项式除法**实现**检错**。

简单奇偶校验因其只有 1 位校验位，检错能力有限。我们希望通过增加更多的校验位，提高检错能力。设信息 bit 串 \vec{m} 可表示为**信息多项式 (message polynomial)**

$$m(x) = m_0 + m_1x + m_2x^2 + \cdots + m_{k-1}x^{k-1}$$

我们选择一个**生成多项式 (generator polynomial)** $g(x)$ ，其次数为**校验位长度** $n - k$ ，即

$$g(x) = g_0 + g_1x + g_2x^2 + \cdots + x^{n-k}$$

编码后的码字 \vec{c} 对应的**码字多项式 (codeword polynomial)** 设为

$$c(x) = \underbrace{m(x)x^{n-k}}_{\alpha(x)} + \underbrace{m(x)x^{n-k} \bmod g(x)}_{\gamma(x)}$$

其中 $\gamma(x)$ 为**校验多项式 (check polynomial)**，即移位后的信息多项式 $\alpha(x)$ 除以生成多项式 $g(x)$ 的**余数**。

由此，存在多项式 $\beta(x)$ 使得

$$\alpha(x) = \beta(x)g(x) + \gamma(x) \implies c(x) = \beta(x)g(x)$$

即**码字多项式 $c(x)$ 可被生成多项式 $g(x)$ 整除**。解码时，收端收到 \vec{y} 对应码字多项式 $y(x) = c(x) + e(x)$ ，只需计算 $y(x) \bmod g(x)$ ，若结果不为零，则认为码字出错。

■ ARQ 传输方案

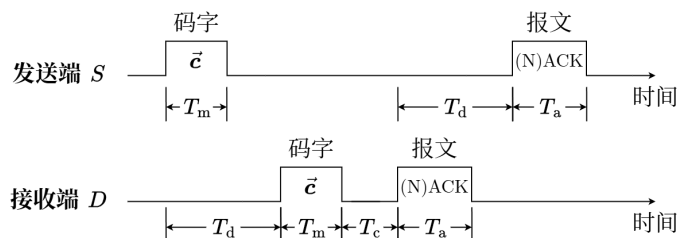
检错重发 (Automatic Repeat reQuest, ARQ) 是一大类典型差错控制方案，其通过**检错码**实现。

► 「停一等」重传机制

在 ARQ 差错控制中，常用的重传机制是**停一等 (stop-and-wait)** 机制，即每发送一个数据包后，必须等待确认收到该数据包的反馈后，才能发送下一个数据包。

不妨设码本将 k bits 信息映射为 n bits 码字，ARQ 过程为：

1. 发端 S 将信息编码为码字 \vec{c} 并发送，**发送用时** $T_m = \frac{n}{R_b}$ ；
2. 信息从发端 S 传输到收端 D ，**传输时延**为 T_d ；
3. 收端 D 经 T_c 时间检测出码字 \vec{c} 是否正确接收，若正确则发送 **ACK (acknowledgement)** 报文，否则发送 **NACK (negative acknowledgement)** 报文，**发送用时** $T_a = \frac{n_a}{R_b}$ ；
4. ACK/NACK 报文从收端 D 经同样的传输时延 T_d 传输到发端 S ，发端据此决定是否重传该码字，决定用时忽略不计。



一个 ARQ 轮次中, 除开发送码字的时间 T_m 外, 其余时间均为等待时间, 总等待时间为 $T_{\text{dca}} = 2T_d + T_c + T_a$ 。因此, 即使一轮次传输成功, 其时间效率也为

$$\xi' = \frac{T_m}{T_m + T_{\text{dca}}} = \frac{T_m}{T_m + 2T_d + T_c + T_a}$$

未检出错即为「成功」, 概率为

$$p_{\text{suc}} = (1 - P_b)^n + p_m$$

其中 p_m 为漏检概率 (missed detection probability), 即发生差错但未被检出的概率。

成功所需轮次服从几何分布 (geometric distribution), 其期望为 $\frac{1}{p_{\text{suc}}}$, 因此实际时间效率为

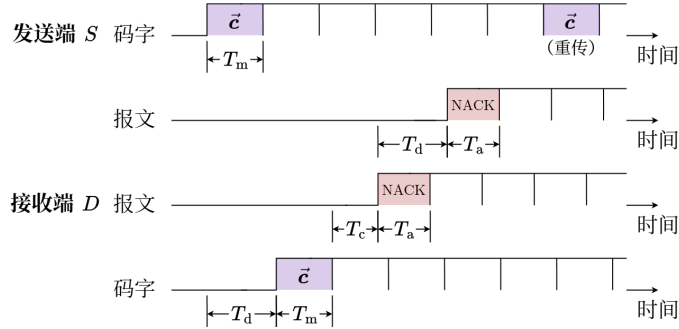
$$\xi = \frac{T_m}{(T_m + T_{\text{dca}}) \cdot \frac{1}{p_{\text{suc}}}} = \xi' p_{\text{suc}}$$

对上层信息 bit 流的整体服务速率即为

$$r = \frac{k}{\frac{n}{R_b} \cdot \frac{1}{\xi}} = \frac{k}{n} \xi R_b = \frac{k((1 - P_b)^n + p_m)}{n} \cdot \frac{R_b T_m}{T_m + 2T_d + T_c + T_a}$$

► 选择重传机制

在「停一等」重传机制下, 发端在等待 ACK/NACK 报文期间处于闲置状态, 时间效率较低。为提高时间效率, 可采用选择重传 (selective repeat) 机制, 即发端在等待 ACK/NACK 报文的同时, 继续发送后续码字。



选择重传机制将码字的传输与 ACK/NACK 报文的传输拆分为独立信道, 从而并行进行, 提高时间效率。对固定、已知的 T_d 、 T_c , 发端可由 NACK 时间反推出错的码字, 收端由自身的报文发送记录也可推知哪个码字是重传码字, 无需序号。

选择重传中每个码字传输的差错性能与「停一等」重传机制相同, 同样有

$$p_{\text{suc}} = (1 - P_b)^n + p_m$$

因为发端持续发送码字, 单位时间内成功发送的消息 bit 数为 $\frac{k}{n} R_b p_{\text{suc}}$, 即对信息 bit 流的整体服务速率为

$$r = \frac{k}{n} ((1 - P_b)^n + p_m) R_b$$

若不忽略 ACK/NACK 报文的传输开销, 则服务速率为

$$r = \frac{T_m}{T_m + T_a} \cdot \frac{k}{n} ((1 - P_b)^n + p_m) R_b$$

► 回溯N重传机制

选择重传需要差错分组的序号判断和乱序重排, 不仅复杂, 而且在 T_d 未知或时变时无法正确工作。为此, 可采用回溯 N 重传 (go-back-N ARQ) 机制, 即发端持续发送码字, 但一旦收到 NACK 报文, 则回溯重传从该码字开始的后续 N 个码字。

Note 重传机制在链路层、传输层的使用

重传机制在链路层 (单跳) 和传输层 (多跳) 都有使用, 但从发出一个分组到接收其 ACK/NACK 的往返时间 (RTT) 的分布特性不同。

- 链路层通常采用停一等 ARQ, 因为链路层传输距离短, 时延小, 且链路层设备通常较简单, 适合使用简单的停一等 ARQ;
- 传输层通常采用选择重传 ARQ 或混合 ARQ 机制, 因为传输层传输距离长, 时延大, 且传输层设备通常较复杂, 适合使用复杂的重传策略。

纠错码的构造与应用

■ FEC 传输方案

除 ARQ 外, 前向纠错 (Forward Error Correction, FEC) 是另一大类型差错控制方案, 其通过纠错码 (又称信道编码) 实现。

依据对信息流的处理方式的不同, FEC 所构造的纠错码可分为两类:

- 分组码 (block code), 即将信息 bit 流划分为若干个固定长度的信息 bit 组, 分别进行编码, 形成固定长度的码字进行传输, 各个码字之间编、译码相互独立。 (k, n) 分组码本质上是从 GF_2^k 到 GF_2^n 的映射, 将信息分散到更大的离散空间中, 从而提高抗差错能力。

分组码又分为:

- 线性分组码 (linear block code), 即编码映射是线性映射。线性分组码存在生成矩阵 (generator matrix) G 使得码字 $\vec{c} = \vec{m} G$ 。

线性分组码的一个重要性质是其码字集合构成 GF_2^n 的子空间 (subspace), 从而可通过校验矩阵 (parity-check matrix) H 定义合法码字的集合, 即 $\vec{c} H^T = \vec{0}$ 。

线性分组码又可进一步分出:

- 线性系统码 (linear systematic code), 其码字 \vec{c} 固定后 $n - k$ 位为校验位, 前 k 位与信息 \vec{m} 相同, 即生成矩阵

$$G_{k \times n} = (I_k \quad P_{k \times (n-k)}), \quad \text{校验矩阵 } H_{(n-k) \times n} = \begin{pmatrix} P_{(n-k) \times k}^T & I_{n-k} \end{pmatrix}$$

。

- 线性非系统码, 如 $G = \begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{pmatrix}$ 。

- 非线性码, 即编码映射不是线性映射, 如「以 1 的计数为校验位」的码字映射。

- 非分组码, 即对信息 bit 流进行连续编码, 形成连续的码字流进行传输, 码字之间编、译码相互关联。

非分组码的例子有:

- 卷积码 (convolutional code), 即通过有限状态机对信息 bit 流进行编码, 将整个信息流与某 bit 串进行卷积运算;
- Turbo 码 (或译为涡轮码), 即通过两个或多个卷积码并行编码, 并通过交织 (interleaver) 实现码字之间的关联, 提高纠错能力。

Remark 题型: 线性系统码

典型题型 给出 (k, n) 线性系统码的若干个 $\vec{m} \mapsto \vec{c}$ 映射对, 求其生成矩阵 G 和校验矩阵 H , 再由未知信息串计算码字, 或由未知码字校验是否为合法码字。

由于 (k, n) 线性系统码的生成矩阵 $G_{k \times n} = (I_k \quad P_{k \times (n-k)})$ 中只有 P 未知, 可由给出的映射对求出 P , 进而求出 G 和 H 。

■ Hamming 码

► Hamming 码的构造

我们希望构造一种简洁的代数运算方式, 从 $\vec{y} = \vec{x} + \vec{e}$ 中在某一条件下确定 \vec{e} , 从而恢复 \vec{x} 。

假定 \vec{e} 中最多只有 1 位错, 即我们讨论的目标就是纠 1 位错, 不考虑多于 1 位错的情况。由线性码性质 $G H^T = \vec{0}$, 有

$$\vec{m} G H^T = \vec{c} H^T = \vec{0}$$

于是, 因合法码字 $\vec{x} \in \{\vec{c} \mid \vec{c} \in C\}$, 有

$$\vec{y} H^T = (\vec{x} + \vec{e}) H^T = \vec{x} H^T + \vec{e} H^T = \vec{0} + \vec{e} H^T = \vec{e} H^T$$

即接收码字的综合 (syndrome) 或称校正子 (syndrome) $\vec{s} = \vec{y} H^T$ 仅与误码 \vec{e} 有关, 与原码字 \vec{x} 无关, 显然有

$$\vec{s} = \vec{e} H^T$$

我们希望通过 \vec{s} 唯一确定 \vec{e} , 从而恢复 \vec{x} 。由于我们假定 \vec{e} 中最多只有 1 位错, \vec{e} 只能是如下形式之一:

- 无错, $\vec{e} = \vec{0}$, $\vec{s} = \vec{0} H^T = \vec{0}$;
- 第 i 位错, $\vec{e} = \hat{u}_i$, $\vec{s} = \hat{u}_i H^T = H^T(i, :)$ 为矩阵 H^T 的第 i 行。

为了使 \vec{s} 能唯一对应 \vec{e} , 矩阵 H^T 的各行必须互不相同, 且不能为零行。注意到, H 是一个 $(n - k) \times n$ 矩阵, 因此 H^T 有 n 行, 每行有 $n - k$ 位, 可表示的非零行共有 $2^{n-k} - 1$ 种, 因此必须满足

$$2^{n-k} - 1 = n \implies n + 1 = 2^{n-k}$$

即 n 和 k 必须满足上述关系，才能构造出能够纠 1 位错的线性码。这类码称为 **Hamming 码**，其校验位个数为 $m = n - k$ 。

$m = n - k$	(n, k) 组合
2	(3, 1)
3	(7, 4)
4	(15, 11)
5	(31, 26)

► Hamming 码的码距

我们讨论 Hamming 码的码距 d_H^{\min} 。任何满足 $\vec{c}H^T = \vec{0}$ 都是**合法码字**，据此知：

- 显然 $\vec{0}$ 是合法码字；
- 不存在**只有 1 位为 1**的合法码字，因为 H^T 的各行均不为零；
- 不存在**只有 2 位为 1**的合法码字，因为任意两行之和均不为零。

另一方面， H^T 的行遍历了 GF_2^{n-k} 中的非零点，任意两行之和也在其中，因此 $\forall i, j, \exists l$ 满足 $H^T(l, :) = H^T(i, :) + H^T(j, :)$ ，即

$$(\hat{u}_i + \hat{u}_j + \hat{u}_l)H^T = H^T(i, :) + H^T(j, :) + H^T(l, :) = \vec{0}$$

故存在 $\hat{u}_i + \hat{u}_j + \hat{u}_l$ 是一个**有 3 位为 1**的合法码字。

因此，任何非零码字与 $\vec{0}$ 的汉明距离最小为 3，又由于 $d_H(\vec{c}_1, \vec{c}_2) = d_H(\vec{0}, \vec{c}_1 + \vec{c}_2)$ ，故任意两个不同码字之间的汉明距离最小也为 3，即 Hamming 码的码距为 3。

▢ *Thm.* Hamming 码的码距

Hamming 码的码距 $d_H^{\min} = 3$ 。

🔗 *Remark* 题型：Hamming 码

典型题型 1 给定信源的消息 bit 速率 R_b ，求 (n, k) Hamming 码编码后的滚降系数 α 和许用电平数 M 。

编码后，需要的 bit 速率提升为 $R'_b = \frac{n}{k}R_b$ ，然后将 R'_b 代入**基带升余弦系统**或**带通升余弦系统**相关公式中，求出相应的滚降系数 α 和许用电平数 M 。

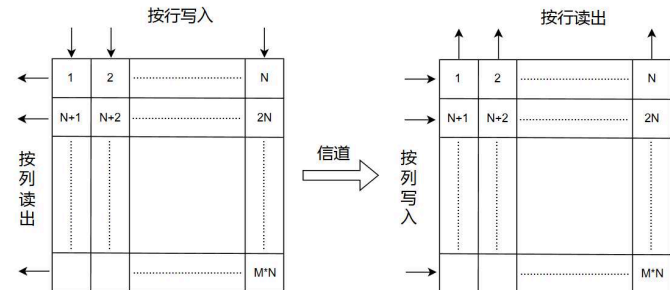
典型题型 2 给定 P_b ，求 (n, k) Hamming 码纠错失败的概率，即误块率 P_e 。

误块率 P_e 即为**发生 2 位或以上差错**的概率

$$P_e = 1 - (1 - P_b)^n - nP_b(1 - P_b)^{n-1} = \sum_{i=2}^n \binom{n}{i} P_b^i (1 - P_b)^{n-i}$$

■ 交织

当出现突发差错时，单个码字中可能出现多位差错，超出纠错码的纠错能力，导致纠错失败。为此，可通过**交织 (interleaving)** 技术将突发差错扩散到多个码字中，从而提高纠错码的有效性。



通常使用**块交织 (block interleaver)**，即将若干个码字**按行填入**一个矩阵，然后**按列读出**形成交织后的码字流。

- 交织块的宽度或列数 N 通常取**分组码的码长 n** ；
- 交织块的高度或行数 M 称为**交织深度 (interleaving depth)**，通常取决于信道的突发差错特性，决定了相邻码元交织后的间隔。

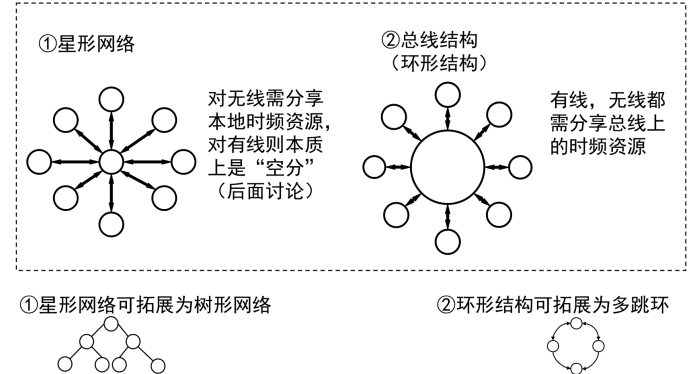
若分组码的纠错能力为 $n = N$ 位中的 t 位错，交织后纠错能力提升为 MN 位中的 Mt 位错。

§5 多用户通信

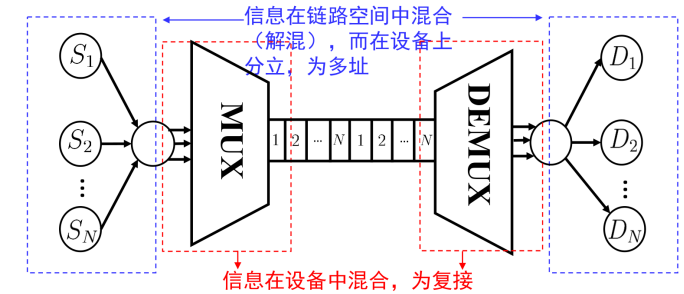
资源切分

多用户通信专注于讨论简单拓扑结构下**多个信息流 (收发对) 共享通信资源 (时, 频等)**的方法，包括多址 (局域网) 和复接 (广域网)。

多址 (multi access) 通信是局域网的核心知识，典型的局域网拓扑包括：



广域网的长距离传输，为提效、降成本，常把多个用户的信息流通过共同的光纤、铜缆、卫星、微波中继传输，称为**复接 (multiplexing)**。



无论是多址，还是复接，本质上都依赖于对通信资源的划分，又称**信道化 (channelization)**。

■ 时分与频分多址

时分多址 (time division multiple access, TDMA) 是将时间划分为多个时隙，不同用户在不同的时隙内发送数据，从而实现多个用户共享同一频带；**频分多址 (frequency division multiple access, FDMA)** 是将频带划分为多个子信道，不同用户在不同的子信道内发送数据，从而实现多个用户同时传输数据。

TD 与 FD 实现简洁，且可以混用，如第 2 代移动通信中的 D-AMPS 系统。

🔗 *Remark* 题型：TDMA、FDMA 与物理层参数

典型题型 给定多址系统的部分网络层参数（用户数 N 、用户数据速率 R_b ）和物理层参数（总带宽 B 、滚降系数 α ），求 TDMA/FDMA 方案（载频 f_c 、调制阶数 M ）。

仍然是欠定问题，不妨先分解 $N = n_T n_F$ ，每个 $\frac{B}{n_F}$ 宽的频带有 n_T 个时隙，需要承载 $n_T R_b$ 的速率。由**带通升余弦系统**的知识可知，调制阶数 M 满足

$$\frac{n_T R_b}{\log M} \leq \frac{B}{n_F} \implies \log M \geq \frac{n_T n_F R_b (1 + \alpha)}{B} = \frac{N R_b (1 + \alpha)}{B}$$

由此可确定调制阶数 M ，且其与 N 的分解方式无关。

M 确定后，对应于 n_T 可得升余弦的实际带宽 $B_{n_T} = \frac{n_T R_b}{\log M} (1 + \alpha)$ ，从而可确定各载频 f_c 使得各频带尽量分散。

反过来，如果已知对各子频带的带宽有要求，则可以据此确定所能支持的用户数 n_T ，从而确定总用户数 N 的分解方式。

■ 时分复接

时分复接 (time division multiplexing, TDM) 是将时间划分为多个时隙，不同用户在不同的时隙内发送数据，从而实现多个用户共享同一频带。TDM 与 TDMA 的区别在于，TDM 强调**多个用户的信息流分时隙接入同一个设备**，而 TDMA 强调**多个用户分时隙共享同一个信道**。

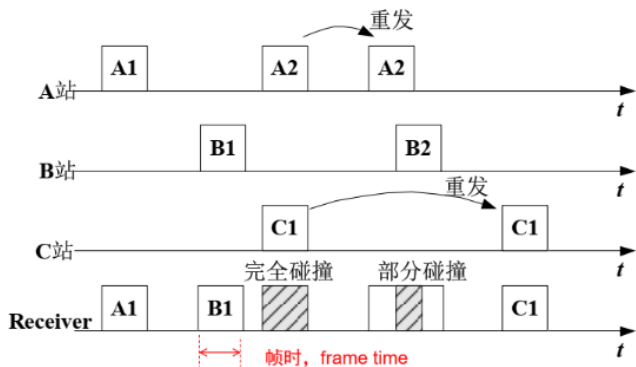
访问控制

Aloha

纯 Aloha

Aloha 是夏威夷大学开发的无线多址协议，基本思想是**随机接入 (random access)**，即用户在有数据发送时**随机选择时间**发送数据。发送后，

- 若无冲突，接收方发送确认 (ACK)，发送方收到 ACK 后结束传输；
- 若有冲突，发送方在**随机等待一段时间**后重新发送数据。



由于 Aloha 中采用**定长分组**，记帧长为 b 个 bit，传输速率为 R_b (bit/s)，则每帧传输时间为 $T_{fr} = \frac{b}{R_b}$ 。定义 Aloha 系统的**供给业务量** G 为**每帧**时间内「尝试传输」的平均次数，则单位时间内「尝试传输」的平均次数即**总帧生成率**为 $\lambda_t = \frac{G}{T_{fr}}$ ；再记**成功传输概率**（无任何碰撞的概率）为 p_{suc} ，则单位时间内成功传输的平均次数为 $\lambda_t p_{suc}$ 。

给出一个引理：大量独立同分布更新过程的叠加为一个 Poisson 过程。因此，对于海量用户，记每个用户单位时间内尝试接入的次数即**用户帧生成率**为 λ_u ，则所有用户的**总帧生成率**为 $\lambda_t = N\lambda_u$ 。换言之，

- 单个用户的尝试（帧）到达率 $\lambda_u = \frac{\lambda_t}{N} \xrightarrow{N \rightarrow +\infty} 0$ ；
- 对任意一个给定的用户来说，其看到的「背景业务」（其他用户的尝试）到达率为 $(N-1)\lambda_u \approx \frac{N-1}{N}\lambda_t \xrightarrow{N \rightarrow +\infty} \lambda_t$ 。

这样，对于任意一个给定的用户来说，其看到的**背景业务到达过程**为**参数为 λ_t 的 Poisson 过程**。本用户在时间 t 发送一帧数据，若在**脆弱时间 (vulnerable time) $[t - T_{fr}, t + T_{fr}]$** 内有其他用户发送数据，则会发生碰撞，导致发送失败。由于背景业务到达过程为参数为 λ_t 的 Poisson 过程，因此在该时间段内没有其他用户发送数据的概率为

$$p_{suc} = \Pr\{N(2T_{fr}) = 0\} = e^{-2\lambda_t T_{fr}} = e^{-2G}$$

Thm. Aloha 系统的吞吐量

海量用户 Aloha 系统在单位时间内成功传输的帧数即**帧吞吐量**为

$$\lambda = \lambda_t e^{-2\lambda_t T_{fr}} = \frac{G}{T_{fr}} \cdot e^{-2G}$$

比特吞吐量为

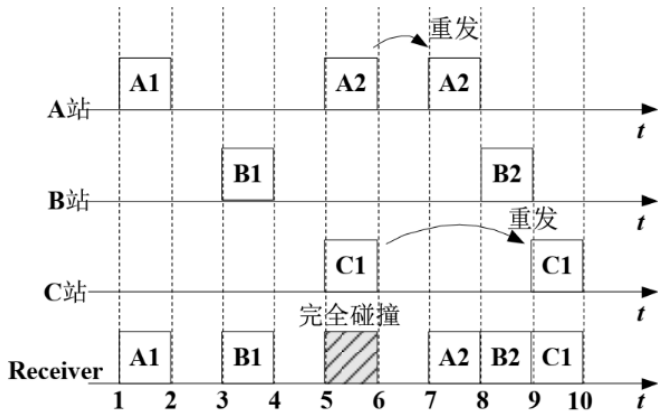
$$\lambda_b = \lambda b = R_b G e^{-2G}$$

其传输速率 R_b 中被有效利用的比例即**归一化吞吐量**为

$$\rho = \frac{\lambda_b}{R_b} = G e^{-2G}$$

时隙 Aloha

时隙 Aloha (slotted Aloha) 是对 Aloha 的改进，假设所有用户**时间同步**，将时间划分为多个时隙，用户只能在**时隙开始**时发送数据。



这样，脆弱时间缩短为 T_{fr} ，成功传输概率变为

$$p_{suc} = \Pr\{N(T_{fr}) = 0\} = e^{-\lambda_t T_{fr}} = e^{-G}$$

或者，每个用户在一个时隙内的发送概率为 $p = \lambda_u T_{fr} = \frac{\lambda_t T_{fr}}{N} = \frac{G}{N}$ ，则在该时隙内有且仅有一个用户发送数据的概率为

$$N p_{u,suc} = N \cdot p(1-p)^{N-1} = G \left(1 - \frac{G}{N}\right)^{N-1} \xrightarrow{N \rightarrow +\infty} G e^{-G}$$

Thm. 时隙 Aloha 系统的吞吐量

海量用户时隙 Aloha 系统的**帧吞吐量**为

$$\lambda = \lambda_t e^{-\lambda_t T_{fr}} = \frac{G}{T_{fr}} \cdot e^{-G}$$

比特吞吐量为

$$\lambda_b = \lambda b = R_b G e^{-G}$$

其**归一化吞吐量**为

$$\rho = \frac{\lambda_b}{R_b} = G e^{-G}$$

Remark 题型：Aloha 与时隙 Aloha

典型题型 1 对于海量用户 Aloha 与时隙 Aloha 系统，求最能有效利用信息速率 R_b 的供给业务量 G 、帧生成率 λ_t 、帧长 b 或帧传输时间 T_{fr} 。

- 纯 Aloha 对 $\rho = G e^{-2G}$ 求导得到 $G^* = \frac{1}{2}$ ，此时 $\rho_{\max} = \frac{1}{2e} \approx 0.184$ 。
- 时隙 Aloha 对 $\rho = G e^{-G}$ 求导得到 $G^* = 1$ ，此时 $\rho_{\max} = \frac{1}{e} \approx 0.368$ 。

确定 G 后，可由约束 $\lambda_t T_{fr} = G$ 、 $T_{fr} = \frac{b}{R_b}$ 求解出相应的参数。

典型题型 2 给定 Aloha 或时隙 Aloha 系统的部分参数（如单用户帧生成率 λ_u 、帧长 b 等）与部分物理层参数（如工作频率），求对通信链路速率 R_b 利用率最高的物理层参数（如滚降系数 α 、调制阶数 M ）。

由 $G^* = 0.5$ 或 1 ，结合 $\lambda_t = N\lambda_u$ 和 $T_{fr} = \frac{b}{R_b}$ 可得

$$R_b^* = \frac{N\lambda_u b}{G^*}$$

结合物理层速率表达式（如带宽 B 、调制阶数 M 、滚降系数 α 等），即可求解出最优物理层参数。

载波侦听 (CSMA)

载波侦听多址 (carrier sense multiple access, CSMA) 是对 Aloha 的进一步改进，用户在发送数据前**先监听信道**，若信道空闲则发送数据，若信道忙碌则等待一段随机时间后重新监听。CSMA 可分为三种类型：

- 非坚持 CSMA (non-persistent CSMA)**：信道忙碌时，用户等待一段随机时间后重新监听信道；
- 1-坚持 CSMA (1-persistent CSMA)**：信道忙碌时，用户持续监听信道，直至信道空闲后立即发送数据；
- p -坚持 CSMA (p -persistent CSMA)**：适用于时隙化信道，信道空闲时，用户以概率 p 发送数据，以概率 $1-p$ 等待下一个时隙。

CSMA/CD 与以太网

以太网的核心协议**带冲突检测的载波侦听 (CSMA with collision detection, CSMA/CD)**是基于 CSMA 的改进。其基本思想是，用户在发送数据时**持续监听信道**，若检测到冲突，则**立即停止发送**。

为了更好地「通知」全网发生了冲突，在检测到冲突时发送一个**冲突信号 (jamming signal)**。以太网**有线媒介的封闭特性**决定了对于网中任意节点发送的信号在其它各节点处侦听的功率近似相同，这使得可以很容易地区分**信道空闲、信道忙碌和信道冲突** 3 种状态，从而实现载波侦听。

设网络中最大传播时间为 τ ，则

- 为了确保发送节点能够检测到冲突，需有**帧长 $T_{fr} > 2\tau$** ；
- 自上一帧传完起，**经过 τ 时间**后最远的节点也**确认信道空闲**，此时才开始竞争；
- 每次竞争耗时 2τ （从发送节点发送信号到最远节点，再从最远节点返回发送节点），可设**经过 k 次竞争**后成功。

这样，传输 1 帧的平均耗时为

$$\mathbb{E}[T] = \mathbb{E}[T_{fr} + \tau + 2\tau k] = T_{fr} + \tau + 2\tau \mathbb{E}[k]$$

可记 $a = \frac{\tau}{T_{fr}}$ ，则（总有用户试图尝试接入时）**CSMA/CD 的效率**为

$$\eta = \frac{T_{fr}}{\mathbb{E}[T]} = \frac{1}{1 + a(1 + 2\mathbb{E}[k])}$$

设用户尝试接入的概率为 q ，则一个竞争时隙内有且仅有一个用户发送数据的概率为

$$\gamma = Nq(1 - q)^{N-1}$$

知 $q = \frac{1}{N}$ 时， γ 取最大值 $\gamma_{\max} = \left(1 - \frac{1}{N}\right)^{N-1} \xrightarrow{N \rightarrow +\infty} \frac{1}{e}$ 。因此，**平均竞争次数**为

$$\mathbb{E}[k] = \sum_{k=1}^{+\infty} k \cdot (1 - \gamma)^{k-1} \gamma = \frac{1}{\gamma} \xrightarrow{N \rightarrow +\infty} e$$

代入效率表达式，得到 **CSMA/CD 的效率极限**为

$$\eta_{\max} = \frac{1}{1 + a(1 + 2e)} \approx \frac{1}{1 + 6.44a}$$

► CSMA/CA 与无线局域网

带冲突避免的载波侦听 (CSMA with collision avoidance, CSMA/CA)是对 CSMA 的改进，适用于无线局域网。由于无线信道的**开放特性**，节点难以区分**信道忙碌和信道冲突**，因此 CSMA/CD 不适用于无线局域网。CSMA/CA 通过引入**确认机制 (acknowledgment)**和**随机退避 (random backoff)**来避免冲突。

具体而言，CSMA/CA 的工作流程如下：

1. 发送方监听信道，若信道忙碌则等待一段随机时间后重新监听，若信道空闲则发送**请求发送 (RTS, Request to Send)** 帧；
2. 接收方收到 RTS 帧后，若信道空闲则发送**允许发送 (CTS, Clear to Send)** 帧，通知发送方可以发送数据；
3. 发送方收到 CTS 帧后，发送数据帧；
4. 接收方收到数据帧后，发送确认 (ACK) 帧；
5. 发送方收到 ACK 帧后结束传输。

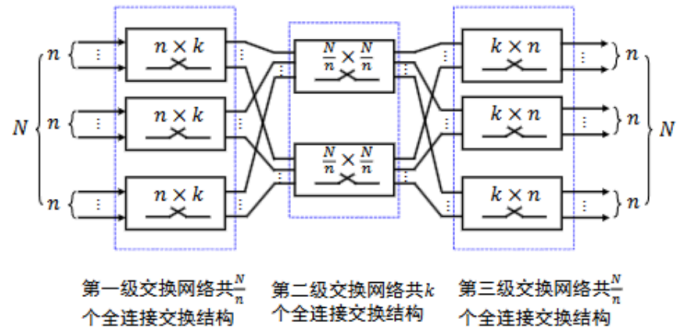
在 CSMA/CA 中，RTS、CTS 和 ACK 帧帮助网络内用户**虚拟侦听 (virtual carrier sensing)**，从而避免冲突的发生。

■ 交换

交换任务是星型网络上、下行链路的核心。为了提高链路利用率，常采用**多级交换结构**，使用多个小规模交叉开关 (crossbar) 组成大规模交换机。

设有一 N 进 N 出的交换需求，则

- **完全交叉开关 (full crossbar switch)** 需要 N^2 个交叉点 (crosspoint)，成本高昂。
- 若 $N = n^2$ ，则可采用**两级交换结构**，使用 n 个 n 进 n 出的交叉开关作为第一阶段，另 n 个 n 进 n 出的交叉开关作为第二阶段，总共需要 $2n \cdot n^2 = 2N^{1.5}$ 个交叉点，成本大幅降低。
- 若 N 被正整数 n 整除，则可采用**三级交换结构**，使用 $\frac{N}{n}$ 个 n 进 k 出交叉开关作为第一阶段， k 个 $\frac{N}{n}$ 进 $\frac{N}{n}$ 出交叉开关作为第二阶段，另 $\frac{N}{n}$ 个 k 进 n 出交叉开关作为第三阶段，总共需要 $2kN + k \frac{N^2}{n^2}$ 个交叉点，成本进一步降低。



广域网

人们希望在大的空间尺度 (Tele-) 上实现广泛 (anywhere, anytime, anyone, anything) 的连接 (communication)，从而形成**广域网 (wide area network, WAN)**。广域网通常由多个局域网通过路由器互联而成，覆盖范围广泛，支持长距离通信。

广域网的拓扑结构相对时域网复杂而多样，可以用有权（甚至有向）图 $G = (V, E)$ 来表示，边 (i, j) 的权重 $C(i, j)$ 表示从节点 i 到节点 j 的**传输开销 (cost)**，如时延、跳数、带宽消耗、经济费用等。

无论是电话网抑或计算机网，因为**鲁棒性需要**和**自组织建设**而常有**环**，于是从信源到宿信可能有大量可行的路径，需要找寻一条最优路径。要求这一寻路算法：

1. 易扩展（面向生长中的大型网络）；
2. 分布式（尽量减少信令交互）；
3. 动态自适应（于网络状态变化）；
4. 收敛快。

■ Bellman-Ford 算法与距离矢量协议

Bellman-Ford 算法是一种分布式的最短路径算法，其基本思想是通过**动态规划**的方式，逐步更新各节点到目标节点的最短路径估计值，直到收敛为止。每个节点仅需**与其直接邻居交换信息**，从而实现分布式计算。基于 Bellman-Ford 算法，设计提出了**距离矢量路由协议 (distance vector routing protocol)**，如 RIP（路由信息协议）。

Algo. 基于 Bellman-Ford 的距离矢量路由

距离矢量协议的基本思想是每个节点 i 维护一个距离矢量 (distance vector) $\vec{D}_i = (D_i(1), D_i(2), \dots, D_i(|V|))$ ，其中 $D_i(j)$ 表示节点 i 到节点 j 的当前最短路径估计值。

GIVEN —

- 图 $G = (V, E)$ 及其边权重 $C(i, j)$ ；
- 本地节点 a 。

INITIALIZE —

1. 对每个节点 $i \in V$ ，设置 $D_a(i) = \begin{cases} 0, & i = a, \\ C(a, i), & (a, i) \in E, \\ +\infty, & \text{otherwise} \end{cases}$
2. 将 \vec{D}_a 发送给所有邻居节点。

REPEAT —

当收到邻居节点 b 发送的距离矢量 \vec{D}_b 时，

1. 对每个节点 $i \in V$ ，计算通过邻居 b 到达节点 i 的路径长度 $D_a^{\text{new}}(i) = C(a, b) + D_b(i)$ ；
2. 若 $D_a^{\text{new}}(i) < D_a(i)$ ，则更新 $D_a(i) = D_a^{\text{new}}(i)$ ，同时对应记录下一跳节点为 b 。

若有变化，将更新后的距离矢量 \vec{D}_a 发送给所有邻居节点。

OUTPUT —

最终的距离矢量 \vec{D}_a ，表示节点 a 到所有节点的最短路径估计值，以及对应目标节点的下一跳节点。

Bellman-Ford 算法的优点在于其**分布式**和**易扩展**的特性，适用于大型动态网络。然而，其缺点也很明显：当网络链路变化甚至断线后，可能出现**计数发散 (count to infinity)**，即距离估计值不断增大，导致收敛缓慢。

■ Dijkstra 算法与链路状态协议

Dijkstra 算法是一种集中式的最短路径算法，其基本思想是通过**贪心策略**，逐步扩展已知最短路径的节点集合，直到找到从源节点到所有节点的最短路径。基于 Dijkstra 算法，设计提出了**链路状态路由协议 (link state routing protocol)**，如 OSPF（开放最短路径优先）。

链路状态协议的基本思想是每个节点 i 维护一个链路状态数据库 (link state database)，包含网络中所有节点及其连接关系和边权重信息。

GIVEN —

- 图 $G = (V, E)$ 及其边权重 $C(i, j)$;
- 本地节点 a 。

INITIALIZE —

1. 设置已知最短路径节点集合 $S = \{a\}$;
2. 对每个节点 $i \in V$ ，设置 $D_a(i) = \begin{cases} 0, & i = a, \\ C(a, i), & (a, i) \in E, \\ +\infty, & \text{otherwise} \end{cases}$ 。

REPEAT —

1. 从 $V \setminus S$ 中选择距离 $D_a(j)$ 最小的节点 j ，将其加入集合 S ;
2. 对每个节点 $i \in V \setminus S$ ，计算通过节点 j 到达节点 i 的路径长度 $D_a^{\text{new}}(i) = D_a(j) + C(j, i)$;
3. 若 $D_a^{\text{new}}(i) < D_a(i)$ ，则更新 $D_a(i) = D_a^{\text{new}}(i)$ ，同时对应记录下一跳节点为 j 。

UNTIL — $S = V$ 。

OUTPUT —

最终的距离矢量 \vec{D}_a ，表示节点 a 到所有节点的最短路径估计值，以及对对应目标节点的下一跳节点。