实验六 线性分组码仿真实现

一、 实验目的

1. 掌握线性分组码的基本原理与编译码算法；

2. 学习使用 MATLAB 软件实现线性分组码编译过程；

3. 学习使用 MATLAB 软件绘制（7，4）汉明码经 AWGN 信道传输后的误码率曲线，并对其进行性能分析。

二、 实验原理

1. 线性分组码的基本原理

分组码的基本思想是对信息序列进行分组编码。对包含 k 个码元的码组 $\boldsymbol{M}=(m_{k-1}, m_{k-2}, \ldots, m_1, m_0)$按照一定的编码规则产生包含 n 个码元的码组 $\boldsymbol{C}=(c_{n-1}, c_{n-2}, \ldots, c_1, c_0)$，编码规则定义为：

$$\begin{cases} c_0 = f_0\left(m_{k-1}, m_{k-2}, \ldots, m_1, m_0\right) \\ c_1 = f_1\left(m_{k-1}, m_{k-2}, \ldots, m_1, m_0\right) \\ \qquad\cdots \\ c_{n-1} = f_{n-1}\left(m_{k-1}, m_{k-2}, \ldots, m_1, m_0\right) \end{cases}$$

若 $f_i(\cdot)(i=0,1,\ldots n-1)$均为线性函数，则称 $\boldsymbol{C}$ 为线性分组码，一般用（n，k）表示，其中 n 表示码组长度，k 为码组中信息码元长度，$r=n-k$ 为码组中监督码元长度。

实际上，（n，k）线性分组码是 q 元有限域 GF(q)上 n 维线性空间 V_n 中的一个 k 维子空间 $V_{n,k}$。若信息码组 $\boldsymbol{M}$ 与码组 $\boldsymbol{C}$ 的所有元素均取自二元有限域 GF(2)（即{0,1}），则称为二元线性分组码。**以下仅讨论二元码的情况。**

2. 线性分组码的编码：生成矩阵与校验矩阵

对于二元线性分组码，其编码过程实际上就是从包含 2^k 个信息码组的 V_k 空间到包含 2^n 个码组的 V_n 空间的映射过程，因此在码空间 $V_{n,k}$ 中一定可以找到一组基底 $g_0, g_1, \cdots g_{k-1}$，使得所有码组都可以写成这 k 个基底的线性组合，即

$$\boldsymbol{C}=m_{k-1}\boldsymbol{g}_{k-1}+m_{k-2}\boldsymbol{g}_{k-2}+\ldots\ldots+m_1\boldsymbol{g}_1+m_0\boldsymbol{g}_0$$

这种线性组合特性正是线性分组码名称的来历。显然，研究线性分组的关键是研究基底、子空间和映射规则，如图 1 所示。

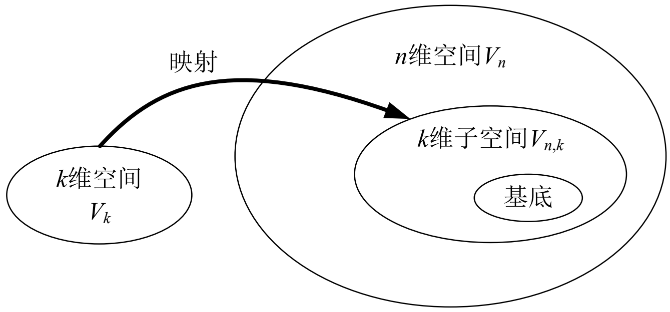

图 1 码空间与映射

用 $\boldsymbol{g}_i=(g_{i,n-1}, g_{i,n-2}, \ldots, g_{i,1}, g_{i,0})$表示第 $i(i=0,1,\ldots k-1)$个基底，再将 k 个基底排列成 k 行 n 列矩阵的形式，即

$$\boldsymbol{G}=\begin{bmatrix}\boldsymbol{g}_{k-1}\\ \vdots\\ \boldsymbol{g}_1\\ \boldsymbol{g}_0\end{bmatrix}=\begin{bmatrix}g_{k-1,n-1} & \cdots & g_{k-1,1} & g_{k-1,0}\\ \vdots & \ddots & \vdots & \vdots\\ g_{1,n-1} & \cdots & g_{1,1} & g_{1,0}\\ g_{0,n-1} & \cdots & g_{0,1} & g_{0,0}\end{bmatrix}$$

由于 k 个基底即 $\boldsymbol{G}$ 的 k 个行矢量线性无关，故矩阵 $\boldsymbol{G}$ 的秩一定等于 k。当信息码组 $\boldsymbol{M}$ 确定后，码组 $\boldsymbol{C}$ 仅由 $\boldsymbol{G}$ 矩阵决定，即（n，k）线性分组码中的任一码组 C 均可由这组基底的线性组合生成：

$$\boldsymbol{C}=\left[m_{k-1}, m_{k-2}, \cdots m_0\right]\begin{bmatrix}\boldsymbol{g}_{k-1}\\ \vdots\\ \boldsymbol{g}_1\\ \boldsymbol{g}_0\end{bmatrix}$$

$$=\boldsymbol{MG}$$

因此称这 $k\times n$ 矩阵 $\boldsymbol{G}$ 为该（n，k）线性分组码的生成矩阵。

基底不是唯一的，生成矩阵也就不是唯一的。事实上，将 k 个基底线性组合后产生另一组 k 个矢量，只要满足线性无关的条件，依然可以作为基底张成另一个码空间。不同的基底也有可能生成同一个码组，但因编码涉及码组和映射两个因素，即使码组相同而映射方法不同也不能认为是同样的码。

基底的线性组合等效于生产矩阵 $\boldsymbol{G}$ 的行运算，能够产生一组新的基底。利用矩阵的行运算可使生产矩阵具有如下的“系统形式”：

$$\boldsymbol{G}=\left[\boldsymbol{I}_k \quad \boldsymbol{P}\right]=\begin{bmatrix}1 & 0 & \cdots & 0 & p_{k-1,n-k-1} & \cdots & p_{k-1,1} & p_{k-1,0}\\ 0 & 1 & \cdots & 0 & p_{k-2,n-k-1} & \cdots & p_{k-2,1} & p_{k-1,0}\\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots & \vdots\\ 0 & 0 & \cdots & 1 & p_{0,n-k-1} & \cdots & p_{0,1} & p_{0,0}\end{bmatrix}$$

其中 $\boldsymbol{P}$ 为 $k\times(n\text{-}k)$矩阵，$\boldsymbol{I}_k$ 为 $k\times k$ 单位矩阵，从而保证矩阵 $\boldsymbol{G}$ 的秩为 k。对于系统码，有

$$\boldsymbol{C}=\boldsymbol{MG}=\left[\boldsymbol{M} \quad \boldsymbol{MP}\right]$$

根据线性代数知识，生成矩阵 $\boldsymbol{G}$ 是由 k 个线性无关的行向量构成的，因此

一定存在一个由 n-k 的线性无关的行向量构成的矩阵 $\boldsymbol{H}$ 与之相交，即 $\boldsymbol{G}\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{0}$。

对于“系统形式”的生成矩阵 $\boldsymbol{G}$，其校验矩阵 $\boldsymbol{H}$ 也是规则的：

$$\boldsymbol{H}=\begin{bmatrix}-\boldsymbol{P}^{\mathrm{T}} & \boldsymbol{I}_{n\text{-}k}\end{bmatrix}$$

对于二元码，由于模二减法等同于模二加法，上式的负号可以省略。

3. 线性分组码的距离与纠检错能力

在分组码中，把码组中“1”的数目称为码组的重量，简称码重；把两个码组中对应位置上数字不同的位数称为码组的距离，简称码距，又称汉明距离，记为 d。某种编码算法中各个码组之间距离的最小值称为最小码距：

$$d_0=\min\left\{d_{\boldsymbol{C}_i\boldsymbol{C}_j}, i\neq j, \boldsymbol{C}_i, \boldsymbol{C}_j\in V_{n,k}\right\}$$

两个码组之间的距离表示它们之间差别的大小：距离越大，两个码组的差别越大，传输时从一个码组错成另一码组的可能性越小，抗干扰能力越强。估算最小码距是纠错码设计的重要步骤，原始方案是逐一计算两两码组之间的距离，找到其中的最小值；然而若每个码集有 2^k 个许用码组, 就需要计算 $2^k(2^k-1)/2$ 个距离，计算量太大。

利用线性分组码的封闭性：任意两个码组之和仍为许用码组，即

$$\boldsymbol{C}_i+\boldsymbol{C}_j=\boldsymbol{C}_m\in\boldsymbol{C}$$

因此任意两个码组之间的码距就是另一码组的码重，表达式如下：

$$d_{\boldsymbol{C}_i\boldsymbol{C}_j}=w\left(\boldsymbol{C}_i+\boldsymbol{C}_j\right)=w\left(\boldsymbol{C}_m\right)$$

式中 $w(\boldsymbol{C}_m)$表示码组 $\boldsymbol{C}_m$ 的码重。最小码距为 d_0 可表示为：

$$d_0=\min\left\{w\left(\boldsymbol{C}_m\right), \boldsymbol{C}_m\in V_{n,k}, \boldsymbol{C}_m\neq\boldsymbol{0}\right\}$$

将计算最小码距问题转化成寻找最轻码字问题，若每个码集有 2^k 个许用码组, 仅需计算 2^k 次。

对于最小码距为 d_0 的线性分组码，其检错能力 $e=d_0-1$，纠错能力：

$$t=\left\lfloor\frac{d_0-1}{2}\right\rfloor$$

式中$\lfloor\cdot\rfloor$表示向下取整。

码的纠错能力取决于码的最小距离，但还需说明的另一点是码的总体纠错能力不仅仅与 d_0 有关。纠错能力 t 只是说明距离 t 以内的差错一定能纠正，并非说距离大于 t 的差错一定不能纠正。事实上，如果每个码集有 2^k 个许用码组, 就存在 $2^k(2^k-1)/2$ 个距离，且每个距离并非是相等的。比如最小距离 $d_0=3$，纠错能力 $t=1$：若许用码组 $\boldsymbol{C}_2$ 与 $\boldsymbol{C}_1$ 之间的距离等于最小距离 3，而 $\boldsymbol{C}_2$ 与 $\boldsymbol{C}_3$ 之间的距离大于最小距离 3（此处假设为 5）；只要 $\boldsymbol{C}_2$ 朝 $\boldsymbol{C}_1$ 方向的偏差大于 1 就会出现译码错误，然而若 $\boldsymbol{C}_2$ 朝 $\boldsymbol{C}_3$ 方向偏差 2，译码时仍可正确地判断为 $\boldsymbol{C}_2$ 而非 $\boldsymbol{C}_3$。可见，总体的、平均的纠错能力不但与最小距离有关，而且与其余码距离或者说与码组

的重量分布特性有关，把码距（码重）的分布特性称为距离（重量）谱，其中最小的重量就是 d_0。正如信息论各符号等概时熵最大一样，从概念上可以想象到：当所有码距相等时（重量谱为线谱）码的性能应该最好；或者退一步说，当各码距相差不大时（重量谱为窄谱）性能应该称得上好。事实证明确实如此，在同样的 d_0 条件下，窄谱的码一般比宽谱的码更优。

纠错重量谱的研究具有理论与现实意义，不仅仅是计算各种译码差错概率的主要依据，也是研究码的结构、改善码集内部关系从而发现新的好码的重要工具。但目前除了少数几类码如汉明码、极长码等的重量分布已知外，还有很多码的重量分布并不知道，距离分布与性能之间确切的定量关系对于大部分码而言尚在进一步研究当中，特别当 n 和 k 较大时，要得出码重分布是非常困难的。

4. 线性分组码的译码：伴随式与错误图样

编码后输出的码组在信道上传输时可能产生一定的码元错误，将这些码元错误称为错误图样。设发送码组为 $\boldsymbol{C}=(c_{n-1}, c_{n-2}, \ldots, c_1, c_0)$，信道产生的错误图样为 $\boldsymbol{E}=(e_{n-1}, e_{n-2}, \ldots, e_1, e_0)$，则接收到的码组为 $\boldsymbol{R}=\boldsymbol{C}+\boldsymbol{E}=(r_{n-1}, r_{n-2}, \ldots, r_1, r_0)$，译码器的作用就是根据接收到的码组 $\boldsymbol{R}$ 来估计错误图样 $\boldsymbol{E}$，进而得到对发送码组 $\boldsymbol{C}$ 的估计。

对于二元码，模二加法等同于模二减法，故错误图样 $\boldsymbol{E}=\boldsymbol{R}-\boldsymbol{C}=\boldsymbol{R}+\boldsymbol{C}$。利用码组与校验矩阵 $\boldsymbol{H}$ 的正交性即 $\boldsymbol{C}\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{M}\boldsymbol{G}\boldsymbol{H}^{\mathrm{T}}=\mathbf{0}$，可检验接收码组 $\boldsymbol{R}$ 是否出错：

$$\boldsymbol{R}\boldsymbol{H}^{\mathrm{T}}=(\boldsymbol{C}+\boldsymbol{E})\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{C}\boldsymbol{H}^{\mathrm{T}}+\boldsymbol{E}\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{E}\boldsymbol{H}^{\mathrm{T}}\begin{cases}=\mathbf{0}\\ \neq\mathbf{0}\end{cases}$$

定义伴随式 $\boldsymbol{S}=\boldsymbol{R}\boldsymbol{H}^{\mathrm{T}}=\boldsymbol{E}\boldsymbol{H}^{\mathrm{T}}$，由上式可知伴随式 $\boldsymbol{S}$ 仅与错误图样 $\boldsymbol{E}$ 有关，与发送码组 $\boldsymbol{C}$ 无关：若在信道传输过程中没有错误发生即 $\boldsymbol{E}=\mathbf{0}$，则 $\boldsymbol{S}=\mathbf{0}$；否则 $\boldsymbol{S}\neq\mathbf{0}$。

伴随式译码算法的基本思想就是根据伴随式 $\boldsymbol{S}$ 的值来估计错误图样 $\boldsymbol{E}$。将检验矩阵 $\boldsymbol{H}$ 写成列向量的形式：

$$\boldsymbol{H}=\left[\boldsymbol{h}_{n-1},\boldsymbol{h}_{n-2},\ldots.,\boldsymbol{h}_1,\boldsymbol{h}_0\right]$$

则有

$$\boldsymbol{S}=\boldsymbol{E}\boldsymbol{H}^{\mathrm{T}}=\sum_{i=0}^{n-1}e_i\boldsymbol{h}_i^{\mathrm{T}}$$

即伴随式 $\boldsymbol{S}$ 是校验矩阵 $\boldsymbol{H}$ 中列向量的线性组合。对于错误图样 $\boldsymbol{E}$，码元中第 j 位发生错误时其值 $e_j=1$，否则 $e_j=0$。因此伴随式 $\boldsymbol{S}$ 的值实际上是出错码元对应的校验矩阵 $\boldsymbol{H}$ 的列向量的模二和。

在线性分组码的纠错能力范围内，如果能够确定伴随式 $\boldsymbol{S}$ 的值是校验矩阵 $\boldsymbol{H}$ 的哪个或哪几个列向量的模二和，就可以确定错误图样 $\boldsymbol{E}$，进而实现译码。

5. 线性分组码编译码示例：汉明码的编译过程

汉明码是最常见的线性分组码，最小距离 $d_0=3$，纠错能力 $t=1$，即能够纠正

单个随机错误。以（7，4）汉明码为例，生成矩阵 $\boldsymbol{G}$ 如下：

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

根据 $\boldsymbol{GH}^{\mathrm{T}}=\boldsymbol{0}$，校验矩阵 $\boldsymbol{H}$ 为：

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

根据 $\boldsymbol{C}=\boldsymbol{MG}=\boldsymbol{M}[\boldsymbol{I}\ \boldsymbol{P}]=[\boldsymbol{M}\ \boldsymbol{MP}]$，信息码元 $\boldsymbol{M}$ 与监督码元 $\boldsymbol{MP}$ 对应关系如表 1 所示。

表 1 信息码元与监督码元对应关系

信息码元 ($c_6c_5c_4c_3$)	监督码元 ($c_2c_1c_0$)	信息码元 ($c_6c_5c_4c_3$)	监督码元 ($c_2c_1c_0$)
0000	000	1000	111
0001	011	1001	100
0010	101	1010	010
0011	110	1011	001
0100	110	1100	001
0101	101	1101	010
0110	011	1110	100
0111	000	1111	111

根据 $\boldsymbol{S}=\boldsymbol{RH}^{\mathrm{T}}=\boldsymbol{EH}^{\mathrm{T}}$，伴随式 $\boldsymbol{S}$ 与错误图样 $\boldsymbol{E}$ 的关系如表 2 所示。

表 2 伴随式与错误图样对应关系

伴随式 ($s_2s_1s_0$)	错误图样 ($e_6e_5e_4e_3e_2\ e_1e_0$)	伴随式 ($s_2s_1s_0$)	错误图样 ($e_6e_5e_4e_3e_2\ e_1e_0$)
000	0000000	011	0001000
001	0000001	101	0010000
010	0000010	110	0100000
100	0000100	111	1000000

设输入信息序列 $\boldsymbol{M}$=(1 0 0 1)，按表 1 进行编码可得：

$$\boldsymbol{C}=\boldsymbol{MG}=(1\ 0\ 0\ 1\ 1\ 0\ 0)$$

得到编码输出 $\boldsymbol{C}$= (1 0 0 1 1 0 0)。

设信号在传输时发生错误，得到译码输入序列 $\boldsymbol{R}$=(1 0 0 0 1 0 0)，计算伴随式 $\boldsymbol{S}=\boldsymbol{RH}^{\mathrm{T}}$=(0 1 1)。根据表 2 可得错误图样为 $\boldsymbol{E}$= (0 0 0 1 0 0 0)，即接收序列第四位发生错误，纠正后得到正确的发送序列 $\boldsymbol{C}=\boldsymbol{R}+\boldsymbol{E}$= (1 0 0 1 1 0 0)，可知信息序列

M=(1 0 0 1)。

6. 实验中可能用到的 Matlab Function（仅供参考）

rem()，reshape()，switch()，randi(), awgn()，sign()。

三、 实验预习

回答以下问题：

1. 给定生成矩阵 ***G*** 如下，请给出信息码元与监督码元对应关系表，并说明该码的最小距离与纠错能力。

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

2. 依据上题中的生成矩阵 ***G*** 计算相应校验矩阵 ***H***，给出伴随式 ***S*** 与错误图样 ***E*** 对应关系表。

四、 实验内容（要求给出结果截图，源代码放在. m 文件中）

给定（7，4）汉明码生成矩阵 ***G*** 如下：

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

1. 编写 MATLAB 函数（function）实现分组编码：函数输入为生成矩阵 ***G*** 与信息序列 ***M***，输出为编码结果 ***C***。（考虑到可能出现输入信息序列无法完整分组的情况，即序列长度不是 k 的整数倍，可在序列末尾适当补零以完整分组。）

   当输入信息序列 ***M***=[1 0 1 1 0 1 0 1 1 0]时，输出编码结果。

2. 编写 MATLAB 函数（function）实现该分组码的译码过程：函数输入为生成矩阵 ***G*** 与接收序列 ***R***，输出为译码结果。（提示：请先完成实验预习 2，得到伴随式与错误图样对应关系。）

   当接收序列 ***R*** 分别为发送序列 ***C*** 在未发生错误、发生一位错误、发生两位错误（同一码组中）的情况下得到时，输出译码结果。

3. 在 AWGN 信道传输与 BPSK 调制的条件下，绘制未编码系统与（7,4）汉明编码系统的误码率曲线（BER v.s. SNR）。信噪比范围：0~5dB，参考流程图见图 2。

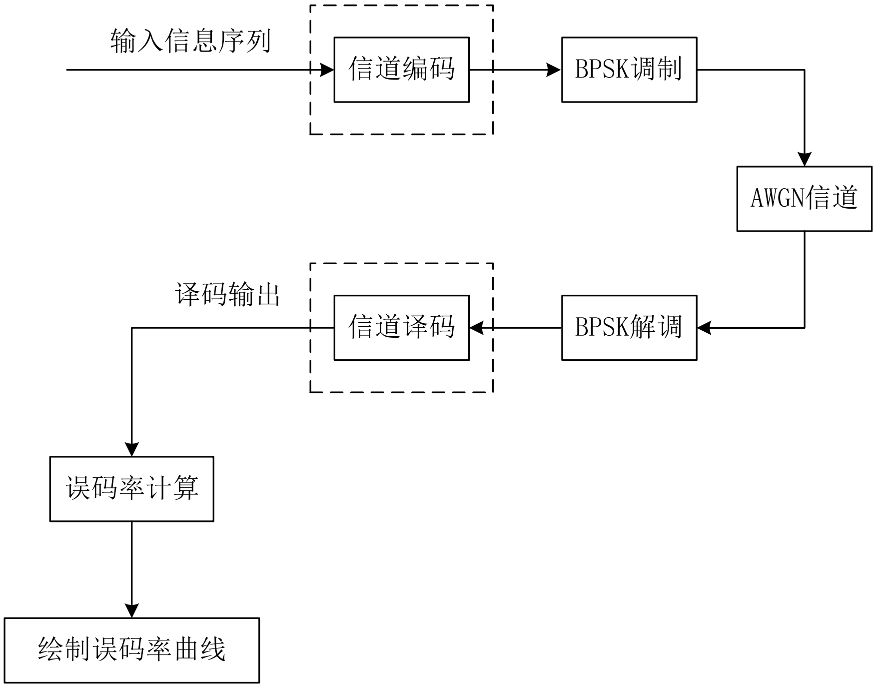

图 2 信道编码系统的实现流程图

五、 实验思考题

1. 对比分析发送序列分别在未发生错误、发生一位错误、发生两位错误（同一码组中）情况下得到的接收序列 ***R*** 的译码结果。
2. 对比分析在 AWGN 信道传输与 BPSK 调制的条件下，未编码系统与（7,4）汉明编码系统的误码性能。