

# 一种可达信道容量的编码方法

王杰林<sup>1,2</sup>

(1 湖南涉外经济学院信息与机电工程学院, 湖南 长沙 410205;

2 湖南遥昇通信技术有限公司, 湖南 长沙 410600)

**摘要:** 本文给出了一种新的信道编码方法, 经证明, 编码码率可达 1 且当码长趋近于无穷, 本文编码方法使得传输速率可达信道容量。码率为 1/2 和二进制输入加白噪声信道 (Binary Input Additive White Gaussian Noise, BIAWGN) 下, 本文方法在仿真实验中表现出了良好的性能。

**关键词:** 信息熵, 信道容量, 数据校验, 前向纠错

**中图分类号:** TN911.22

**文献标志码:** A

## A Coding Method to Achieve Channel Capacity

WANG Jielin<sup>1,2</sup>

1 School of information and mechanical and electrical engineering, Hunan International Economics University, Changsha, 410205, China

2 Hunan YESINE Communication Technology Co., Ltd., Changsha, 410600, China

**Abstract:** This paper presents a new channel coding method. It has been proved that the coding rate can reach 1 and when the code length approaches infinity, the coding method in this paper enables the transmission rate to approximate the channel capacity. Under the code rate 1/2 and the Binary Input Additive White Gaussian Noise (BIAWGN), the method in this paper has shown good performance in the simulation experiment.

**Keywords:** Information entropy, Channel capacity, Data validation, Forward error correction

## 1 引言

为了构造逼近信道容量的编码方法, 专家学者们付出了不懈努力。2009 年, Arikan 教授提出了基于信道极化现象, 在码长趋近于无限时被证明容量可达的编码方法, 称为极化码 (Polar Code)<sup>[1]</sup>。LDPC 码<sup>[2]</sup>, Turbo 码<sup>[3]</sup>也可逼近香农限。为了构造新的信道编码算法, 本文通过分析二进制信源序列的特征, 找出可用于数据校验的方法, 从而构造出新的编码方法和前向纠错方法。我们列举三种二进制信源序列:

(1) 随机生成长度为  $n$  的二进制信源序列  $X$ , 设  $X = (0, 1, 1, 0, 1, 0, 0, 1, 1, 1, 0, 1, 0, 0, 0, 0 \dots)$ , 序列  $X$  中符号等概率。

(2) 将  $X$  中符号 1 替换为 “10” 得到序列  $Y$ , 则  $Y = (0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 0 \dots)$ 。

(3) 将  $X$  中符号 1 替换为 “101” 且符号 0 替换为 “01” 得到序列  $Z$ , 则  $Z = (0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 0, 1 \dots)$ 。将  $Y$  或  $Z$  的长度记为  $l$ , 序列转换目的是使得  $Y$  和  $Z$  满足数据校验的条件。  $Y$  的数据校验条件为:

序列中连续符号 1 的个数最多为 1 (1-1)

$Z$  的数据校验条件为:

序列中连续符号 0 的个数最多为 1, 且连续符号 1 的个数最多为 2 (1-2)

$Z$  比  $Y$  多一个数据校验条件, 不同的转换方法使得序列具备更多的数据校验条件。本文将二进制序列转换方法定义为信源处理。信源处理使得  $Y$  和  $Z$  中存在大量冗余信息, 可采用熵编码对  $Y$  和  $Z$  进行无损压缩编码, 将编码结果记为二进制序列  $V$ 。算术编码<sup>[4]</sup><sup>[5]</sup>具有较差的容错性<sup>[6]</sup><sup>[7]</sup>, 如存在比特错误的序列  $V$  译码出错误的二进制序列  $Y'$  或  $Z'$ , 当  $Y'$  不满足 (1-1) 时必然有  $Y' \neq Y$ , 或  $Z'$  不满足 (1-2) 时必然有  $Z' \neq Z$ , 可实现数据校验。

因  $Y$  和  $Z$  中符号存在明显的上下文关系, 以序列  $Y$  为例, 可选择上下文二进制算术编码 (Context-Adaptive Binary Arithmetic Coding, CABAC)<sup>[8]</sup><sup>[9]</sup> 算法对序列  $Y = (y_1, y_2, \dots, y_l)$  编码, 设  $R_0 = 1$ ,  $L_0 = 0$ , 则第  $i$  ( $i = 1, 2, 3, \dots$ ) 个符号  $y_i$  的二进制算术编码运算式为:

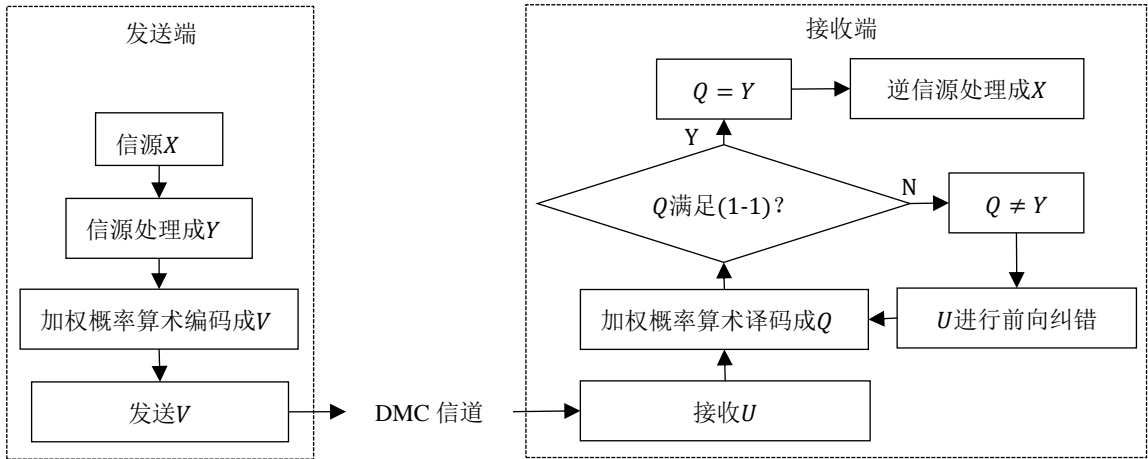
$$\begin{aligned}
R_i &= R_{i-1}p(y_i) \\
L_i &= L_{i-1} + R_{i-1}F(y_i - 1) \\
H_i &= L_i + R_i
\end{aligned} \tag{1-3}$$

因 $x_i \in \{0,1\}$ , 所以 $y_i = 0$ 时 $-1 \notin \{0,1\}$ , 则 $p(-1) = 0$ ,  $F(-1) = 0$  ( $F$ 为累积分布函数<sup>[10][11]</sup>)。约定 $y_0 = 0$ , 则 $P(y_i = 0)$ 存在两个概率 $p(0|0)$ 和 $p(0|1)$ ;  $y_i = 1$ 时 $F(0) = P(y_i = 0)$ 且 $P(y_i = 1)$ 存在概率 $p(1|0)$ 。当 $y_{i-1} = 1$ 且 $y_i = 0$ 时第 $i$ 个符号 0 的概率 $P(y_i = 0) = p(0|1) = 1$ , 由(1-3)可得 $L_i = L_{i-1}$ ,  $R_i = R_{i-1}$ 则 $H_i = H_{i-1}$ , 于是当前符号 0 被无效编码 (或跳过编码),  $L_i$ 为编码结果。

序列 $X$ 的信息熵为 $H(X) = -p(0)\log_2 p(0) - p(1)\log_2 p(1)$  ( $p(0) = p$ ,  $p(1) = 1 - p$ ,  $p$ 为符号 0 的概率且 $0 \leq p \leq 1$ ), 序列 $Y$ 的信息熵为 $H(Y) = -p(0|0)\log_2 p(0|0) - p(1|0)\log_2 p(1|0) - p(0|1)\log_2 p(0|1)$ , 因 $p(0|1) = 1$ 且 $p(0|0) = p$ ,  $p(1|0) = 1 - p$ , 所以 $H(Y) = H(X)$ 。于是 $L_i = L_n$ , 即序列 $Y$ 编码结果 $L_i$ 等同于序列 $X$ 编码结果 $L_n$ 。

将序列 $Y$ 的编码结果 $L_i$ 转换为长度为 $m$ 的二进制序列 $V = (v_1, v_2, \dots, v_m)$ ,  $V$ 通过 DMC 信道传输,  $U = (u_1, u_2, \dots, u_m)$ 为接收到的二进制序列, 如图 1 所示。

图 1 基于加权概率信道编译码示意图



若 $U = V$ ,  $U$ 译码出序列 $Q = X$ ; 若 $U \neq V$ , 因 $U$ 译码出错误的序列 $X'$ ,  $X'$ 不具有数据校验条件, 所以无法校验 $U$ 的错误。同理, 序列 $Z$ 也无法基于上下文二进制算术编码实现数据校验, 基于条件概率 (或马尔科夫链) 无法构造序列 $Y$ 或 $Z$ 的数据校验编码方法。

当 $y_{i-1} = 1$ ,  $y_i = 0$ 且第 $i$ 个符号 0 的概率 $P(y_i = 0) = p(0|1) = r$  ( $0 < r < 1$ ) 时,  $U = V$ , 则 $U$ 译码得序列 $Q = Y$ ;  $U \neq V$ , 因 $Q$ 必须满足(1-1), 可有效判断 $U$ 发生错误。

$P(y_i = 0) = p(0|1) = r$ 可被理解为 $P(y_i = 0) = rp(0|1) = r$ , 本文将 $r$ 称为概率的权系数, 于是 $p(0|0)$ 和 $p(1|0)$ 的权系数为 1。当 $r < 1$ 且 $r \rightarrow 1$ 时, 序列 $Y$ 的编码结果 $L_i$ 趋近于序列 $X$ 编码结果 $L_n$ , 基于信息熵很容易证明这一结论。令 $\varphi(y_i) = rp(y_i)$ ,  $\varphi(y_i)$ 被定义为加权概率。最简单的加权概率为所有符号概率的权系数均为同一实数 $r$ , 令 $y_i \in A = \{0,1,\dots,s\}$ , 加权累积分布函数被定义为:

$$F(y_i, r) = \sum_{y \leq y_i} rp(y) = r \sum_{y \leq y_i} p(y) = rF(y_i) \tag{1-4}$$

若 $y_i \in \{0,1\}$ ,  $y_i = 0$ 时 $F(y_i - 1, r) = rF(y_i - 1) = rp(-1) = 0$ ;  $y_i = 1$ 时 $F(y_i - 1, r) = rF(y_i - 1) = rp(0)$ 。设 $R_0 = 1$ ,  $L_0 = 0$ , 则第 $i$  ( $i = 1, 2, 3, \dots$ ) 个符号 $y_i$ 的加权概率算术编码运算式为:

$$\begin{aligned}
R_i &= R_{i-1}rp(y_i) \\
L_i &= L_{i-1} + R_{i-1}rF(y_i - 1) \\
H_i &= L_i + R_i
\end{aligned} \tag{1-5}$$

图 1 中 DMC 信道的传输速率为 $I(V; U) = H(V) - H(V|U)$ 。常见 DMC 信道有  $BSC(\xi)^{[12][13]}$  和  $BSC(\varepsilon)^{[14][15]}$ 。BSC( $\xi$ )信道使得序列 $V$ 中符号 0 接收成 1 的概率为 $\xi$ , 符号 1 接收成 0 的概率为 $\xi$ 。于是,  $p(u = 0|v = 0) = p(u = 1|v = 1) = 1 - \xi$ ,  $p(u = 1|v = 0) = p(u = 0|v = 1) = \xi$  ( $u, v \in \{0,1\}$ ), 可得 $H(V|U) = H(\xi) = -\xi \log_2 \xi - (1 - \xi) \log_2 (1 - \xi)$ 。BSC( $\xi$ )信道的传输速率 $R_{BSC} = H(V) - H(\xi)$ 。当 $H(V) = 1$ 时传输速率最大, 信道容量为:

$$C_{BSC} = 1 - H(\xi) \tag{1-6}$$

BEC( $\varepsilon$ )信道发送符号 $v \in \{0,1\}$ , 接收符号 $u \in \{0,1,\hat{e}\}$ ,  $\hat{e}$ 为擦除符号, 擦除概率为 $\varepsilon$ 。于是 $p(u = 0|v = 0) = p(u = 1|v = 1) = 1 - \varepsilon$ ,  $p(u = \hat{e}|v = 0) = p(u = \hat{e}|v = 1) = \varepsilon$ , 可得 BEC( $\varepsilon$ )信道的传输速率 $R_{BEC} = H(V) - \varepsilon$ , 信道容量为:

$$C_{BEC} = 1 - \varepsilon \tag{1-7}$$

接收端通过加权概率算术译码得二进制序列 $Q = (q_1, q_2, \dots, q_l)$ , 当序列 $Q$ 满足(1-1)时存在概率 $P_{err}$ 使得 $Q \neq Y$ , 则 $P_{err}$ 为本文方

法的平均译码错误概率<sup>[16]</sup>。当 $Q$ 不满足(1-1)时必然有 $Q \neq Y$ ，本文中根据不同的信道模型构造不同的前向纠错方法。

通过上述分析，本文编译码方法区别于 BCH<sup>[17]</sup>、Hamming Code<sup>[18]</sup>、RS<sup>[19]</sup>、CRC<sup>[20]</sup>、LDPC<sup>[21][22]</sup>、Turbo<sup>[23][24]</sup>和 Polar Code<sup>[25][26][27]</sup>等编码方法，我们不可能从这些编码方法中找到直接比较方法。

本文在第二部分中给出加权概率模型信道编码方法，并给出 BSC 和 BEC 信道下的数据校验和前向纠错译码方法。第三部分中分析加权概率模型信息熵、编码率和译码错误概率，证明  $\lim_{n \rightarrow \infty} P_{err} = 0$  且  $H(V) = 1$ 。第四部分通过仿真实验得出本文方法的性能。

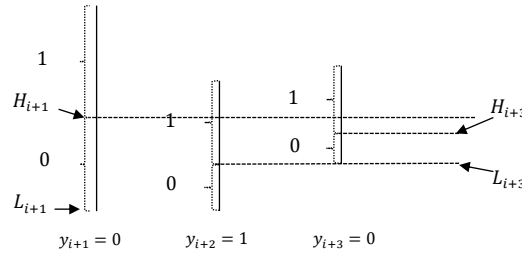
## 2 加权概率算术编译码

**定义 2.1** 设离散随机变量 $y, y \in A = \{0, 1, \dots, k\}$ ,  $P\{y = a\} = p(a) (a \in A)$ , 加权概率质量函数为  $\varphi(a) = rP\{y = a\} = rp(a)$ ,  $p(a)$  为的概率质量函数,  $0 \leq p(a) \leq 1$ ,  $r$  为权系数, 且

$$F(a) = \sum_{y \leq a} p(y) \quad (2-1)$$

若  $F(a, r)$  满足  $F(a, r) = rF(a)$ , 则称  $F(a, r)$  为加权累积分布函数, 简称加权分布函数。显然, 所有符号的加权概率之和为  $\sum_{a=0}^k \varphi(a) = r$ 。权系数  $r$  存在三种基本情形:  $0 < r < 1$ ;  $r = 1$ ;  $r > 1$  且此情形下无损编译码必然存在极大值  $r_{max}$ 。令  $r > 1$ ,  $A = \{0, 1\}$ , 序列  $Y$  中符号 0 和符号 1 的概率分别为  $p(0) = p$ ,  $p(1) = 1 - p$ 。根据(1-5)编码  $y_{i+1} = 0$ ,  $y_{i+2} = 1$  和  $y_{i+3} = 0$  的过程如图 2。

图 2 加权模型编码 010 的过程示意图



根据图 2,  $L_{i+3} = L_i + R_i r^2 p^2$ ,  $R_{i+3} = R_i r^3 p^2 (1 - p)$ ,  $H_{i+3} = L_i + R_i r^2 p^2 + R_i r^3 p^2 (1 - p)$ 。因  $y_{i+1} = 0$ ,  $F(-1) = 0$ , 所以  $L_{i+1} = L_i$ ,  $R_{i+1} = R_i r p$ ,  $H_{i+1} = L_i + R_i r p$ 。令  $H_{i+3} \leq H_{i+1}$  可得:

$$\varphi(0)\varphi(1) + \varphi(0) = r^2 p(1 - p) + r p \leq 1 \quad (2-2)$$

设方程  $ar^2 + br + c = 0$ , 其中  $a = p(1 - p)$ ,  $b = p$ ,  $c = -1$ , 且  $r > 0$ 。满足方程的正实数根为  $\frac{\sqrt{p^2 + 4(1-p)p} - p}{2(1-p)p}$ , 化简:

$$r \leq \frac{\sqrt{4/p - 3} - 1}{2 - 2p} \quad (2-3)$$

全 0 或全 1 的二进制序列不作考虑, 也无需纠错, 所以  $0 < p < 1$ 。译码时, 根据(1-5)分配符号  $y_{i+1}$  对应的区间, 当  $y_{i+1} = 0$  时区间为  $[L_i, L_i + R_i r p)$ , 当  $y_{i+1} = 1$  时区间为  $[L_i + R_i r p, L_i + R_i r)$ 。于是, 当  $L_{i+3} < L_i + R_i r p$  时  $y_{i+1} = 0$ ; 当  $L_{i+3} \geq L_i + R_i r p$  时  $y_{i+1} = 1$ 。因  $1 - p > 0$ , 所以  $4(1 - p)^2 > 0$ ,  $4p^2 - 8p + 4 > 0$ ,  $p^2 - 4p + 4 > 4p - 3p^2$ ,  $(2 - p)^2 > 4p - 3p^2$ ,  $2 - p > \sqrt{4p - 3p^2}$ ,  $2 - 2p > p \sqrt{\frac{4}{p} - 3} - p$ , 可得  $\frac{1}{p} > \frac{\sqrt{\frac{4}{p} - 3} - 1}{2 - 2p}$ , 即  $r_{max} < \frac{1}{p}$ 。假设  $L_{i+3} = L_i + R_i r^2 p^2 \geq L_i + R_i r p$ , 可得  $r \geq \frac{1}{p}$ 。因  $0 < r \leq r_{max}$  且  $r_{max} < \frac{1}{p}$ , 假设不成立, 可得  $L_{i+3} = L_i + R_i r^2 p^2 < L_i + R_i r p$ , 于是  $y_{i+1} = 0$ 。

因  $y_{i+1} = 0$ , 所以  $L_{i+1} = L_i$ ,  $R_{i+1} = R_i r p$ 。当  $L_{i+3} < L_{i+1} + R_{i+1} r p$  时  $y_{i+2} = 0$ , 当  $L_{i+3} \geq L_{i+1} + R_{i+1} r p$  时  $y_{i+2} = 1$ 。又因  $L_{i+1} + R_{i+1} r p = L_i + R_i r^2 p^2$ , 所以  $L_{i+3} = L_{i+1} + R_{i+1} r p$ , 于是  $y_{i+2} = 1$ 。

因  $y_{i+2} = 1$ , 所以  $L_{i+2} = L_{i+1} + R_{i+1} r F(0) = L_i + R_i r^2 p^2$ ,  $R_{i+2} = R_{i+1} r (1 - p) = R_i r^2 p (1 - p)$ 。当  $L_{i+3} < L_{i+2} + R_{i+2} r p$  时  $y_{i+3} = 0$ , 当  $L_{i+3} \geq L_{i+2} + R_{i+2} r p$  时  $y_{i+3} = 1$ 。因  $L_{i+2} + R_{i+2} r p = L_i + R_i r^2 p^2 + R_i r^3 p^2 (1 - p)$  且  $R_i r^3 p^2 (1 - p) > 0$ , 所以  $L_{i+3} = L_i + R_i r^2 p^2 < L_{i+2} + R_{i+2} r p$ , 于是  $y_{i+3} = 0$ 。

**定理 2.2** 令  $r_{max} = \frac{\sqrt{4/p - 3} - 1}{2 - 2p}$ ,  $p$  为序列  $Y$  中符号 0 的概率, 当  $0 < r \leq r_{max} < \frac{1}{p}$  时加权概率算术编码通过  $V$  可无损译码序列  $Y$ 。

**证明:** 假设  $L_i$  ( $i \geq 1, i \in \mathbb{Z}$ ) 可无损译码  $y_1, y_2, \dots, y_i$ , 因序列  $Y$  满足(1-1), 需分两种情形证明。

(1)  $y_i = 1$ , 不存在  $y_{i+1} = 1$ , 当  $y_{i+1} = 0$  根据(1-5)编码可得  $L_{i+1} = L_i$ ,  $R_{i+1} = R_i r p$ , 因为  $L_{i+1} = L_i$ , 所以  $L_{i+1}$  可无损译码  $y_1, y_2, \dots, y_i$ 。又因  $R_i r p > 0$ ,  $L_{i+1} < L_i + R_i r p$ , 所以译码得  $y_{i+1} = 0$ 。

(2)  $y_i = 0$ , 根据(1-5)编码可得  $L_i = L_{i-1}$ ,  $R_i = R_{i-1} r p$ ,  $L_{i-1}$  可无损译码  $y_1, y_2, \dots, y_i$ 。当  $y_{i+1} = 0$  根据(1-5)编码可得  $L_{i+1} = L_i$ ,  $R_{i+1} = R_i r p$ , 因为  $L_{i+1} = L_i$ , 所以  $L_{i+1}$  可无损译码  $y_1, y_2, \dots, y_i$ 。又因  $L_{i+1} < L_i + R_i r p$ , 所以译码得  $y_{i+1} = 0$ 。

当  $y_{i+1} = 1$  根据(1-5)编码可得  $L_{i+1} = L_i + R_i r^2 p^2$ ,  $R_{i+1} = R_{i-1} r^2 p (1 - p)$ 。因  $r p < 1$ , 所以  $L_{i+1} < L_i + R_i r p$ , 译码得  $y_i = 0$ 。因

$y_i = 0$ , 且  $L_{i+1} + R_{i+1}rp = L_i + R_i r^2 p^2$ , 所以  $L_{i+1} = L_{i+1} + R_{i+1}rp$ , 译码得  $y_{i+1} = 1$ 。当  $y_i = 0$ ,  $y_{i+1} = 1$  时, 若能证明  $y_{i-1}$  可被正确译码, 则由归纳法可得  $L_{i+1}$  可无损译码  $y_1, y_2, \dots, y_i, y_{i+1}$ 。

当  $y_{i-1} = 0$ ,  $y_i = 0$ ,  $y_{i+1} = 1$  时编码可得  $L_i = L_{i-1} = L_{i-2}$ ,  $L_{i+1} = L_{i-2} + R_{i-2}r^3p^3$ 。因为  $rp < 1$ , 所以  $L_{i+1} < L_{i-2} + R_{i-2}rp$ , 译码得  $y_{i-1} = 0$ 。当  $y_{i-1} = 1$ ,  $y_i = 0$ ,  $y_{i+1} = 1$  时编码可得  $L_{i-1} = L_{i-2} + R_{i-2}rp$ ,  $L_i = L_{i-1}$ ,  $L_{i+1} = L_i + R_i rp = L_{i-2} + R_{i-2}rp + R_{i-1}r^3p^2(1-p)$ 。显然,  $L_{i+1} > L_{i-2} + R_{i-2}rp$ , 译码得  $y_{i-1} = 1$ 。即  $y_{i-1}$  可被正确译码。

设序列  $Y$  中第  $i+1$  个位置起有  $t+2$  ( $t=1,2,3,\dots$ ) 个符号为  $0,1,\dots,1,0$ , 其中符号 1 的连续个数为  $t$ 。根据(1-5)有:

$$H_{i+t+2} = L_i + R_i\varphi(0)^2 + R_i\varphi(0)^2\varphi(1) + R_i\varphi(0)^2\varphi(1)^2 + R_i\varphi(0)^2\varphi(1)^3 + \dots + R_i\varphi(0)^2\varphi(1)^{t-1} + R_i\varphi(0)^3\varphi(1)^t$$

$$H_{i+1} = L_i + R_i\varphi(0)$$

由  $H_{i+t+2} \leq H_{i+1}$  可得:

$$\varphi(0) + \varphi(0)\varphi(1) + \varphi(0)\varphi(1)^2 + \dots + \varphi(0)\varphi(1)^t \leq 1 \quad (2-4)$$

当  $\varphi(1) = 1$  时:

$$\varphi(0) \leq \frac{1}{t+1} \quad (2-5)$$

**定理 2.3** 设序列  $Y$  中连续符号 1 的个数最多为  $t$ , 当  $\varphi(1) = 1$  且  $\varphi(0) \leq \frac{1}{t+1}$  时加权概率算术编码通过  $V$  可无损译码序列  $Y$ 。

**证明** 设  $d$  为序列  $Y$  中连续符号 1 的个数,  $0 \leq d \leq t$ 。当  $d = 0$  时根据(1-5)可得  $L_{i+t+2} = L_{i+t+1} = L_{i+t} = \dots = L_i$ , 因  $L_{i+t+2} < L_i + R_i \frac{1}{(t+1)^{t+2}} < L_i + R_i \frac{1}{(t+1)^{t+1}} < \dots < L_i + R_i \frac{1}{t+1}$ , 所以  $y_{i+1} = y_{i+2} = \dots = y_{i+t+2} = 0$ 。当  $1 \leq d \leq t$ ,  $\varphi(1) = 1$ ,  $\varphi(0) \leq \frac{1}{t+1}$ 。根据(1-5)可得  $L_{i+d+2} = L_i + R_i \frac{d}{(t+1)^2} \leq L_i + R_i \frac{t}{(t+1)^2}$ 。译码时, 因  $L_i + R_i \frac{t}{(t+1)^2} < L_i + R_i \frac{1}{t+1}$ , 所以  $L_{i+d+2}$  可准确译码  $y_{i+1} = 0$ 。因  $y_{i+1} = 0$ , 所以  $L_{i+1} + R_{i+1}\varphi(0) = L_i + R_i\varphi(0)^2 = L_i + R_i \frac{1}{(t+1)^2}$ , 因  $\frac{d}{(t+1)^2} \geq \frac{1}{(t+1)^2}$ , 所以  $L_{i+d+2} \geq L_i + R_i \frac{1}{(t+1)^2}$ ,  $L_{i+d+2}$  可准确译码  $y_{i+2} = 1$ 。当  $d \geq 2$  时  $L_i + R_i \frac{d}{(t+1)^2} \geq L_i + R_i \frac{d}{(t+1)^2}$ ,  $L_{i+d+2}$  可准确译码  $y_{i+d+1} = 1$ 。又因  $L_{i+d+2} = L_{i+d+1}$  且  $L_{i+d+2} < L_i + R_i \frac{d+1}{(t+1)^2}$ , 所以  $L_{i+d+2}$  可准确译码  $y_{i+d+2} = 0$ 。当  $d = t+1$  时  $L_i + R_i \frac{t+1}{(t+1)^2} = L_i + R_i \frac{1}{t+1} > L_i + R_i \frac{t}{(t+1)^2}$ , 所以  $y_{i+1} = 1$ , 译码错误; 因  $d = t+1+c$  ( $c \geq 1$ ) 时  $L_i + R_i \frac{t+1+c}{(t+1)^2} = L_i + R_i \frac{1}{t+1} + R_i \frac{c}{(t+1)^2} > L_i + R_i \frac{1}{t+1} > L_i + R_i \frac{t}{(t+1)^2}$ , 所以  $y_{i+1} = 1$ , 译码错误。可得  $0 \leq d \leq t$  时  $V$  可无损译码序列  $Y$ 。当  $\varphi(1) = 1$  且  $\varphi(0) < \frac{1}{t+1}$  时根据上述证明步骤可得  $0 \leq d \leq t$  时  $V$  可无损译码序列  $Y$ 。

## 2.1 加权概率算术编码

编码过程中可将二进制信源序列  $X$  经信源处理成  $Y$  和序列  $Y$  进行加权概率算术编码合并, 根据图 1, 发送端对长度为  $n$  序列  $X$  的编码步骤如下:

- (1): 初始化参数,  $R \leftarrow 1$ ,  $L \leftarrow 0$ ,  $p \leftarrow 0$ ,  $i \leftarrow 1$ ;
- (2): 统计序列  $X$  中符号 0 的个数记为  $c$ ;
- (3):  $p \leftarrow \frac{n}{2n-c}$ ;
- (4):  $r_{\max} \leftarrow \frac{\sqrt{4/p-3}-1}{2-2p}$ ;
- (5): 获取序列  $X$  中第  $i$  个符号  $x_i$ ;
- (6): 若  $x_i = 0$ ,  $R \leftarrow Rr_{\max}p$ ;
- (7): 若  $x_i = 1$ ,  $R \leftarrow Rr_{\max}^2p(1-p)$  且  $L \leftarrow L + Rr_{\max}p$ ;
- (8):  $i \leftarrow i + 1$ ;
- (9): 若  $i \leq n$  则重复(5)到(9);
- (10): 将  $L$  转换为长度为  $m$  的二进制序列  $V$ ;
- (11): 发送  $V$ ,  $c$  和  $n$ 。

## 2.2 错误校验和译码

接收端获取  $V$ ,  $c$  和  $n$ , 并将  $V$  转换为实数  $v$ 。根据图 1, 设  $q_i$  为序列  $Q$  中第  $i$  个符号,  $l = 2n - c$ 。接收端译码步骤如下:

- (1): 初始化参数,  $R \leftarrow 1, L \leftarrow 0, i \leftarrow 1, H \leftarrow 0, s \leftarrow 0, p \leftarrow 0$ ;
- (2):  $p \leftarrow \frac{n}{2n-c}$ ;
- (3):  $r_{\max} \leftarrow \frac{\sqrt{4/p-3}-1}{2-2p}$ ;
- (4):  $H \leftarrow L + Rr_{\max}p$ ;
- (5): 当  $v \geq H$  且  $s = 1$  时,  $Q \leftarrow \text{null}$ , 结束译码;
- (6): 当  $L \geq H$  且  $s = 0$  时,  $q_i \leftarrow 1, s \leftarrow 1, R \leftarrow Rr_{\max}(1-p), L \leftarrow L + Rr_{\max}p$ ;
- (7): 当  $L < H$  且  $s = 1$  时,  $s \leftarrow 0, R \leftarrow Rr_{\max}p$ ;
- (8): 当  $L < H$  且  $s = 0$  时,  $q_i \leftarrow 0, s \leftarrow 0, R \leftarrow Rr_{\max}p$ ;
- (9):  $i \leftarrow i + 1$ ;
- (10): 如果  $i \leq l$  则重复(4)到步骤(10);
- (11): 返回序列  $Q$ 。

因序列  $Y$  满足(1-1), 所以步骤(5)中  $v \geq H$  时  $Q_i = 1$ , 又因  $s = 1$ , 所以译码出了“11”, 则  $U \neq V$ , 数据出现错误。当  $Q \neq \text{null}$  时存在概率  $P_{err}$  使得  $Q \neq Y$ 。文中将在第三部分中分析  $P_{err}$ 。当  $Q = \text{null}$  时, 可根据 DMC 信道模型分别给出不同的前向纠错方法。

### 2.3 BSC( $\xi$ )前向纠错译码

BSC( $\xi$ )信道模型, 存在概率  $\xi$  使得符号 0 被接收为符号 1, 或符号 1 被接收为符号 0。当  $Q = \text{null}$  时序列  $U$  中部分符号 0 或符号 1 错误, 设序列  $U = \{u_1, u_2, \dots, u_m\}$  且存在  $e$  个比特错误, 当  $e = 1$  时, 前向纠错译码步骤如下。

- (1): 初始化参数,  $i \leftarrow m$ ;
- (2): 将  $U$  的第  $i$  符号非运算, 即  $u_i \leftarrow \bar{u}_i$  (当  $u_i = 0$  时  $u_i \leftarrow 1$ , 当  $u_i = 1$  时  $u_i \leftarrow 0$ , 下同);
- (3):  $U$  代入 2.2 节译码;
- (4): 若  $i = 1$  且  $Q = \text{null}$ , 结束;
- (5): 若  $i > 1$  且  $Q \neq \text{null}$ , 译码完成, 结束;
- (6): 若  $i > 1$  且  $Q = \text{null}$ ,  $u_i \leftarrow \bar{u}_i, i \leftarrow i - 1$ , 重复(2)到(6)。

上述步骤遍历  $C_m^1$  次后  $Q = \text{null}$ , 则  $e > 1$ , 随后进行  $e = 2$  的前向纠错, 步骤如下。

- (1): 初始化参数,  $i \leftarrow j \leftarrow m$ ;
- (2):  $u_i \leftarrow \bar{u}_i$ ;
- (3):  $j \leftarrow i - 1$ ;
- (4):  $u_j \leftarrow \bar{u}_j$ ;
- (5):  $U$  代入 2.2 节译码;
- (6): 若  $j > 1$  且  $Q \neq \text{null}$ , 译码完成, 结束;
- (7): 若  $j > 1$  且  $Q = \text{null}$ ,  $u_j \leftarrow \bar{u}_j, j \leftarrow j - 1$ , 重复(4)到(8);
- (8): 若  $j = 1$  且  $Q = \text{null}$ ,  $u_j \leftarrow \bar{u}_j$ ;
- (9): 若  $i > 1$  且  $Q = \text{null}$ , 译码完成, 结束;
- (10): 若  $i > 1$  且  $Q = \text{null}$ ,  $u_i \leftarrow \bar{u}_i, i \leftarrow i - 1$ , 重复(2)到(10);
- (11): 若  $i = 1$  且  $Q = \text{null}$ , 结束;

$e = 2$  时遍历  $C_m^2$  (组合数) 次数  $Q = \text{null}$  则  $e > 2$ , 随后进行  $e = 2$  的前向纠错。当  $e \leq \tau$  ( $\tau$  为自定义的整数,  $1 \leq \tau \leq m$ ) 时, 需进行  $\sum_{i=1}^{\tau} C_m^i$  次遍历。当  $\tau = m$  时  $P_{err}$  是 BSC( $\xi$ ) 前向纠错译码时唯一的译码错误概率。

### 2.4 BEC( $\epsilon$ )前向纠错译码

设序列  $U$  中存在  $e$  个擦除符号  $\hat{e}$ ,  $\hat{e}$  取值于  $\{0, 1\}$ , 按顺序将  $e$  个擦除符号记为  $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_e$ 。设  $i = 1, 2, \dots, 2^e$ , 得表 1。

表 1  $\hat{e}_1, \hat{e}_2, \dots, \hat{e}_e$  存在  $2^e$  种排列

$i$	$\hat{e}_1, \hat{e}_2, \dots, \hat{e}_e$
1	0, 0, 0, $\dots$ , 0
2	1, 0, 0, $\dots$ , 0
3	0, 1, 0, $\dots$ , 0

...	...
$e$	$1, 1, 0, \dots, 0$
$e + 1$	$1, 0, 1, \dots, 0$
...	...
$2^e$	$1, 1, 1, \dots, 1$

按照表 1 第  $i$  种方式赋值给序列  $U$  中  $e$  个擦除符号，如  $i = 1$  时  $\hat{e}_1 = 0, \hat{e}_2 = 0, \dots, \hat{e}_e = 0$ ； $i = 2$  时  $\hat{e}_1 = 1, \hat{e}_2 = 0, \dots, \hat{e}_e = 0$ 。随后将序列  $U$  代入 2.2 节译码，若  $Q \neq \text{null}$ ，前向纠错译码完成。步骤如下：

- (1): 初始化参数， $i \leftarrow 1$ ；
- (2): 按照表 1 第  $i$  种方式赋值给序列  $U$  中  $e$  个擦除符号；
- (3):  $U$  代入 2.2 节译码；
- (4): 若  $i \leq 2^e$  且  $Q \neq \text{null}$ ，译码完成，结束；
- (5): 若  $i < 2^e$  且  $Q = \text{null}$ ， $i \leftarrow i + 1$ ，重复(2)到(5)；

因为  $2^e$  包含了所有可能性，所以  $2^e$  种可能性中至少存在一种可能性使得序列  $Q$  满足(1-1)。显然， $P_{err}$  是 BEC( $\epsilon$ ) 前向纠错译码时唯一的译码错误概率。

### 3 加权概率模型信息熵、编码率和平均译码错误概率

#### 3.1 加权概率模型信息熵

设二进制离散无记忆信源序列  $Y = (y_1, y_2, \dots, y_l) (y_i \in A = \{0, 1\})$ ，当  $r = 1$  时， $\varphi(y_i) = p(y_i)$ 。由香农信息熵定义， $Y$  的熵为：

$$H(X) = -p(0) \log_2 p(0) - p(1) \log_2 p(1) \quad (3-1)$$

当  $r \neq 1$  时，定义具有概率  $\varphi(y_i)$  的随机变量  $y_i$  的自信息量为：

$$I(y_i) = -\log_2 p(y_i) \quad (3-2)$$

设集合  $\{y_i = a\} (i = 1, 2, \dots, l, a \in A)$  中有  $c_a$  个符号  $a$ 。 $r$  已知时信源序列  $Y$  的总信息量为：

$$-c_0 \log_2 p(0) - c_1 \log_2 p(1)$$

于是平均每个符号的信息量为：

$$-\frac{1}{l} \sum_{a=0}^1 c_a \log_2 p(a) = -\sum_{a=0}^1 p(a) \log_2 p(a)$$

**定义 3.1** 令  $H(Y, r)$  为：

$$\begin{aligned} H(Y, r) &= -\sum_{a=0}^1 p(a) \log_2 \varphi(a) \\ &= -\log_2 r - \sum_{a=0}^1 p(a) \log_2 p(a) \\ &= -\log_2 r + H(Y) \end{aligned} \quad (3-3)$$

根据定义 3.1， $r$  确定后，通过加权概率算术编码后  $V$  的长度为  $nH(Y, r)$  (bit)。于是加权概率算术无损编码最小极限为：

$$\begin{aligned} H(Y, r_{\max}) &= -\sum_{y_i=0}^1 p(y_i) \log_2 \varphi(y_i) \\ &= -\log_2 r_{\max} - \sum_{y_i=0}^1 p(y_i) \log_2 p(y_i) \\ &= -\log_2 r_{\max} + H(Y) \end{aligned} \quad (3-4)$$

**证明** 根据定理 2.2， $r_{\max}$  是加权概率算术无损编码的最大值，且  $r_{\max} > 1$ 。当  $r > r_{\max}$ ， $V$  无法还原序列  $Y$ ，所以  $H(Y, r_{\max})$  为加权概率算术无损编码最小极限。

#### 3.2 编码码率

序列  $Y$  中平均每个比特所携带的信息量为  $H(Y, r_{\max})$  (bit/符号)，总信息量为  $lH(Y, r_{\max})$  (bit)。信源序列  $X$  的总信息量为  $nH(X)$  (bit)，可得加权概率算术编码的编码率：

$$R = \frac{nH(X)}{lH(Y, r_{\max})} \quad (3-5)$$

设长度为 $n$ 的二进制信源序列 $X$ 中符号 0 的概率为 $q$  ( $0 \leq q \leq 1$ )。根据(3-1),  $nH(X) = -qn \log_2 q - (1-q)n \log_2(1-q)$ 。序列 $Y$ 的长度为 $l = (2-q)n$ , 则 $\frac{n}{l} = \frac{1}{2-q}$ 。

**定理 3.2** 当 $\varphi(0) = \frac{1}{2}, \varphi(1) = 1$ 时, 加权概率算术编码率 $R = 1$ 。

**证明**  $q = \frac{1}{2}$ 时 $nH(X) = n$  (bit)。根据定理 2.3 有  $H(Y, \varphi(0) = \frac{1}{2}, \varphi(1) = 1) = -\frac{1}{2-q} \log_2 \frac{1}{2} = \frac{1}{2-q}$ 。于是 $lH(Y, \varphi(0) = \frac{1}{2}, \varphi(1) = 1) = \frac{(2-q)n}{2-q} = n$  (bit)。由(3-5)可得:

$$R = \frac{nH(X)}{lH(Y, \varphi(0) = \frac{1}{2}, \varphi(1) = 1)} = 1$$

序列 $Y$ 中符号 0 和符号 1 的概率质量函数 $p(0) = p = \frac{1}{2-q}$ ,  $p(1) = 1-p = \frac{1-q}{2-q}$ , 且 $r_{\max} = \frac{\sqrt{4/p-3}-1}{2-2p}$ 。

$$\begin{aligned} lH(Y, r_{\max}) &= -n \log_2 \frac{\sqrt{5-4q}-1}{2-2q} - (1-q)n \log_2 \frac{\sqrt{5-4q}-1}{2} \\ &= (2-q)n \log_2 \frac{2}{\sqrt{5-4q}-1} + n \log_2(1-q) \end{aligned} \quad (3-6)$$

**定理 3.3** 当 $\varphi(0) = r_{\max}p, \varphi(1) = r_{\max}(1-p)$ 时, 加权概率算术编码率可达 1。

**证明** 根据 (3-6),  $lH(Y, r_{\max}) - nH(X) = (2-q)n \log_2 \frac{2}{\sqrt{5-4q}-1} + (2-q)n \log_2(1-q) + qn \log_2 q = (2-q)n \log_2 \frac{2-2q}{\sqrt{5-4q}-1} + qn \log_2 q = n \log_2 \left( \left( \frac{2-2q}{\sqrt{5-4q}-1} \right)^{2-q} * q^q \right)$ 。因 $0 \leq q \leq 1$ , 所以 $4(1-q)^2 \geq 0$ , 则 $4-8q+4q^2 \geq 0$ 。因 $4-8q+4q^2 = (3-2q)^2 - (5-4q) \geq 0$ , 所以 $2-2q \geq \sqrt{5-4q}-1$ 。因 $(2-2q)^2 \geq (\sqrt{5-4q}-1)^2$ , 可得 $2q-2q^2 \leq \sqrt{5-4q}-1$ , 则 $\frac{\sqrt{5-4q}-1}{2-2q} \geq q$ 。因为 $\left( \frac{2-2q}{\sqrt{5-4q}-1} \right)^{2-q} * p^p = \left( \frac{\sqrt{5-4q}-1}{2-2q} \right)^{q-2} * q^q \geq q^{2q-2} = \left( \frac{1}{q} \right)^{2-2q}$ 且 $2-2q \geq 0, \frac{1}{q} \geq 1$ , 所以 $\left( \frac{1}{q} \right)^{2-2q} \geq 1$ , 即 $lH(Y, r_{\max}) - nH(X) \geq 0$ , 可得 $R = \frac{nH(X)}{lH(Y, r_{\max})} \leq 1$ , 所以加权概率算术编码率可达 1。

### 3.3 平均译码错误概率

设事件 $E$ 表示满足(1-1)的序列 $Q$ 的集合, 事件 $E$ 有 $f(l)$ 个序列 $Y$ 。当 $l=1$ 时,  $E = (0,1)$ ,  $f(l=1) = 2$ , 互补事件为 $\bar{E} = \emptyset$ 。当 $l=2$ 时,  $E = (00,01,10)$ ,  $f(l=2) = 3$ ,  $\bar{E} = (11)$ 。当 $l=3$ 时,  $E = (000,001,010,100,101)$ ,  $f(l=3) = 5$ ,  $\bar{E} = (011,110,111)$ 。当 $l \geq 3$ 时:

$$f(l) = f(l-1) + f(l-2) \quad (3-7)$$

可得事件 $E$ 的概率为:

$$P(Q \in E) = \frac{f(l)}{2^l} \quad (3-8)$$

令事件 $E$ 中 $f(l)$ 个序列 $Q$ 服从均匀分布, 则:

$$P(Q = Y) = \frac{1}{f(l)}, P(Q \neq Y) = \frac{f(l)-1}{f(l)} \quad (3-9)$$

于是,  $Q \in E$ 且 $Q \neq Y$ 的概率为:

$$P(Q \neq Y | Q \in E) = P(Q \in E)P(Q \neq Y) = \frac{f(l)-1}{2^l} \quad (3-10)$$

$P(Q \neq Y | Q \in E)$ 为平均译码错误概率, 于是 $P_{err} = P(Q \neq Y | Q \in E)$ 。

**定理 3.4**  $\lim_{l \rightarrow \infty} P_{err} = \lim_{l \rightarrow \infty} P(Q \neq Y | Q \in E) = 0$ 。

**证明**  $l \rightarrow \infty$ 则 $P(Y \neq Q) \rightarrow 1$ , 得出 $P(Y \neq Q | Y \in E) \rightarrow P(E)$ 。根据斐波那契数列<sup>[24]</sup>, 令 $F(0) = 0, F(1) = 1$ , 且当 $l \geq 2, l \in N^*$ 时

$F(l) = F(l-1) + F(l-2)$ 。于是 $l \geq 1, l \in N^*$ 时,  $f(l) = F(l) + F(l+1)$ 。由斐波那契数列通项式得:

$$f(l) = \frac{1}{\sqrt{5}} \left( \left[ \left( \frac{1+\sqrt{5}}{2} \right)^l - \left( \frac{1-\sqrt{5}}{2} \right)^l \right] + \left[ \left( \frac{1+\sqrt{5}}{2} \right)^{l+1} - \left( \frac{1-\sqrt{5}}{2} \right)^{l+1} \right] \right)$$

可得:

$$P(Q \in E) = \frac{1}{\sqrt{5}} \left( \left[ \left( \frac{1+\sqrt{5}}{4} \right)^l - \left( \frac{1-\sqrt{5}}{4} \right)^l \right] + 2 \left[ \left( \frac{1+\sqrt{5}}{4} \right)^{l+1} - \left( \frac{1-\sqrt{5}}{4} \right)^{l+1} \right] \right)$$

因 $\frac{1+\sqrt{5}}{4} < 1$ ,  $\frac{1-\sqrt{5}}{4} < 1$ , 所以 $l \rightarrow \infty$ 时 $P(E) \rightarrow 0$ , 即 $\lim_{l \rightarrow \infty} P(Q \neq Y|Q \in E) = 0$ 。

设事件 $F$ 表示满足(1-2)的序列 $Q$ 的集合, 事件 $F$ 有 $g(l)$ 个序列 $Q$ 。当 $l=1$ 时,  $F = (0,1)$ ,  $g(l=1) = 2$ , 互补事件为 $\bar{F} = \emptyset$ 。当 $l=2$ 时,  $F = (01,10,11)$ ,  $g(l=2) = 3$ ,  $\bar{F} = (00)$ 。当 $l=3$ 时,  $F = (010,101,011,110)$ ,  $f(l=3) = 4$ ,  $\bar{F} = (000,001,100,111)$ 。当 $l \geq 4$ 时:

$$g(l) = g(l-2) + g(l-3) \quad (3-11)$$

$P(Z \in F) = \frac{g(l)}{2^l}$ , 得:

$$P(Q \neq Z|Q \in F) = P(F)P(Q \neq Z) = \frac{g(l)-1}{2^l} \quad (3-12)$$

$f(l)$ 和 $g(l)$ 均为单调递增, 且 $g(l) \leq f(l)$ , 即 $\frac{g(l)-1}{2^l} \leq \frac{f(l)-1}{2^l}$ 。根据定理 3.4,  $l \rightarrow \infty$ 时 $\frac{f(l)-1}{2^l} \rightarrow 0$ ,  $P_{err} = P(Q \neq Z|Q \in F)$ , 可得 $\lim_{l \rightarrow \infty} P_{err} = 0$ 。根据(3-10)和(3-12)可计算出 $P_{err}$ 如表 2。

表 2 根据 $l$ 计算 $P(Q \neq Y|Q \in E)$ 和 $P(Q \neq Z|Q \in F)$

$l$ (bit)	$P(Q \neq Y Q \in E)$	$P(Q \neq Z Q \in F)$
20	0.016890526	0.000443459
64	$1.50584 \times 10^{-6}$	$5.9561 \times 10^{-12}$
112	$5.75104 \times 10^{-11}$	$1.53974 \times 10^{-20}$
256	$3.20367 \times 10^{-24}$	$2.66011 \times 10^{-46}$

根据 2.1 节编码过程, 接收端已知 $V$ ,  $c$ 和 $n$ 。定理 3.3,  $lH(Y, r_{max}) = m$ ,  $m$ 为序列 $V$ 的比特长度, 序列 $Y$ 中符号 0 的概率 $p = \frac{n}{2n-c}$ 且 $r_{max} = \frac{\sqrt{4/p-3}-1}{2-2p}$ , 所以 $H(Y, r_{max})$ 已知, 于是:

$$l = \frac{m}{H(Y, r_{max})} \quad (3-13)$$

将 $l$ 代入(3-10)和(3-12)可得 $P(Q \neq Y|Q \in E)$ 和 $P(Q \neq Z|Q \in F)$ , 当 $l$ 的值确定时 $P(Q \neq Y|Q \in E)$ 和 $P(Q \neq Z|Q \in F)$ 确定, 此时 $n$ 的值无法确定。

### 3.5 信道容量可达

根据图 1, 序列 $Y$ 经加权概率算术编码得二进制序列 $V$ , 且 DMC 信道的传输速率 $I(V; U) = H(V) - H(V|U)$ 。于是 $R_{BSC} = H(V) - H(\xi)$ ,  $R_{BEC} = H(V) - \varepsilon$ 。

根据定理 3.2, 因编码码率 $R = 1$ , 所以 $lH(Y, \varphi(0) = \frac{1}{2}, \varphi(1) = 1) = nH(X) = m$  (bit), 于是 $H(X) = \frac{m}{n}$ 。序列 $X$ 中符号概率均等, 所以 $H(X) = 1$ , 则 $m = n$ 。根据定理 2.3 序列 $V$ 可无损译码序列 $Y$ , 且序列 $Y$ 中将“10”替换为“1”可无损还原序列 $X$ , 所以序列 $V$ 可无损还原序列 $X$ 。 $0 \leq H(V) \leq 1$ , 假设 $H(V) < 1$ , 则 $mH(V) < nH(X)$ , 不符合无失真信源编码定理<sup>[28][29]</sup>, 仅 $mH(V) \geq nH(X)$ 时方可无损译码, 所以 $H(V) = 1$ , 即 $V$ 中符号概率均等。于是 $\varphi(0) = \frac{1}{2}, \varphi(1) = 1$ 时 $R_{BSC} = 1 - H(\xi) = C_{BSC}$ ,  $R_{BEC} = 1 - \varepsilon = C_{BEC}$ 。

根据定理 3.3 和定理 2.2, 因为 $R \leq 1$ , 所以 $lH(Y, r_{max}) = m \geq nH(X)$ , 即 $m \geq n$ 。当 $m > n$ , 序列 $V$ 中存在冗余信息。此时, 将序列 $V$ 再进行无损编码得到序列 $V'$ , 随后将 $V'$ 经 DMC 信道传输,  $U'$ 为接收到的二进制序列。接收端首先将 $U'$ 译码出 $U$ , 然后将 $U$ 译码出序列 $Q$ , 若 $Q$ 不满足(1-1), 则 $Q \neq Y$ , 可根据 2.3 和 2.4 节方法对 $U'$ 进行前向纠错译码。因为 $V'$ 可无损译码 $X$ , 所以 $H(V') = H(X) = 1$ , 即 $V'$ 中符号概率均等。于是 DMC 信道的传输速率 $I(V'; U') = H(V') - H(V'|U')$ , 则 $R_{BSC} = 1 - H(\xi) = C_{BSC}$ ,  $R_{BEC} = 1 - \varepsilon = C_{BEC}$ 。

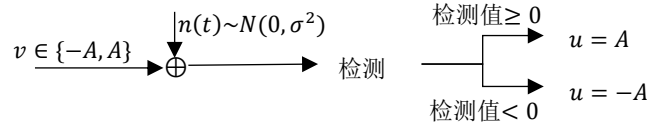
## 4 仿真实验设计和实验结果



#### 4.1 BPSK 信号 BSC( $\xi$ )信道

设 BSC( $\xi$ )信道分析模型如图 3,  $\sigma^2$  为高斯白噪声双边功率谱密度,  $\sigma^2 = N_0/2$ ,  $A = \sqrt{E_s}$ ,  $E_s = RE_b$ ,  $R$  为编码码率,  $E_b$  为信息比特能量,  $N_0$  是噪声单边功率谱密度,  $E_b/N_0$  为信噪比。

图 3 BSC( $\xi$ )信道分析模型



判决门限为 0 时判决错误概率  $\xi$  最小:

$$\xi = \frac{1}{2} \left( 1 - \operatorname{erf} \left( \sqrt{\frac{RE_b}{N_0}} \right) \right) \quad (4-1)$$

其中  $\operatorname{erf}$  为高斯误差函数:

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy \quad (4-2)$$

于是:

$$C_{BSC} = 1 - H(\xi) = 1 + \xi \log_2 \xi + (1 - \xi) \log_2 (1 - \xi) \quad (4-3)$$

将(4-1)代入(4-3)可得  $C_{BSC}$ ,  $R$  和  $E_b/N_0$  的方程式。

设  $\tau = m$ , 序列  $U$  经 2.3 节纠错译码后仍存在误比特率  $P_e$ , 因为  $Q \in E$  且  $Q \neq Y$  时本文算法无法进行错误校验, 使得部分错误无法纠正, 所以  $P_e = P(Q \neq Y | Q \in E)$  或  $P_e = P(Q \neq Z | Q \in F)$ , 即  $P_e = P_{err}$ 。于是:

$$R' = R(1 - H(P_{err})) \quad (4-4)$$

根据定理 3.4 有  $\lim_{l \rightarrow \infty} P_{err} = 0$ , 于是  $\lim_{l \rightarrow \infty} H(P_{err}) = \lim_{l \rightarrow \infty} (-P_{err} \log_2 P_{err} - (1 - P_{err}) \log_2 (1 - P_{err})) = 0$ , 可得  $l \rightarrow \infty$  时  $R' = R$ 。当  $\tau < m$ , 序列  $U$  中  $\tau < e \leq m$  ( $e$  为序列  $U$  中实际错误比特个数) 的概率为:

$$P(\tau < e \leq m) = \sum_{i=\tau+1}^m C_m^i \xi^i (1 - \xi)^{m-i} \quad (4-5)$$

设  $\tau < m$ , 当  $l$  为已知确定值或者  $\tau < m$  时使得部分错误无法纠正, 所以  $P_e = P_{err} + P(\tau < e \leq m) - P_{err} P(\tau < e \leq m)$ ,  $R' = R(1 - H(P_e))$ , 因  $l \rightarrow \infty$  则  $m \rightarrow \infty$ , 于是  $\lim_{l \rightarrow \infty} P_e = \sum_{i=\tau+1}^m C_m^i \xi^i (1 - \xi)^{m-i}$ 。显然,  $P(\tau < e \leq m)$  可根据信噪比 SNR(dB) 计算求得, 设  $m = 32$ ,  $R = 0.5$ ,  $\tau = 6$ , SNR = 1, 2, ..., 6 (dB) 代入(4-6)得  $E_b/N_0$ , 随后将  $E_b/N_0$  代入(4-1)(4-5)得表 3。

$$\text{SNR} = 10 \log_{10} \frac{E_b}{N_0} \quad (4-6)$$

表 3 由 SNR(dB) 计算出  $P(\tau < e \leq m)$

SNR(dB)	$P(\tau < e \leq m)$
1	0.1168603691
2	0.0435367147
3	0.0114630033
4	0.0017649586
5	0.0001578748
6	0.0000069219

当  $P_e = 0$  时, 将(4-1)代入  $R = 1 + \xi \log_2 \xi + (1 - \xi) \log_2 (1 - \xi)$  中可得  $R$  和  $E_b/N_0$  的方程, 此时信息无误传输所需的信噪比最小, 记为  $(E_b/N_0)_{\min}$ ,  $(E_b/N_0)_{\min}$  为香农限。当  $P_e > 0$  时, 将(4-1)和(4-5)代入  $R(1 - H(P_e)) = 1 + \xi \log_2 \xi + (1 - \xi) \log_2 (1 - \xi)$ , 可求得  $(E_b/N_0)_{\min}$ 。因信源处理方法和加权概率算术编码已知, 所以  $R$  和  $P_{err}$  已知,  $(E_b/N_0)_{\min}$  可直接求解, 如表 3 和表 4。

表 4 BPSK 信号 BSC( $\xi$ )信道当  $P_e = P(Q \neq Y | Q \in E)$  时最小的 SNR(dB)

$R \backslash P_e$	0	$3.20367 \times 10^{-24}$	$5.75104 \times 10^{-11}$	$1.50584 \times 10^{-6}$	0.001327807
0.01	0.3910	0.3910	0.3910	0.3909	0.3269
0.3	1.1129	1.1129	1.1129	1.1128	1.0365
0.5	1.7725	1.7725	1.7725	1.7724	1.6819

0.8	3.3701	3.3701	3.3701	3.3698	3.2154
0.99	6.9572	6.9572	6.9572	6.9549	6.0335
表 5 BPSK 信号 BSC( $\xi$ )信道当 $P_e = P(Q \neq Z Q \in F)$ 时最小的 SNR(dB)					
$P_e \backslash R$	0	$2.66011 * 10^{-46}$	$1.53974 * 10^{-20}$	$5.9561 * 10^{-12}$	$3.16184 * 10^{-6}$
0.01	0.3910	0.3910	0.3910	0.3919	0.3909
0.3	1.1129	1.1129	1.1129	1.1130	1.1128
0.5	1.7725	1.7725	1.7725	1.7726	1.7724
0.630929	2.3335	2.3335	2.3335	2.3335	2.3330

序列 $Y$ 经加权概率算术编码, 根据定理 2.3 和(3-4), 当 $\varphi(1) = 1$ 且 $\varphi(0) \rightarrow \frac{1}{2}$ 时 $R \rightarrow 1$ , 当 $\varphi(1) = 1$ 且 $\varphi(0) \rightarrow 0$ 时 $R \rightarrow 0$ 。序列 $Z$ 经加权概率算术编码, 根据定理 2.3 和(3-4), 当 $\varphi(1) = 1$ 且 $\varphi(0) \leq \frac{1}{3}$ 时可无损编译码, 于是:

$$0 < R \leq 1 / -\log_2 \frac{1}{3} \quad (4-7)$$

因 $\lim_{t \rightarrow \infty} P_{err} = 0$ , 所以信息无误传输所需的信噪比可达香农限。

## 4.2 仿真编码码率和数据校验代码

仿真实验采用1/2编码码率, 可选择 $Y$ 和 $Z$ 。当 $\varphi(1) = 1$ ,  $\varphi(0) = \frac{1}{4}$ 则 $R = \frac{1}{2}$ 。因序列 $Z$ 比 $Y$ 多一个数据校验条件, 所以本文仿真实验选择 $Z$ , 并采用 $\varphi(1) = 1$ ,  $\varphi(0) = \frac{1}{4}$ 进行加权概率算术编码。

### (1) 编码

---

Algorithm(1): 基于 $Z$ 的加权概率算术编码

---

输入: 长度为 $n$ 的序列 $X$ 的比特数组 $XBitArray$

---

输出: 序列 $V$ 的比特数组 $VBitArray$

---

```

1:  $R_0 \leftarrow 1; L_0 \leftarrow 0;$ 
2:  $\varphi(0) \leftarrow \frac{1}{4}; \varphi(1) \leftarrow 1;$ 
3: for  $i \leftarrow 1$  to  $n$ 
4:   if  $XBitArray[i - 1] = 0$  then
5:      $R_i = R_{i-1} \varphi(0) \varphi(1);$ 
6:      $L_i = L_{i-1} + R_{i-1} \varphi(0)^2;$ 
7:   else
8:      $R_i = R_{i-1} \varphi(0) \varphi(1)^2;$ 
9:      $L_i = L_{i-1} + R_{i-1} \varphi(0)^2 \varphi(1);$ 
10:  end if
11: end for
12:  $VBitArray \leftarrow L_n;$ 
13: return  $VBitArray;$ 
```

---

Algorithm(1)中 $VBitArray \leftarrow L_n$ 是将实数 $L_n$ 经进制转换为二进制序列。

### (2) 数据校验译码

---

Algorithm(2): 基于 $Z$ 的错误校验译码

---

输入: 序列 $U$ 的比特数组 $UBitArray$ 和 $c$

---

输出: 序列 $Z$ 的比特数组 $ZBitArray$ 或null

---

```

1:  $R_0 \leftarrow 1; L_0 \leftarrow 0; i \leftarrow 1; j \leftarrow 1; H \leftarrow 0; s_1 = s_2 \leftarrow -1;$ 
2:  $U \leftarrow UBitArray;$ 
3:  $\varphi(0) \leftarrow \frac{1}{4}; \varphi(1) \leftarrow 1;$ 
4: while  $i < n$ 
5:    $H \leftarrow L_{j-1} + r p(0) R_{j-1};$ 
```

---

---

```

6:   if  $U < H$  then
7:       if  $s_2 = s_1 = -1$  then
8:            $s_1 \leftarrow 0$ ;
9:            $R_j = R_{j-1}\varphi(0)$ ;
10:      else if  $s_2 = -1$  and  $s_1 = 1$  then
11:           $s_2 \leftarrow 1$ ;  $s_1 \leftarrow 0$ ;
12:           $R_j = R_{j-1}\varphi(0)$ ;
13:      else if  $s_2 = 1$  and  $s_1 = 1$  then
14:          return null;
15:      else if  $s_1 = 0$  then
16:          return null;
17:      end if
18:       $j \leftarrow j + 1$ ;
19:  else
20:      if  $s_2 = s_1 = -1$  then
21:           $R_j \leftarrow R_{j-1}\varphi(1)$ ;
22:           $L_j = L_{j-1} + R_{j-1}\varphi(0)$ ;
23:      else if  $s_2 = -1$  and  $s_1 = 0$  then
24:           $R_j \leftarrow R_{j-1}\varphi(1)$ ;
25:           $L_j = L_{j-1} + R_{j-1}\varphi(0)$ ;
26:           $ZBitArray[i] = 0$ ;
27:           $i \leftarrow i + 1$ ;
28:           $s_2 \leftarrow -1$ ;  $s_1 \leftarrow -1$ ;
29:      else if  $s_2 = -1$  and  $s_1 = 1$  then
30:           $R_j \leftarrow R_{j-1}\varphi(1)$ ;
31:           $L_j = L_{j-1} + R_{j-1}\varphi(0)$ ;
32:           $s_2 \leftarrow 1$ ;  $s_1 \leftarrow 1$ ;
33:      else if  $s_2 = 1$  and  $s_1 = 0$  then
34:           $R_j \leftarrow R_{j-1}\varphi(1)$ ;
35:           $L_j = L_{j-1} + R_{j-1}\varphi(0)$ ;
36:           $ZBitArray[i] = 1$ ;
37:           $i \leftarrow i + 1$ ;
38:           $s_2 \leftarrow -1$ ;  $s_1 \leftarrow -1$ ;
39:      else if  $s_2 = 1$  and  $s_1 = 1$  then
40:          return null;
41:      end if
42:       $j \leftarrow j + 1$ ;
43:  end if
44: end while
45: return  $ZBitArray$ ;

```

---

当 $ZBitArray = \text{null}$ 时说明 $U$ 错误，采用 2.3 或 2.4 节方法实现前向纠错译码。

### 4.3 BPSK 信号 BSC( $\xi$ )信道仿真实验

随机生成 $n = 32$ 位的二进制信源序列 $X$ ， $X$ 中符号概率均等，将 $X$ 中符号 1 替换为“101”且符号 0 替换为“01”得到序列 $Z$ ，使得序列 $Z$ 满足(1-2)，则序列 $Z$ 的比特长度为 $l = 80$ 。实验中仅需将 $X$ 代入Algorithm(1)编码。

$l = 80$ 时，根据(3-11)(3-12)可直接求解出 $P_{err} = 8.17441 \times 10^{-15}$ 。设 $\varphi(0) = \frac{1}{4}$ ， $\varphi(1) = 1$ ，于是序列 $Z$ 中符号 1 不携带信息量，序列 $Z$ 经加权概率模型编码。于是序列 $V$ 的长度为 $m = lH\left(Z, \varphi(0) = \frac{1}{4}, \varphi(1) = 1\right) = -32 \log_2 \frac{1}{4} - 48 \log_2 1 = 64$  (bit)，编码

码率  $R = 0.5$ 。

设  $\tau = 4, 6, 8, 12, 16$ ,  $SNR = 0, 1, 2, \dots, 8$  (dB), 因  $m = 64$  且  $P_{err} = 8.17441 \times 10^{-15}$ , 所以可根据(4-1)(4-2)(4-5)(4-6)计算出  $P_e = P_{err} + P(\tau < e \leq m) - P_{err}P(\tau < e \leq m)$ ,  $P_e$  可求解, 如图 3 所示。

图 3 基于  $E_b/N_0$  和  $\tau$  计算得出  $P_e$

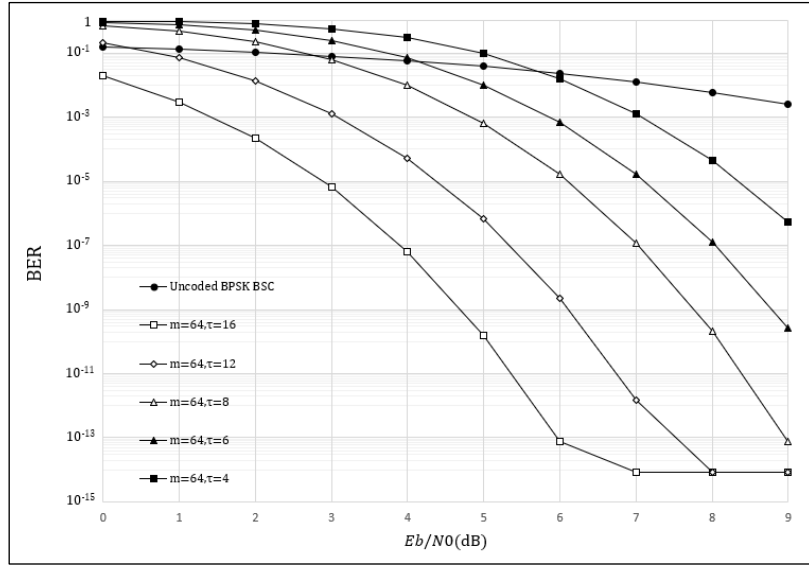


图 3, 当  $\tau = 12, 16$  且  $SNR = 8$  dB 时  $P(\tau < e \leq m) = 1.80973 \times 10^{-23}$ ,  $P_{err} = 8.17441 \times 10^{-15}$ , 于是  $P_e \approx P_{err}$ 。当  $\tau = 12$  且  $SNR = 4$  dB 时本文方法性能可达  $10^{-5}$  量级。当  $\tau = 16$  且  $SNR = 4$  dB 时本文方法性能可达  $10^{-8}$  量级。 $\tau = 16$  时, 基于 2.3 节纠错译码最多需进行  $\sum_{i=1}^{16} C_{64}^i = 713250450657108$  次校验运算,  $m$  和  $\tau$  越大校验运算的次数越多。

仿真实验  $\tau = 4, 6, 8$ ,  $m = 64$ ,  $SNR = 5, 6, 7, 8$  (dB), 帧数大于  $10^5$ , 得出仿真结果与理论计算如图 4 所示。

图 4 仿真测试和理论计算比较

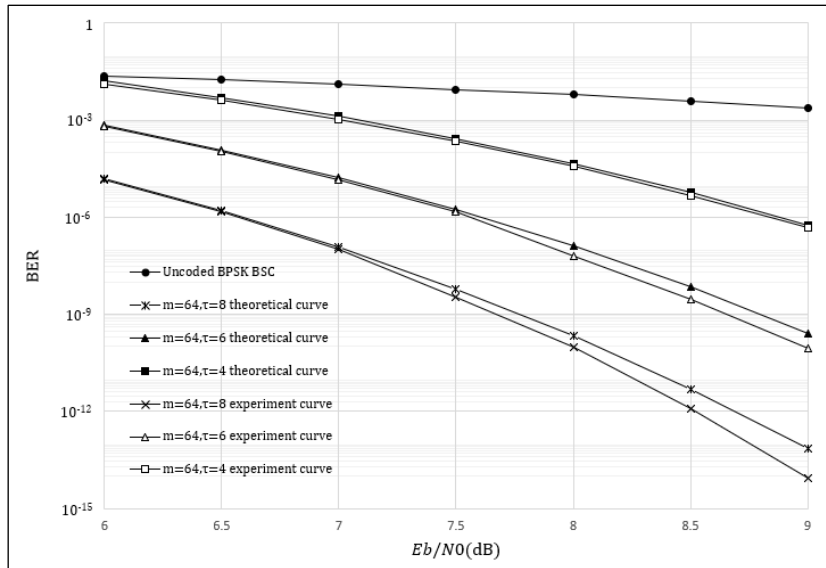


图 4 中仿真性能基本符合理论计算。

## 5 结束语

本文提出了一种新的信道检错纠错思路, 并给出具体方法。因方法种类众多, 本文举例了两种方法, 分别对应于不同的码率和译码错误概率, 通过证明可达到信道容量。方法简单, 易于软硬件实现。可自适应于信道的干扰情况, 通过增大  $\tau$  的值提高纠错率。控制  $\tau$  的大小后, 可构造前向纠错与数据校验重传一体的信道编码方法。 $\tau$  越大则运算量成指数增长, 因比特错误位置与数据校验位置存在关联, 未来可根据误比特位置细化纠错范围从而提高运算效率。实验基于 C 语言开发仿真程序, 源代码已经发布在 GitHub 网站, 下载地址为: <https://github.com/Jielin-Code/Wj1ChannelCodeAlgorithm>。

## 参考文献:

- [1] Erdal Arkan. Channel Polarization: A Method for Constructing Capacity-Achieving Codes for Symmetric Binary-Input Memoryless Channels. IEEE Transactions on Information Theory, Volume:55 , Issue:7, 3051 - 3073, July 2009.
- [2] R. G. Gallager, "Low-Density Parity-Check Codes," M.I.T. Press, Cambridge, Massachusetts, 1963.
- [3] Axel Huebner, Kamil Sh. Zigangirov, Daniel J. Costello. Laminated turbo codes: A new class of block-convolutional codes. IEEE Transactions on Information Theory, Volume: 54 , Issue: 7 , July 2008.
- [4] Ian H. Witten, Radford M. Neal, John G. Cleary. Arithmetic Coding for Data Compression. Communications of the ACM. 1987, 30(6):520~539.
- [5] G. N. N. Martin, Range encoding: an algorithm for removing redundancy from a digitised message. Video & Data Recording Conference, held in Southampton July 24-27 1979.
- [6] 夏志进, 杨铭, 崔慧娟, 等. 多禁止符号算术编码高效错误检测算法[J]. 清华大学学报:自然科学版, 2005, 045(007):935-938.
- [7] 徐向明, 彭坦, 崔慧娟, 等. 带禁止符号的算术码序列译码算法[J]. 通信技术, 2009(04):154-155.
- [8] Auli-Llinas F . Context-Adaptive Binary Arithmetic Coding With Fixed-Length Codewords[J]. IEEE Transactions on Multimedia, 2015, 17(8):1-1.
- [9] 卓德珊, 傅晨, 冯冠雄. 一种低延迟的 HEVC 熵编码方法设计[J]. 广播电视网络, 2020, v.27;No.363(03):103-108.
- [10] 刘轩黄. 关于随机过程一阶概率分布函数的遍历性[J]. 大学数学, 1989.
- [11] Cao X P . Deconvolution of Cumulative Distribution Function with Unknown Noise Distribution[J]. Acta Applicandae Mathematicae, 2020, 170(1):483-514.
- [12] 隋厚棠. 二进对称信道的最佳帧同步和性能[J]. 电子与信息学报, 1992, 14(2):127-133.
- [13] 胡恒铭, 刘尉悦. 基于二进制对称信道的极化码研究[J]. 计算机工程, 2019, 45(04):84-87.
- [14] 王继伟, 王学东, 李斌, 等. 极化码在 BEC 信道下性质研究[J]. 通信技术, 2012, 000(009):33-35.
- [15] 黄志亮, 张施怡, 周水红. BEC 信道下高维核矩阵极化码的精确设计[J]. 无线电通信技术, 2019, 45(003):271-275.
- [16] 杨义先. 纠错编码的译码错误概率[J]. 通信技术, 1992, 000(002):9-12.
- [17] 江宝安. 基于循环长除法的 BCH(31,16)译码算法[J]. 通信与信息技术, 2020, No.243(01):59-62.
- [18] Harrison R R , Kim W J . Analog MAP Decoder for (8, 4) Hamming Code in Subthreshold CMOS. IEEE Computer Society, 2001.
- [19] 鲁芳旭, 刘翠海. RS 码的性能分析与仿真[J]. 数字技术与应用, 2020(8):25-27.
- [20] 王晓玲. CRC 交织级联编码在 QPSK 通信系统中的应用研究[J]. 微电子学与计算机, 2018, v.35; No.408(05):90-94.
- [21] 何善宝, 赵春明, 姜明. LDPC 码的一种循环差集构造方法[J]. 通信学报, 2004, 25(011):112-118.
- [22] 梁宇, 杨卫华, 李玉瑛. 一种基于有限域的二进制和多进制空间耦合 LDPC 码构造方法[J]. 应用数学进展, 2021, 10(2).
- [23] 邵璐. Turbo 编码 SC-FDMA 技术研究[J]. 数字通信世界, 2020, No.190(10):13-14+43.
- [24] 王东, 李秀朋. Turbo 译码器基于组合逻辑电路的低复杂度 Log-MAP 算法[J]. 无线电通信技术, 2018, 044(003):263-267.
- [25] 李斌, 王学东, 王继伟. Theory and Application of Polar Code[J]. 通信技术, 2012, 000(010):21-23.
- [26] Oliveira R M , Lamare R . Design of Rate-Compatible Polar Codes Based on Non-Uniform Channel Polarization[J]. IEEE Access, 2021, PP(99):1-1.
- [27] Wu W , Siegel P H . Generalized Partial Orders for Polar Code Bit-Channels[J]. IEEE Transactions on Information Theory, 2019, 65(11):7114-7130.
- [28] C. E. Shannon. A mathematical theory of communication. Bell Syst. Tech. J., 27:379-423, 623-656, 1948.
- [29] 周炯崧, 丁晓明. 信源编码原理[M]. 人民邮电出版社, 1996.
- [30] 陈达, 余曼, 戴美权, 等. 基于可变比特率编码无环图的图像无损压缩[J]. 控制工程, 2020, v.27;No.185(05):44-50.