文章编号：1672-2892(2011)05-0591-05

二进制 BCH 码的一种盲识别方法

王甲峰，岳 旸，权友波

(中国工程物理研究院 电子工程研究所，四川 绵阳 621900)

摘 要：提出二进制 BCH 码的一种盲识别方法。该算法适用于本原和非本原二进制 BCH 码。首先，在帧长度已知的条件下，根据循环特性，给出一种分组长度的统计识别方法；然后，根据循环特性及各种约束条件得到备选多项式；再根据校正子权重和最小原则，得到最优多项式；最后通过因式分解得到生成多项式的最终估计表达式。仿真表明，本文算法具有较强的抗随机误码能力，而且其识别性能随着参加统计的码字数增多而提高。该算法不涉及矩阵运算，因此非常适合硬件实现。

关键词：二进制 BCH 码；盲识别；分组长度；生成多项式

中图分类号：TN911 **文献标识码**：A

Blind recognition method of binary BCH code

WANG Jia-feng, YUE Yang, QUAN You-bo

(Institute of Electronic Engineering, China Academy of Engineering Physics, Mianyang Sichuan 621900, China)

Abstract: A blind recognition of the binary BCH code was proposed. The method was applied to both primitive and non-primitive binary BCH codes. A statistic recognition method was first proposed based on the cyclic feature under the condition of having known the frame length. And then candidate polynomials were achieved based on the cyclic feature and other restrictions. From the candidate polynomials, the best polynomial was selected according to the minimum rule of the weights sum of the syndromes. Finally, the best polynomial was factorized to get the recognition result of the generator polynomial. The simulation results show that the method possesses the capability of anti-random bit error, and the algorithm involved in this method is very simple, so it is suitable to be used in practice.

Key words: binary BCH code; blind recognition; code length; generator polynomial

随着通信对抗技术的发展，对抗的手段逐渐由信号层深入到信息层；而信道编码的盲识别是获取通信原始信息的前提和基础，已经成为通信对抗领域的一项关键技术，从而日益受到重视。信道编码可分为线性分组码和卷积码两大类。从公开的资料来看，目前的研究主要集中于卷积码的盲识别[1-6]，而涉及线性分组码盲识别的公开文献很少。文献[7]给出了一种针对低码率线性分组码的盲识别方法，其基本思路是通过解方程识别校验矩阵，然后得到生成矩阵；文献[8]讨论了一种本原 RS 码的盲识别方法，其基本思路是通过基于矩阵行向量化简(Reduced Row Echelon Form of the matrix，RREF)、容错矩阵分解(Fault-Tolerant Matrix Decomposing，FTMD)和有限域傅里叶变换(Galois Field Fourier Transform，GFFT)识别编码参数及生成矩阵。这两篇文献所述方法都涉及有限域的矩阵运算，在长码情况下所需内存较大，对平台的要求较高，而且不适于硬件实现，误码性能也不够理想。本文针对二进制 BCH 码这一循环码的重要子类，在帧长度已知的条件下，根据其循环特性，提出一种统计识别方法，原理简单，可以获得较好的误码性能，而且不需要矩阵运算，非常适合硬件实现。

1 二进制 BCH 码的原理

对于 1 个定义在 GF(2)上的分组长度为 n，消息长度为 k 的(n,k)BCH 码[9-10]，设 $m=(m_0,m_1,\cdots,m_{k-1})$ 为编码前的消息字，$c=(c_0,c_1,\cdots,c_{n-1})$ 为编码后的码字，由于是循环码，消息字和码字分别对应 1 个 GF(2)上的消息多项式和码多项式：

收稿日期：2010-08-31；修回日期：2010-11-22

$$m(x)=m_0x^{k-1}+m_1x^{k-2}+\cdots+m_{k-2}x+m_{k-1} \tag{1}$$

$$c(x)=c_0x^{n-1}+c_1x^{k-2}+\cdots+c_{k-2}x+c_{n-1} \tag{2}$$

由参考文献[9]和[10]可知这个(n,k)二进制 BCH 码有如下特性：

1) 分组长度 n 等于 2^m-1 或为其因数($m\geqslant 3$，为正整数)，n 等于 2^m-1，对应本原二进制 BCH 码，n 为 2^m-1 因数，对应非本原二进制 BCH 码；

2) 码字 c 满足循环特性，即将码字 c 循环左移 j 次后所得到的码字 $c'=(c_{j-1},c_j,\cdots,c_{k-1},c_0,\cdots,c_{j-3},c_{j-2})$，仍然是同一$(n, k)$二进制 BCH 码集合中的码字；

3) $m(x)$ 与 $c(x)$ 之间满足如下关系：

$$c(x)=m(x)g(x) \tag{3}$$

式中 $g(x)$ 是一个定义在 GF(2)上的多项式，称为生成多项式，可表示为：

$$g(x)=x^{n-k}+g_1x^{n-k-1}+\cdots+g_{n-k-1}x+1 \tag{4}$$

4) 存在定义在 GF(2)上的校验多项式 $h(x)$，可表示为：

$$h(x)=x^k+h_1x^{k-1}+\cdots+h_{k-1}x+1 \tag{5}$$

并且有如下关系式：

$$h(x)c(x)=0 \bmod (x^n+1) \tag{6}$$

$$h(x)g(x)=x^n+1 \tag{7}$$

2 分组长度的识别方法

二进制(n, k)BCH 码的识别涉及 3 个参数的识别：分组长度 n，信息码长度 k，以及生成多项式 $g(x)$。但是，由式(4)可知 $g(x)$ 的阶数为 $n-k$，因此只要识别出分组长度 n 及生成多项式 $g(x)$，则 k 值随之确定。因此，只需要识别分组长度 n 及生成多项式 $g(x)$。本节介绍分组长度 n 的识别方法，下一节介绍生成多项式 $g(x)$ 的识别方法。

本文识别算法是建立在帧长度 f_l 已知这一基本假设前提下的。这一假设是合理的，因为在实际工程中，一般帧头是不进行编码的，所以比较容易获得帧长度这一参数。

由于帧长度 f_l 已知，假设 1 帧内至少有 2 个码字，因此可以得到如下 2 个结论：1) n 在 $[3,\left\lceil \frac{f_l}{2} \right\rceil]$ 内取值，其中 $\lceil\cdot\rceil$ 表示上限取整；2) n 整除 f_l，即 n 是 f_l 的 1 个因数。

在 $[3,\left\lceil \frac{f_l}{2} \right\rceil]$ 内 f_l 的因数可能不只 1 个，假设 i 是 f_l 的 1 个因数，则有 2 种情况：

1) $i=n$

此时如果以 i 作为分组长度进行分组，得到 N_i 个码字。设 $c_p(x)$ 为其中第 p 个码字所对应的码多项式，则通过 j 循环左移可以得到 j 个码多项式 $c_{p1}(x), c_{p2}(x),\cdots, c_{pj}(x)$，而 $1\leqslant j\leqslant i-1$。

由于二进制 BCH 码满足循环特性，如果 c_p 中不存在误码，那么 $c_{p1}(x), c_{p2}(x),\cdots, c_{pj}(x)$ 所对应的码字必然与 $c_p(x)$ 所对应的码字同属一个(n, k)二进制 BCH 集合，即其生成多项式是相同的。根据码多项式与生成多项式的关系式(3)可知，$c_p(x), c_{p1}(x), c_{p2}(x),\cdots, c_{pj}(x)$ 之间必然存在公因式。如果设 $c_{p0}(x)=c_p(x)$，则有如下关系存在，

$$\gcd[c_{p0}(x),c_{p1}(x),\cdots,c_{pj}(x)]\neq 1,\ 1\leqslant j\leqslant i-1 \tag{8}$$

称满足式(8)的码字为有效码字，并设在 N_i 个码字中有效码字的个数为 N_{ic}。显然，在无误码的情况下，$N_{ic}=N_i$，即有效码字在所有码字中所占的比例 f_{ic} 为

$$f_{ic}=\frac{N_{ic}}{N_i}=1 \tag{9}$$

2) $i\neq n$

此时，由于没有实现正确分组，则以 i 作为分组长度进行分组所得到 N_i 个码字中必然存在不满足式(8)的码字，因此有 $N_{ic}<N_i$，即

$$f_{ic}=\frac{N_{ic}}{N_i}<1 \tag{10}$$

综上所述，在无误码的条件下，如果以式(8)作为判断码字是否有效的准则，则在所有可能的分组长度中，当 $i=n$ 时，有效码字比例 $f_{ic}=1$；当 $i\neq n$ 时，有效码字比例 $f_{ic}<1$。当然，在存在误码的情况下，即使 $i=n$，f_{ic} 仍然会小于 1；但是，可以预期此时 $i=n$ 使得 f_{ic} 取最大值。

另外由二进制 BCH 码的性质 1)可知 n 值必为奇数。由此可以得到分组长度 n 的识别式为：

$$n=\underset{i,i\in[3,\lceil f_l/2\rceil],rem(i,2)=1,rem(f_l,i)=0}{\arg\max}(f_{ic}) \tag{11}$$

其中 $\underset{i}{\arg}$ 表示以 i 为变量；$\max(\cdot)$ 表示求最大值运算；$rem(f_l,i)$ 表示 f_l 除以 i 取余数。码字有效的判据为式(8)。由此得到二进制 BCH 码分组长度 n 的识别流程如下：

1) 初始化 $i=3$ 及 j 值；
2) 如果 i 不能够整除 f_l，则转向 6)；
3) 以 i 为分组长度进行分组，得到 N_i 个码字；
4) 按照式(8)统计 N_i 个码字中的有效码字数 N_{ic}；
5) 计算 $f_{ic}=\dfrac{N_{ic}}{N_i}$ 并保存；
6) $i=i+2$；
7) 如果 $i\leqslant\lceil f_l/2\rceil$，转向 2)；
8) 比较所存储的 f_{ic}，使 f_{ic} 值最大的 i 值即为分组长度的估计值 n；
9) 识别结束。

由式(11)及式(8)可看出，分组长度 n 的识别性能与 N_i 及 j 取值有关，直观来看 N_i 及 j 的取值越大，识别的性能越好；但是取值越大也就意味这需要接收的码流序列长度越长，计算量越大，因此要根据实际需求折中选择。

3　生成多项式的识别方法

本节讨论在分组长度 n 已经正确识别的条件下，如何识别生成多项式 $g(x)$。

设接收到 N 个码字，$c_p(x)$ 为第 p 个码字所对应的码多项式，并且假设码字中不存在误码。由第 2 节可知，由 $c_p(x)$ 经过循环左移可得到的 $n-1$ 个码多项式 $c_{p1}(x),c_{p2}(x),\ldots,c_{pn-1}(x)$，并且这 $n-1$ 个码多项式与 $c_p(x)$ 的生成多项式是相同的，换言之生成多项式 $g(x)$ 是 $c_p(x),c_{p1}(x),c_{p2}(x),\ldots,c_{pn-1}(x)$ 的 1 个公约式。

同样令 $c_{p0}(x)=c_p(x)$，并设，

$$f_p(x)=\gcd[c_{p0}(x),c_{p1}(x),\ldots,c_{pj}(x)] \tag{12}$$

则 $f_p(x)$ 必然是 $g(x)$ 的倍式。

对接收到的 N 个码字分别按式(12)计算，则可以得到 $M(1\leqslant M\leqslant N)$ 个不同的多项式，考虑存在误码的情况，这 M 个多项式必属于以下 4 种情况之一：a) 等于 1；b) 不等于 1，但不等于 $g(x)$，也不是 $g(x)$ 的倍式；c) 等于 $g(x)$；d) 等于 $g(x)$ 的倍式。

前 2 种情况说明码字中存在误码，后 2 种情况说明码字中没有误码，或者加入误码后构成同一码子集合中的另一个码字。

可按如下 3 个步骤从 M 个备选多项式中识别出生成多项式。

1) 根据 $g(x)$ 所需满足的约束条件，剔出备选多项式中不符合条件的多项式

由二进制循环码的性质可得，$g(x)$ 需满足条件[9]：a) $g(x)\neq 1$；b) $g(x)\neq x^q+1$，其中 q 为正整数，且 $1\leqslant q<n$；c) $g(x)$ 整除 x^n+1。根据这 3 个条件可以剔出情况 a)下的多项式及情况 b)的部分多项式。

2) 按照校正子重量和最小准则选择最优多项式

假设通过过程 1)还余下 L 个备选多项式，$g_i(x)\ (i=1,2,..,L)$ 是其中第 i 个多项式，则对应的校验多项式为：

$$h_i(x)=(x^n+1)/g_i(x) \tag{13}$$

第 j 个码字的校正子，

$$r_{ij}=[h_i(x)c_j(x)]\bmod(x^n+1) \tag{14}$$

其码重为

$$w_{ij}=weight(r_{ij}) \tag{15}$$

对所有接收到的 N 个码字计算校正子码重和：

$$w_i = \sum_{j=1}^{N} w_{ij}\ ,\ \ i = 1,2,\cdots,L \tag{16}$$

使 w_i 最小的备选多项式就是最优多项式，记为 $g_b(x)$，即

$$g_b(x) = \underset{g_i(x),i=1,2,\cdots,L}{\arg\min}\ (w_i) \tag{17}$$

3) 由最优多项式 $g_b(x)$ 得到估计的生成多项式 $g_0(x)$

在正确识别的情况下，最优多项式 $g_b(x)$ 并不一定等于 $g(x)$，但一定是 $g(x)$ 的倍式，因此还需要进一步由 $g_b(x)$ 估计生成多项式。

首先根据文献[11]中给出的准则，判断 $g_b(x)$ 是否为不可约；如果是不可约的，则

$$g_0(x) = g_b(x) \tag{18}$$

如果是可约的，求出 $g_b(x)$ 所有满足 1)中条件的不可约因式，并按照步骤 2)求出最优多项式即为估计的生成多项式 $g_0(x)$。因式分解的方法，可根据文献[11]中准则间接得出，限于篇幅这里不再赘述。

从以上识别过程可以看出，理论上，只要在接收的码字中存在一个无误码的码字，则可以以很高的概率正确识别出生成多项式，因此这种识别方法有较强的抗误码能力。

4 识别性能仿真

二进制 BCH 码的仿真分 3 步进行：首先进行分组长度的识别性能仿真；然后，进行生成多项式识别性能的仿真；最后，根据前 2 步的仿真结果给出综合的识别性能。

在仿真中，选取(15,11)本原二进制 BCH 码和(21,12)非本原二进制 BCH 码进行识别性能仿真，分别进行 1000 次试验统计识别正确率。其中(15,11)本原二进制 BCH 码的生成多项式为 $g_1(x) = x^4 + x + 1$；(21,12)非本原二进制 BCH 码的生成多项式为 $g_1(x) = x^9 + x^3 + 1$；假设 5 个码字构成 1 个数据帧，并且在识别分组长度时选取循环左位次数为 1。

4.1 仿真结果

按照 2,3 节中所述方法进行识别。仿真结果如图 1~图 6 所示。其中综合识别性能是根据分组长度识别性能和生成长度识别性能计算得到的，计算公式如下：

$$p = p_1 p_2 \tag{19}$$

式中：p 为综合识别正确率；p_1 为分组长度识别正确率；p_2 为生成多项式识别正确率。

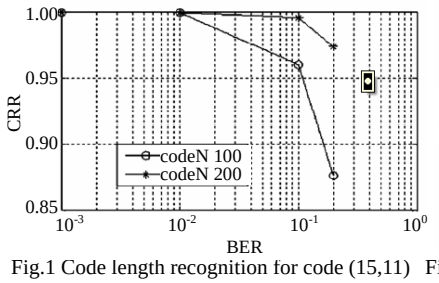

Fig.1 Code length recognition for code (15,11)
图 1 (15,11)码分组长度识别性能图

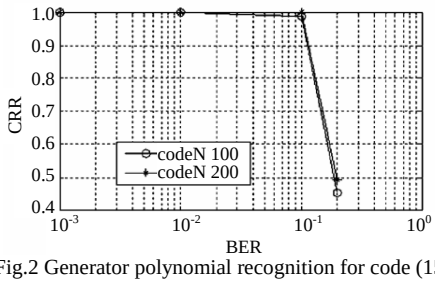

Fig.2 Generator polynomial recognition for code (15,11)
图 2 (15,11)码生成多项式识别性能

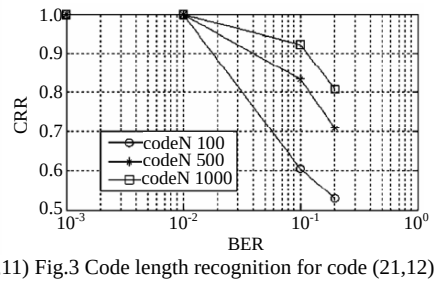

Fig.3 Code length recognition for code (21,12)
图 3 (21,12)码分组长度识别性能图

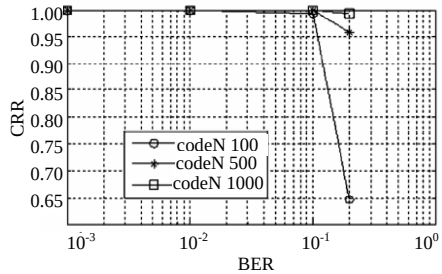

Fig.4 Generator polynomial recognition for code (21,12)
图 4 (21,12)码生成多项式识别性能

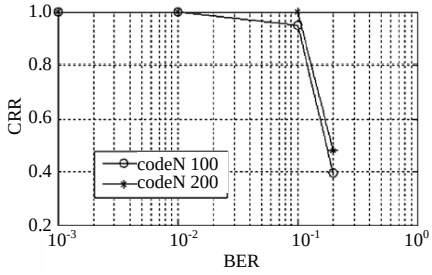

Fig.5 Total recognition capability for code (15,11)
图 5 (15,11)码综合识别性能

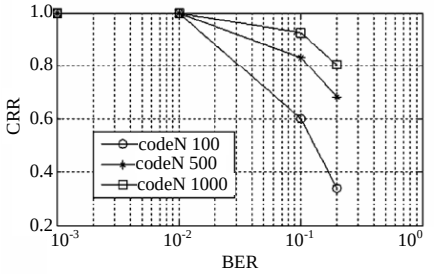

Fig.6 Total recognition capability for code (21,12)
图 6 (21,12)码综合识别性能

4.2 仿真结果分析

从仿真结果可以看出，在循环移位的次数一定的条件下，无论是分组长度的识别性能、生成多项式的识别性能，还是综合识别性能，都随着参与识别的码字数目的增加而提高。

通过对比可以看出，在码字数相同的条件下，2 种码的识别性能差别较大。对于(15,11)码，误码率为 10%时，取 100 个码字用于识别，其综合识别性能可达到 90%以上；而对于(21,12)码，在相同的误码率下，要达到相同的性能，所需要的码字数为 1 000。造成这种差异的原因是多方面的：

1) (21,12)码的码长比(15,11)码长，这样，在相同码字数的条件下，经过 1 次循环移位，满足循环特性的比例会高于(15,11)码，即使没有实现正确分组；

2) (21,12)码的帧长为 105,有 6 个奇因数为 3,5,7,15,21,35;而(15,11)码的帧长为 75,有 4 个奇因数为 3,5,15,25;显然，因数越多，在码字数相同的条件下，识别性能就会越差。

当然,识别性能的差异还与码的结构有关,(15,11)码是本原二进制 BCH 码,而(21,12)码是非本原二进制 BCH 码。不过，从仿真结果可以看出，虽然在码字数相同时，(21,12)码的识别性能比(15,11)码差，但可以通过增加码字达到相同的识别性能。

5　结论

本文所提出的二进制 BCH 码盲识别方法，是一种统计识别方法。正因如此，识别性能与用于统计的比特数直接相关；用于统计的比特数越多，识别性能越好，然而，相应的计算量也就越大，尤其是在长码的情况下，这是本文算法的一个重要缺陷。但是，与其诱人的识别性能相比，大的计算量还是可以接受的。另外，本文算法不涉及复杂运算，而且二进制特性与硬件的处理机理有着非常自然的契合性，因此用硬件实现非常方便，从而可以解决运算速度问题。

参考文献：

[1] Wang Fenghua,Huang Zhitao,Zhou Yiyu. A Method for Blind Recognition of Convolution Code Based on Euclidean Algorithm[C]// IEEE International Conference on Wireless Communication Networking and Mobile Computing. Shanghai:[s.n.], 2007:1414–1417.

[2] 陆佩忠,沈利,邹艳. 删除卷积码的盲识别[J]. 中国科学 E 辑:信息科学, 2005,35(2):173–185.

[3] 沈利. 删除卷积码的性质及其盲识别[D]. 郑州:解放军信息工程大学, 2004.

[4] 邹艳. 信息接获与处理的容错技术研究[D]. 上海:复旦大学, 2006:8–26.

[5] 韩国宾. 删除卷积码的识别技术[D]. 成都:电子科技大学, 2006.

[6] 宋镜业. 信道编码识别技术研究[D]. 西安:西安电子科技大学, 2009.

[7] 俊君,李艳斌. 低码率二进制线性分组码的盲识别[J]. 无线电工程, 2009, 39(1):19–24.

[8] 刘健,谢锘,周希元. RS 码的盲识别方法[J]. 电子科技大学学报, 2009,38(3):363–367.

[9] Shu Lin,Daniel J Costello Jr. 差错控制编码[M]. 北京:机械工业出版社, 2007:91–179.

[10] 王新梅,肖国镇. 纠错码—原理与方法[M]. 西安:西安电子科技大学出版社, 2001:242–318.

[11] 王鑫,王新梅,韦宝典. 判定有限域上不可约多项式及本原多项式的一种高效算法[J]. 中山大学学报(自然科学版), 2009,48(1):6–9.

作者简介：

王甲峰(1974–)，男，吉林省镇赉县人，工程硕士，副研究员，主要研究方向为通信对抗. email:hugeghost@hotmail.com.

岳　旸(1982–)，女，四川省绵阳市人，工程硕士，助理工程师，主要研究方向为通信信号处理.

权友波(1980–)，男，四川省广元市人，工程硕士，助理工程师，主要研究方向为通信信号处理.