基于概率逼近的本原 BCH 码编码参数的盲识别方法

阔永红* 曾伟涛 陈 健
(西安电子科技大学通信工程学院 西安 710071)

摘 要：针对本原 BCH 码编码参数的盲识别问题，该文提出了一种基于概率逼近的盲识别方法。首先，利用 Gauss 分布和 Poisson 分布逼近随机码字的根概率特性，确定了搜索 BCH 码长的门限；然后，通过分析本原域元素的检错能力及同构对域的影响，应用临近域对的方法确定编码域，提高了其识别能力；最后，给出识别生成多项式时的共轭根系表，从而减少了计算量。仿真结果表明，在较高的误码率下，该方法能快速地识别出 BCH 码编码所采用的编码参数。

关键词：信道编码；BCH 码；共轭根系；盲识别
中图分类号： TP391;TN911.22 **文献标识码**： A **文章编号**：1009-5896(2014)02-0332-08
DOI: 10.3724/SP.J.1146.2013.00584

Blind Identification of Primitive BCH Codes Parameters Based on Probability Approximation

Kuo Yong-hong Zeng Wei-tao Chen Jian
(*School of Telecommunications Engineering, Xidian University, Xi'an* 710071, *China*)

Abstract: To solve the issues of blind identification of primitive BCH codes encoding parameters, a novel identification algorithm with probability approximation is presented. Frist, by taking advantage of the approximation of random code words' root probability character which uses Gaussian distribution and Poisson distribution, the thresholds for searching code length are structured. Second, though analyzing the checking ability of the primitive element and the impact of isomorphism on searching, the coding filed is determined by using the method of nearby fields pair which improves the performace of identification. Finally, the calculation is reduced by creating and using the conjugate roots table in the recognition of generator polynomial. Simulation results show that, the proposed algorithm achieves a significant improvement in identification probability even if in high BER situation.

Key words: Channel coding; BCH codes; Conjugate roots; Blind identification

1 引言

在电子战中，通信情报系统(COMmunication INTelligence,COMINT)[1]在通信侦察中扮演着重要的角色。COMINT 系统接收未知信号，并试图解调、解码，然后从中获取信息，而在这一过程中，信道编码参数的盲识别问题，占据着重要的地位。

目前关于编码盲识别的研究主要涉及卷积码[2−6]，Turbo 码[7−9]，BCH 码[10,11]，RS 码[12]等线性分组码[13−16]。 BCH 码是一种很好的线性纠错码类，具有纠错能力强、构造方便、编码简单等特点，特别是在中短码长下，其性能接近理论值，是目前应用最为广泛的码类之一。针对 BCH 码盲识别问题，公开发表文献并不多见。文献[10]提出了一种利用码根差熵和码根统计的盲识别方法，但在高码率情况下，码根差熵抵抗误码能力变差，且算法计算量较大。针对文献[10]计算量较大的问题，文献[11]尝试利用域的同构性质加以改进，但算法以无法准确确定编码所在域为代价，且也未能提高算法的识别性能。

针对以上方法的不足，本文提出了一种利用 Gauss 分布和 Poisson 分布逼近码字组以本原域元素为根的满足率 η 的概率分布，在静态的第 1 个满足率门限和动态的第 2 个满足率门限下，搜寻待分析数据所采用编码参数的盲识别方法。该方法一方面借助概率门限以容许更多的误码；另一方面，通过舍去检测出的含误码码字，提升无误码码字比例，提高生成多项式识别的可靠性。仿真结果表明，在误码率为 1.05×10^{-2} 的情况下，对于码长不大于 511 的各种本原 BCH 码，该方法的识别率达 90%以上。

2013-04-27 收到，2013-07-27 改回
国家自然科学基金(60972072)和高等学校学科创新引智计划(B08038)资助课题
*通信作者：阔永红 yhkuo@mail.xidian.edu.cn

2　BCH 码定义及分析

2.1 BCH 码定义及分析

定义 1[17]　设 α 是域 GF(2^m) 中的本原域元素，有 t 个误码纠错能力的二进制本原 BCH 码的生成多项式 $g(X)$ 是以 GF(2^m) 中 $\alpha,\alpha^2,\alpha^3,\cdots,\alpha^{2t}$ 为根的次数最小的多项式。

命题 1[18]　设 $f(X)$ 是系数取自 GF(2) 中的多项式，β 是 GF(2) 扩域中一个元素，若 $f(\beta)=0$，则对任意 $l\geq 0$，都有 $f(\beta^{2^l})=0$。

命题 2[18]　BCH 码字多项式 $c(X)$ 与其生成多项式 $g(X)$ 满足：$c(X)=g(X)q(X)$，其中 $q(X)$ 是一个与信息位相关的多项式。

以上是关于多项式理论及码字多项式的一些基本规律，为了区分码字多项式和随机码字以及 $g(X)$ 根元素和非根元素，本文提出命题 3 和命题 4，以研究随机码字和非根元素的一些性质。

命题 3　若 β 是域 GF(2^m) 中 d 级非零元素，设 $m*$ 为 2 对模 d 的方次数[18]，则随机码字以 β 为根的概率为 $1/2^{m*}$。

证明　设随机码字的码字多项式 $r(X)=r_{2^m-2}X^{2^m-2}+r_{2^m-3}X^{2^m-3}+\cdots+r_1X+r_0$，其中 r_i 服从二项分布：$r_i\sim\mathfrak{B}(1,0.5)$, $i=0,1\cdots,2^m-2$。由元素 β 的级为 d 以及级的定义[18]可知，对于任意 $k>d-1$，都有 $\beta^k=\beta^{dt}\beta^{k-dt}=e\beta^{k-dt}=\beta^{k-dt}$，其中 $t\in Z^*$，$0\leq k-dt\leq d-1$，故对 $\forall\xi\in Z^*$，$\beta^\xi\in\{\beta,\beta^2,\beta^3,\cdots,\beta^{d-1},e\}$。将 β 代入码字多项式，得 $r(\beta)=r_{2^m-2}\beta^{2^m-2}+r_{2^m-3}\beta^{2^m-3}+\cdots+r_1\beta+r_0=\gamma_{d-1}\beta^{d-1}+\gamma_{d-2}\beta^{d-2}+\cdots+\gamma_2\beta^2+\gamma_1\beta+\gamma_0 e$，其中 $\gamma_i\sim\mathfrak{B}(1,0.5)$，$i=0,1,\cdots,d-1$。

记 $\Omega(X)=\gamma_{d-1}X^{d-1}+\gamma_{d-2}X^{d-2}+\cdots+\gamma_2X^2+\gamma_1X+\gamma_0 e$，设以 β 为根的最小多项式为 $M(X)$，很显然该最小多项式的最高幂次为 $m*$。若 $r(\beta)=0$，则 $\Omega(X)=M(X)H(X)$，其中 $H(X)$ 为 $n-m*-1$ 次多项式。那么一个系数取自 GF(2) 的 $d-1$ 次随机多项式 $\Omega(X)$ 是多项式 $M(X)$ 的倍式的概率为

$$2^{d-m^*-1}/2^{d-1}=1/2^{m^*} \tag{1}$$

命题 4　若 BCH$[n,k]$ 的码字多项式为 $c(X)$，元素 β 的级为 d，2 对模 d 的方次数为 $m*$，且 β 不是生成多项式 $g(X)$ 的根，则 $c(\beta)=0$ 的概率为 $1/2^{m^*}$。

命题 3 指出了随机码字以 α 以及域中任意其它元素为根的概率与该元素所在根系中元素个数有关；而对于 BCH 码字 $c(X)$，若 $g(\beta)=0$，则 $c(\beta)=0$。由命题 4 可知，若 $g(\beta)\neq 0$，则 $p(c(\beta)=0)=1/2^{m^*}$。

2.2 概率分布特性的逼近

根据 BCH 码定义，对于任何本原 BCH 码，其生成多项式必然以本原域元素 α 为根。而根据级和方次数的概念可知，α 所在共轭根系中的元素个数必然为 m (码长 $n=2^m-1$)，因此随机码字以 α 为根的概率为 $\lambda_m=1/2^m$。设待分析数据的码字个数为 N (码长不正确时)，那么满足 $c(\alpha)=0$ 码字个数 $\mathrm{N_s}$ 服从 Bernoulli 分布：$N_s\sim\mathfrak{B}(N,\lambda_m)$。

设 η 是随机码字组满足 $c(\alpha)=0$ 的满足率，即 $\eta=N_s/N$。若 $N\lambda_m(1-\lambda_m)\geq 9$，由棣莫弗-拉普拉斯中心极限定理可知，随机变量 η 近似服从 Gauss 分布：$\eta\sim\mathcal{N}(\lambda_m,\delta_1^{\,2})$，方差为

$$\delta_1^{\,2}=\lambda_m(1-\lambda_m)/N \tag{2}$$

若 $N\lambda_m(1-\lambda_m)<9$，则 η 近似服从 Poisson 分布[19]：

$$p(\eta=\kappa/N)=(\lambda_m)^\kappa e^{-\lambda_m}/\kappa! \tag{3}$$

而对于无误码的码长为 n 的本原 BCH 码字组，由命题 2 可知，在编码所在域内，码字一定以本原域元素 α 为根，即码字组以 α 为根的满足率 η 满足 $p(\eta=1)=1$，这与上述分布有着很大的差异。考虑到误码的存在，并设信道误码率为 p_e，误码率较低时码字组以 α 为根的满足率 η 也近似服从 Gauss 分布：$\eta\sim\mathcal{N}(\upsilon,\delta_2^2)$，方差为

$$\delta_2^2=\upsilon(1-\upsilon)/N \tag{4}$$

其中 υ 是误码率为 p_e 的一个码字以 α 为根的概率，易知

$$\upsilon=(1-p_\mathrm{e})^n+(1-(1-p_\mathrm{e})^n)\lambda_m \tag{5}$$

3　BCH 编码盲识别的实现

在帧同步的情况下，待识别的 BCH 码参数包括码长，本原多项式生成的编码所在域以及生成多项式的根，其中每个域与唯一本原多项式对应。文献[11]识别出正确的生成多项式，但是所得的若干根并非一定连续，因而识别出的本原多项式不一定是构造 BCH 码编码所在域的本原多项式。

本文提出的方法通过利用满足率门限，在不同的域间搜索并确定码长，利用生成多项式根的连续性以及共轭根系的性质，确定编码所在域以及生成多项式的根，完成 BCH 码编码参数盲识别。下面首先给出用于域间搜索的门限定义，再进行本原域元素的检错分析，最后给出 BCH 码编码参数盲识别方案。

3.1 用于域间搜索的门限定义

设帧同步后待分析的数据 $\widehat{B}=(r_1,r_2,\cdots,r_{Nn})$，其中 $r_j\in\{0,1\}$, $1\leq j\leq Nn$，并假设在某次搜索中 BCH 码长为 n，且 $m=\log_2(n+1)$，则数据 $\widehat{B}$ 可表示成码字组 $\widehat{W}=(W_1,W_2,\cdots,W_N)$，其中码字 $W_i=(r_{(i-1)n+1},r_{(i-1)n+2},\cdots,r_{in})$, $1\leq i\leq N$, BCH 码码字多项式 $W_i(X)=r_{(i-1)n+1}e+r_{(i-1)n+2}X+\cdots+r_{in}X^{n-1}$。若 W_i 是 BCH 码字，编码所在域中的本原域元素为 α，则有

$$W_i(\alpha)=r_{(i-1)n+1}e+r_{(i-1)n+2}\alpha+\cdots+r_{\text{in}}\alpha^{n-1}=0 \quad (6)$$

设满足 $W_i(\alpha)=0$ 的码字集合 $\widehat{S}=\{W_v \mid v\in\Delta\}$，且对 $\forall v\in\Delta, W_v(\alpha)=0$，故码字组在某域(设对应码长为 n)中以其本原域元素 α 为根的满足率：

$$\eta=\text{card}(\Delta)/N=N_{\text{s}}/N \quad (7)$$

其中 Δ 为集合 $\{1,2,\cdots,N\}$ 的子集，$\text{card}(\Delta)$ 表示集合 Δ 中元素个数。对于属于线性分组码类的BCH码，若以非真实的码长为单位划分BCH码字，则得到码字组 $\widehat{W}$ 将是类似于随机码字组的随机码组。根据上述码字分析和概率逼近，若某域中本原域元素 α 对应的满足率 η 满足

$$\Phi(\eta)<p_1 \quad (8)$$

则称搜索算法以概率 p_1 排除了该域(本原多项式)，其中 $\Phi(x)$ 有如式(9)的定义

$$\Phi(x)=\begin{cases}\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-\infty}^{(x-\lambda_m)/\delta_1}e^{-t^2/2}\text{d}t, & \\ & x\neq 0, N\lambda_m(1-\lambda_m)\geq 9\\ 0, & x=0\\ \displaystyle\sum_{\kappa=0}^{Nx-1}(\lambda_m^{\kappa}e^{-\lambda_m}/\kappa!), & \text{其它}\end{cases} \quad (9)$$

因此概率门限 $p_1\in(0,1)$ 对应的满足率门限

$$H_1=\begin{cases}\lambda_m+t_{p_1}\sqrt{\lambda_m(1-\lambda_m)/N}, & \\ \quad N\lambda_m(1-\lambda_m)\geq 9 & \\ \left\{\min(K/N)\left|\displaystyle\sum_{\kappa=0}^{K}(\lambda_m^{\kappa}e^{-\lambda_m}/\kappa!)>p_1\right.\right\}, & \\ \quad \text{其它} & \end{cases} \quad (10)$$

其中 t_{p_1} 满足等式 $\Phi(t_{p_1})=p_1$。若在某搜索的当前域中，满足率 $\eta\geq H_1$，则认为该域是 p_1 概率可接受的，即接受该域中的本原域元素 α 作为码字组 $\widehat{W}$ 的公共根。

若 $g(X)$ 仅以一个共轭根系为全部根，则该纠错码的纠错能力 $t=1$，但实际使用的纠错码纠错能力多不止1个，此时 $g(X)$ 必然又以第 2 个根系中的元素为根。由于第 1 根系的元素集合 $\{\alpha,\alpha^2,\alpha^4,\cdots,\alpha^{2^{(m-1)}}\}$ 中并不包含 α^3，根据BCH码的定义可知，第 2 个根系中必然包含 α^3，并设该共轭根系中互不相同的元素个数为 m^*。对于含误码的BCH码字组，在编码所在域内无误码字以及少部分含误码码字满足 $W_i(\alpha)=0$；而使 $W_i(\alpha)\neq 0$ 的码字则必然含有误码，因此略除该部分码字可以提高剩余码字组无误码码字比例，并假设对任意 $W_i\in\widehat{S}=\{W_v \mid v\in\Delta\}$，均满足 $W_i(\alpha)=0$，则有 $N_{\text{s}}=\text{card}(\Delta)$。因此在不同的域间搜索时，若 α 的满足率使得 $\eta\geq H_1$，进而可利用 $W_i(\alpha^3)=0$，以概率 p_2 排除非编码所在域，其对应的满足率门限

$$H_2(m^*)=\begin{cases}\lambda_{m^*}+t_{p_2}\sqrt{\lambda_{m^*}(1-\lambda_{m^*})/N}, & \\ \quad N\lambda_{m^*}(1-\lambda_{m^*})\geq 9 & \\ \left\{\min(K/N)\left|\displaystyle\sum_{\kappa=0}^{K}(\lambda_{m^*}^{\kappa}e^{-\lambda_{m^*}}/\kappa!)>p_2\right.\right\}, & \\ \quad \text{其它} & \end{cases} \quad (11)$$

其中 t_{p_2} 满足等式 $\Phi(t_{p_2})=p_2$，$\lambda_{m^*}=1/2^{m^*}$，并令 $p_3=1-(1-p_1)(1-p_2)$，$H_3=\Phi^{-1}(p_3)$。

设信道误码率为 p_{e}，对编码参数为 $[n,k]$ 的BCH码，令 $\varepsilon=2^{-(n-k-m)}$，则能通过本原域元素校验但是不能通过 $g(X)$ 其它根元素校验的概率为

$$p_{\text{r}}=\frac{(1-R)(1-\varepsilon)}{(2^m-(1-\varepsilon))R+1-\varepsilon} \quad (12)$$

其中 R 为误码率为 p_{e} 时无误码字在所有码字中的比率，易知 $R=(1-p_{\text{e}})^n$。对比经过本原域元素校验略除处理后，码字组 $\widehat{S}$ 中的含误码码字比率和 $\widehat{W}$ 中含误码码字比率，有

$$\theta=\frac{p_{\text{r}}}{1-R}=\frac{1-\varepsilon}{(2^m-(1-\varepsilon))R+1-\varepsilon} \quad (13)$$

可以看出当误码率较低时，θ 接近 $1/2^m$，表明在误码率不是很高的情况下经过本原域元素的校验略除处理，含误码码字的比例大致下降到原来的 $1/2^m$。图 1 为误码率 p_{e} 从 $1\times10^{-3}\sim1\times10^{-2}$ 时，BCH(15,7)和BCH(255,199)码经过本原域元素处理后的含误码码字比例 p_r 和未经处理的含误码码字组比例的对比图。由图可知，处理后的含误码码字组的比例大约下降 1 个数量级甚至更多，因此概率 p_2 可以设为一个接近 1 的较高的概率门限。

3.2 算法描述

3.2.1 识别码长 根据上述定义及分析，算法依次在不同的本原多项式生成的域中，利用该域中的本原域元素 α 以及 α^3 进行搜索，直至在某个域 F 中，码字组以 p_1 概率门限接受 α 为码字组 $\widehat{W}$ 的公共根，并且以 p_2 概率门限接受通过 α^3 为码字组 $\widehat{W}$ 的公共

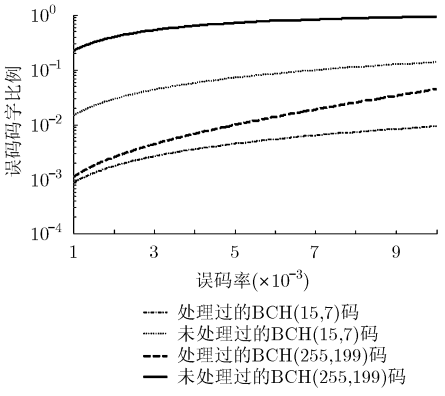

图 1 误码率与误码码字比例的关系

根，或者 α 的满足率 η 使不等式 $\Phi(\eta)\ge p_3$ 成立，则接受域 F 对应的码长 $n=2^h-1$ 为识别码长。

对于纠错能力 $t>2$ 的 BCH 码，设 ω_i 表示 $m=i$ 时所有本原多项式的个数[11]，并令 $\lambda_h=1/2^h$，$\varpi_h=\sum_{i=3}^{i=h-1}\omega_i$，定义满足率 H 在误码率为 p_{e} 的码字组中的出现的概率函数

$$p_{\mathrm{t}}(H)=p_{\mathrm{t}}(H;\upsilon)=\begin{cases}\displaystyle\int_{-\infty}^{H}\frac{1}{\sqrt{2\pi}\delta_2}\exp\left(-\frac{(x-\upsilon)^2}{2\delta_2^2}\right)\mathrm{d}\,x,\\ \quad N\upsilon(1-\upsilon)\ge 9\\ \displaystyle\sum_{\kappa=0}^{[HN]}\lambda_h^{\kappa}\exp(-\lambda_h)/\kappa!,\\ \quad \text{其它}\end{cases}\tag{14}$$

其中 υ 是式(5)所示的概率。算法的虚警概率 p_{fc} 和漏警概率 p_{mc} 以及误判概率 p_{c} 满足

$$\begin{aligned}p_{\mathrm{c}}&=p_{\mathrm{fc}}+p_{\mathrm{mc}}\approx[(1-p_3^{\varpi_h})]\\&+[p_{\mathrm{t}}(H_3)+(1-p_{\mathrm{t}}(H_3))(1-p_2^{t-1})]\end{aligned}\tag{15}$$

由于存在同构现象，即对域 F 和 $\overline{F}$，若 $\oplus$ 为域的加法“+”，则对 $\forall\beta_1,\beta_2\in F$，若 $\oplus$ 为域的乘法“•”，则对 $\forall\beta_1,\beta_2\in F\setminus\{0\}$，且对 $\forall k\in\mathrm{GF}(2)$，存在一个双射 σ 满足

$$\sigma(\beta_1\oplus\beta_2)=\sigma(\beta_1)\oplus\sigma(\beta_2)\tag{16}$$

$$\sigma(k\beta_1)=k\sigma(\beta_1)\tag{17}$$

对编码所在域 F，那么可能会存在另外一个域中也满足码字组 $\widehat{W}$ 以该域中的 α 或 α 和 α^3 为公共根，于是同构带来的影响是不能立即确定编码所在域，为此本文提出命题 5。

命题 5　若 $g(X)$ 根元素集合 R 中在域 F 中的元素 φ 或 φ 和 ψ 与域 $\overline{F}$ 中的本原域元素 $\overline{\alpha}$ 或 $\overline{\alpha}$ 和 $\overline{\alpha}^3$ 在某双射 σ 下满足 $\sigma(\varphi)=\overline{\alpha}$ 或 $\sigma(\varphi)=\overline{\alpha}$ 且 $\sigma(\psi)=\overline{\alpha}^3$，那么在域 $\overline{F}$ 中，码字多项式 $c(X)=c_{n-1}X^{n-1}+c_{n-2}X^{n-2}+\cdots+c_1X+c_0e$ 满足 $c(\overline{\alpha})=\overline{0}$ 或 $c(\overline{\alpha})=\overline{0}$ 且 $c(\overline{\alpha}^3)=\overline{0}$，其中 $\overline{0}$ 为域 $\overline{F}$ 相对于加法的单位元，e 为域 F 相对于乘法的单位元。

根据式(16)和式(17)不难证明命题 5。该命题指出，与域 F 同构的域 $\overline{F}$，若其满足命题 5 所提条件，则域 $\overline{F}$ 也必然满足终止码长长度搜索的条件。同构使得真实编码所在域 F 不能立即被确定，但域 $\overline{F}$ 和域 F 所对应的码长相同，因此同构并没影响对码长的搜索。现只需要搜寻编码所在域 F 和 $g(X)$ 在域 F 中的根。

3.2.2 识别域和生成多项式　因为 $g(X)$ 若以任一元素为根，则必也以该元素所在共轭根系中的其它所有元素为根，并且由 BCH 码的定义及命题 1 可知，生成多项式 $g(X)$ 具有若干连续但根集合 R 并不是全部连续的若干根。为了提高搜索速度和精简运算，算法在某确定域中采用的搜索方法是依次搜索顺序的共轭根系表，直至某共轭根系不被接受为码字组的公共根，若判定为非编码所在域，则在下一个域中继续搜索。

设域 F 是编码所在域，生成多项式 $g(X)$ 的根元素在域 F 中构成的根集合 $R=\{\alpha,\alpha^2,\cdots\}$，并设本原域元素 α 和 α^3 所在的共轭根系集合分别为 $\hat{\alpha}$, $\hat{\alpha}_3$。由定义 1, BCH 码生成多项式有 $2t$ 个连续的根，那么生成多项式的全部根是涵盖这 $2t$ 个连续根的最少共轭根系组中的全部元素，即有根集合 $R=\{\alpha,\alpha^2,\cdots\}=\{\hat{\alpha},\hat{\alpha}^3,\cdots\}$，因此建立涵盖 $2t$ 个连续根的顺序的共轭根系分组并在表中搜索，可以简化计算。

顺序的共轭根系表 $\mathcal{T}=(G_1,G_2,\cdots)$ 的构造方法：从 α 开始，以集合 $G_1=\{\alpha^1,\alpha^2,\alpha^{2^2},\cdots,\alpha^{2^{m-1}}\}\pmod{\alpha^{2^m-1}}$ 构成第 1 个共轭根系分组。$\beta=\alpha^3\in G=\{\alpha,\alpha^2,\alpha^3,\cdots,\alpha^{2^m-2}\}$ 是尚未出现且 α 次幂最小的元素，则包含 β 在内的共轭根系集合 $G_2=\{\beta^1,\beta^2,\beta^{2^2},\cdots,\beta^{2^{m-1}}\}(\mathrm{mod}\ \alpha^{2^m-1})$ 构成第 2 个共轭根系分组。依此类推，可得下一个共轭根系分组，直至集合 G 中元素全部出现。由于经过求模运算，一个共轭根系分组中的元素个数可能会小于 m 个，根据级的相关理论，任意元素 $\alpha^i\in G$ 的级 $d^*=\min\{d\mid\alpha^{id}=\alpha^{(2^m-1)j},j=1,2,\cdots\}$，则 α^i 所在共轭根系分组内的元素个数 m^* 等于 2 对模 d^* 方次数，即有 $2^{m^*}=1(\mathrm{mod}\ d^*)$。当 $m=5$ 时，集合 $G_1=\{\alpha^1,\alpha^2,\alpha^4,\alpha^8,\alpha^{16}\}$，集合 $G_2=\{\alpha^3,\alpha^6,\alpha^{12},\alpha^{24},\alpha^{17}\}$，此时 α^5 是集合 G_1,G_2 中尚未出现且次幂最小的元素，于是集合 $G_3=\{\alpha^5,\alpha^{10},\alpha^{20},\alpha^9,\alpha^{18}\}$。不难得出，集合 $G_4=\{\alpha^7,\alpha^{14},\alpha^{28},\alpha^{25},\alpha^{19}\}$，集合 $G_5=\{\alpha^{11},\alpha^{22},\alpha^{13},\alpha^{26},\alpha^{21}\}$，集合 $G_6=\{\alpha^{15},\alpha^{30},\alpha^{29},\alpha^{27},\alpha^{23}\}$。

根元素的搜索顺序：在顺序的共轭根系表 $\mathcal{T}$ 中，按 $G_1,G_2,\cdots$ 的顺序，在每个集合中任意挑选一个元素并验证是否为根，若某元素被验证为非根元素，则停止下一个集合中元素的验证。事实上，由命题 1 可知，在每个共轭根系中都只需任意验证其中任意一个元素，这样可以减少运算，加快搜索速度。设 BCH 信息位长度为 k，相比于按元素指数完全顺序搜索，该搜索计算量大约只有其 ρ 倍，其中 ρ 满足

$$\begin{aligned}&(2^m-2-k)/(m\times(2^m-2))\le\rho\\&<(2^m-2-k)/(2\times(2^m-2))\end{aligned}\tag{18}$$

由于域 $\mathrm{GF}(2^m)$ 中的绝大部分根系中的元素个数都是 m，因此一般情况下 $\rho<1/m$ 均成立。

从上述同构现象可知，生成多项式在编码所在域 F 中的连续根系的个数最多。设域 F 与域 $\overline{F}$ 同构，

且 $g(X)$ 在域 F 中有 μ 个根系：$R=\{\Gamma_1,\Gamma_2,\cdots,\Gamma_\mu\}\in F$，则根据BCH码的定义，该 μ 个根系一定是顺序的共轭根系表中的前 μ 个根系。对应地，设域 $\overline{F}$ 的前 μ 个根系为 $\overline{R}=\{\overline{\Gamma}_1,\overline{\Gamma}_2,\cdots,\overline{\Gamma}_\mu\}\in\overline{F}$。若存在某元素 $\overline{\beta}\in\overline{R}$，使得 $\sigma^{-1}(\overline{\beta})\notin R$，那么 $g(X)$ 在域 $\overline{F}$ 中的所有根元素集合 R^* 满足：$\mathrm{card}(R^*)<\mathrm{card}(R)$。如果不存在元素 $\overline{\beta}$，就会存在不止 1 个域，且各域中可被接受的连续共轭根系数目是相等的，那么这些域均是编码所在域。一般地，对于两个同构的域，若根元素映射关系满足一定条件，则对应的多项式相同，根据式(16)，式(17)和命题 5，不难证明如下命题：

命题 6 设 $g(X)$ 的根集合 $R=\{\beta_1,\beta_2,\cdots,\beta_q\}\in F$，域 $\overline{F}$ 中根集合 $\overline{R}=\{\overline{\beta}_1,\overline{\beta}_2,\cdots,\overline{\beta}_q\}$，若对 $\forall\beta_v\in R$，都有 $\sigma(\beta_v)=\overline{\beta}_v,1\le v\le q$，那么系数都在域GF(2)中的多项式 $g(X)=\overline{g}(X)$。

命题 6 指出，如果不存在元素 $\overline{\beta}$，那么分别以域 F 和域 $\overline{F}$ 中的元素集合 R 和 $\overline{R}$ 为生成多项式的根集合，所得生成多项式是一致的。在这种情况下，域 F 和 $\overline{F}$ 均是编码所在域。

经计算机仿真表明，码长为 $n=7\sim511(m=3\sim9)$，$\mu=1$ 或 2 时，存在上述元素 $\overline{\beta}$，即不等式 $\mathrm{card}(R^*)<\mathrm{card}(R)$ 成立。虽然尚不能证明对 $\mu\ge3$ 元素 $\overline{\beta}$ 的必然存在性，但仿真表明一般 $\mathrm{card}(R^*)<\mathrm{card}(R)$ 均成立，即编码所在域可以不止 1 个的情况是极少或根本不存在的。

虽然 $\mu=1,2$ 时一定存在元素 $\{\alpha^{21},\alpha^{42}\}$，但由于同构，若非编码所在域 $\overline{F}$ 中的元素 $\overline{\alpha},\overline{\alpha}^3$ 满足 $g(\overline{\alpha})\ne\overline{0}$ 且 $g(\overline{\alpha}^3)=\overline{0}$，这时若码组以小概率 $(1-p_1)$ 接受了 $\overline{\alpha}$ 满足：$g(\overline{\alpha})=\overline{0}$，那么在域 $\overline{F}$ 中必然会有 $g(\overline{\alpha}^3)=\overline{0}$，进而导致识别的出错。记满足以上性质的 $(F,\overline{F})$ 为临近域对，并记上述所有 F 的集合为 $\mathbb{F}$。为减小此情况下的出错概率，可提高在域 $\overline{F}$ 中搜索的概率门限 p_1。表 1 列举了部分出现这一现象的临近域对 $(F,\overline{F})$，其本原多项式以十进制表示。比如 $607=(1001011111)_2$ 对应的本原多项式为 $D^9+D^6+D^4+D^3+D^2+D^1+1$。

纠错能力 $t=2$ 的BCH码，算法的误判概率 p_b 与虚警概率 $p_{\mathrm{fb}1},p_{\mathrm{fb}2}$ 和漏警概率 $p_{\mathrm{mb}1},p_{\mathrm{mb}2}$ 满足

$$p_\mathrm{b}=\begin{cases}p_{\mathrm{fb}1}+p_{\mathrm{mb}1}\approx[1-p_3^{\varpi_h}]+[p_t(H_3)\\ \qquad+(1-p_\mathrm{t}(H_3))(1-p_2)], & F\notin\mathbb{F}\\ p_{\mathrm{fb}2}+p_{\mathrm{mb}2}\approx[(1-p_\mathrm{t}(H_3))p_2(1-(p_1\\ \qquad+(1-p_1)p_2))+p_{\mathrm{fb}1}]+p_{\mathrm{mb}1}, & F\in\mathbb{F}\end{cases}\tag{19}$$

对于 $t=1$ 的BCH码，算法的误判概率 p_a 与虚警概率 $p_{\mathrm{fa}1},p_{\mathrm{fa}2}$ 和漏警概率 $p_{\mathrm{ma}1},p_{\mathrm{ma}2}$ 满足

表 1 部分临近域对

$\mu=1$	(203,253)	(607,895)	(701,647)	(827,607)
$\mu=2$	(425,333)	(761,539)	(859,761)	(911,787)

$$p_\mathrm{a}=\begin{cases}p_{\mathrm{fa}1}+p_{\mathrm{ma}1}\approx[1-p_3^{\varpi_h}]+p_\mathrm{t}(H_3), & F\notin\mathbb{F}\\ p_{\mathrm{fa}2}+p_{\mathrm{ma}2}\approx[(1-p_\mathrm{t}(H_3))p_2^2(1-p_1)\\ \qquad+p_{\mathrm{fa}1}]+p_{\mathrm{ma}1}, & F\in\mathbb{F}\end{cases}\tag{20}$$

当 $p_2\to1$ 时，若 $F\notin\mathbb{F}$，$p_\mathrm{a},p_\mathrm{b}\to p_\mathrm{t}(H_3)$；若 $F\in\mathbb{F}$，$p_\mathrm{a},p_\mathrm{b}\to1-p_1+p_1p_\mathrm{t}(H_3)$；$p_\mathrm{c}\to p_\mathrm{t}(H_3)$。这表明 $F\in\mathbb{F}$ 且 $t=1,2$ 时，受 p_1 影响的虚警概率对算法误判概率影响较大。

在确定码长 n 后，根据在 $m=\log_2(n+1)$ 对应的所有域中，以 $H_2(m^*)$ 门限依次验证顺序的共轭根系表中各共轭根系中的某一个元素，得到各个域中能够被码字组接受的生成多项式的根，取根元素最多的域为编码所在域 F，并验证该域中单位元 e 是否能被接受为 $g(X)$ 的根。取域 F 中所有的根元素，并依据域 F 的结构计算出生成多项式 $g(X)$。若单位元 e 也是生成多项式的根，识别出的将是增余删信码[18]。

若搜索至最后一个域，若并不存在 α 的满足率 η 使不等式 $\Phi(\eta)\ge p_3$ 成立或同时通过 H_1 和 $H_2(m^*)$ 满足率门限的域，设 $m_{\max}=9$，p_s 为一个接近 1 的实数，所有使不等式

$$p_3>\Phi(\eta)\ge p_\mathrm{s}\tag{21}$$

成立的域为集合 F_s，定义 $\varsigma=1-p_\mathrm{s}^{\varpi_{m_{\max}+1}}$ 为接受 $t=1$ 的最大误判概率，即为非 BCH 编码被识别为能纠正 1 个误码的 BCH 编码的最大概率。取域集合 F_s 中使得 $\Phi(\eta)$ 最大的域为编码所在域，对应生成多项式仅以一个共轭根系为根。

对于码长不大于 $n=2^{m_{\max}}-1$ 的本原 BCH 码，本文所提识别方法步骤如下：

步骤 1 初始化：码长 $n=2^m-1$，$m=3$，域(本原多项式)为 11；

步骤 2 在当前域中验证本原域元素 α 的满足率 η 是否满足 $\eta<H_1$，若成立则转到下一个域中(假如还有，否则 $m=m+1$)，并继续验证 α 的满足率 η 是否满足 $\eta<H_1$（若 $m=m_{\max}+1$ 也未搜索满足 $\eta\ge H_1$ 的域，则待分析数据非本原 BCH 编码，结束程序），否则转步骤 3；

步骤 3 略除 $c(\alpha)\ne0$ 的码字，并在当前域中验证 α^3 的满足率 η 是否不小于 H_2。若 $\eta\ge H_2$，则转步骤 4，否则转步骤 5；

步骤 4 以 $H_2(m^*)$ 为满足率门限，在当前码长 n 对应的剩余的所有域内，按照顺序的共轭根系表，

验证并统计识别出来的根个数，选择根个数最多的域为编码所在域(码长为 n)，并转步骤 8；

步骤 5　比较 $\Phi(\eta)$ 与 p_3 的大小，若 $\Phi(\eta)\ge p_3$，则回转步骤 4，否则：转步骤 6，并判断 $\Phi(\eta)\ge p_s$ 是否成立，若成立则记录当前域和对应的 $\Phi(\eta)$，转到下一个域中(假如还有，否则 $m=m+1$)，若不成立则不记录；

步骤 6　判断 $m=m_{\max}+1$ 是否成立，若不成立回转步骤 2，否则转步骤 7；

步骤 7　判断是否存在被记录的域，若存在取 $\Phi(\eta)$ 最大的域及对应的码长为编码所在域和码长，转步骤 8，否则：待分析数据非本原 BCH 编码，结束程序；

步骤 8　以 $H_2(m^*=1)$ 为满足率门限，在编码所在域内验证单位元 e 是否为 $g(X)$ 的根，按照域的加法和乘法运算，计算 $g(X)$ 表达式，结束程序。

4　仿真分析

4.1 仿真实验与复杂度分析

图 2 显示的是搜索误码率为 0.05 的 BCH(63,18) 码生成多项式根时，在编码所在域内共轭根系中的一种搜索线路以及对应的码根满足率。除本原域元素 α 外，在验证其它元素时，使用的码字组均是能够通过本原域元素校验的码字，因此 $\alpha^3,\alpha^5,\cdots$ 比 α 的码根满足率高。从域间搜索的方法来看，算法以高概率 p_1 快速排除非编码所在域，没有重复迭代验证，因而搜索速度较快；从图中可以看出，$\alpha^3,\alpha^5,\cdots$ 等码根满足率都能在较高误码率的情况下保持较高且相对稳定的数值，说明算法对误码率不敏感，能够可靠地搜寻出所有生成多项式的根，具有较高可靠性；算法并未按元素指数完全顺序搜索，而是按顺序共轭根系表搜索，验证失败则停止，提高了搜索速度。注意到 α^{21} 的码根满足率大致接近 0.25，虽然远高于 α 的满足率，但因为 α^{21} 所在共轭根系集合 $\{\alpha^{21},\alpha^{42}\}$ 中仅仅只有 2 个元素，即 $m^*=2$，码根满足率的期望等于 $1/2^{m^*}=0.25$，所以 α^{21} 并不被接受为 $g(X)$ 的根。

表 2 列举了已有算法和本文所提算法的复杂度，其中包括 $m=h,3\le h\le m_{\max}$ 时各种算法的最坏情况，以及 $m=9$ 的最坏情形和 $m=3$ 时的最好情形。式中 N_i 表示 $m=i$ 时待分析数据 $\widehat{B}$ 可划分的最大完整码字数；$n_i=2^i-1$，等于码长，且 $n_i\gg\omega_i=\phi(n_i)/i$，其中 $\phi(n_i)$ 为欧拉函数，表示 $\{0,1,\cdots,n_i-1\}$ 中与 n_i 互素的元素个数；c_i 表示 $m=i$ 时验证 1 个码字是否以某个元素为根的计算复杂度，是一常量，$3\le i\le m_{\max}$。

由表 2 可得，本文算法不仅在最好的情形下有 $5N_3c_3\ll\sum_{i=3}^{9}N_in_ic_i<\sum_{i=3}^{9}N_in_ic_i+N_3n_3c_3$，而且在 $m=9$ 时的最坏情形有 $\sum_{i=3}^{8}N_i\omega_ic_i\ll\sum_{i=3}^{8}N_in_ic_i$，且 $(N_9n_9c_9/9)+N_9\omega_9c_9\ll N_9n_9c_9$，s.t. $p_1\to1$, $p_2\to1$，因此算法的计算复杂度优于文献[11]，更优于文献[10]。

4.2 仿真性能

随着码长的增加，BCH 码的性能就会逐渐变坏[18]，因而在实际中应用的多是中短码。仿真实验选择的码长变化范围为 $n=7\sim511(m=3\sim9)$，随机选择所有可选的信息位长度 k，随机选择码长对应的所有可选域为编码所在域，每种编码对应的分析比特数均为 2×10^5。参数设置：$p_1=1-10^{-2}$，$p_2=1-10^{-6}$，$p_s=1-10^{-5}$，依据 $m=3\sim9$ 中本原多项式的个数计算可得 $\varsigma=9.8\times10^{-4}$。图 3 显示了本文所提算法的识别概率与误码率的关系。从图中可知，识别概率随码长的增加而变差。图 3 还给出了在相同数据量下文献[11]所提算法在 $m=7$ 且第 4 步参数为 0.8 时的误码率-识别率右限，由图可见本文所提算法的抗误码能力有较大提升。

由式(9)，式(14)，式(15)，式(19)和式(20)可得，误判概率随误码率增加而减小。若 $F\notin\mathbb{F}$，误判概率会随 p_1 的增大而增大，但若误码率较小，其增大可以忽略不计；若 $F\in\mathbb{F}$，误判概率会随 p_1 的增大而减小。若 p_3 不变，降低 p_1 可以获得更好的抗误码性能，但计算量增加。若 $p_1=p_2=1$，对应的误判概率为零，但此情况只适用于无误码或极低误码。

据式(2)及式(4)可知，方差 δ_1^2 及 δ_2^2 随码字数 N

表 2 算法的复杂度分析

	文献[10]	文献[11]	本文所提算法
m=h	$(\omega_h-1)N_hn_hc_h+\sum_{i=3}^{m_{\max}}N_in_ic_i$	$\sum_{i=3}^{m_{\max}}N_in_ic_i$	$(N_hn_hc_h/h)+(3+p_1p_2-p_2-2p_1)\sum_{i=3}^{h}N_i\omega_ic_i$
m=9	$(\omega_9-1)N_9n_9c_9+\sum_{i=3}^{9}N_in_ic_i$	$\sum_{i=3}^{9}N_in_ic_i$	$(N_9n_9c_9/9)+(3+p_1p_2-p_2-2p_1)\sum_{i=3}^{9}N_i\omega_ic_i$
m=3	$N_3n_3c_3+\sum_{i=3}^{9}N_in_ic_i$	$\sum_{i=3}^{9}N_in_ic_i$	$5N_3c_3$

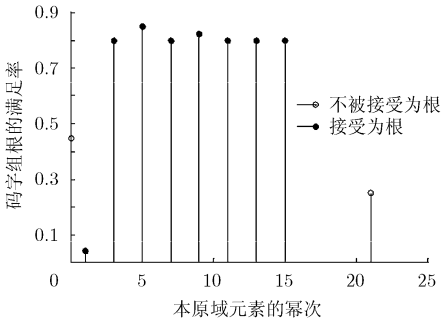

图 2 BCH(63,18)根的搜索过程及码字根满足率

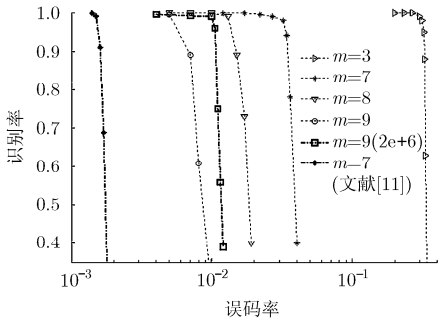

图 3 不同码长在各种误码率下的识别率

的增加反比例减小，因而算法性能会随数据量的增加而变好。图 3 中还给出了比特数为 2×10^6，$m=9$ 时算法对应的识别性能，相应 90%识别率对应的误码率上限达到了 1.05×10^{-2}；可见，算法的识别性能随数据量的增加而提升，验证了理论分析结果。不难得出，$m=3\sim8$ 时 BCH 码的识别性能也都会随数据量的增加而得到提升。

对于某确定但未知的 BCH 码，设 $m=h$，若数据量增加 ζ 倍，近似地 $N_i^{'}=\zeta N_i$，$i=3\sim h$。由表 3 可得，算法的复杂度增加 ζ 倍。由此可见，算法复杂度随输入规模的增加而线性增加，实际应用中可根据需要来平衡计算复杂度和抗误码性能。

5 结束语

本文主要研究了较为通用的本原 BCH 码的参数识别问题，所提方法能准确识别码长，编码所在域，生成多项式等参数。方法利用概率门限容许更多的误码，利用本原域元素的校验作用排除误码对生成多项式的干扰，提升识别可靠性；搜索到正确参数则停止验证，避免了各种迭代验证，因而计算速度快；对某确定 BCH 码型，方法的复杂度随输入数据规模的增加而线性增加，且增加数据量可使性能变好。后续的研究主要针对缩短 BCH 码，拓展本原 BCH 码，非本原 BCH 码等码型的参数识别，以及完善 BCH 码的参数盲识别。

参 考 文 献

[1] Naseri Ali, Azmoon Omid, and Fazeli Samad. Blind recognition algorithm of Turbo codes for communication intelligence systems[J]. *International Journal of Computer Science Issues,* 2011, 8(6): 68–72.

[2] Jia Yong-qiang, Li Li-ping, Li You-zhu, *et al*.. Blind estimation of communication emitter features parameters[C]. IEEE 12th International Conference on Computer and Information Technology, Chengdu, 2012: 281–285.

[3] 刘建成, 杨晓静. 基于求解校验序列的(n,1,m)卷积码盲识别[J]. 电子与信息学报, 2012, 34(10): 2363–2368.
Liu Jian-cheng and Yang Xiao-jing. Blind recognition of (n,1,m) convolutional code based on solving check-sequence[J]. *Journal of Electronics & Information Technology*, 2012, 34(10): 2363–2368.

[4] Lu Pei-zhong, Li Shen, Zou Yan, *et al*.. Blind recognition of punctured convolutional codes[J]. *Science in China Series F: Information Sciences*, 2005, 48(4): 484–498.

[5] 于沛东, 李静, 彭华. 一种利用软判决的信道编码识别新算法[J]. 电子学报, 2013, 41(2): 301–306.
Yu Pei-dong, Li Jing, and Peng Hua. A novel algorithm for channel coding recognition using soft-decision[J]. *Acta Electronica Sinica*, 2013, 41(2): 301–306.

[6] Dingel Janis and Hagenauer Joachim. Parameter estimation of convolutional encoder from noisy observations[C]. IEEE International Symposium on Information Theory, Nice, 2007: 1776–1780.

[7] Cote Maxime and Sendrier Nicolas. Reconstruction of a Turbo-code interleaver from noisy observation[C]. IEEE International Symposium on Information Theory, Austin, 2010: 2003–2007.

[8] Barbier Johann. Reconstruction of Turbo-code encoders[C]. Proceedings of Denfense and Security Symposium, Space Communication Technologies Conference, Orlando, 2005: 463–473.

[9] Cluzeau Mathieu, Finiasz Matthieu, and Jean-Pierre Tillich. Methods for the reconstruction of parallel Turbo codes[C]. IEEE International Symposium on Information Theory, Texas, 2010: 2008–2012.

[10] 杨晓静, 闻年成. 基于码根信息差熵和码根统计的 BCH 码识别方法[J]. 探测与控制学报, 2010, 32(3): 69–73.
Yang Xiao-jing and Wen Nian-cheng. Recognition method of BCH codes based on roots information dispersion entropy and roots statistic[J]. *Journal of Detection & Control*, 2010, 32(3): 69–73.

[11] 吕喜在, 黄芝平, 苏绍璟. BCH 码生成多项式快速识别方法[J]. 西安电子科技大学学报, 2011, 38(6): 159–172.
Lü Xi-zai, Huang Zhi-ping, and Su Shao-jing. Fast recognition method for generator polynomial of BCH codes[J].

Journal of Xidian University, 2011, 38(6): 159–172.

[12] 甘露, 周攀. 基于中国剩余定理分解的 RS 码快速盲识别算法[J]. 电子与信息学报, 2012, 34(12): 2838–2842.
Gan Lu and Zhou Pan. Fast blind recognition method of RS codes based on chinese remainder theorem decomposition[J]. *Journal of Electronics & Information Technology*, 2012, 34(12): 2838–2842.

[13] Cluzeau Mathieu and Finiasz Matthieu. Recovering a code's length and synchronization from a noisy intercepted bitstream[C]. IEEE International Symposium on Information Theory, Seoul, 2009: 2737–2741.

[14] Cluzeau Mathieu. Block code reconstruction using iterative decoding techniques[C]. IEEE International Symposium on Information Theory, Seattle, 2006: 2269–2273.

[15] 杨晓炜, 甘露. 基于 Walsh-Hadamard 变换的线性分组码参数盲估计算法[J]. 电子与信息学报, 2012, 34(7): 1642–1646.
Yang Xiao-wei and Gan Lu. Blind estimation algorithm of the linear block codes parameters based on WHT[J]. *Journal of Electronics & Information Technology*, 2012, 34(7): 1642–1646.

[16] Liu Xiao-bei, Koh Soo Ngee, Wu Xin-wen, *et al*.. Reconstructing a linear scrambler with improved detection capability and in the presence of noise[J]. *IEEE Transactions on Information Forensics and Security*, 2012, 7(1): 208–218.

[17] Lin Shu and Costello D J. Error Control Coding: Fundamentals and Applications[M]. Upper Saddle River: Pearson Prentice Hall, 2004: 197, 234–238.

[18] 王新梅, 肖国镇. 纠错码 — 原理与方法(修订版)[M]. 西安: 西安电子科技大学出版社, 2011: 117,124,147,254.
Wang Xin-mei and Xiao Guo-zhen. Error Correcting Code——Theory and Method(Revised Edition)[M]. Xi'an: Xidian University Publishing Company, 2011: 117, 124, 147, 254.

[19] Soong T T. Fundamentals of Probability and Statistics for Engineers[M]. Chichester: John Wiley & Sons, 2004: 182–183.

阔永红：　女，1967 年生，博士，教授，研究方向为智能信号处理.
曾伟涛：　男，1988 年生，硕士生，研究方向为通信编码识别技术.
陈　健：　男，1968 年生，教授，博士生导师，研究方向为通信对抗、认知网络.