doi:10.3969/j.issn.1001-2400.2013.01.019

一种利用本原元的快速 RS 码盲识别算法

王　平[1]，曾伟涛[1]，陈　健[1]，陆继翔[2]

(1. 西安电子科技大学 通信工程学院,陕西 西安　710071;
2. 西北工业大学 机电学院,陕西 西安　710072)

摘要: 建立了一种 RS 码编码参数的盲识别新方法.该方法利用本原元的校检作用并行搜索码长和域,提高了盲识别的效率;略掉不符合本原元校检的码字,增强了码根搜寻的可靠性;利用码根的连续性采用前进-倒退法搜索生成多项式,简化计算,提高了搜索速度.仿真结果表明,新算法在 90%识别率的误比特率上限上有明显提高.

关键词: RS 码;信道编码;盲识别;本原域元素

中图分类号:TN919　　**文献标识码**:A　　**文章编号**:1001-2400(2013)01-0105-06

Fast blind recognition algorithm for RS codes by primitive element

WANG Ping[1], *ZENG Weitao*[1], *CHEN Jian*[1], *LU Jixiang*[2]

(1. School of Telecommunication Engineering, Xidian Univ., Xi'an　710071, China;
2. School of Mechanical Eng., Northwestern Polytechnical Univ., Xi'an　710072, China)

Abstract: A new algorithm is presented to solve RS code blind recognition problems. By searching the code length and the field parallelly based on the parith check function of the primitive element, the recognition efficiency is improved. The reliability of roots' searching is promoted by omitting error contained code words which do not accord with the primitive element parith check. The generator polynomial is forward-backward searched using the continuity of roots to simplify calculations and accelerate the search speed. Simulation results indicate that the new algorithm's upper limit of BER when correct recognition rate is 90% has a significant increase.

Key Words: Reed-Solomon codes; channel coding; blind recognition; primitive element

在非合作通信侦察领域和智能通信中,编码识别为信息解码或解译提供了重要的支撑和保障,因而编码的盲识别问题日显重要[1]. RS 码是一类纠错能力很强的多进制 BCH 码,是线性分组码中较为重要的一类[2]. RS 码现已广泛应用在深空通信、移动通信、扩频数据通信、军用通信中.

目前 RS 码的盲识别鲜有国外的研究成果公布,国内近几年陆续出现了一些研究成果.文献[3]最早给出 RS 码盲识别的算法,但其分为无误码、低误码和高误码 3 种情况,这在实际中无法预先得知,并且该算法的矩阵运算量巨大,抗误码能力较弱.文献[4-5]利用无误码字与其 m 位移位码字的最大公约式是生成多项式的倍式这一特点,从而识别出一系列参数,但是其发现误码的能力有限.文献[6-7]利用了编码冗余和生成多项式的根是码字根的子集的特性,但在分析码长时遍历各种码长并多次大规模的矩阵运算限制了该算法的识别速度,且具有较高码率的编码方式,因其冗余度低,加之误码影响,降低了识别的准确性.文献[8]利用同构的原理来减少搜索量,但是无法保证域(本原多项式)的识别,且化为二进制后矩阵规模增大,必然导致计算量的相应增加.

收稿日期:2012-08-29　　**网络出版时间**:2012-09-25
基金项目:高等学校学科创新引智计划资助项目(B08038)
作者简介:王　平(1978-),女,讲师,E-mail: pingwang@mail.xidian.edu.cn.
网络出版地址:http://www.cnki.net/kcms/detail/61.1076.TN.20120925.1056.201301.134_019.html

针对以上提及的RS编码盲识别算法的各种不足,笔者进行了改进,利用本原域元素在RS码中的校检作用来并行搜索码长和域,并继续利用码字根和生成多项式根的包含关系来识别生成多项式.利用本原元约束,首先舍弃不满足以本原元为根的含误码码字,提升了生成多项式的识别能力.

1 RS码基础及识别原理

1.1 RS码定义

RS码属于多进制BCH码中一种重要的码类,其根域和符号域相同.对于多进制本原BCH码有如下定义.

定义1 设α是$GF(q^{m_0})$域中的本原域元素,有t个误码纠错能力的q进制本原BCH码的生成多项式$g(x)$是以$GF(q)$中$\alpha^1,\alpha^2,\cdots,\alpha^{2t}$为根的次数最小的多项式[9].

$GF(q)$为有限域,则$g(x)=g_{\mathrm{LCM}}(m_1(x),m_2(x),\cdots,m_{2t}(x))$. BCH码码长$n=n_{\mathrm{LCM}}(e_1,e_2,\cdots,e_{2t})$,其中$e_i$是$\alpha^i$的级.$m_i(x)$是以$\alpha^i$为根的最小多项式,其次幂最高为$m_0$,因而$g(x)$的次幂最高是$2m_0t$.如果$m_0=1$,得到的码就是RS码,其码长为$q-1$,$\alpha^i$的最小多项式是$x-\alpha^i$,生成多项式的次幂是$2t$.可以证明,$g(x)$的系数均不为零,因而最小码距是$2t+1$[9].

定义2 $GF(q)(q\neq 2)$上,码长$N=q-1$的本原BCH码称为RS码[3].

从RS码的定义可知,其生成多项式必以本原域元素(本原元)为根[9].

1.2 RS码盲识别原理

定义3 若α是域$GF(q)$中的n级元素,则称α是n次单位原根.若$GF(q)$中的某一元素α的级是$q-1$,则称α为本原域元素[2].

在RS码编码所在域$GF(q)\mid q=2^m$中,若码字多项式$C(x)=c_{2m-2}x^{2m-2}+\cdots+c_1x+c_0$,则该码字以本原元$\alpha$为根的多项式表达式为$C(\alpha)=c_{2m-2}\alpha^{2m-2}+\cdots+c_1\alpha+c_0=0$.若用矩阵表达,要先利用$\alpha^{2^m-1}=1$化简,然后用等价的$m$重表达,最终化为矩阵运算.

定理1 码字多项式$C(x)=q(x)g(x)$,则在特定域中,$g(x)$的根必然是$C(x)$的根[2].

该定理为从截获的大量码字中提取出公共根奠定了最基本的理论基础.但如何判别公共根和非公共根的问题还未得到解决.针对这一问题,提出如下的定理2和定理3,以区分出公共根和非公共根.

定理2 对于编码参数为$[n,k]$的RS码,任意一个长度为n的随机码字,以本原元为根的概率$p=(1/2)^m$,其中$n=2^m-1$.

证明 因为码字是长度为n的随机码字,即$C_\xi(x)=c_{2m-2}x^{2m-2}+\cdots+c_1x+c_0$,其码字向量$(c_{2m-2},c_{2m-3},\cdots,c_1,c_0)$是取自域$GF(2^m)$上的多进制随机向量,用$\alpha$的幂次表示为$\xi_i\in\{0,1,\cdots,2m-2\}$,$i=0,1,2,\cdots,2m-2$.该随机码字多项式以某域中本原元为根,因此$C(\alpha)=0$,则有

$$\alpha^{\xi_{2m-2}}\alpha^{2m-2}+\alpha^{\xi_{2m-3}}\alpha^{2m-3}+\cdots+\alpha^{\xi_1}\alpha^1+\alpha^{\xi_0}=\alpha^{\xi_{2m-2}+2m-2}+\alpha^{\xi_{2m-3}+2m-3}+\cdots+\alpha^{\xi_1+1}+\alpha^{\xi_0}=$$
$$\zeta_{m-1}\alpha^{m-1}+\zeta_{m-2}\alpha^{m-2}+\cdots+\zeta_1\alpha+\zeta_0 \quad . \tag{1}$$

ξ_i是$\{0,1,\cdots,2m-2\}$内的随机变量,则$\zeta_i(i=0,\cdots,m-1)$服从等概的$(0,1)$分布.因此$C(\alpha)=0$,随机码字以本原元为根的概率等于$p(\zeta_i=0\mid i=0,\cdots,m-1)=(1/2)^m$.命题得证.

RS码必然以本原元为根,在编码域内,RS码字必然以概率1满足$C(\alpha)=0$.对比随机码字可知,只有在正确的域和码长下,码字以本原元为根才会有较高的概率.不难看出,误码率越低,两者区分度越大.RS码长长度$n=2^m-1$,其中m一般取3~8,BCH码的性能会随码长的增加而变坏[2],它达不到香农编码定理所要求的能力,因此一般使用的都是中短码.

定理3 对于编码参数为$[n,k]$的RS码,不是$g(x)$根的元素是$C(x)$的根的概率$p=(1/2)^m$,其中$n=2^m-1$.

定理3的证明与定理2的类似.定理2适用于利用本原元搜索码长和本原多项式.定理3适用于在搜索到可能的码长和域时,利用公共根和非公共根之间的概率差别搜索出所有的公共根.

2　RS 编码盲识别的实现

RS 码的识别，其搜索空间可以概括为 3 个维度：码长，域，根. 在 3 个维度全部确立后，RS 编码方式也被惟一确定. 文献[3-5，7]单独搜索每个维度，而新算法采用并行方式进行前两维的搜索，有效地避免了因前两维独立识别而降低了整体识别概率的不利影响. 文献[8]利用域的同构这一特性，省略了第 2 个维度的搜索，在一定程度上简化了运算，但并未提升识别性能.

为了尽可能地降低误码的干扰，新算法还提出了利用本原元约束，舍弃不满足以本原元为根的含误码码字. 仿真结果表明，这一改进有效地提升了生成多项式的识别性能.

2.1　码长和域(本原多项式)

当 m 小于(不等于)真实 m^* 时，在依据 m 构造的域中，真正的码字被分解成不规则的多段，致使当前假设长度的码字就类似于随机的码字.

运用定理 2，N 个随机码字中，本原元满足的码字个数比率 x 的概率分布以及方差分别为

$$\begin{cases} p(x)=C_N^{Nx}(1/2)^{Nxm}(1-(1/2)^m)^{N-Nx} \quad , \\ \delta^2=(1/2)^m(1-(1/2)^m)/N \quad . \end{cases} \tag{2}$$

由二项分布的性质，可知

$$E(x)=(1/2)^m \quad . \tag{3}$$

当 m 等于真实 m^* 且搜寻到编码所采用的域时，若无误码，则符合率为 1. 在真实的码长和域及非真实的码长和域中，符合率概率分布差异较大.

据此可以设定一个符合率门限：大于该门限时，即认为搜寻到了可能正确的码长和域. 首先设置一个较低的门限用来抗误码，以防漏掉真实的解；后续再设置一个门限，以防接受错误的解. 图 1 是 $m=3$，$N=100$ 时的随机码字和 RS 码误码率 $P_e=0.05$ 时的 RS 码字分别以本原元为根的符合率概率分布图. 由图 1 可知，随机码字和 RS 码字在该指标上有巨大的分布特性差异，据此可以利用分布差异设定门限，在一定误码率下用来搜索目标码长和域. 同样地，误码率为 P_e 下的 RS 码，单个码字满足以本原元为根的概率 P_{es} 和满足以本原元为根的码字比率 x_c 的概率分布以及方差为

$$\begin{cases} P_{es}=(1-P_e)^{2^m-1}+[1-(1-P_e)^{2^m-1}](1/2)^m \quad , \\ p(x_c)=C_N^{Nx_c}(P_{es})^{Nx_c}(1-P_{es})^{N-Nx_c} \quad , \\ \delta^2=P_{es}(1-P_{es})/N \quad . \end{cases} \tag{4}$$

由二项分布，易知

$$E(x_c)=P_{es} \quad . \tag{5}$$

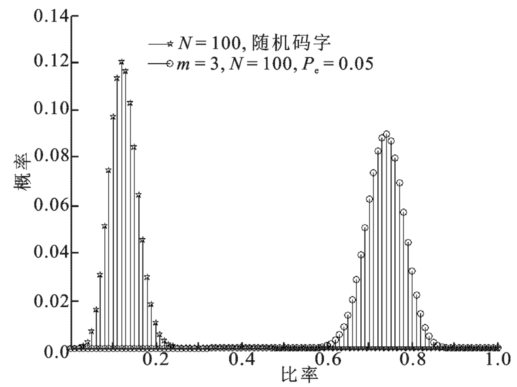

图 1　随机码字和有误码 RS 码以本原元为根的符合率

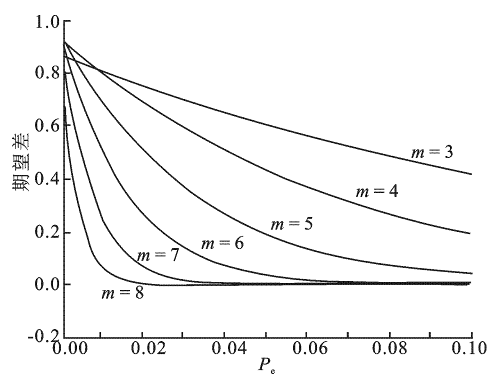

图 2　含误码 RS 码与随机码字的本原元符合率期望差

图 2 是 m 为 3～8，P_e 为 0.001～0.10 下含误码 RS 码与随机码字的以本原元为根的符合率期望之差. 随着 m 和 P_e 的增大，两种概率分布相距越近，算法性能将因此受限.

根据二项分布的极限是正态分布，即随机变量 X 服从参数为 n 和 p 的二项分布，则

$$\lim_{N\to\infty}\frac{X-np}{\left(Np(1-p)\right)^{1/2}}\sim N(0,1)\quad . \tag{6}$$

若截获的码字数目充分多,可以认为随机变量 x 和 x_c 分别服从

$$\lim_{N\to\infty}\frac{x-(1/2)^m}{\left((1/2)^m(1-(1/2)^m)/N\right)^{1/2}}\sim N(0,1)\quad , \tag{7}$$

$$\lim_{N\to\infty}\frac{x_c-P_{\mathrm{es}}}{\left(P_{\mathrm{es}}(1-P_{\mathrm{es}})/N\right)^{1/2}}\sim N(0,1)\quad . \tag{8}$$

查标准正态分布函数数值表可得,若要以概率 P_m 排除随机码字的干扰,当码字数 N 为1 000时 x 的取值上限(门限) T_1 满足

$$T_1=(1/2)^m+t_{P_m}\left((1/2)^m(1-(1/2)^m)/N\right)^{1/2}\quad , \tag{9}$$

其中,t_{P_m} 指在正态分布中以概率 P_m 排除随机码字干扰对应的横轴值.

在当前搜索域和码长中,算法以 $1-P_m$ 的概率接受信息序列以本原元为根.假设搜索到的当前域就是目标域,略掉那些不符合本原元为根的码字,即使当前域并不是目标域,但剩下的"码字"依然不会改变随机特性.因此不会从根本上改变"码字"规律,进而影响跳出错误的码长和域所构成的空间.因为本原元检错的能力有限,所以并不能略掉所有含有误码的码字.

一般地,RS 码的监督矩阵[2]为

$$\boldsymbol{H}=\begin{pmatrix}\alpha^{n-2} & \cdots & \alpha^2 & \alpha & 1\\ (\alpha^{i_2})^{n-2} & \cdots & (\alpha^{i_2})^2 & \alpha^{i_2} & 1\\ \vdots & & \vdots & \vdots & \vdots\\ (\alpha^{i_{n-k}})^{n-2} & \cdots & (\alpha^{i_{n-k}})^2 & \alpha^{i_{n-k}} & 1\end{pmatrix}\quad , \tag{10}$$

其中,α 是本原域元素.从上面的过程可以看出,该方法只用到了监督矩阵的第 1 个约束.随机码字、含误码码字能通过本原元的约束,但不能通过监督矩阵其他约束的概率不大于 p_c,$p_c=(1/2)^m(1-(1/2)^{m(n-k-1)})$.由于这个概率较低,所以绝大部分含有误码的码字都会被略掉.

由 $\overline{p_c}=(1/2)^m$,取最坏情形 $p_c=(1/2)^m$,则略掉部分误码后仍然含误码码字的比率为

$$R_{\mathrm{cor}}=(1-R_{p_{\mathrm{e}}})/(1-R_{p_{\mathrm{e}}}+2^mR_{p_{\mathrm{e}}})\quad , \tag{11}$$

其中,$R_{p_{\mathrm{e}}}$ 指误码率 p_{e} 下码字无误码的比率.在误码率较低的情况下,$R_{p_{\mathrm{e}}}$ 接近 1.

容易得到处理后的含误码字比例和未处理过含误码字比例的比率,可以表达为

$$\gamma=R_{\mathrm{cor}}/(1-R_{p_{\mathrm{e}}})=1/(1-R_{p_{\mathrm{e}}}+2^mR_{p_{\mathrm{e}}})\quad . \tag{12}$$

可以看出,在低误码率时,处理后的误码字比例大约下降到原误码字比例的 $1/2^m$.因此,相对于不略掉误码码字的算法,新算法可靠性较高.

由 RS 码的编码定义可知,RS 码是一种根域和符号域一致的 BCH 码,根的个数和检验符号个数是相同的,都是 $n-k=2t$[2,9].这就表明了以本原元 α 为根,就必然也以 α^2 为根.

因为存在随机码字,也以一个较小的概率 $1-P_m$ 满足以本原元和其他元素为根,所以为了排除错误的码长和域,P_m 的选取应有如下概率限制:

$$P_{\mathrm{r}}=\prod_{m=3}^{m^*}\left(1-(1-P_m)P_{\mathrm{sec}}\right)^{n_m}\quad , \tag{13}$$

其中,P_{r} 指的是在搜索到正确的码长和域之前排除随机码字干扰的概率;P_{sec} 指的是错误接受第 2 个元素为根的概率;n_m 是指 m 时本原多项式的数目或 $m=m^*$ 时,目标域之前的所有 m^* 阶域的个数.这里为简化问题复杂度,舍弃其他元素的微小影响.为了保证识别的准确性,P_{r} 必须接近 1.

因为略掉了部分码字,留下的码字数目将会小于 N.考虑到长码长和较高误码的情形,假设仅剩 N_{re} 组码字没有被略去,则由式(8)可得

$$T_2=(1/2)^m+t_{P_{\mathrm{sec}}}\left((1/2)^m(1-(1/2)^m)/N_{\mathrm{re}}\right)^{1/2}\quad , \tag{14}$$

其中，T_2 是域中第 2 个及其他元素根验证时的门限，$t_{P_{sec}}$ 指在正态分布中以概率 P_{sec} 错误接收第 2 个根对应的横轴值. 由于在验证第 2 个元素的时候已经略掉了大部分含误码的根，因此可以通过提高接受第 2 个元素为根的门限，使 P_{sec} 尽可能得小. 若考虑到最坏时的情形，则有 P_{sec} 和 P_m 满足

$$P_r = \left(1-(1-P_m)P_{sec}\right)^{157} \quad , \tag{15}$$

即意味着目标域是 $m=8$ 时的最后一个域.

2.2　生成多项式

由定理 2 可知，$g(x)$ 的根必然是 $C(x)$ 的根. 又由定理 3 可知，不是 $g(x)$ 根的元素是 $C(x)$ 的根的概率为 $(1/2)^m$. 类似地，在验证 $e,\alpha^2,\cdots,\alpha^{2^m-2}$ 是否为根时，门限取值可参考式(13)～(15).

得到所有可能的根后，根据 RS 码根的特点，若根的个数小于 2 或不是偶数[9]，则继续在下一个域中搜索，即如前所述的第 2 道门限. 得到 $g(x)$ 的根后，则其表达式 $g(x)=(x-\alpha)\cdots(x-\alpha^{n-k})$，利用所在的域结构化成整数域的表达形式即可. 图 3 是 RS 码盲识别算法流程图.

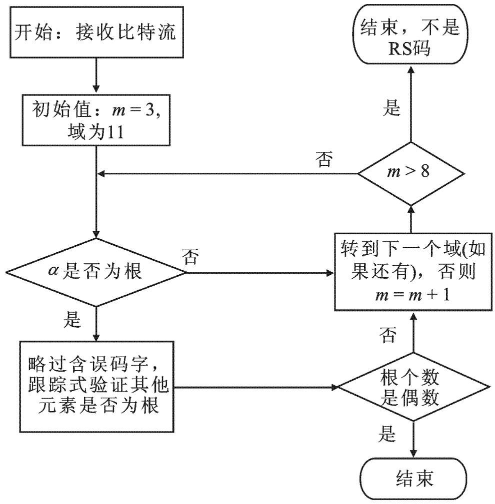

图 3　RS 码盲识别算法流程图

2.3　生成多项式识别的改进

文献[2]提到了 GFFT 运算[3]，但实际上并不需要进行完整的 GFFT 运算. 由定义 1 知，其根本身是连续的，且 GFFT 运算的实质约束了只需寻找到若干连续的码根，即只需得到码根的个数. 在进行搜索时，笔者采用前进-倒退搜索法，而非依 $\alpha^2,\alpha^3,\alpha^4,\cdots,\alpha^{2^m-2}$ 次序遍历搜索. 若搜索发现其不是码根，则倒退. 例如，$\alpha,\alpha^2,\alpha^3,\alpha^4,\cdots,\alpha^{18}$ 是生成多项式的根，则只需搜索 $e,\alpha,\alpha^2,\alpha^4,\alpha^8,\alpha^{16},\alpha^{30},\alpha^{23},\alpha^{19},\alpha^{17},,\alpha^{18}$，这样搜索就减少了近一半运算量. $n-k$ 越大，运算量减小得就越明显.

3　仿真实验

仿真实验以 RS(31，13) 为例，然后生成 1 000 组码字. 截获的码字是借助帧同步而找到的码字的起始点，而且是等价的加过高斯白噪声的二进制比特流. 首先，依次增大可能的码长，$n=2^m-1$，并相应地生成对应阶的各种域(利用本原多项式)，本原多项式个数[8]见表 1. 再验证域中本原元是否是各个假设码长的码字的根，并统计以本原元为根的码字符合率. 仿真中使用的门限是当 $P_m=0.99$ 时，对应的门限. 验证其他元素时的门限取 $P_{sec}=1\times10^{-4}$，$N_{re}=10$ 对应的各种门限. 考虑运算速度，每次搜索只需验证 1 000 组码字. 图 4 是 RS(31,13)码在应用前进-倒退法跳跃验证码根时的统计频次图. 由图可看出，一方面通过本原元约束而通不过 α^2 约束的码字个数很小，另一方面在目标域中仅需要 11 次搜索就找到了全部码根 $\alpha,\alpha^2,\alpha^3,\alpha^4,\cdots,\alpha^{17},\alpha^{18}$，而不必搜索 31 次才确定所有的根. 经过计算，RS(31,13)编码所在的域为 $37(D^5+D^2+1)$，码长为 31 ($m=5$)，进而得到 $g(x)$ 多项式的系数为(次幂从高到低)：1，31，7，25，14，22，24，9，5，4，30，5，27，7，29，31，12，14，27. 可以看出这些系数全部不为零，所得结果与 RS 码的极大最小距离的特点相符. 若要达到 90%以上识别率 ($N=1\,000$)，RS 码误

表 1　本原多项式的个数与其阶数的关系

m	3	4	5	6	7	8
本原多项式个数	2	6	6	18	16	48

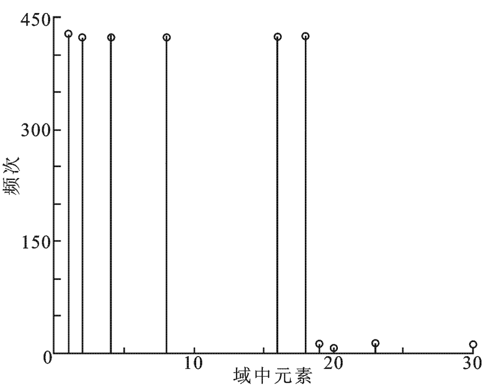

图 4　RS (31，13)码码根搜索过程元素频次图

比特率上限值见图 5. 其中文献[3，6]使用的数据量分别为 $N=1\,000$，$N=3\,000$，文献[4]使用的数据量为 2.04×10^{6} bit ($m=3$ 时，$N=97\,142$；$m=8$ 时，$N=1\,000$). 由图可知，在较少截获数据量的情况下，算法也能有较高的识别性能.

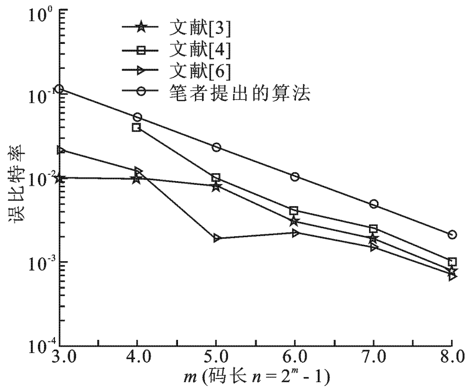

图 5 几种主要算法的 90%识别率误比特上限

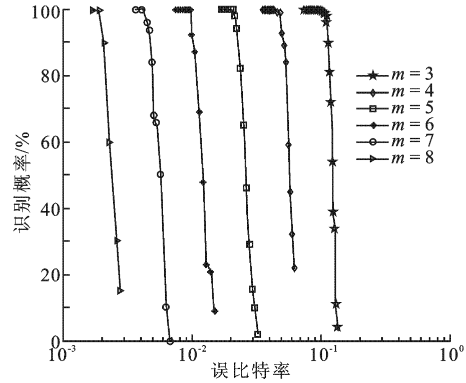

图 6 m 为 3～8 时算法识别性能

新算法采用并行搜索，仅仅利用本原元进行搜索(一般地，在非目标域中绝大部分只验证本原元，最多验证两个元素)，并且采用前进-倒退法跳跃寻根，在可能的搜索域中减少了运算量. 相对于其他算法中的多次大规模的矩阵求秩运算，新算法能够快速、有效地进行 RS 码盲识别.

图 6 是 m 为 3～8 时在各种误比特率下的识别概率($N=1\,000$)，其中选取的码型分别是：RS(7,3)，RS(15,11)，RS(31,13)，RS(63,57)，RS(127,111)，RS(255,223). 当 m 不同时，性能下降区间相邻较远，这与码长指数增长有关.

4 结　　语

在误码率较高的情况下，新算法在识别速度和准确度两方面都得到了改进. 通过概率期望的分析，利用本原元的约束限制，首先搜索到码长和域，再略掉含错码字，寻找到正确码根，进而完成全部识别. 与已有算法相比，不仅速度更快，而且 90%识别率的误码率上限也有较高的提升.

参考文献：

[1] 张永光. 一种 Turbo 码编码参数的盲识别方法[J]. 西安电子科技大学学报, 2011, 38(2): 167-172.
Zhang Yongguang. Blind Recognition Method for the Turbo Coding Parameter[J]. Journal of Xidian University, 2011, 38(2): 167-172.

[2] 王新梅, 肖国镇. 纠错码-原理与方法[M]. 修订版. 西安: 西安电子科技大学出版社, 2002.

[3] 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.
Liu Jian, Xie Nuo, Zhou Xiyuan. Blind Recognition of RS Code[J]. Journal of University of Electronic Science and Technology of China, 2009, 38(3): 363-367.

[4] 戚林, 郝士崎, 王磊, 等. 一种 RS 码快速盲识别方法[J]. 电路与系统学报, 2011, 16(2): 71-76.
Qi Lin, Hao Shiqi, Wang Lei, et al. A Fast Blind Recognition Method of RS Codes[J]. Journal of Circuits and Systems, 2011,16(2): 71-76.

[5] 戚林, 郝士崎, 李今山. 基于有限域欧几里德算法的 RS 码识别[J]. 探测与控制学报, 2011, 33(2): 63-67.
Qi Lin, Hao Shiqi, Li Jinshan. Recognition Method of RS Codes Based on Euclidean Algorithmin Galois Field[J]. Journal of Detection &Control, 2011, 33(2): 63-67.

[6] 闻年成, 杨晓静. RS 码的盲参数识别[J]. 计算机工程与应用, 2011, 47(19): 136-139.
Wen Niancheng, Yang Xiaojing. Blind Recognition of RS Codes Parameters[J]. Computer Engineering and Applications, 2011, 47(19): 136-139.

(下转第 168 页)