一种构造GC常重量DNA码的方法

梁 静*[①] 李红菊[①] 赵 凤[①] 丁 健[②]

[①](安徽新华学院通识教育部 合肥 230088)

[②](福建师范大学数学与信息学院 福州 350117)

摘 要：GC重量是DNA码的一个重要参数，如何构造满足GC常重量约束的DNA码是一个有趣的问题。该文通过在DNA码与四元码之间建立一个双射，将构造满足GC常重量约束的DNA码转化为构造GC常重量四元码。通过代数的方法，构造了3类满足GC常重量约束的DNA码。

关键词：四元码；DNA码；GC重量

中图分类号：O157.4 **文献标识码**：A **文章编号**：1009-5896(2019)10-2423-05

DOI: 10.11999/JEIT190070

A Method for Constructing GC Constant Weight DNA Codes

LIANG Jing[①] LI Hongju[①] ZHAO Feng[①] DING Jian[②]

[①](*Ministry of General Education, Anhui Xinhua University, Hefei* 230088, *China*)

[②](*College of Mathematics and Informatics, Fujian Normal University, Fuzhou* 350117, *China*)

Abstract: GC weight is an important parameter of DNA code, and how to meet GC constant weight constraint DNA code is an interesting problem. In this paper, by establishing a bijection between DNA code and quaternion code, the DNA code that satisfies the GC constant weight constraint is converted into a GC constant weight quaternary code. Through the algebraic method, three types of DNA codes that meet the constant weight constraints of GC are constructed.

Key words: Quaternary code; DNA code; GC weight

1 引言

脱氧核糖核酸(DNA)是一种长链聚合物。通常DNA是由两条DNA链反向平行盘绕构成，具有双螺旋结构。每条DNA链的组成单元为4种脱氧核苷酸，即腺嘌呤脱氧核苷酸A，胸腺嘧啶脱氧核苷酸T，胞嘧啶脱氧核苷酸G，鸟嘧啶脱氧核苷酸C。两条DNA链按照Watson-Crick模型链接：A与T相连，C与G相连，这种配对是反向平行的。设$\Xi=\{A,G,T,C\}$，定义$\Xi^n=\{(c_0,c_1,\cdots,c_{n-1}):c_0,c_1,\cdots,c_{n-1}\in\Xi\}$，其中$n$是一个正整数。设$a\in\Xi$，与$a$配对的元素称为$a$的补，记作$\bar{a}$，即$\overline{A}=T$, $\overline{C}=G$, $\overline{T}=A$, $\overline{G}=C$。

对于Ξ^n中的每个n–元组$\boldsymbol{x}=(x_0,x_1,\cdots,x_{n-1})$，定义$\boldsymbol{x}$的补、倒换和倒换补分别为$\bar{\boldsymbol{x}}=(\bar{x}_0,\bar{x}_1,\cdots,\bar{x}_{n-1})$, $\boldsymbol{x}^r=(x_{n-1},x_{n-2},\cdots,x_0)$ 和$\bar{\boldsymbol{x}}^r=(\bar{x}_{n-1},\bar{x}_{n-2},\cdots,\bar{x}_0)$ 。因此，对于一个长度为n的DNA单链$\boldsymbol{x}$，与之配对的另一条DNA单链为$\bar{\boldsymbol{x}}^r$。例如，设$AGTTC$是一条DNA链，则它的补为$TCAAG$，而与这条链配对的另外一条DNA链为$GAACT$。

基于DNA分子的特殊结构，Adleman[1]通过生化方法求解了7个顶点的哈密顿回路问题，显示了DNA计算的可行性和高效性。Ξ^n的每个非空子集C都叫做一个码长n的DNA码。在DNA计算中，关键在于设计满足特定约束的DNA码。设d是一个固定的整数，针对特定的目的，通常考虑如下约束：

(1) 汉明距离约束：对任意$\boldsymbol{x}=(x_0,x_1,\cdots,x_{n-1})$, $\boldsymbol{y}=(y_0,y_1,\cdots,y_{n-1})\in C$ 且$\boldsymbol{x}\neq\boldsymbol{y}$ ，$|\{i:x_i\neq y_i,0\le i\le n-1\}|=:d(\boldsymbol{x},\boldsymbol{y})\ge d$ ；

(2) 倒换约束：对任意$\boldsymbol{x},\boldsymbol{y}\in C$, $d(\boldsymbol{x}^r,\boldsymbol{y})\ge d$；

(3) 倒换补约束：对任意$\boldsymbol{x},\boldsymbol{y}\in C$, $d(\bar{\boldsymbol{x}}^r,\boldsymbol{y})\ge d$；

(4) GC常重量约束：码C中的每个码字含有G和C的总位数是一个常值。

前3个约束的目的是为了减少非特定配对的概率，GC常重量约束用于获得相似的熔化温度[2,3]。

收稿日期：2019-01-24；改回日期：2019-08-15；网络出版：2019-08-29

*通信作者： 梁静 beaulj8607@163.com

基金项目：安徽省高校自然科学研究项目(KJ2017A623，KJ2018A0584)，安徽新华学院自然科学重点项目(2018zr001)

Foundation Items: Anhui University Natural Science Research Project (KJ2017A623, KJ2018A0584), Anhui Xinhua University Natural Science Key Project (2018zr001)

基于各种经典的线性码，学者们构造了满足各种特定约束的DNA码。文献[4]利用域$\mathbb{F}_4$上的循环码构造了满足倒换约束的DNA码。随后，通过加性和线性的四元码，构造了满足各种约束的DNA码[5]，特别地，以四元汉明码和$\mathbb{Z}_4$Kerdock码为工具构造了两类满足GC常重量约束的DNA码。文献[6]用域$\mathbb{F}_4$上的循环码构造了满足倒换补约束的DNA码，并获得了一些参数好的DNA码，自此，循环DNA码受到学者们的关注。学者们进一步研究了环$\mathbb{F}_2[u]/(u^2-1)$上的循环码并构造了满足倒换补约束的DNA码[7]，同时，基于擦除距离，研究了满足GC常重量约束的循环DNA码。文献[8]基于环$\mathbb{F}_2[u]/(u^2)$研究了长度为奇数的循环码，构造了满足各种约束的DNA码。文献[9]将文献[8]的结果推广到一般情形，构造了任意长度的满足各种约束的DNA码。对于满足倒换约束和倒换补约束的任意长度的循环DNA码的结构，文献[10]利用Gray映射进行了研究。Dinh等人[11]建立了多项式环$\mathbb{F}_2[u,v]/(u^2,v^3=v)$中元素与64密码子之间的联系，从而构造了满足倒换约束和倒换补约束的DNA码。接着Shi等人[12]研究了环$\mathbb{F}_2[u,v]/(u^3,v^2=v,uv-vu)$上循环DNA码的结构，Singh等人[13]研究了一类64元环上的循环码的对偶码，并由此构造了DNA码。目前，满足GC常重量约束的DNA码的研究相对较少。2017年Oztas等人[14]基于不可约循环码构造了一类满足GC常重量约束的DNA码，受此启发，本文通过可约循环码构造满足GC常重量约束的DNA码。

2 预备知识

设$\mathbb{F}_4=\{0,1,\omega,\bar{\omega}\}$是4元有限域，其中$\bar{\omega}=\omega^2=\omega+1$。设$\mathbb{F}_4^n$是$\mathbb{F}_4$上的$n$维行向量空间，其中，$n$是一个正奇数。向量空间$\mathbb{F}_4^n$的每个非空子集$C$都称为一个码长$n$的四元码。对于任意的$a\in\mathbb{F}_4$和向量$\boldsymbol{x}=(x_0,x_1,\cdots,x_{n-1})\in\mathbb{F}_4^n$，定义

$$N_a(\boldsymbol{x}):=|\{i:x_i=a,0\le i\le n-1\}| \tag{1}$$

码C的最小汉明距离定义为

$$d(C)=n-\max\{N_0(\boldsymbol{x}-\boldsymbol{y}):\boldsymbol{x},\boldsymbol{y}\in C,\boldsymbol{x}\ne\boldsymbol{y}\} \tag{2}$$

向量$\boldsymbol{x}$的GC重量定义为

$$N_{\mathrm{GC}}(\boldsymbol{x})=N_\omega(\boldsymbol{x})+N_{\bar{\omega}}(\boldsymbol{x}) \tag{3}$$

码C的GC重量分布定义为$N(C)=\{N_{\mathrm{GC}}(\boldsymbol{x}):\boldsymbol{x}\in C\}$。如果$N(C)$只有一个值，则称$C$为GC常重量码。用$(n,M,d_1,d_2)$表示一个码长为$n$码字个数为$M$最小汉明距离为$d_1$GC重量为$d_2$的GC常重量码。定义映射$\varphi:\mathbb{F}_4\to\Xi$，$0\mapsto A$，$\omega\mapsto C$，$\bar{\omega}\mapsto G$且$1\mapsto T$。将映射$\varphi$延拓可得

$$\begin{aligned}\Phi:\mathbb{F}_4^n\to\Xi^n,&(c_0,c_1,\cdots,c_{n-1})\\ &\mapsto(\varphi(c_0),\varphi(c_1),\cdots,\varphi(c_{n-1}))\end{aligned} \tag{4}$$

显然，映射Φ是一个双射。设C是一个码长为n的四元码，定义$\Phi(C)=\{\Phi(\boldsymbol{c}):\boldsymbol{c}\in C\}$，则$\Phi(C)$是一个码长为$n$的DNA码。如果码$C$是一个码长为$n$的GC常重量码，由$\Phi$的定义，$\Phi(C)$是一个满足GC常重量约束的DNA码。于是构造满足GC常重量约束的DNA码等价于构造一个GC常重量码。

设C是一个码长n的四元码，如果C是$\mathbb{F}_4^n$的线性子空间，则称C是一个码长n的四元线性码。四元线性码C中，对任意$(c_0,c_1,\cdots,c_{n-1})\in C$，有$(c_{n-1},c_0,c_1,\cdots,c_{n-2})\in C$，称四元线性码$C$为循环码。设$R=\mathbb{F}_4[x]/(x^n-1)$表示多项式环$\mathbb{F}_4[x]$关于理想$(x^n-1)$的商环，则环$R$是一个主理想环。定义映射$\phi:\mathbb{F}_4^n\to R$，$(c_0,c_1,\cdots,c_{n-1})\mapsto c_0+c_1x+\cdots+c_{n-1}x^{n-1}$，则码$C$是一个码长$n$的四元循环码当且仅当$\phi(C)$是环$R$的理想，其中$\phi(C)$表示码$C$在映射$\phi$下的像。在不引起混淆的情况下，本文不区分$C$与$\phi(C)$。由于$R$是主理想环，所以存在$g(x)\in\mathbb{F}_4[x]$使得$C=(g(x))$且$g(x)|(x^n-1)$。通常，$g(x)$称为$C$的生成多项式，$h(x)=(x^n-1)/g(x)$称为$C$的校验多项式。如果$h(x)$是不可约的，则称$C$为不可约循环码，如果$h(x)$是可约的，则称$C$为可约循环码。

设m是一个正整数，用$\mathbb{F}_{4^m}$表示域$\mathbb{F}_4$的m次扩域。域$\mathbb{F}_{4^m}$到域$\mathbb{F}_4$的迹函数定义为$\mathrm{Tr}_2^m(x)=x+x^4+\cdots+x^{4^{m-1}}$，对任意的$x\in\mathbb{F}_{4^m}$。可以验证，$Tr_2^m$是一个$\mathbb{F}_{4^m}$-线性映射。类似地，域$\mathbb{F}_{4^m}$到域$\mathbb{F}_2$的绝对迹定义为$Tr_1^m(x)=x+x^2+\cdots+x^{2^{m-1}}$。关于迹函数的更多性质可参阅文献[15]。设$g$是域$\mathbb{F}_{4^m}$的一个本原元，$t$是一个正整数，$s$是使得

$$t4^s\equiv t\ (\mathrm{mod}\ 4^m-1) \tag{5}$$

成立的最小正整数，显然，s整除m。

定义码$C=\{\boldsymbol{c}(a,b)=(c_0(a,b),c_1(a,b),\cdots,c_{4^m-2}(a,b)):a\in\mathbb{F}_{4^s},b\in\mathbb{F}_{4^m}\}$，其中$c_l(a,b)=Tr_2^s(ag^{lt})+Tr_2^m(bg^l)$，$0\le l\le 4^m-2$，则$C$是一个码长为$4^m-1$维数为$m+s$的四元可约循环码。下文通过码$C$构造3类码长为$n$的GC常重量码，并由此构造3类满足GC常重量约束的DNA码。

3 主要结果

情形1 $t=0$，即$s=1$，定义$\tilde{C}=\{\boldsymbol{c}(a,b)=(c_0(a,b),c_1(a,b),\cdots,c_{4^m-2}(a,b)):a\in\{0,1\},\ b\in\mathbb{F}_{4^m}^*\}$则码$\tilde{C}$的参数如下。

定理1 码$\tilde{C}$是一个参数为$(4^m-1,2(4^m-1),3\times4^{m-1}-1,2^{2m-1})$的GC常重量码。

证明　码长和码字的个数是显然的。下面讨论码$\tilde{C}$的最小汉明距离和GC重量。由定义，码$\tilde{C}$的最小汉明距离为

$$4^m-1-\max\{N_0(\boldsymbol{c}(a,b)-\boldsymbol{c}(a',b')):(a,b)\neq(a',b')\} \tag{6}$$

其中$a,a'\in\{0,1\},b,b'\in\mathbb{F}_{4^m}^*$。对任意的$a\in\{0,1\}$，$b\in\mathbb{F}_{4^m}^*$和$\beta\in\mathbb{F}_4$，定义

$$\begin{aligned}T_{a,b}(\beta)&=\left|\{l:a+\mathrm{Tr}_2^m(bg^l)=\beta,0\le l\le 4^m-2\}\right|\\&=|\{x\in\mathbb{F}_{4^m}^*:\mathrm{Tr}_2^m(bx)=\beta-a\}|\end{aligned} \tag{7}$$

因为$\boldsymbol{c}(a,b)-\boldsymbol{c}(a',b')=\boldsymbol{c}(a-a',b-b')$，所以$N_0(\boldsymbol{c}(a,b)-\boldsymbol{c}(a',b'))=T_{a-a',b-b'}(0)$。

当 $b\neq 0$ 时，$T_{a,b}(\beta)=|\{x\in\mathbb{F}_{4^m}^*:\mathrm{Tr}_2^m(x)=\beta-a\}|$。由于$\mathrm{Tr}_2^m$是$\mathbb{F}_{4^m}$到$\mathbb{F}_4$的$4^{m-1}$对1的满映射，故当$\beta\neq a$时，$T_{a,b}(\beta)=4^{m-1}$；当$\beta=a$时，$T_{a,b}(\beta)=4^{m-1}-1$。由此推出，

$$\max\{N_0(c(a,b)-c(a',b')):(a,b)\neq(a',b')\}=4^{m-1} \tag{8}$$

即码$\tilde{C}$的最小汉明距离为$3\times4^{m-1}-1$。

接下来，证明码$\tilde{C}$的GC重量为2^{2m-1}。对任意$b\in\mathbb{F}_{4^m}^*$，由GC重量的定义，

$$\begin{aligned}N_{\mathrm{GC}}(c(a,b))&=N_\omega(c(a,b))+N_{\bar{\omega}}(c(a,b))\\&=T_{a,b}(\omega)+T_{a,b}(\bar{\omega})\\&=4^{m-1}+4^{m-1}\end{aligned} \tag{9}$$

综上所述，结论成立。　证毕

情形2　$m\ge4$是一个偶数且$t=2^m+1$。容易验证，$s=m/2$。对任意的$\alpha\in\mathbb{F}_4$，设

$$\begin{aligned}L_\alpha=\{(a,b):a\in\mathbb{F}_{2^m}^*,b\in\mathbb{F}_{4^m},\\\mathrm{Tr}_2^{m/2}(a^{-1}b^{2^m+1})=\alpha\}\end{aligned} \tag{10}$$

显然$\{(a,b):a\in\mathbb{F}_{2^m}^*,b\in\mathbb{F}_{4^m}\}=L_0\cup L_1\cup L_\omega\cup L_{\bar{\omega}}$。定义

$$\widehat{C}_1=\{\boldsymbol{c}(a,b):(a,b)\in L_0\cup L_1\}$$
$$\text{和}\widehat{C}_2=\{\boldsymbol{c}(a,b):(a,b)\in L_\omega\cup L_{\bar{\omega}}\} \tag{11}$$

下面，证明码$\widehat{C}_1$和$\widehat{C}_2$是GC常重量码。为此，首先给出两个必要的引理。

引理1 设L_α定义如上，则

$$|L_\alpha|=\begin{cases}(4^m-1)2^{m-2}, & \alpha\in\mathbb{F}_4^*\\(4^m-2^{m+2}+3)2^{m-2}, & \alpha=0\end{cases} \tag{12}$$

证明　对任意的$\beta\in\mathbb{F}_{4^{m/2}}$，令$\mathcal{M}_\alpha=\{x\in\mathbb{F}_{4^m}:x^{2^m+1}=\beta\}$。显然$|\mathcal{M}_0|=1$。当$\beta\neq0$时，容易验证，$x^{2^m+1}$是$\mathbb{F}_{4^m}^*$到$\mathbb{F}_{4^{m/2}}^*$的$2^m+1$对1满映射，故$|\mathcal{M}_\beta|=2^m+1$。又因为$\mathrm{Tr}_2^{m/2}$是$\mathbb{F}_{4^{m/2}}$到$\mathbb{F}_4$的$2^{m-2}$对1满映射，于是当$\alpha\neq0$时，

$$\begin{aligned}|L_\alpha|&=\Big|\{(a,b)\in\mathbb{F}_{2^m}^*\times\mathbb{F}_{4^m}:\mathrm{Tr}_2^{m/2}(\beta)=\alpha,\\&\qquad a^{-1}b^{2^m+1}=\beta\}\Big|\\&=(2^m-1)(2^m+1)2^{m-2}\end{aligned} \tag{13}$$

当$\alpha=0$时，

$$\begin{aligned}|L_\alpha|&=\Big|\{(a,b)\in\mathbb{F}_{2^m}^*\times\mathbb{F}_{4^m}:\mathrm{Tr}_2^{m/2}(\beta)=0,\\&\qquad a^{-1}b^{2^m+1}=\beta\neq0\}\Big|\\&\quad+\Big|\{(a,b)\in\mathbb{F}_{2^m}^*\times\mathbb{F}_{4^m}:a^{-1}b^{2^m+1}=0\}\Big|\\&=(2^m-1)(2^m+1)(2^{m-2}-1)+(2^m-1)\\&=(2^{2m}-2^{m+2}+3)2^{m-2}\end{aligned} \tag{14}$$

综上所述，结论成立。　证毕

引理2　设$a\in\mathbb{F}_{2^m}^*$和$b\in\mathbb{F}_{4^m}$，则

$$\begin{aligned}S(a,b)&=\sum_{x\in\mathbb{F}_{4^m}}(-1)^{\mathrm{Tr}_1^{m/2}(ax^{2^m+1})+\mathrm{Tr}_1^m(bx)}\\&=-2^m(-1)^{\mathrm{Tr}_1^{m/2}(a^{-1}b^{2^m+1})}\end{aligned} \tag{15}$$

证明　首先，x^{2^m+1}是$\mathbb{F}_{4^m}^*$到$\mathbb{F}_{2^m}^*$的2^m+1对1满映射，所以，

$$\begin{aligned}S(a,0)&=\sum_{x\in\mathbb{F}_{4^m}}(-1)^{\mathrm{Tr}_1^{m/2}(ax^{2^m+1})}\\&=1+\sum_{x\in\mathbb{F}_{4^m}^*}(-1)^{\mathrm{Tr}_1^{m/2}(ax^{2^m+1})}\\&=1+(2^m+1)\sum_{x\in\mathbb{F}_{2^m}^*}(-1)^{\mathrm{Tr}_1^{m/2}(ax)}\\&=1+(2^m+1)(-1)=-2^m\end{aligned} \tag{16}$$

接下来，计算

$$\begin{aligned}&S(a,b)S(a,0)\\&=\sum_{z\in\mathbb{F}_{4^m}}(-1)^{\mathrm{Tr}_1^{m/2}(az^{2^m+1})+\mathrm{Tr}_1^m(bz)}\sum_{y\in\mathbb{F}_{4^m}}(-1)^{\mathrm{Tr}_1^{m/2}(ay^{2^m+1})}\\&=\sum_{x,y\in\mathbb{F}_{4^m}}(-1)^{\mathrm{Tr}_1^{m/2}(a[(x+y)^{2^m+1}+y^{2^m+1}])+\mathrm{Tr}_1^m(b(x+y))}\\&=\sum_{x\in\mathbb{F}_{4^m}}(-1)^{\mathrm{Tr}_1^{m/2}(ax^{2^m+1})+\mathrm{Tr}_1^m(bx)}\sum_{y\in\mathbb{F}_{4^m}}(-1)^{\mathrm{Tr}_1^m((ax^{2^m}+b)y)}\\&=4^m\sum_{x\in\mathbb{F}_{4^m},x^{2^m}=a^{-1}b}(-1)^{\mathrm{Tr}_1^{m/2}(ax^{2^m+1})+\mathrm{Tr}_1^m(bx)}\\&=4^m(-1)^{\mathrm{Tr}_1^{m/2}(a^{-1}b^{2^m+1})}\end{aligned} \tag{17}$$

因此，$S(a,b)=-2^m(-1)^{\mathrm{Tr}_1^{m/2}(a^{-1}b^{2^m+1})}$。　证毕

定理2　(1) 码$\widehat{C}_1$是一个参数为$(4^m-1,(4^m-$

$2^{m+1}+1)2^{m-1}, 3\times 4^{m-1}-2^{m-2}, 2(4^{m-1}+2^{m-2}))$的GC常重量码。

(2) 码$\widehat{C}_2$是一个参数为$(4^m-1, (4^m-1)2^{m-1}, 3\times 4^{m-1}-2^{m-2}, 2(4^{m-1}-2^{m-2}))$的GC常重量码。

证明 下面证明定理2(1)成立。定理2(2)的证明是类似的，略去。

由引理1和$\widehat{C}_1$的定义知，$\left|\widehat{C}_1\right| = |L_0|+|L_1| = (4^m-2^{m+1}+1)2^{m-1}$。下面讨论码$\widehat{C}_1$的最小汉明距离和GC重量。由定义，码$\widehat{C}_1$的最小汉明距离为

$$4^m-1-\max\{N_0(\boldsymbol{c}(a,b)-\boldsymbol{c}(a',b')) : (a,b) \neq (a',b') \in L_0\cup L_1\} \tag{18}$$

因为，$\boldsymbol{c}(a,b)-\boldsymbol{c}(a',b') = \boldsymbol{c}(a-a', b-b')$，所以

$$\begin{aligned} &N_0(\boldsymbol{c}(a,b)-\boldsymbol{c}(a',b')) \\ &= N_0(\boldsymbol{c}(a-a',b-b')) \\ &= \Big|\{x\in\mathbb{F}_{4^m}^* : Tr_2^{m/2}((a-a')x^{2^m+1}) \\ &\quad + Tr_2^m((b-b')x)=0\}\Big| \end{aligned} \tag{19}$$

由特征的正交关系[15]，

$$\begin{aligned} &N_0(\boldsymbol{c}(a,b)-\boldsymbol{c}(a',b')) \\ &= \sum_{x\in\mathbb{F}_{4^m}^*}\frac{1}{4}\sum_{y\in\mathbb{F}_4}(-1)^{\mathrm{Tr}(y\mathrm{Tr}_2^{m/2}((a-a')x^{2^m+1})+y\mathrm{Tr}_2^m((b-b')x))} \\ &= -1+\sum_{x\in\mathbb{F}_{4^m}}\frac{1}{4} \\ &\quad\cdot\sum_{y\in\mathbb{F}_4}(-1)^{\mathrm{Tr}_1^{m/2}((a-a')yx^{2^m+1})+\mathrm{Tr}_1^m((b-b')yx)} \\ &= 4^{m-1}-1+\frac{1}{4} \\ &\quad\cdot\sum_{y\in\mathbb{F}_4^*}\sum_{x\in\mathbb{F}_{4^m}}(-1)^{\mathrm{Tr}_1^{m/2}((a-a')yx^{2^m+1})+\mathrm{Tr}_2^m((b-b')yx)} \end{aligned} \tag{20}$$

其中Tr表示$\mathbb{F}_4$到$\mathbb{F}_2$的迹映射。

当$a-a'=0$, $b-b'\neq 0$, $N_0(\boldsymbol{c}(a,b)-\boldsymbol{c}(a',b')) = 4^{m-1}-1$。

当$a-a'\neq 0$，由引理2，

$$\begin{aligned} &N_0(\boldsymbol{c}(a,b)-\boldsymbol{c}(a',b')) \\ &= 4^{m-1}-1-2^{m-2} \\ &\quad\cdot\sum_{y\in\mathbb{F}_4^*}(-1)^{\mathrm{Tr}_1^{m/2}((a-a')^{-1}(b-b')^{2^m+1}y^{2^m})} \\ &= 4^{m-1}-1-2^{m-2} \\ &\quad\cdot\sum_{y\in\mathbb{F}_4^*}(-1)^{\mathrm{Tr}(y(\mathrm{Tr}_2^{m/2}(a-a')^{-1}(b-b')^{2^m+1})))} \end{aligned} \tag{21}$$

所以，

$$\begin{aligned} &N_0(\boldsymbol{c}(a,b)-\boldsymbol{c}(a',b')) \\ &= \begin{cases} 4^{m-1}-1+2^{m-2}, \\ \qquad \mathrm{Tr}_2^{m/2}((a-a')^{-1}(b-b')^{2^m+1})\neq 0 \\ 4^{m-1}-1-3\times 2^{m-1}, \\ \qquad \mathrm{Tr}_2^{m/2}((a-a')^{-1}(b-b')^{2^m+1})=0 \end{cases} \end{aligned} \tag{22}$$

注意当$(a,b)\neq(a',b')\in L_\alpha$时，$(a,b)-(a',b')$取遍$\mathbb{F}_{2^m}\times\mathbb{F}_{4^m}$中非零元，故码$\widehat{C}_1$的最小汉明距离为$3\times 4^{m-1}-2^{m-2}$。

由GC重量的定义得

$$N_{\mathrm{GC}}(\boldsymbol{c}(a,b)) = N_\omega(\boldsymbol{c}(a,b))+N_{\bar{\omega}}(\boldsymbol{c}(a,b)) \tag{23}$$

对任意的$\beta\in\mathbb{F}_4^*$，$N_\beta(\boldsymbol{c}(a,b)) = |\{x\in\mathbb{F}_{4^m}^* : \mathrm{Tr}_2^{m/2}(ax^{2^m+1})+\mathrm{Tr}_2^m(bx)=\beta\}|$。由于$a\in\mathbb{F}_{2^m}^*$和$b\in\mathbb{F}_{4^m}$，则

$$\begin{aligned} &4N_\beta(\boldsymbol{c}(a,b)) \\ &= \sum_{x\in\mathbb{F}_{4^m}^*}\sum_{y\in\mathbb{F}_4}(-1)^{\mathrm{Tr}(y[\mathrm{Tr}_2^{m/2}(ax^{2^m+1})+\mathrm{Tr}_2^m(bx)-\beta])} \\ &= \sum_{x\in\mathbb{F}_{4^m}}\sum_{y\in\mathbb{F}_4}(-1)^{\mathrm{Tr}(y[\mathrm{Tr}_2^{m/2}(ax^{2^m+1})+\mathrm{Tr}_2^m(bx)-\beta])} \\ &= 4^m+\sum_{x\in\mathbb{F}_{4^m}}\sum_{y\in\mathbb{F}_4^*}(-1)^{\mathrm{Tr}(y[\mathrm{Tr}_2^{m/2}(ax^{2^m+1})+\mathrm{Tr}_2^m(bx)-\beta])} \\ &= 4^m+\sum_{y\in\mathbb{F}_4^*}(-1)^{\mathrm{Tr}(\beta y)} \\ &\quad\cdot\sum_{x\in\mathbb{F}_{4^m}}(-1)^{\mathrm{Tr_1}^{m/2}(ayx^{2^m+1})+\mathrm{Tr_1}^m(byx)} \end{aligned} \tag{24}$$

由引理1

$$\begin{aligned} &4N_\beta(\boldsymbol{c}(a,b)) \\ &= 4^m-2^m\sum_{y\in\mathbb{F}_4^*}(-1)^{\mathrm{Tr}(\beta y)}(-1)^{\mathrm{Tr}_2^m(a^{-1}b^{2^m+1}y)} \\ &= 4^m-2^m\sum_{y\in\mathbb{F}_4^*}(-1)^{\mathrm{Tr}(\beta y)}(-1)^{\mathrm{Tr}(y\mathrm{Tr}_2^m(a^{-1}b^{2^m+1}))} \end{aligned} \tag{25}$$

由于$(a,b)\in L_\alpha$，所以

$$4N_\beta(\boldsymbol{c}(a,b)) = 4^m-2^m\sum_{y\in\mathbb{F}_4^*}(-1)^{\mathrm{Tr}((\alpha-\beta)y)} \tag{26}$$

即

$$N_\beta(\boldsymbol{c}(a,b)) = \begin{cases} 4^{m-1}-3\times 2^{m-2}, & \beta=\alpha \\ 4^{m-1}+2^{m-2}, & \beta\neq\alpha \end{cases} \tag{27}$$

由此推出，当$(a,b)\in L_0\cup L_1$时，

$$\begin{aligned} N_{\mathrm{GC}}(\boldsymbol{c}(a,b)) &= N_\omega(\boldsymbol{c}(a,b))+N_{\bar{\omega}}(\boldsymbol{c}(a,b)) \\ &= 2(4^{m-1}+2^{m-2}) \end{aligned} \tag{28}$$

综上所述，结论成立。 证毕

由定理1和定理2，得出如下推论。

推论 (1) 设m是一个正整数，则存在参数为

$(4^m-1, 2(4^m-1), 3\times 4^{m-1}-1, 2^{2m-1})$ 且满足GC常重量约束的DNA码。

(2) 设m为偶数，且$m\geq 4$，则存在参数为$(4^m-1, (4^m-2^{m+1}+1)2^{m-1}, 3\times 4^{m-1}-2^{m-2}, 2(4^{m-1}+2^{m-2}))$ 的满足GC常重量约束的DNA码。

(3) 设$m\geq 4$是偶数，则存在参数为$(4^m-1, (4^m-1)2^{m-1}, 3\times 4^{m-1}-2^{m-2}, 2(4^{m-1}-2^{m-2}))$的满足GC常重量约束的DNA码。

4 结束语

本文基于四元可约的循环码构造了3类满足GC常重量的DNA码。通过建立四元码和DNA码的一个对应关系，将构造满足GC常重量约束的DNA码转化为构造四元GC常重量码。利用循环码的迹表示，选取循环码的特定子码，构造了3类四元GC常重量码，由此获得3类满足GC常重量约束的DNA码。

参 考 文 献

[1] ADLEMAN L M. Molecular computation of solutions to combinatorial problems[J]. *Science*, 1994, 266(5187): 1021–1024. doi: 10.1126/science.7973651.

[2] FRUTOS A G, LIU Qinghua, THIEL A J, *et al.* Demonstration of a word design strategy for DNA computing on surfaces[J]. *Nucleic Acids Research*, 1997, 25(23): 4748–4757. doi: 10.1093/nar/25.23.4748.

[3] MARATHE A, CONDON A E, and CORN R M. On combinatorial DNA word design[J]. *Journal of Computational Biology*, 2001, 8(3): 201–220. doi: 10.1089/10665270152530818.

[4] RYKO V V, MACULA A J, TORNEY D C, *et al.* DNA sequences and quaternary cyclic codes[C]. 2001 IEEE International Symposium on Information. Washington, USA, 2001: 248–248.

[5] GABORIT P and KING O D. Linear constructions for DNA codes[J]. *Theoretical Computer Science*, 2005, 334(1/3): 99–113. doi: 10.1016/j.tcs.2004.11.004.

[6] ABUALRUB T, GHRAYEB A, and ZENG Xiangnian. Construction of cyclic codes over GF(4) for DNA computing[J]. *Journal of the Franklin Institute*, 2006, 343(4/5): 448–457. doi: 10.1016/j.jfranklin.2006.02.009.

[7] SIAP I, ABUALRUB T, and GHRAYEB A. Cyclic DNA codes over the ring $F_2[u]/(u^2-1)$ based on the deletion distance[J]. *Journal of the Franklin Institute*, 2009, 346(8): 731–740. doi: 10.1016/j.jfranklin.2009.07.002.

[8] GUENDA K and GULLIVER T A. Construction of cyclic codes over $\mathbb{F}_2+u\mathbb{F}_2$ for DNA computing[J]. *Applicable Algebra in Engineering, Communication and Computing*, 2013, 24(6): 445–459. doi: 10.1007/s00200-013-0188-x.

[9] LIANG Jing and WANG Liqi. On cyclic DNA codes over $\mathbb{F}_2+u\mathbb{F}_2$[J]. *Journal of Applied Mathematics and Computing*, 2016, 51(1/2): 81–91. doi: 10.1007/s12190-015-0892-8.

[10] ZHU Shixin and CHEN Xiaojing. Cyclic DNA codes over $\mathbb{F}_2+u\mathbb{F}_2+v\mathbb{F}_2+uv\mathbb{F}_2$ and their applications[J]. *Journal of Applied Mathematics and Computing*, 2017, 55(1/2): 479–493. doi: 10.1007/s12190-016-1046-3.

[11] DINH H Q, SINGH A K, PATTANAYAK S, *et al.* Cyclic DNA codes over the ring $\mathbb{F}_2+u\mathbb{F}_2+v\mathbb{F}_2+uv\mathbb{F}_2+v^2\mathbb{F}_2+uv^2\mathbb{F}_2$[J]. *Designs, Codes and Cryptography*, 2018, 86(7): 1451–1467. doi: 10.1007/s10623-017-0405-x.

[12] SHI Minjia and LU Yaqi. Cyclic DNA codes over $\mathbb{F}_2[u,v]/<u^3, v^2-v, vu-uv>$[J]. *Advances in Mathematics of Communications*, 2019, 13(1): 157–164. doi: 10.3934/amc.2019009.

[13] SINGH A K, KUMAR N, MISHRA P, *et al.* Construction of dual cyclic codes over $\mathbb{F}_2[u,v]/\langle u^2, v^2-v, uv-vu\rangle$ for DNA Computation[J]. *Defence Science Journal*, 2018, 68(5): 467–472. doi: 10.14429/dsj.68.12344.

[14] OZTAS E S, YILDIZ B, and SIAP I. A novel approach for constructing reversible codes and applications to DNA codes over the ring $\mathbb{F}_2[u]/<u^{2k}-1>$[J]. *Finite Fields and Their Applications*, 2017, 46: 217–234. doi: 10.1016/j.ffa.2017.04.001.

[15] LIDL R and NIEDERREITE H. Finite Fields[M]. New York: Addison-Wesley Publishing Company, 1983.

梁 静：女，1986年生，讲师，硕士，研究方向为代数编码与密码.

李红菊：女，1982年生，副教授，硕士，研究方向为统计学.

赵 凤：女，1985年生，讲师，硕士，研究方向密码学.

丁 健：男，1982年生，副教授，硕士，研究方向为代数编码与密码.