ON CYCLIC DNA CODES OVER $F_4[u]/(u^2+1)$

Fanghui Ma [1], Yonglin Cao[2,*] & Jian Gao[3]

[1, 2] School of Science, Shandong University of Technology, Zibo, 255091 P. R. China

[3] Chern Institute of Mathematics and LPMC, Nankai University, Tianjin, 300071 P. R. China

ABSTRACT

In this paper, we study the construction of cyclic DNA codes by cyclic codes over the finite chain ring $F_4[u]/(u^2+1)$. First, we establish a 1-1 correspondence φ between DNA pairs and the 16 elements of the ring $F_4[u]/(u^2+1)$. Considering the biology features of DNA codes, we investigate the structure and properties of self-reciprocal complement cyclic codes. Hamming minimum distances and GC-content are also studied. Finally, by use of the result above, we construct concrete cyclic DNA codes as the image of self-reciprocal complement cyclic codes over $F_4[u]/(u^2+1)$ under the map φ, and list the best cyclic DNA codes of length 6.

Keywords: *Cyclic DNA codes, Rings, WCC pairing.*

1. INTRODUCTION

DNA plays an important role as the carrier of genetic information in all living species. DNA molecules stored blueprints of life in all species. DNA sequences consist of four bases nucleotides: adenine (A), guanine (G), thymine (T) and cytosine(C). A single DNA strand is an ordered quaternary sequence of the letters A, C, G and T with chemically distinct polar terminals known as the 5-and 3-ends. DNA has two strands that are arranged in an order with a rule that is named as Watson Crick complement (WCC). The WCC of A is T and vice versa and the WCC of C is G and vice versa. We describe this as $\overline{A}=T$, $\overline{T}=A$, $\overline{G}=C$ and $\overline{C}=G$.

Since the disorder of DNA computing and we can have following similar result. A single DNA strand $X=x_1x_2\cdots x_n$ can pair with its WCC or its reverse complement strand $X^{RC}=\overline{x}_n\overline{x}_{n-1}\cdots\overline{x}_1$ where $\overline{x}_i$ is the WCC of x_i for all $i=1,2,\cdots,n$.

Adelman [4] described an experiment involving the use of DNA to solve a seven node instance of the famous directed salesman problem. Adelman's approach was based on the WCC property of DNA strands. In [6], Boneh et al. and independently Adelman et al. in [5] presented a molecular program for breaking the Data Encryption Standard (DES) cryptographic system. In [12], it is shown that DNA molecules can be used as a storage medium. For the last 20 years, cyclic codes over finite rings played a very important role in the area of error-correcting codes [3,9,13,14,15].

Several papers have proposed different techniques to construct a set of DNA codewords that are unlikely to form undesirable bonds with each other by hybridization. For example, in [7,8,10,11], four different constraints on DNA codes are considered. These are the Hamming distance constraint, the reverse-complement constraint, the reverse constraint and the fixed GC-content. The purpose of the first three constraints is to make non-desirable hybridizations difficult to occur. The fixed GC-content constraints is used to make sure that similar melting temperatures are obtained [7]. The four constraints are defined as follows:

1. Let $H(x,y)$ denote the Hamming distance between two codewords. The Hamming distance constraint is that $H(x,y)\geq d$ for all $x,y\in C$ with $x\neq y$, for some prescribed minimum distance d.
2. The reverse constraint is that $H(x^R,y)\geq d$ for all $x,y\in C$, where x^R is the reverse of a codeword x. Note that $x=y$ is included.
3. The reverse-complement constraint is that $H(x^{RC},y)\geq d$ for all $x,y\in C$. Again $x=y$ is included.
4. The GC-content constraint is that each codeword $x\in C$ has the same GC-content. The GC-content of a DNA word is defined to be the number of positions in which the word has coordinate C or G. For the GC-content, we consider $\lfloor n/2\rfloor$.

In [14] the authors designed cyclic DNA codes over the ring $F_2[u]/(u^2-1) = F_2 + uF_2 = \{0,1,u,u+1\}$ where $u^2 = 1$. The ring $F_2[u]/(u^2-1)$ is in a one to one corresponding with the DNA bases A, C, G and T where we have $A \to 0$, $T \to (1+u)$, $C \to u$, $G \to 1$. In [15], the authors designed the cyclic DNA codes over the ring $F_2[u]/(u^4-1)$ with 16 elements.

In [14,15], cyclic DNA codes depend on the existence of self-reciprocal polynomials that are divisors of $x^n - 1$ over F_2. As from the factorization of $x^n - 1$ over F_4 one can get more self-reciprocal divisors than that of $x^n - 1$ over F_2 in general, we study the construction of cyclic DNA codes by cyclic codes over the finite chain ring $F_4[u]/(u^2+1)$ that satisfy the first three constraints mentioned above. Once such codes are constructed, the GC-content constraint can be easily incorporated by removing the code words that violate this constraint.

The rest of the paper is organized as follows. In section 2 we establish a 1-1 correspondence φ between DNA pairs and the 16 elements of the ring $F_4[u]/(u^2+1)$, and present a description and the definition of cyclic DNA codes over the ring R. In section 3, we investigate the structure and properties of self-reciprocal complement cyclic codes, find a unique set of generators for these codes as ideals of the ring $\Re_n = R[x]/(x^n-1)$. We show that the generators of cyclic DNA codes are related to self-reciprocal divisors of $x^n - 1$ in $F_4[x]$. By use of the result above, we construct concrete cyclic DNA codes as the image of self-reciprocal complement cyclic codes over $F_4[u]/(u^2+1)$ under the map φ, and obtain DNA codes of length 6 satisfying Hamming distance and GC-content.

2. CYCLIC DNA CODES OVER THE RING R

Let $F_4 = \{0,1,\alpha,\alpha+1\}$ be the quaternary finite field, where α is a root of the primitive polynomial $x^2 + x + 1$ over F_2. We consider the ring $F_4[u]/(u^2+1)$ where $u^2 = 1$. Clearly, the set $\{1, 1+u\}$ forms an F_4 - basis of R and every element of R can be expressed uniquely as $a + b(1+u)$ where $a, b \in F_4$. The finite field F_4 can be viewed as a subring of the ring R. Therefore the factorization of $x^n - 1$ over F_4 will be still valid over the ring R.

Since there are four basic nucleotides A, T, G, and C, we have 16 pairs such as AA, TT, GT, CA, AG, TC, AT, TA, TG, AC, GA, CT, GC, CG, CC , GG . We define the map φ which gives a one to one correspondence between R and $\{A,C,G,T\}^2$ given by Table 1.

Table1: Identifying nucleotide pairs with elements of the ring

AA	0	TT	$\alpha+1+(\alpha+1)u$	GT	1	CA	$\alpha+(\alpha+1)u$
AG	α	TC	$1+(\alpha+1)u$	AT	$\alpha+1$	TA	$(\alpha+1)u$
TG	u	AC	$\alpha+1+\alpha u$	GA	αu	CT	$1+\alpha+u$
GC	$1+\alpha u$	CG	$\alpha+u$	CC	$1+u$	GG	$\alpha+\alpha u$

This map φ gives the following WCC pairing:
$\bar{0} = \alpha+1+(\alpha+1)u$, $\bar{1} = \alpha+(\alpha+1)u$, $\bar{\alpha} = 1+(\alpha+1)u$, $\overline{\alpha+1} = (\alpha+1)u$, $\bar{u} = \alpha+1+\alpha u$, $\overline{\alpha u} = 1+\alpha+u$,
$\overline{1+\alpha u} = \alpha+u$, $\overline{1+u} = \alpha+\alpha u$.

The correspondence satisfies the complement property by simply adding $\alpha+1+(\alpha+1)u$ to an element of the ring. For instance if we take AG which is identified by α and want to find its complement we just add $\alpha+1+(\alpha+1)u$ to α so we get $1+(\alpha+1)u$ which is TC. This is true for all other identifications. The WCC complement is defined as

in the papers mentioned above. Another observation from the above map φ is that multiplying an element a of R by u reverses the DNA pair corresponding to a.

A cyclic code of length n over R is a nonzero R-submodule of R^n invariant with respect to the cyclic shift operator that maps a codeword $a=(a_0,a_1,\ldots,a_{n-1})\in C$ to another codeword $(a_{n-1},a_0,\ldots,a_{n-2})$ in C. If $a=(a_0,a_1,\ldots,a_{n-1})$ is represented by the polynomial $f(x)=a_0+a_1x+\cdots+a_{n-1}x^{n-1}$ then C is a cyclic code over R if and only if C is an ideal of the ring $\Re_n=R[x]/(x^n-1)$.

For each polynomial $f(x)=a_0+a_1x+\cdots+a_rx^r$ with $a_r\neq 0$, we define the reciprocal of $f(x)$ to be the polynomial $f^*(x)=x^r f(1/x)=a_r+a_{r-1}x+\cdots+a_0x^r$. Note that $\deg(f^*(x))\le\deg(f(x))$ and if $a_0\neq 0$, then $\deg(f^*(x))=\deg(f(x))$. $f(x)$ is called a self-reciprocal polynomial if there is a constant m such that $f^*(x)=mf(x)$.

Definition 1. A cyclic code C of length n is called DNA code over R if

(1) C is a cyclic code, i.e. C is an ideal of $R[x]/(x^n-1)$;

(2) For any codeword $X\in C$, $X\neq X^{RC}$ and $X^{RC}\in C$.

3. GENERATORS OF CYCLIC DNA CODES

The elements of the ring R can be mapped to the elements of $F_4=\{0,1,\alpha,\alpha+1\}$ via the map Φ as $\Phi(a+b(1+u))=a$ for any element $a+b(1+u)$ of R. Let C be a cyclic code of length n over R, i.e. an ideal of $\Re^n$. Let $R_n=F_4[x]/(x^n-1)$. We extend the map Φ to a map

$$\phi: \quad C\to R_n$$
$$a_0+a_1x+\cdots+a_{n-1}x^{n-1}\mapsto\Phi(a_0)+\Phi(a_1)x+\ldots+\Phi(a_{n-1})x^{n-1}$$

Let $J=\{r(x):(1+u)r(x)\in\ker\phi\}$. It is easy to check that J is an ideal of R_n. So $J=\langle a(x)\rangle$ where $a(x)\mid(x^n-1)$. This implies that $\ker\phi=\langle(1+u)a(x)\rangle$ where $a(x)\mid(x^n-1)$. The image of ϕ is also an ideal and hence a quaternary cyclic code that has a generator $g(x)$ where $g(x)\mid(x^n-1)$. This implies that $C=\langle g(x)+(1+u)p(x),(1+u)a(x)\rangle$ where $p(x)\in F_4[x]$.

The following lemma 1 will be an immediate result following from the above discussion and similar arguments used in the proof of Theorem 1 in [3].

Lemma 1. Let C be a cyclic code of $\Re_n$.

(1) If n is odd, then $\Re_n$ is a principal ideal ring and $C=\langle g(x),(1+u)a(x)\rangle$, where $g(x)$ and $a(x)\in F_4[x]$ with $a(x)\mid g(x)\mid(x^n-1)$;

(2) If n is even then

(a) If $g(x)=a(x)$, then $C=\langle g(x)+(1+u)p(x)\rangle$, where $g(x)$ and $a(x)\in F_4[x]$ with $g(x)\mid(x^n-1)$, $(g(x)+(1+u)p(x))\mid(x^n-1)$ in $R[x]$;

(b) $C=\langle g(x)+(1+u)p(x),(1+u)a(x)\rangle$ where $g(x)$, $a(x)$ and $p(x)\in F_4[x]$ with $a(x)\mid g(x)\mid(x^n-1)$ and $\deg(g(x))>\deg(a(x))>\deg(p(x))$.

Lemma 2. For any $a\in R$, we have $a+\bar{a}=\alpha+1+(\alpha+1)u$.

Proof .The proof process follows from trying all elements in R.

Lemma 3. If $a \in \{0,1,\alpha,\alpha+1\}$, then we have $\alpha+1+(\alpha+1)u+\overline{\alpha+1+(\alpha+1)u} = (\alpha+1+(\alpha+1)u)a$

Proof. If $a=0$, then

$$\begin{aligned}&\alpha+1+(\alpha+1)u+\overline{\alpha+1+(\alpha+1)u}\\&=\alpha+1+(\alpha+1)u+\bar{0}\\&=\alpha+1+(\alpha+1)u+\alpha+1+(\alpha+1)u=0\\&=(\alpha+1+(\alpha+1)u)a\end{aligned}$$

If $a=1$, then

$$\begin{aligned}&\alpha+1+(\alpha+1)u+\overline{(\alpha+1+(\alpha+1)u)a}\\&=\alpha+1+(\alpha+1)u+\overline{\alpha+1+(\alpha+1)u}\\&=\alpha+1+(\alpha+1)u+0\\&=(\alpha+1+(\alpha+1)u)a\end{aligned}$$

Similarly, one can also check the cases of $a=\alpha$ and $a=\alpha+1$ respectively. Here, we omit them for the interesting of space.

Lemma 4. Let $f(x)$ and $g(x)$ be polynomials in $R(x)$. Suppose $\deg(f(x))-\deg(g(x))=m$, then

(1) $(f(x)g(x))^* = f(x)^* g(x)^*$;

(2) $(f(x)+g(x))^* = f(x)^* + x^m g(x)^*$.

Proof. The proof process is similar to that of Lemma 19 in [2].

Theorem 1. Let $C=\langle g(x),(1+u)a(x)\rangle$ be a cyclic code of length n of $\Re_n$. If $f(x)^{RC} \in C$ for any $f(x)\in C$, then $(1+u)\dfrac{x^n-1}{x-1}\in C$, and there are two constants $c,d \in F_4 \setminus \{0\}$ such that $g^*(x)=cg(x)$ and $a^*(x)=da(x)$.

Proof. Suppose $C=\langle g(x),(1+u)a(x)\rangle$ where $a(x)\,|\,g(x)\,|\,(x^n-1) \in F_4[x]$. Since the zero codeword is an element in C, then its WCC is also in C, i.e.

$$\begin{aligned}&\overline{(0,0,\cdots,0)}\\&=(\alpha+1+(\alpha+1)u,\alpha+1+(\alpha+1)u,\cdots,\alpha+1+(\alpha+1)u)\\&=(\alpha+1)(1+u)(1,1,\cdots,1)\\&=(\alpha+1)(1+u)(1+x+x^2+\cdots+x^{n-1})\\&=(\alpha+1)(1+u)\frac{x^n-1}{x-1}\in C.\end{aligned}$$

Since $0\neq\alpha+1\in F_4$, then $(1+u)\dfrac{x^n-1}{x-1}\in C$.

Note that

$$\begin{aligned}g(x)^{RC} &= \alpha+1+(\alpha+1)u+(\alpha+1+(\alpha+1)u)x+\cdots+(\alpha+1+(\alpha+1)u)x^{n-r-2}\\&\quad+\bar{g}_r x^{n-r-1}+\bar{g}_{r-1}x^{n-r}+\cdots+\bar{g}_1 x^{n-2}+\bar{g}_0 x^{n-1}\in C.\end{aligned}$$

Since C is a linear code, then

$$g(x)^{RC}+(\alpha+1)(1+u)\frac{x^n-1}{x-1}\in C,$$

Which implies that

$$x^{n-r-1}\left(\left(\overline{g}_r+\alpha+1+(\alpha+1)u\right)+\left(\overline{g}_{r-1}+\alpha+1+(\alpha+1)u\right)x+\cdots+\left(\overline{g}_0+\alpha+1+(\alpha+1)u\right)x^r\right)\in C.$$

By Lemma 2, $\overline{a}+\alpha+1+(\alpha+1)u=a$. This implies that

$$x^{n-r-1}\left(g_r+g_{r-1}x+\cdots+g_0x^r\right)=x^{n-r-1}g^*(x)\in C.$$

So we have

$$g^*(x)=g(x)l(x)+(1+u)a(x)k(x).$$

where $l(x)$, $k(x)\in F_4[x]$. Since $g_i\in\{0,1,\alpha,\alpha+1\}$, for any $i=0,1,\cdots,r$, we get that $k(x)=0$ .Since $\deg\left(g^*(x)\right)=\deg\left(g(x)\right)$, then $l(x)\in F_4\setminus\{0\}$. So there is a constant $c\in F_4\setminus\{0\}$ such that $g^*(x)=cg(x)$. Therefore, $g(x)$ is a self-reciprocal polynomial.

Now suppose that

$$\begin{aligned}(\alpha+1)(1+u)a(x)=&(\alpha+1)(1+u)a_0+(\alpha+1)(1+u)a_1(x)+\cdots\\&+(\alpha+1)(1+u)a_{t-1}x^{t-1}+(\alpha+1)(1+u)a_tx^t.\end{aligned}\tag{1}$$

Then

$$\begin{aligned}\left((\alpha+1)(1+u)a(x)\right)^{RC}=&(\alpha+1)(1+u)+(\alpha+1)(1+u)x+\cdots\overline{(\alpha+1)(1+u)a_t}x^{n-t-1}\\&+\overline{(\alpha+1)(1+u)a_{t-1}}x^{n-t}+\cdots+\overline{(\alpha+1)(1+u)a_1}x^{n-2}\\&+\overline{(\alpha+1)(1+u)a_0}x^{n-1}\in C.\end{aligned}\tag{2}$$

Since $(\alpha+1)(1+u)\dfrac{x^n-1}{x-1}\in C$ and C is a linear code, then

$$\left((\alpha+1)(1+u)a(x)\right)^{RC}+(\alpha+1)(1+u)\frac{x^n-1}{x-1}\in C.$$

Hence

$$x^{n-t-1}\left(\begin{array}{l}\left(\overline{(\alpha+1)(1+u)a_t}+(\alpha+1)(1+u)\right)+\left(\overline{(\alpha+1)(1+u)a_{t-1}}+(\alpha+1)(1+u)\right)x+\\\cdots+\left(\overline{(\alpha+1)(1+u)a_1}+(\alpha+1)(1+u)\right)x^{t-1}+\left(\overline{(\alpha+1)(1+u)a_0}+(\alpha+1)(1+u)\right)x^t\end{array}\right)\in C.\tag{3}$$

By Lemma 3, we have

$$\alpha+1+(\alpha+1)u+\overline{\left(\alpha+1+(\alpha+1)u\right)a}=\left(\alpha+1+(\alpha+1)u\right)a.$$

Therefore

$$\begin{aligned}&x^{n-t-1}\left((\alpha+1)(1+u)a_t+(\alpha+1)(1+u)a_{t-1}x+\cdots+(\alpha+1)(1+u)a_1x^{t-1}+(\alpha+1)(1+u)a_0x^t\right)\\&=x^{n-t-1}(\alpha+1)(1+u)a^*(x)\in C.\end{aligned}\tag{4}$$

Since $0\neq\alpha+1\in F_4$, then $(1+u)a^*(x)\in C$. So we have

$$(1+u)a^*(x)=g(x)h(x)+(1+u)a(x)s(x).$$

Since $1+u$ dose not appear in $g(x)$, it follows that $h(x)=0$ and $a^*(x)=a(x)s(x)$. Further, since $\deg\left(a^*(x)\right)=\deg\left(a(x)\right)$, then $s(x)\in F_4\setminus\{0\}$. It means that there is a constant $d\in F_4\setminus\{0\}$ such that $a^*(x)=da(x)$.

Theorem 2. Let $C=\left\langle g(x),(1+u)a(x)\right\rangle$ be a cyclic code of length n of $\Re_n$. If $(1+u)\dfrac{x^n-1}{x-1}\in C$ and there are two constants $c,d\in F_4\setminus\{0\}$ such that $g^*(x)=cg(x)$ and $a^*(x)=da(x)$, then $f(x)^{RC}\in C$ for any $f(x)\in C$.

Proof. Let $C=\left\langle g(x),(1+u)a(x)\right\rangle$, $g^*(x)=cg(x)$ and $a^*(x)=da(x)$. Let $g(x)=g_0+g_1x+\cdots+g_{r-1}x^{r-1}+g_rx^r$ and $c(x)\in C$. Then there exist $l(x)$ and $k(x)$ in $R(x)$ such that $c(x)=g(x)l(x)+(1+u)a(x)k(x)$. Since $g^*(x)=cg(x)$ and $a^*(x)=da(x)$, by Lemma 4, we have

$$\begin{aligned} c^*(x) &= \left(g(x)l(x)+(1+u)a(x)k(x)\right)^* \\ &= \left(g(x)l(x)\right)^* + x^m\left((1+u)a(x)k(x)\right)^* \\ &= g^*(x)l^*(x)+(1+u)a^*(x)\left(x^m k^*(x)\right) \\ &= cg(x)l^*(x)+d(1+u)a(x)\left(x^m k^*(x)\right). \end{aligned} \tag{5}$$

Therefore, $c^*(x)\in C$. Since $0\neq\alpha+1\in F_4$, then $(\alpha+1)(1+u)\dfrac{x^n-1}{x-1}\in C$. Therefore, we have

$$(\alpha+1)(1+u)\frac{x^n-1}{x-1}=(\alpha+1)(1+u)+(\alpha+1)(1+u)x+\cdots+(\alpha+1)(1+u)x^{n-1}\in C. \tag{6}$$

Let $c(x)=c_0+c_1x+\cdots+c_tx^t\in C$. Since C is a cyclic code, then we have

$$x^{n-t-1}c(x)=c_0x^{n-t-1}+c_1x^{n-t}+\cdots+c_tx^{n-1}\in C. \tag{7}$$

Summing Eq. (6) and (7) gives

$$\begin{aligned} &(\alpha+1)(1+u)+(\alpha+1)(1+u)x+\cdots+(\alpha+1)(1+u)x^{n-t-2}+\left(c_0+(\alpha+1)(1+u)\right)x^{n-t-1} \\ &\qquad+\left(c_1+(\alpha+1)(1+u)\right)x^{n-t}+\cdots+\left(c_t+(\alpha+1)(1+u)\right)x^{n-1}\in C. \end{aligned} \tag{8}$$

By Lemma 2, we know that $\alpha+1+(\alpha+1)u+a=\bar{a}$. Thus, Eq. (8) becomes

$$\begin{aligned} &(\alpha+1)(1+u)+(\alpha+1)(1+u)x+\cdots+(\alpha+1)(1+u)x^{n-t-2}+\bar{c}_0x^{n-t-1} \\ &\qquad+\bar{c}_1x^{n-t}+\cdots+\bar{c}_tx^{n-1}\in C. \end{aligned} \tag{9}$$

Note that the polynomials in Eq. (9) is $c^*(x)^{RC}$. Therefore $\left(c^*(x)^{RC}\right)^*=\left(x^{n-t-1}c(x)\right)^{RC}$.

Now, suppose C is a cyclic DNA codes over R of length n, then $\varphi(C)$ is a set of DNA strands of length 2n. If $c=c(x)\in C$ then $c^{RC}(x)\in C$, which means $\varphi(C)$ has the reverse complement property on pairs of nucleotides. However since C is linear, $u^2c\in C$ as well, and we know that $\varphi\left(u^2c^{RC}\right)$ is exactly the WCC complement of $\varphi(c)$. Thus, we obtain

Corollary 1. Let C be a cyclic DNA code of length n over the ring R and minimum Hamming distance d. Then, $\varphi(C)$ is a DNA code of length 2n over the alphabet A,T,G,C with minimum Hamming distance at least d.

At the end of this section, we give a example to illustrate the main work in this paper.

Example 1. Let $x^3-1=(x+1)(x+\alpha)\left(x+\alpha^2\right)\in F_4[x]$.. Let $C=\langle g(x),(1+u)a(x)\rangle$ be a cyclic code of length 3 over R ,where $a(x)=x+\alpha$ and $g(x)=(x+\alpha)\left(x+\alpha^2\right)$. The image of C under the map φ is a DNA code of length 6. This code has 64 codewords. These codewords are listed in the Table 2. These codewords that satisfy the four constraints mentioned above we get 576 kinds of results. We have chosen 6 kinds of results with minimum Hamming distance 3 that are listed in the Table 3.

Table 2: All 64 codewords of C

AAAAAA	AAGGAA	AATTAA	AACCAA	GGAAAA	GGGGAA	GGTTAA
GGCCAA	TTAAAA	TTGGAA	TTTTAA	TTCCAA	CCAAAA	CCGGAA
CCTTAA	CCCCAA	GTGTGT	GTACGT	GTCAGT	GTTGGT	ACGTGT
ACACGT	ACCAGT	ACTGGT	CAGTGT	CAACGT	CACAGT	CATGGT
TGGTGT	TGACGT	TGCAGT	TGTGGT	AGAGAG	AGGAAG	AGTCAG
AGCTAG	GAAGAG	GAGAAG	GATCAG	GACTAG	TCAGAG	TCGAAG
TCTCAG	TCCTAG	CTAGAG	CTGAAG	CTTCAG	CTCTAG	ATATAT
ATGCAT	ATTAAT	ATCGAT	GCATAT	GCGCAT	GCTAAT	GCCGAT
TAATAT	TAGCAT	TATAAT	TACGAT	CGATAT	CGGCAT	CGTAAT
CGCGAT						

Table 3: 6 kinds of results with minimum Hamming distance 3

1	2	3	4	5	6
GTGTGT	GTGTGT	GTGTGT	GTGTGT	GTGTGT	GTGTGT
ACCAGT	ACCAGT	CATGGT	ACCAGT	ACCAGT	ACCAGT
CAACGT	CAACGT	TGACGT	CAACGT	CAACGT	CAACGT
AGAGAG	AGAGAG	AGTCAG	AGAGAG	AGAGAG	AGAGAG
GATCAG	GACTAG	GAGAAG	GACTAG	TCCTAG	TCCTAG
TCGAAG	TCGAAG	CTCTAG	CTGAAG	CTGAAG	CTTCAG

4. OVERALL CONCLUSIONS

In this paper cyclic DNA codes over $F_4[u]/(u^2+1)$ have been studied. Our work of cyclic DNA codes over R is motivated by the simple and less complexity structure of these codes over R. We found a unique set of generators of these codes as ideals of the ring $\Re_n = R[x]/(x^n-1)$. We included an example of cyclic DNA codes of length 6. Comparing with some previous work such as [1,15], we can get more codewords.

5. ACKNOWLEDGEMENTS

This research is supported by the National Key Basic Research Program of China (Grant No. 2013CB834204), and the National Natural Science Foundation of China (Grant Nos. 11471255, 61171082 and 61301137).

6. REFERENCES

[1] T. Abualrub, A. Ghrayeb, X. Zeng: Construction of cyclic codes over F_4 for DNA computing, Journal of the Frankin Institute **343,** (4) 448-457(2006).

[2] T. Abualrub, R. Oehmke: On the generators of □ 4 codes of length 2^e, IEEE Transactions on Information Theory **49,** (9) 2126-2133 (2003).

[3] T. Abualrub, I. Siap: Cyclic codes over □ 2 + u□ 2 and □ 2 + u□ 2 + u²□ 2, Designs, Codes and Cryptography **42,** (3) 273-287(2007).

[4] L. Adleman: Molecular computation of solutions to combinatorial problems, Science **266,** 1021-1024(1994).

[5] L. Adleman, P. Rothemund, S. Roweis, E. Winfree: On applying molecular computation to the data encryption standard, Joural of Computational Biology **6,** (1) 53-63(1999).

[6] D. Boneh, C. Dunworth, R. Lipton: Breaking DES using molecular computer, Princeton CS Tech-Report Number CS-TR-489-95(1995).

[7] A. G. Frutos, Q. Liu, A. J. Thiel, A. M. W. Sanner, A. E. Condon, L. M. Smith, R. M. Corn: Demonstration of a word design strategy for DNA computing on surfaces, Nucleic Acids Res. (25) 4748-4757(1997).

[8] P. Gaborit, O. D. King: Linear construction for DNA codes. Theoret. Comput. Sci. (334) 99-13 (2005).

[9] A. Hammons, P. Kumar, A. Calderbank, N. Sloane, P. Sole: The □ 4 -Linearity of Kerdock, Preparata, Goethals and related codes, IEEE Transactions on Information Theory **40,** (2) 301-319(1994).

[10] O. King: Bounds for DNA codes with constant GC-content, J. Combin. (10) 1-13(2003).

[11] A. Marathe, A. E. Condon, R. M. Corn: On combinatorial DNA word design, J. Comput. Biol. (8) 201-220(2001).

[12] M. Mansuripur, P. Khulbe, S. Kuebler, J. Perry, M. Giridhar, N. Peyghambarian: Information storage and retrieval using macromolecules as storage media, University of Arizona Technical Report (2003).

[13] V. Pless, Z. Qian: Cyclic codes and quadratic residue codes over □ 4 , IEEE Transactions on Information Theory **42,** (5) 1594-1600(1996).

[14] I. Siap, T. Abualrub, A. Ahrayeb: Cyclic DNA codes over the ring $F_2[u]/(u^2-1)$ based on the deletion distance, Journal of the Franklin Institute **346,** (8) 731-740 (2009).

[15] B. Yildiz, I. Siap: Cyclic codes over $F_2[u]/(u^4-1)$ and applications to DNA codes, Computers and Mathematics with Applications 63**, (**7) 1169-1176(2012).