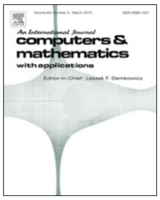
Computers and Mathematics with Applications 63 (2012) 1169–1176

Contents lists available at SciVerse ScienceDirect

Computers and Mathematics with Applications

journal homepage: www.elsevier.com/locate/camwa



Cyclic codes over $\mathbb{F}_2[u]/(u^4 - 1)$ and applications to DNA codes

Bahattin Yildiz [a,*], Irfan Siap [b]

[a] Department of Mathematics, Fatih University, 34500, Istanbul, Turkey

[b] Faculty of Science and Letters, Department of Mathematics, Yıldız Technical University, Istanbul, Turkey

ARTICLE INFO

Article history:
Received 11 January 2011
Accepted 9 December 2011

Keywords:
Cyclic codes
DNA-cyclic codes
Watson–Crick model
The *RC*-property
Reversible codes

ABSTRACT

The structure of DNA is used as a model for constructing good error correcting codes and conversely error correcting codes that enjoy similar properties with DNA structure are also used to understand DNA itself. Recently, naturally four element sets are used to model DNA by some families of error correcting codes. Hence the structure of such codes has been studied. In this paper, the authors first relate DNA pairs with a special 16 element ring. Then, the so-called cyclic DNA codes of odd length that enjoy some of the properties of DNA are studied. Their algebraic structure is determined. Further, by introducing a map, a family of cyclic codes over this ring is mapped to DNA codes. Hamming minimum distances are also studied. The paper concludes with some DNA examples obtained via this family of cyclic codes.

1. Introduction

DNA contains the instructions for the structure and function of cells. It contains the information on how the cell runs, reproduces, builds and repairs itself, and every other function necessary for the cell life. It is the long-term storage of information in all living species.

DNA sequences consists 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. These strands are paired with each other as a double helix. This pairing is done by obeying the Watson–Crick model. According to this model, A and T bound to each other and G and C bound to each other. A and G are called the complements of T and C, respectively or vice versa. The complement of a base say X will be denoted by $\overline{X}$, for instance, the complement of A is $\overline{A} = T$.

This pairing is done in reverse order with opposite direction. For instance the strand $5' - ACCTGAGT - 3'$ is paired with $3' - \overline{T}\,\overline{G}\,\overline{A}\,\overline{G}\,\overline{T}\,\overline{C}\,\overline{C}\,\overline{A} - 5'$ which is $3' - ACTCAGGA - 5'$ (Fig. 1).

Adelman [1] described an experiment involving the use of DNA molecules 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 [2], Boneh et al. and independently Adelman et al. in [3] presented a molecular program for breaking the Data Encryption *(DES)* cryptographic system. Another important application of bimolecular computing is the design of DNA chips for mutational analysis and for sequences. In [4], it is shown that DNA molecules can be used as a storage medium.

In all of the above applications the design of DNA strands that is suitable for such applications is an important problem.

Since the activities in cells are still not well understood and complicated, especially DNA plays an important role in cell division through Watson–Crick complements.

[*] Corresponding author. Tel.: +90 543 7791818.
E-mail addresses: byildiz@fatih.edu.tr (B. Yildiz), isiap@yildiz.edu.tr (I. Siap).

0898-1221/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2011.12.029

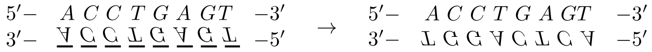


Fig. 1. A WCC pairing.

Codon are formed by three letter words, while genes are formed by codons. A system of genes directs the production of sensitive DNA repair enzymes, which monitor for genetic damage and fix most errors. In [5], the authors study error correction in DNA and develop an efficient procedure to determine whether such an error correcting code is present in the base sequences. Recently, the error correction capability of DNA has found interest in the error correcting theory. Basically, codes with similar properties as DNA are studied. This is natural since DNA has a complicated structure and also has an excellent error correcting capability. On the other hand, studying error correcting codes with similar properties as DNA helps on understanding the structure of DNA combinatorially contrary to chemical essays which are quite expensive.

The Hamming distance constraint, the reverse-complement constraint, the reverse constraint and the fixed GC-content constraint are the most common constraints used in DNA codes [6–9].

In [10,11,6,7], the authors focused on constructing large sets of DNA codewords with prescribed minimum Hamming distance.

Among many constraints we focus on Hamming distance constraint in this paper. The DNA alphabet consists of four letters. Most of the work in the literature is done over four element sets with algebraic structure. Linear codes that enjoy some of the properties of DNA are called DNA codes. Basically, linear codes with complement (C) or reverse (R) properties or both (C) and (R) properties and some constraints such as Hamming, GC-content, etc. are studied.

For instance, in [10] DNA codes over finite fields with four elements are studied. Later, in [12], DNA codes over the finite ring $\mathbb{F}_2[u]/(u^2-1)$ with four elements are studied. In [12], a motivation of studying DNA codes over this ring is also presented.

Error correcting codes have been extensively studied over finite binary fields and later this study has been extended to non binary finite fields and some good optimal codes are obtained [13,14]. Recently, error correcting codes over finite rings have been of great interest since in some special cases they produce "good" codes over finite fields [15,16].

Here, in this paper, we study a family of cyclic codes which are modeled by the properties of DNA and so called DNA cyclic codes over a finite ring with 16 elements. This ring contains the ring $\mathbb{F}_2[u]/(u^2-1)$, which was studied by Abualrub et al. [12], hence our work covers the DNA codes studied in that paper.

The sequel of the paper is structured as follows: The second section studies the structure of the ring $\mathbb{F}_2[u]/(u^4-1)$. Also the structure of linear codes over this ring and their generator matrix is presented here. In the third section, cyclic codes over the ring $\mathbb{F}_2[u]/(u^4-1)$ and their algebraic structure is studied. In the fourth section, a one to one correspondence between DNA double nucleotides and the elements of the ring $\mathbb{F}_2[u]/(u^4-1)$ is proposed. A family of cyclic codes over the ring $\mathbb{F}_2[u]/(u^4-1)$ with some constraints whose image under a special map gives DNA strands with required constraints is studied. It is shown that this family of DNA codes is a special ideal in this ring with generators having some constraints on them. So, the existence and the construction of such families is completely answered. In the fifth section, by applying the theory established in the previous sections, we present some cyclic DNA codes over the ring $\mathbb{F}_2[u]/(u^4-1)$ together with their images. In the sixth and the last section, the work in this paper is summarized and some further possible directions for study are pointed out.

2. Basics

The ring we consider in this work is

$$R=\mathbb{F}_2[u]/(u^4-1)=\{a+bu+cu^2+du^3 \mid a,b,c,d\in\mathbb{F}_2,\ u^4=1\},$$

which is a commutative, characteristic 2 ring of size 16. It is a finite chain ring and all its ideals can be listed as

$$\{0\}=(1+u)^4R\subset(1+u)^3R\subset(1+u)^2R\subset(1+u)R\subset R.$$

It is also a local ring with the unique maximal ideal being $\langle 1+u\rangle$ and $R/\langle 1+u\rangle\simeq\mathbb{F}_2$.

The group of units of R, is given by

$$\mathfrak{U}(R)=\{1,u,u^2,u^3,1+u+u^2,1+u^2+u^3,1+u+u^3,u+u^2+u^3\}=\langle u,1+u+u^3\rangle \tag{2.1}$$

which means the remaining 8 elements are zero and zero-divisors.

A code C over R of length n is a subset of R^n, while a linear code C over R of length n is an R-submodule of R^n. Because of the ideal structure of the ring R, we know that every linear code C over R of length n is equivalent to a code with the generating matrix

$$G=\begin{bmatrix} I_{k_1} & A_1 & A_2 & A_3 & A_4\\ 0 & (1+u)I_{k_2} & (1+u)B_1 & (1+u)B_2 & (1+u)B_3\\ 0 & 0 & (1+u)^2I_{k_3} & (1+u)^2C_1 & (1+u)^2C_2\\ 0 & 0 & 0 & (1+u)^3I_{k_4} & (1+u)^3D_1\end{bmatrix}$$

where A_i's B_j's C_k's and D_1 are matrices over R. A linear code that has such a generating matrix is said to be of *type* (k_1,k_2,k_3,k_4) and consequently has size $2^{4k_1+3k_2+2k_3+k_4}$.

3. Cyclic codes over the ring $\mathbb{F}_2[u]/(u^4-1)$

Let C be a code over $R=\mathbb{F}_2[u]/(u^4-1)$ of length n. A codeword $c=(c_0,c_1,\ldots,c_{n-1})$ of C can be viewed as a polynomial $c_0+c_1x+\cdots+c_{n-1}x^{n-1}\in R[x]$. Let τ be the cyclic shift acting on the codewords of C in the following way:

$$\tau(c_0,c_1,\ldots,c_{n-1})=(c_{n-1},c_0,c_1,\ldots,c_{n-2}). \tag{3.1}$$

We say the linear code C is cyclic if C is invariant under τ. Note that the action of τ on a polynomial $c(x)$ is $xc(x)$ modulo x^n-1. This leads to the observation that cyclic codes of length n over R are just ideals of $R[x]/(x^n-1)$. That is why it is necessary to study the ideal structure of the ring $R[x]/(x^n-1)$. The ideal structure of similar rings have been studied in the literature and we will make use of the same ideas. In particular, the results obtained in [17] seem to fit our particular case as one can see that the rings $\mathbb{F}_2[u]/(u^4)$ and $\mathbb{F}_2[u]/(u^4-1)$ are isomorphic.

Note that x^n-1 is a regular polynomial in $R[x]$. A factorization of x^n-1 in $\mathbb{F}_2[x]$ can be lifted uniquely to the same factorization in $R[x]$ since R is a local ring [18]. Hence we have the following lemma:

Lemma 3.1 (*[18]*)**.** *Let $x^n-1=f_1(x)f_2(x)\cdots f_l(x)$ be the factorization into irreducible monic polynomials $f_i(x)$ in $\mathbb{F}_2[x]$. Then x^n-1 has the same factorization and irreducible polynomials in $R[x]$.*

The following theorem can be easily obtained by following similar steps as in [19,17].

Theorem 3.2. *Assume that n is odd. There exist $F_i(0\le i\le 4)\in\mathbb{F}_2[x]$ basic irreducible and pairwise coprime polynomials where $x^n-1=F_0(x)F_1(x)\cdots F_4(x)$ such that any ideal in $R[x]/(x^n-1)$ is generated by $\{\hat{F}_1,(1+u)\hat{F}_2,(1+u)^2\hat{F}_3,(1+u)^3\hat{F}_4\}$; i.e. if C is a cyclic code of length n over R, then there exists F_i like above such that*

$$C=\left(\hat{F}_1,(1+u)\hat{F}_2,(1+u)^2\hat{F}_3,(1+u)^3\hat{F}_4\right).$$

Here, $\hat{F}_i$ denote the product of all F_j except F_i, in other words $\hat{F}_i=(x^n-1)/F_i$. Moreover, $|C|=2^k$, where $k=\sum_{i=0}^{3}(4-i)\deg F_{i+1}$.

Corollary 3.3. *Suppose n is odd, and C is a cyclic code over R of length n. Then there exists polynomials f_0,f_1,f_2,f_3 such that*

$$f_3\mid f_2\mid f_1\mid f_0\mid x^n-1$$

and

$$C=(f_0,(1+u)f_1,(1+u)^2f_2,(1+u)^3f_3).$$

Another corollary of Theorem 3.2 is that the ring $R[x]/(x^n-1)$ is a principal ideal ring when n is odd.

Corollary 3.4. *If C is a cyclic code over R of odd length n, then there exists $g(x)\in R[x]/(x^n-1)$ such that $C=(g)$.*

Proof. If C is a cyclic code over R of length n, for odd n, then by Theorem 3.2, $C=(\hat{F}_1,(1+u)\hat{F}_2,(1+u)^2\hat{F}_3,(1+u)^3\hat{F}_4)$, where $x^n-1=F_0(x)F_1(x)\ldots F_4(x)$, and the polynomials $F_i(x)$ are pairwise coprime. Let us take

$$g(x)=\hat{F}_1+(1+u)\hat{F}_2+(1+u)^2\hat{F}_3+(1+u)^3\hat{F}_4.$$

We claim that $C=(g)$. It is clear that $(g)\subset(\hat{F}_1,(1+u)\hat{F}_2,(1+u)^2\hat{F}_3,(1+u)^3\hat{F}_4)$.

To show the reverse inclusion, note that in $R[x]/(x^n-1)$, $\hat{F}_i\hat{F}_j=0$ whenever $i\neq j$. Also for any i, $\hat{F}_i$ and F_i are coprime in $\mathbb{F}_2[x]$ and so there exists polynomials $b_i,c_i\in\mathbb{F}_2[x]$ such that $b_i\hat{F}_i+c_iF_i=1$. This means for example that

$$(b_1\hat{F}_1+c_1F_1)(b_2\hat{F}_2+c_2F_2)(b_2\hat{F}_2+c_2F_2)=1$$

or that

$$a_0F_1F_2F_3+a_1\hat{F}_1F_2F_3+a_2F_1\hat{F}_2F_3+a_3F_1F_2\hat{F}_3=1$$

for some $a_i\in\mathbb{F}_2[x]$. Multiplying the above equation by $(1+u)^3\hat{F}_4$ we obtain

$$(1+u)^3\hat{F}_4=(1+u)^3a_0F_1F_2F_3\hat{F}_4.$$

But note that $F_1F_2F_3g=(1+u)^3F_1F_2F_3\hat{F}_4$ which means $(1+u)^3\hat{F}_4\in(g)$. Similarly it can be shown that $(1+u)^2\hat{F}_3$, $(1+u)\hat{F}_2$ and $\hat{F}_1$ are all in (g). This completes the proof. □

Table 4.1
Identifying nuclotide pairs with the elements of the ring.

AA	0	AT	$1+u$	GT	1	CT	$1+u+u^2$
TT	$1+u+u^2+u^3$	TA	u^2+u^3	TG	u^2	TC	$1+u^2+u^3$
GG	$1+u^2$	GC	$u+u^2$	AC	$1+u+u^3$	AG	u
CC	$u+u^3$	CG	$1+u^3$	CA	$u+u^2+u^3$	GA	u^3

4. DNA-cyclic codes over $\mathbb{F}_2[u]/(u^4-1)$

In previous papers [10,12,20], pairing of the nucleotides (A, T, G and C) is done by using a finite field or a ring with four elements. Here we pair doubles such as "AT, GC" with a ring element presented in Table 4.1. Since there are four basic nucleotides A, T, G and C, we have 16 pairs such as

$AA, AT, AC, AG, CA, CT, CG, CC, TA, TC, TT, TG, GA, GC, GT$ and GG.

Now we identify these 16 pairs with the elements of the ring $\mathbb{F}_2[u]/(u^4-1)=\{a+bu+cu^2+du^3 \mid a,b,c,d\in\mathbb{F}_2$ and $u^4=1\}$.

We define the map Φ which gives a one to one correspondence between $\mathbb{F}_2[u]/(u^4-1)^n$ and $\{A,T,G,C\}^{2n}$ given by Table 4.1.

This identification satisfies the complement property by simply adding $1+u+u^2+u^3$ to an element of the ring. For instance, if we take CA which is identified by $u+u^2+u^3$ and want to find its complement we just add $1+u+u^2+u^3$ so we get 1 which is GT. These is true for all other identifications. The WCC complement is defined as in the papers mentioned above. Another observation from the table is that multiplying an element a of R by u^2 reverses the DNA pair corresponding to a.

Now, by a DNA-cyclic code over R of length n, we mean a cyclic code that has the reverse complement property, i.e., C is DNA cyclic if C is cyclic and $f(x)=c_0+c_1x+\cdots c_{n-1}x^{n-1}\in C$ implies $f(x)^{RC}=\overline{c_{n-1}}+\overline{c_{n-2}}x+\cdots+\overline{c_1}x^{n-2}+\overline{c_0}x^{n-1}\in C$, where $\bar{a}$ stands for the complement of a in R given in Table 4.1.

Our goal is to classify such cyclic codes. Note that we can easily observe from Table 4.1 that the following lemma is true:

Lemma 4.1. *For any $a\in R$, we have $a+\bar{a}=u^3+u^2+u+1$.*

For $f(x)\in R[x]$, let $f(x)^*$ denote its reciprocal, i.e.

$$f(x)^*=x^{\deg(f)}f(1/x). \tag{4.1}$$

The following lemma is taken from [10] and helps classify DNA-cyclic codes:

Lemma 4.2. *Let $f(x)$ and $g(x)$ be polynomials in $R[x]$. Suppose $\deg(f)\geq\deg(g)$ and $m=\deg(f)-\deg(g)$. Then*

(i) $[f(x)g(x)]^*=f(x)^*g(x)^*$ *and*
(ii) $[f(x)+g(x)]^*=f(x)^*+x^mg(x)^*$.

Theorem 4.3. *Let $C=(f_0,(1+u)f_1,(1+u)^2f_2,(1+u)^3f_3)$ be a cyclic code over R of odd length n, with $f_3\mid f_2\mid f_1\mid f_0\mid x^n-1$ in $\mathbb{F}_2[x]$, given by Corollary 3.3. If $f(x)^{RC}\in C$ for any $f(x)\in C$, then $(1+u+u^2+u^3)((x^n-1)/(x-1))\in C$ and all f_i's are self-reciprocal polynomials.*

Proof. Let C be the cyclic code given as in the theorem. Since the zero codeword must be in C, by hypothesis, the WCC of it should also be in C. But by Lemma 4.1, we have

$$\overline{(0,0,\ldots,0)}=(1+u+u^2+u^3,\ldots,1+u+u^2+u^3)=(1+u+u^2+u^3)((x^n-1)/(x-1))\in C.$$

Now, let $f_0(x)=g_0+g_1x+\cdots g_rx^r$. Since $f_0|x^n-1$ in $\mathbb{F}_2[x]$, we can assume $g_0=g_r=1$, so that

$$f_0(x)=1+g_1x+g_2x^2+\cdots+g_{r-1}x^{r-1}+x^r.$$

Note that $f_0(x)$ corresponds to the vector $(1,g_1,\ldots,g_{r-1},1,0,\ldots,0)$, and since the complement of 0 in R is $1+u+u^2+u^3$, we see that

$$\begin{aligned}f_0(x)^{RC} &= (1+u+u^2+u^3)(1+x+\cdots+x^{n-r-2})+(u+u^2+u^3)x^{n-r-1}+\overline{g_{r-1}}x^{n-r}+\cdots+\overline{g_1}x^{n-2}\\ &\quad+(u+u^2+u^3)x^{n-1}\in C.\end{aligned}$$

Since C is a linear code, we must have

$$f_0(x)^{RC}+(1+u+u^2+u^3)\left(\frac{x^n-1}{x-1}\right)\in C.$$

But this implies that

$$\begin{aligned}&x^{n-r-1} + (\overline{g_{r-1}} + 1 + u + u^2 + u^3)x^{n-r} + \cdots + (\overline{g_1} + 1 + u + u^2 + u^3)x^{n-2} + x^{n-1}\\ &\quad = x^{n-r-1}[1 + (\overline{g_{r-1}} + 1 + u + u^2 + u^3)x + \cdots + (\overline{g_1} + 1 + u + u^2 + u^3)x^{r-1} + x^r] \in C.\end{aligned}$$

Multiplying this last polynomial by x^{r+1} and remembering that we are doing operations in $R[x]/(x^n - 1)$, we see that we must have

$$1 + (\overline{g_{r-1}} + 1 + u + u^2 + u^3)x + \cdots + (\overline{g_1} + 1 + u + u^2 + u^3)x^{r-1} + x^r \in C.$$

But by Lemma 4.1, we know that $\bar{a} + 1 + u + u^2 + u^3 = a$, so we obtain

$$f_0(x)^* = 1 + g_{r-1}x + \cdots + g_1x^{r-1} + x^r \in C.$$

But this means that we have

$$f_0(x)^* = f_0k_0 + (1+u)f_1k_1 + (1+u)^2f_2k_2 + (1+u)^3f_3k_3,$$

where f_i and k_i are all in $\mathbb{F}_2[x]$. But, note that $f_0(x)^* \in \mathbb{F}_2[x]$, and so we must have $k_1 = k_2 = k_3 = 0$. This means $f_0(x)^* = f_0(x)k_0(x)$. But note that both the leading coefficient and the constant terms of f_0 are 1. This means $\deg(f_0) = \deg(f_0^*)$, combine this with the fact that they have the same leading and constant terms, we obtain $f_0(x) = f_0(x)^*$, which means $f_0(x)$ is self reciprocal.

Now, let

$$(1+u)f_1(x) = (1+u)(1 + a_1x + \cdots + a_{t-1}x^{t-1} + x^t).$$

Then

$$\begin{aligned}(1+u)f_1(x)^{RC} &= (1 + u + u^2 + u^3)[1 + x + \cdots + x^{n-t-2}] + \overline{(1+u)}x^{n-t-1}\\ &\quad + \overline{(1+u)a_{t-1}}x^{n-t} + \cdots + \overline{(1+u)a_1}x^{n-2} + \overline{(1+u)}x^{n-1} \in C.\end{aligned}$$

Again, since C is linear, we must have

$$(1+u)f_1(x)^{RC} + (1 + u + u^2 + u^3)\left(\frac{x^n - 1}{x - 1}\right) \in C.$$

Now, when we add, we see that we get the following

$$\begin{aligned}&(1+u)x^{n-t-1} + (\overline{(1+u)a_{t-1}} + 1 + u + u^2 + u^3)x^{n-t} + \cdots + (\overline{(1+u)a_1} + 1 + u + u^2 + u^3)x^{n-2} + (1+u)x^{n-1}\\ &\quad = x^{n-t-1}[(1+u) + (\overline{(1+u)a_{t-1}} + 1 + u + u^2 + u^3)x + \cdots + (\overline{(1+u)a_1} + 1 + u + u^2 + u^3)x^{t-1}\\ &\qquad + (1+u)x^t] \in C.\end{aligned}$$

But note that $\overline{(1+u)a} + 1 + u + u^2 + u^3 = 0$ if $a = 0$ and it is $(1+u)$ if $a = 1$. So we get $\overline{(1+u)a} + 1 + u + u^2 + u^3 = (1+u)a$ for all $a \in \mathbb{F}_2$. But now, putting this above we see that we have obtained $(1+u)f_1(x)^* \in C$. But proceeding in the same way we did for the first case we get $f_1(x)^* = f_1(x)$. The same can be done for f_2 and f_3 as well. □

Now that we have obtained the necessary conditions for DNA cyclic codes over R, we must try to prove that these conditions are sufficient.

Theorem 4.4. *Suppose $C = (f_0, (1+u)f_1, (1+u)^2f_2, (1+u)^3f_3)$ is a cyclic code over R of odd length n, with $f_3 \mid f_2 \mid f_1 \mid f_0 \mid x^n - 1$ in $\mathbb{F}_2[x]$, given by Corollary 3.3. If $(1 + u + u^2 + u^3)((x^n - 1)/(x - 1)) \in C$ and $f_i(x)$'s are all self-reciprocal, then $f(x)^{RC} \in C$ for all $f(x) \in C$.*

Proof. Let $f_0(x) = 1 + g_1x + g_2x^2 + \cdots + g_{r-1}x^{r-1} + x^r$. Suppose $c(x) \in C$, say

$$c(x) = f_0k_0 + (1+u)f_1k_1 + (1+u)^2f_2k_2 + (1+u)^3f_3k_3 \tag{4.2}$$

where k_i are polynomials in $\mathbb{F}_2[x]$. Taking the reciprocals and by a repeated use of Lemma 4.2, and using the fact that f_i's are all self-reciprocal, we see that there exists integers m_1, m_2, m_3 such that

$$\begin{aligned}c(x)^* &= [f_0k_0 + (1+u)f_1k_1 + (1+u)^2f_2k_2 + (1+u)^3f_3k_3]^*\\ &= (f_0k_0)^* + x^{m_1}((1+u)f_1k_1)^* + x^{m_2}((1+u^2)f_2k_2)^* + x^{m_3}((1+u)^3f_3k_3)^*\\ &= f_0k_0^* + (1+u)f_1x^{m_1}(k_1)^* + (1+u)^2f_2x^{m_2}(k_2)^* + (1+u)^3f_3x^{m_3}(k_3)^*,\end{aligned}$$

which means that $c(x)^* \in C$ for all $c(x) \in C$.

Since $(1 + u + u^2 + u^3)((x^n - 1)/(x - 1)) \in C$, we know

$$(1 + u + u^2 + u^3) + (1 + u + u^2 + u^3)x + \cdots + (1 + u + u^2 + u^3)x^{n-1} \in C. \tag{4.3}$$

Now, let $c(x) = c_0 + c_1x + \cdots + c_tx^t \in C$. Since C is cyclic, we know that

$$x^{n-t-1}c(x) = c_0x^{n-t-1} + c_1x^{n-t} + \cdots + c_tx^{n-1} \in C. \tag{4.4}$$

Adding Eqs. (4.3) and (4.4) we get

$$\begin{bmatrix} (1+u+u^2+u^3) + (1+u+u^2+u^3)x + \cdots + (1+u+u^2+u^3)x^{n-t-2} \\ + (c_0+1+u+u^2+u^3)x^{n-t-1} + \cdots + (c_t+1+u+u^2+u^3)x^{n-1} \end{bmatrix} \in C. \tag{4.5}$$

But, by Lemma 4.1, we know that $a + (1+u+u^2+u^3) = \overline{a}$. Thus Eq. (4.5) becomes

$$\begin{aligned} &(1+u+u^2+u^3) + (1+u+u^2+u^3)x + \cdots + (1+u+u^2+u^3)x^{n-t-2} \\ &\quad + \overline{c_0}x^{n-t-1} + \overline{c_1}x^{n-t} + \cdots + \overline{c_{t-1}}x^{n-2} + \overline{c_t}x^{n-1} \end{aligned}$$

which is precisely $(c(x)^*)^{RC}$. Thus we obtain $(c^*(x)^{RC})^* = c(x)^{RC} \in C$. □

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

Corollary 4.5. *Let C be a cyclic DNA code of length n over the ring R and minimum Hamming distance d. Then, $\Phi(C)$ is a DNA code of length $2n$ over the alphabet $\{A, T, G, C\}$ with minimum Hamming distance at least d.*

Corollary 4.6. *If $C = \langle(1+u^2)g(x)\rangle$ with $g(x)|x^n - 1$ is a cyclic DNA code of length n over the ring R with minimum Hamming distance d, then, $\Phi(C)$ is a DNA linear code of length $2n$ over the alphabet $\{AA, TT, CC, GG\}$ with minimum Hamming distance at least d. If $C = \langle(1+u^2)g(x)\rangle$ with $g(x)|x^n - 1$ is a cyclic DNA code of length n, size M, with minimum distance d over the ring $\mathbb{F}_2[u]/(u^4-1)$, then there exists a DNA linear code of length n, size M, with minimum distance code d.*

Proof. Due to the algebra in $\mathbb{F}_2[u]/(u^4-1)$, $(1+u^2)R = \{0, 1+u^2, u+u^3, (1+u)^3\}$. Note that the images of elements in $(1+u^2)R$ under Φ are precisely AA, GG, CC and TT respectively. So, this means if c is any codeword in $C = \langle(1+u^2)g(x)\rangle$, then $\Phi(c)$ consists of these pairs of DNA nucleotides. Thus, if two codewords in C differ at a coordinate, this corresponds to exactly 2 coordinates in $\Phi(C)$ where they differ, which means the distances in the image are twice that of the pre-image. This tells us that the Hamming distance of the image is equal to $2d$, but the size and the length are the same. Now, if $C = \langle(1+u^2)g(x)\rangle$ with $g(x)|x^n - 1$ is a cyclic DNA code of length n, size M, with minimum distance code $2d$ over the ring $\mathbb{F}_2[u]/(u^4-1)$, then by puncturing the linear code over the quotient ring $R/(1+u^2)R$ obtained above in the odd indexed coordinates we obtain the desired linear code (An example of this is presented in 5.3). □

5. Examples

While searching for codes of particular length n over the ring $\mathbb{F}_2[u]/(u^4-1)$ we apply Theorem 4.4. So we need to have the factorization of $x^n - 1$ over $\mathbb{F}_2$ and determine all possible f_0, f_1, f_2 and f_3 with particular requirements. All f_i must be self-reciprocal, not divisible by $x - 1$ and naturally $f_3 \mid f_2 \mid f_1 \mid f_0 \mid x^n - 1$ in $\mathbb{F}_2[x]$.

Example 5.1. Since $x^3 - 1 = (x-1)(x^2+x+1) \in \mathbb{F}_2[x]$, if $C = \langle g(x)\rangle$ where $g(x) = x^2+x+1$, then C is a cyclic DNA code of length 3 with R-C property and minimum Hamming distance 3. The image of C under the map Φ is a DNA code of length 6, size 16 and minimum Hamming distance 3 (please see Table 5.1). As expected this is a repetition code.

The only possible cyclic DNA codes over $\mathbb{F}_2[u]/(u^4-1)$ are ideals generated by $g(x) = x^2+x+1$, $(1+u)g(x)$, $(1+u^2)g(x)$ and $(1+u)^3g(x)$. Hence, there are only four DNA codes of length 3 over the ring $\mathbb{F}_2[u]/(u^4-1)$.

Example 5.2. Since $x^5 - 1 = (x-1)(x^4+x^3+x^2+x+1) \in \mathbb{F}_2[x]$, if $C = \langle g(x)\rangle$ where $g(x) = x^4+x^3+x^2+x+1$, then C is a cyclic DNA code of length 5 with R-C property and minimum Hamming distance 5. The image of C under the map Φ is a DNA code of length 5, size 16 and minimum Hamming distance 5. Again, here, the only possible cyclic DNA codes of length 5 over $\mathbb{F}_2[u]/(u^4-1)$ are ideals generated by $g(x) = x^4+x^3+x^2+x+1$, $(1+u)g(x)$, $(1+u^2)g(x)$ and $(1+u)^3g(x)$. Hence, there are only four DNA codes of length 5 over the ring $\mathbb{F}_2[u]/(u^4-1)$. All non trivial cyclic DNA codes over the ring $\mathbb{F}_2[u]/(u^4-1)$ of length 5 and minimum distance of its image DNA code is listed in Table 5.4.

Example 5.3. Since $x^9 - 1 = (x-1)(x^2+x+1)(x^6+x^3+1) \in F_2[x]$, if $C = \langle(1+u^2)g(x)\rangle$ where $g(x) = x^6+x^3+1$, then C is a cyclic DNA code of length 9 with R-C property. By puncturing the odd indices of the image code $\Phi(C)$ we obtain a DNA code of length 9 minimum distance 3 given in Table 5.2.

We give all non trivial DNA cyclic codes over $\mathbb{F}_2[u]/(u^4-1)$ up to some moderate lengths:
Length 3: See Table 5.3.
Length 5: See Table 5.4.

Table 5.1
A DNA code of length 6 obtained from $C = \langle x^2 + x + 1\rangle$.

AAAAAA	*TTTTTT*	*CCCCCC*	*GGGGGG*
GAGAGA	*CTCTCT*	*AGAGAG*	*TCTCTC*
TGTGTG	*ACACAC*	*GTGTGT*	*CACACA*
TATATA	*GCGCGC*	*ATATAT*	*CGCGCG*

Table 5.2
A punctured DNA code of length 9 obtained from $C = \langle x^6 + x^3 + 1\rangle$.

AAAAAAAAA	*CCCCCCCCC*	*ACAACAACA*	*AGAAGAAGA*
TTTTTTTTT	*GGGGGGGGG*	*TGTTGTTGT*	*TCTTCTTCT*
ATAATAATA	*CACCACCAC*	*CGCCGCCGC*	*CTCCTCCTC*
TATTATTAT	*GTGGTGGTG*	*GCGGCGGCG*	*GAGGAGGAG*
CTTCTTCTT	*TTCTTCTTC*	*CCTCCTCCT*	*TCCTCCTCC*
GAAGAAGAA	*AAGAAGAAG*	*GGAGGAGGA*	*AGGAGGAGG*
AACAACAAC	*CAACAACAA*	*AGCAGCAGC*	*CGACGACGA*
TTGTTGTTG	*GTTGTTGTT*	*TCGTCGTCG*	*GCTGCTGCT*
AATAATAAT	*TAATAATAA*	*ACCACCACC*	*CCACCACCA*
TTATTATTA	*ATTATTATT*	*TGGTGGTGG*	*GGTGGTGGT*
ACGACGACG	*GCAGCAGCA*	*ACTACTACT*	*TCATCATCA*
TGCTGCTGC	*CGTCGTCGT*	*TGATGATGA*	*AGTAGTAGT*
ATCATCATC	*CTACTACTA*	*ATGATGATG*	*GTAGTAGTA*
TAGTAGTAG	*GATGATGAT*	*TACTACTAC*	*CATCATCAT*
CAGCAGCAG	*GACGACGAC*	*CCGCCGCCG*	*GCCGCCGCC*
GTCGTCGTC	*CTGCTGCTG*	*GGCGGCGGC*	*CGGCGGCGG*

Table 5.3
($g(x) = x^2 + x + 1$.) Non trivial cyclic DNA codes over the ring $\mathbb{F}_2[u]/(u^4 - 1)$ of length $n = 3$ and its DNA code image.

No	The code	Length	Type	$d_H(C)$	$d(\Phi(C))$
1.	$\langle g(x)\rangle$	3	(1,0,0,0)	3	3
2.	$\langle (1+u)g(x)\rangle$	3	(0,1,0,0)	3	3
3.	$\langle (1+u)^2g(x)\rangle$	3	(0,0,1,0)	3	6
4.	$\langle (1+u)^3g(x)\rangle$	3	(0,0,0,1)	3	6

Table 5.4
($g(x) = x^4 + x^3 + x^2 + x + 1$.) Non trivial cyclic DNA codes over the ring $\mathbb{F}_2[u]/(u^4 - 1)$ of length $n = 5$ and its DNA code image.

No	The code	Length	Type	$d_H(C)$	$d(\Phi(C))$
1.	$\langle g(x)\rangle$	5	(1,0,0,0)	5	5
2.	$\langle (1+u)g(x)\rangle$	5	(0,1,0,0)	5	5
3.	$\langle (1+u)^2g(x)\rangle$	5	(0,0,1,0)	5	10
4.	$\langle (1+u)^3g(x)\rangle$	5	(0,0,0,1)	5	10

Table 5.5
Some non trivial cyclic DNA codes over the ring $\mathbb{F}_2[u]/(u^4 - 1)$ of length $n = 9$ and the DNA code images.

No	The code	Length	Type	$d_H(C)$	$d(\Phi(C))$
1.	C_{02}	9	(3,0,0,0)	3	3
2.	C_{12}	9	(0,3,0,0)	3	3
3.	C_{22}	9	(0,0,3,0)	3	6
4.	C_{32}	9	(0,0,0,3)	3	6

P.S. There are more cyclic DNA codes of length 9, but they are not listed here in order to save space.

Length 7: Since $x^7 - 1 = (x + 1)(x^3 + x + 1)(x^3 + x^2 + 1), x^7 - 1$ has the trivial reciprocal divisor $x + 1$ and $x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$, so there are cyclic DNA codes of length 7 over the ring $\mathbb{F}_2[u]/(u^4 - 1)$.

Length 9:

Let $g_1(x) = x^2 + x + 1$ and $g_2(x) = x^6 + x^3 + 1$. Since $x^n - 1 = (x + 1)g_1(x)g_2(x)$, cyclic DNA codes of length 9 are:

$C_{ji} = \langle (1 + u)^j g_i(x)\rangle$ where $0 \leq j \leq 3$ and $1 \leq i \leq 2$ and $C_{kli} = \langle (1 + u)^k g_1(x)g_2(x), (1 + u)^l g_i(x)\rangle$ where $i = 1, 2$ and $0 \leq k < l \leq 3$. (See Table 5.5).

6. Conclusion

The algebraic structure of the ring $\mathbb{F}_2[u]/(u^4 - 1)$ and a special family of cyclic codes of odd length over this ring are studied. This family of cyclic codes is related to DNA codes and their relation is also studied. Necessary and sufficient conditions for cyclic codes to have the DNA properties have been explored. Some examples of such families are presented together with their Hamming distances. Note that as $\mathbb{F}_2[u]/(u^2-1)$ is a subring of $R=\mathbb{F}_2[u]/(u^4-1)$, the codes we have obtained cover those obtained in [12]. Although we have characterized DNA-cyclic codes over R of odd lengths, the DNA images of these codes have even lengths.

For future study, the algebraic structure of cyclic codes of even length and their relation to DNA codes is still an open problem. Next, the study of cyclic DNA codes over the ring $\mathbb{F}_2[u]/(u^4-1)$ with different metrics(such as deletion, etc.) rather then Hamming metric is still an open and potentially fruitful direction. Some recently introduced new metrics related to studying DNA codes that may be of future interest are presented in [21,12]. Another possible research direction is considering DNA codes over more general rings than $\mathbb{F}_2[u]/(u^4-1)$.

References

[1] L. Adleman, Molecular computation of solutions to combinatorial problems, Science 266 (1994) 1021–1024.

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

[3] L. Adleman, P.W.K. Rothemund, S. Roweis, E. Winfree, On applying molecular computation to the data encryption standard, Journal of Computational Biology 6 (1999) 53–63.

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

[5] Larry S. Liebovitch, Yi Tao, Angelo T. Todorov, Leo Levine, Is there an error correcting code in the base sequence in DNA? Biophysical Journal 71 (1996) 1539–1544.

[6] 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 Research 25 (1997) 4748–4757.

[7] O.D. King, Bounds for DNA codes with constant GC-content, Electronic Journal of Combinatorics 10 (2003) R33. pp. 1–13.

[8] M. Li, H.J. Lee, A.E. Condon, R.M. Corn, DNA word design strategy for creating sets of non-interacting oligonucleotides for DNA microarrays, Langmuir 18 (2002) 805–812.

[9] A. Marathe, A.E. Condon, R.M. Corn, On combinatorial DNA word design, Journal of Computational Biology 8 (2001) 201–220.

[10] Taher Abulraub, Ali Ghrayeb, Xiang Nian Zeng, Construction of cyclic codes over $GF(4)$ for DNA computing, Journal of the Franklin Institute 343 (4–5) (2006) 448–457.

[11] P. Gaborit, O.D. King, Linear construction for DNA codes, Theoretical Computer Science 334 (1–3) (2005) 99–113.

[12] Irfan Siap, Taher Abulraub, Ali Ghrayeb, Cyclic DNA codes over the ring $\mathbb{F}_2[u]/(u^2-1)$ based on the deletion distance, Journal of the Franklin Institute 346 (8) (2009) 731–740.

[13] N. Aydin, I. Siap, New quasi-cyclic codes over F_5, Applied Mathematics Letters 15 (7) (2002) 833–836.

[14] T.A. Gulliver, V.K. Bhargava, New linear codes over $GF(8)$, Applied Mathematics Letters 13 (2) (2000) 17–19.

[15] Jian-Fa Qian, Li-Na Zhang, Shi-Xin Zhu, $(1+u)$-constacyclic and cyclic codes over F_2+uF_2, Applied Mathematics Letters 19 (8) (2006) 820–823.

[16] Maria Carmen V. Amarra, Fidel R. Nemenzo, On $(1-u)$-cyclic codes over $F_{p^k}+uF_{p^k}$, Applied Mathematics Letters 21 (11) (2008) 1129–1133.

[17] Mehmet Ozen, Irfan Siap, Linear codes over $\mathbb{F}_q[u]/(u^s)$ with respect to the Rosenbloom–Tsfasman metric, Direct Client-to-Client 38 (2006) 17–29.

[18] Bernard R. McDonald, Finite Rings with Identity, Dekker, New York, 1974.

[19] Promad Konmar, Sergio Ki Lopez-Permouth, Cyclic codes over integers modulo p^m, Finite Fields and their Applications 3 (1997) 334–352.

[20] Irfan Siap, Taher Abulraub, Ali Ghayreb, Similarity cyclic DNA codes over rings, in: International Conference on Bioinformatics and Biomedical Engineering, in Shanghai, PRC, May 16–18th 2008.

[21] I. Siap, M. Ozen, The complete weight enumerator for codes over $M_{n\times s}(R)$, Applied Mathematics Letters 17 (1) (2004) 65–69.