

Watson-Crick Conjugate and Commutative Words

Lila Kari and Kalpana Mahalingam

University of Western Ontario,
Department of Computer Science,
London, ON, Canada N6A 5B7
{lila, kalpana}@csd.uwo.ca

Abstract. This paper is a theoretical study of notions in combinatorics of words motivated by information being encoded as DNA strands in DNA computing. We generalize the classical notions of conjugacy and commutativity of words to incorporate the notion of an involution function, a formalization of the Watson-Crick complementarity of DNA single-strands. We define and study properties of Watson-Crick conjugate and commutative words, as well as Watson-Crick palindromes. We obtain, for example, a complete characterization of the set of all words that are not Watson-Crick palindromes. Our results hold for more general functions, such as arbitrary morphic and antimorphic involutions. They generalize classical results in combinatorics of words, while formalizing concepts meaningful for DNA computing experiments.

1 Introduction

Theoretical DNA Computing is an area of biomolecular computing that has seen a surge of activity in recent years. It loosely encompasses contributions to fundamental research in computer science originated in or motivated by research in DNA computing. Examples are numerous and they include theoretical aspects of self-assembly [1], [20], DNA sequence design [11], [17], and mathematical properties of DNA-encoded information [10], [8].

This paper constitutes a contribution to the field of theoretical DNA computing by investigating a generalization of the classical notions of conjugacy and commutativity of words motivated by DNA-encoded information. The main idea is that information-encoding strings that are used in DNA computing experiments have an important property that differentiates them from their electronic computing counterparts. This property is the Watson-Crick complementarity between DNA single-strands that allows information-encoding strands to potentially interact. Mathematically, this translates into generalizing the identity function, which is the only one operating in the electronic realm, to an arbitrary involution function. An involution is a function θ such that θ^2 equals the identity. Given an alphabet Σ , an antimorphic involution, i.e., an involution θ with the additional property that $\theta(uv) = \theta(v)\theta(u)$ for all strings $u, v \in \Sigma^*$, is the mathematical notion that formalizes the Watson-Crick complementarity. Indeed,

an antimorphic involution captures the two main properties of the Watson-Crick complement of a DNA strand, namely its being the reverse (antimorphic property) complement (involution property) of the original strand. Replacing identity with involutions paves thus the way to concepts that are both meaningful formalizations of information-encoding DNA strands, and mathematically interesting generalizations of classical concepts in formal language theory, coding theory and combinatorics of words.

For example, using the concept of involutions one obtains generalizations of the classical notions of prefix codes, suffix codes and comma-free codes [12], [13]. In addition to being of theoretical interest, these notions prove to be meaningful in the context of DNA computing experiments. Indeed, if θ is the Watson-Crick involution, then a θ -sticky-free, or θ -overhang-free code is a set of words where no unwanted hybridizations of a certain type occur between DNA codewords. More recently, in [14] we extended the concept of bordered and unbordered words to involution-bordered and involution-unbordered words.

In this paper we extend the notions of conjugate and commutative words to Watson-Crick conjugate and Watson-Crick commutative words. Our results hold in a more general context where the function θ involved is an arbitrary morphic or antimorphic involution. To put these results in context, they augment studies of combinatorial properties of words which have meaningful applications in numerous other fields. For example, word properties such as periodicity and borderedness play a role in many areas including string searching algorithms [4,5,6], data compression [7,21] and in the study of coding properties of sets of words [2,19] as well as sequence assembly in computational biology [18]. Relevant to this paper, there are several classical results about conjugacy of words and words that commute [19]. In addition, in [3] the authors extend certain combinatorial properties of conjugacy of words to partial words with an arbitrary number of holes. An authoritative text on the study of combinatorial properties of strings would be [16].

The paper is organized as follows. We begin by reviewing basic concepts of combinatorics of words and the definition of θ -bordered and θ -unbordered words for an arbitrary morphic or antimorphic involution θ . In Section 2, we also define the concept of θ -conjugacy on words. If θ is the antimorphic Watson-Crick involution, this gives rise to the notion of Watson-Crick conjugate words. Figure 1 illustrates the interaction between two DNA strands u and v over the DNA alphabet $\Delta = \{A, C, G, T\}$ that are Watson-Crick conjugates to each other. We show that for a morphic involution θ , the θ -conjugacy on words is reflexive, symmetric and transitive. We also obtain several properties of θ -conjugate words including a general characterization of the words that are θ -conjugate in Proposition 1. These results generalize well-known properties of conjugate words [19].

In Section 3, we introduce the concept of θ -commutativity on words for an arbitrary morphic or antimorphic involution θ , and its particular case of Watson-Crick commutativity. Figure 3 illustrates the interaction between two DNA strands u and v that Watson-Crick commute. We obtain several properties of words that θ -commute, including their characterization (Proposition

3), and properties of the set $C_\theta(1)$ of words that cannot be written as a concatenation of two non-empty words x, y such that x θ -commutes with y . These properties generalize classical properties of words that commute, [19]. We define the notion of θ -palindrome that was obtained independently in [9]. Note that if θ is the Watson-Crick involution, then the notion of Watson-Crick palindromes coincides with the term “palindrome” as used by molecular biologists. We define a relation on words using the θ -commutativity and show that, for an antimorphic involution θ , the set of all θ -palindromes can be characterized using this relation.

2 Watson-Crick Conjugate Words

Before introducing the notion of Watson-Crick conjugate words, we review some basic concepts of combinatorics of words. An alphabet Σ is a finite non-empty set of symbols. A word u over Σ is a finite sequence of symbols in Σ . We denote by Σ^* the set of all words over Σ , including the empty word λ and, by Σ^+ , the set of all non-empty words over Σ . We note that with the concatenation operation on words, Σ^* is the free monoid and Σ^+ is the free semigroup generated by Σ . For a word $w \in \Sigma^*$, the length of w is the number of symbols in w and is denoted by $|w|$. For a word w , the set of its prefixes/ suffixes are defined as follows: $Pref(w) = \{u \in \Sigma^+ | \exists v \in \Sigma^*, w = uv\}$ and $Suff(w) = \{u \in \Sigma^+ | \exists v \in \Sigma^*, w = vu\}$.

Bordered words were initially called “overlapping words” and unbordered words were called “non-overlapping words”. For properties of bordered and unbordered words we refer the reader to [19,22]. In [14], we extended the concept of bordered words to involution bordered words. We now recall some notions defined and used in [22] and [14].

Definition 1. *Let θ be either a morphic or an antimorphic involution on Σ^* .*

1. *For $v, w \in \Sigma^*$, $w \leq_p v$ iff $v \in w\Sigma^*$.*
2. *For $v, w \in \Sigma^*$, $w \leq_s^\theta v$ iff $v \in \Sigma^*\theta(w)$.*
3. *$\leq_d^\theta = \leq_p \cap \leq_s^\theta$.*
4. *For $u \in \Sigma^*$, $v \in \Sigma^*$ is said to be a θ -border of u if $v \leq_d^\theta u$, i.e., $u = vx = y\theta(v)$.*
5. *For $w, v \in \Sigma^*$, $w <_p v$ iff $v \in w\Sigma^+$.*
6. *For $w, v \in \Sigma^*$, $w <_s^\theta v$ iff $v \in \Sigma^+\theta(w)$.*
7. *$<_d^\theta = <_p \cap <_s^\theta$.*
8. *For $u \in \Sigma^*$, $v \in \Sigma^*$ is said to be a proper θ -border of u if $v <_d^\theta u$.*
9. *For $u \in \Sigma^+$, define $L_d^\theta(u) = \{v | v \in \Sigma^*, v <_d^\theta u\}$.*
10. *$\nu_\theta(u) = |L_d^\theta(u)|$.*
11. *$D_\theta(i) = \{u | u \in \Sigma^+, \nu_\theta(u) = i\}$.*
12. *A word $u \in \Sigma^+$ is said to be θ -bordered if there exists $v \in \Sigma^+$ such that $v <_d^\theta u$, i.e., $u = vx = y\theta(v)$ for some $x, y \in \Sigma^+$.*
13. *A non-empty word which is not θ -bordered is called θ -unbordered.*

A word u in Σ^* is a conjugate of w in Σ^* if there exists $v \in \Sigma^*$ such that $uv = vw$. Note that conjugacy on words is an equivalence relation. In [3], the authors showed that conjugacy on partial words is reflexive and symmetric but not transitive. In this section we extend the concept of conjugacy of words to incorporate the notion of an involution function and show that θ -conjugacy on words is reflexive. We also show that θ -conjugacy on words is symmetric and transitive when θ is a morphic involution.

Definition 2. Let θ be either a morphic or an antimorphic involution. A word u is a θ -conjugate of another word w if $uv = \theta(v)w$ for some $v \in \Sigma^*$.

Example 1. Let $\Sigma = \{a, b\}$ and θ be an antimorphic involution which maps a to b and vice versa. Let $u = aba$ and $w = bab$. Then u is a θ -conjugate of w since $aba \cdot b = \theta(b) \cdot bab$.

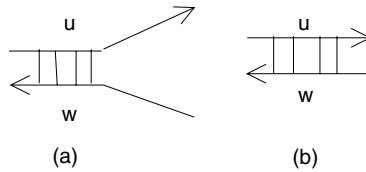


Fig. 1. If u is Watson-Crick conjugate of w , then u and the Watson-Crick complement of w overlap, resulting thus in one of the two intermolecular hybridizations shown above

For any DNA string u over the DNA alphabet $\Delta = \{A, G, C, T\}$, the Watson-Crick conjugates of u are defined as the DNA strings w such that $uv = \theta(v)w$ for some $v \in \Delta^*$. In this case, θ is the Watson-Crick involution which maps $A \mapsto T$, $C \mapsto G$ and viceversa such that θ is an antimorphic involution. In the following example we find all the Watson-Crick conjugates of a given DNA string.

Example 2. Let $\Delta = \{A, G, C, T\}$ be the DNA alphabet and let $u = ATAG$. Then the Watson-Crick Conjugates of u are given by $Conj_\theta(u) = \{ATAG, TAGT, AGAT, GTAT, CTAT\}$. For all $w \in Conj_\theta(u)$, there exists a $v \in \Sigma^*$ such that $uv = \theta(v)w$. These words v respectively are $T, AT, TAT, CTAT$.

The characterization of θ -conjugate words in Proposition 2 will show that if u and w are Watson-Crick conjugates, then u and the Watson-Crick complement of w overlap, hence forming the hybridization in Fig 1.

Note that for all $u \in \Sigma^*$, u is a θ -conjugate of u since $u\lambda = \theta(\lambda)u$. Also u is a θ -conjugate of $\theta(u)$ since $u\theta(u) = \theta(\theta(u))\theta(u)$ and hence for all $u, v \in \Sigma^*$, uv is a θ -conjugate to $v\theta(u)$ since $uv\theta(u) = \theta(\theta(u))v\theta(u)$. Even though we concentrate on Watson-Crick conjugates, we provide results that hold for any general morphic or an antimorphic involution. In the next lemma we show that the θ -conjugacy of words is transitive when θ is a morphic involution.

Lemma 1. *Let $u, v, w \in \Sigma^+$ such that u is a θ -conjugate of w and w is a θ -conjugate of v .*

1. *If θ is a morphic involution then u is a θ -conjugate of v .*
2. *If θ is an antimorphic involution then u is not necessarily a θ -conjugate of v .*

Proof. 1. Let θ be a morphic involution. Since u is a θ -conjugate of w and w is a θ -conjugate of v then there exists $r, s \in \Sigma^*$ such that $ur = \theta(r)w$ and $ws = \theta(s)v$ which implies that $urs = \theta(r)\theta(s)v$. Hence $urs = \theta(rs)v$ and u is a θ -conjugate of v .

2. Let θ be an antimorphic involution. Then u is not necessarily a θ -conjugate of v . For example let $\Sigma = \{a, b\}$ and $\theta(a) = b$ and let $u = aba$, $w = bab$ and $v = bba$. Note that aba is a θ -conjugate of bab since $aba \cdot b = \theta(b) \cdot bab$. Also bab is a θ -conjugate of bba since $bab \cdot ba = \theta(ba) \cdot bba$. Suppose there exist a $y \in \Sigma^*$ such that $aba \cdot y = \theta(y) \cdot bba$ then $\theta(y) = ax$ for some $x \in \Sigma^*$ which implies that $y = \theta(x)b$ which is not possible since y has to be of the form za . Hence the θ -conjugacy relation is not transitive for an antimorphic involution θ .

□

Lemma 2. *Let $x, y \in \Sigma^*$ such that x is a θ -conjugate of y .*

1. *If θ is an antimorphic involution then for all $u \in \Sigma^*$ ux is a θ -conjugate of $y\theta(u)$.*
2. *If θ is a morphic involution then there exists a $u \in \Sigma^*$ such that ux is not a θ -conjugate of $y\theta(u)$.*

Proof. 1. Let θ be an antimorphic involution. Since x is a θ -conjugate of y there exists $v \in \Sigma^*$ such that $xv = \theta(v)y$ and hence $uxv\theta(u) = u\theta(v)y\theta(u)$. Take $r = v\theta(u)$, then $\theta(r) = \theta(v\theta(u)) = u\theta(v)$ which implies that $uxr = \theta(r)y\theta(u)$ hence ux is a θ -conjugate of $y\theta(u)$.

2. Let θ be a morphic involution and let $\Sigma = \{a, b\}$ such that $\theta(a) = b$. Note that for $x = abb$ and $y = bbb$, x is a θ -conjugate of y since $x \cdot b = \theta(b) \cdot bbb$. But for $u = ab$ ux is not a θ -conjugate of $y\theta(u)$. Also for $w = ux = ababb$, the set of all θ -conjugates is $C = \{babaa, bbaba, bbbab, abbba, babbb, ababb\}$ and clearly $y\theta(u) = bbbba \notin C$.

□

Proposition 1. *Let u be a θ -conjugate of w such that $uv = \theta(v)w$ for some $v \in \Sigma^*$. Then for a morphic involution θ there exists $x, y \in \Sigma^*$ such that $u = xy$ and one of the following hold:*

1. *$w = y\theta(x)$ and $v = (\theta(x)\theta(y)xy)^i\theta(x)$ for some $i \geq 0$.*
2. *$w = \theta(y)x$ and $v = (\theta(x)\theta(y)xy)^i\theta(x)\theta(y)x$ for some $i \geq 0$.*

Proof. Let θ be a morphic involution. Given $uv = \theta(v)w$ for some $v \in \Sigma^*$. Then we either have $|u| < |v|$ or $|v| \leq |u|$. Suppose $|u| \geq |v|$ then $u = \theta(v)\alpha$ and $w = \alpha v$

for some $\alpha \in \Sigma^*$. Hence for $v = \theta(x)$, $u = xy$ and $w = y\theta(x)$. Assume that $|u| < |v|$. Then there exists $p_1 \in \Sigma^+$ such that $\theta(v) = up_1$ and $v = p_1w$. Hence $v = p_1w = \theta(u)\theta(p_1)$. Suppose $|u| < |p_1|$ then there exists $p_2 \in \Sigma^+$ such that $p_1 = \theta(u)p_2$ and $\theta(p_1) = p_2w$ and hence $u\theta(p_2) = p_2w$ and $v = \theta(p_2)\theta(w)w = \theta(u)u\theta(p_2)$. Continuing this way we can find a $p_n \in \Sigma^+$ such that $|u| > |p_n|$ and $v = a_j^n\theta(x_n)$ for $a_j = \theta(u)$ when j is odd and $a_j = u$ when j is even. When n is even, we have $n = 2k$ and $v = (\theta(u)u)^k\theta(x_{2k})$ with $u\theta(x_{2k}) = x_{2k}w$ which implies $u = x_{2k}r = xy$ and $w = r\theta(x_{2k}) = y\theta(x)$ and $v = (\theta(x)\theta(y)xy)^k\theta(x)$. When n is odd $n = 2k + 1$ for some k and $v = (\theta(u)u)^{2k-1}\theta(u)\theta(x_{2k+1})$ with $\theta(u)\theta(x_{2k+1}) = x_{2k+1}w$. Then we have $\theta(u) = x_{2k+1}r = \theta(x)\theta(y)$ and $w = r\theta(x_{2k+1}) = \theta(y)x$ and $v = (\theta(x)\theta(y)xy)^{2k-1}\theta(x)\theta(y)x$. \square

Corollary 1. *For a morphic involution θ on Σ^* , θ -conjugacy on words is a symmetric relation.*

Example 3. Let $\Sigma = \{a, b\}$ and let θ be a morphic involution which maps a to b and viceversa. From Proposition 1 for $w = ux = ababb$, the set of all θ -conjugates are $C = \{babaa, bbaba, bbbab, abbba, babbb, ababb\}$.

Proposition 2. *Let u be a θ -conjugate of w . Then for an antimorphic involution θ , there exists $x, y \in \Sigma^*$ such that either $u = xy$ and $w = y\theta(x)$ (Figure 1, (a)) or $w = \theta(u)$ (Figure 1, (b)).*

Corollary 2. *Let θ be either a morphic or an antimorphic involution and let u be a θ -conjugate of w for $u, w \in \Sigma^+$. Then either uw or wu is θ -bordered.*

Let u be a θ -conjugate of w . Then for an antimorphic involution θ , either uw or wu precisely form a hairpin-like structure. For example, choose a DNA string $u = ATAGCT$ and one of its Watson-Crick conjugates $w = GCTTAT$. Then $uw = ATAGCTGCTTAT = (ATA)GCTGCT\theta(ATA)$, as illustrated in Fig. 2.



Fig. 2. The DNA string $GCTTAT$ is a Watson-Crick conjugate of $ATAGCT$, and their catenation $ATAGCTGCTTAT$ forms a hairpin

3 Watson-Crick Commutative Words

Two words x and y are said to commute when $xy = yx$, [19]. In this section we define the concept of θ -commutative words and show that commutative words are a special case of θ -commutative words when θ is identity. We also introduce the θ -commutativity order and characterize the set of all θ -palindromes for an antimorphic involution θ .

Definition 3. Let θ be either a morphic or an antimorphic involution.

1. For $x, y \in \Sigma^*$, x is said to θ -commute with y if $xy = \theta(y)x$.
2. We define the θ -commutativity order as $v \leq_c^\theta u$ iff $u = vx = \theta(x)v$ for some $x \in \Sigma^*$.
3. $L_c^\theta(u) = \{v | v \in \Sigma^*, v \leq_c^\theta u\}$.
4. $\nu_c^\theta(u) = |L_c^\theta(u)|$.
5. For $i \geq 1$, define $C_\theta(i) = \{u | u \in \Sigma^+, \nu_c^\theta(u) = i\}$.
6. A word $x \in \Sigma^*$ is called a θ -palindrome if $x = \theta(x)$.

Suppose $uv = \theta(v)u$ holds. Then, if $v = \lambda$, then u is a θ -conjugate of u . (This also implies that u θ -commutes with λ .) Otherwise, it means that u θ -commutes with v . For any non-empty DNA strings u and v over the DNA alphabet $\Delta = \{A, G, C, T\}$, we say that u Watson-Crick commutes with v if $uv = \theta(v)u$ where θ is the Watson-Crick involution. The word $u \in \Delta^*$ is called a Watson-Crick palindrome if $u = \theta(u)$ where θ is the Watson-Crick involution. In what follows, we will show that for the Watson-Crick involution θ if u θ -commutes with v , then u is a Watson-Crick palindrome and either u is a prefix of $\theta(v)$ or $\theta(v)$ is a prefix of u .

Example 4. Consider a string $u = AGCT$ over the DNA alphabet Δ . Let θ be the Watson-Crick involution and $v = CTAGAGCT$. Then u θ -commutes with v since $uv = AGCT \cdot CTAGAGCT = \theta(CTAGAGCT) \cdot AGCT = \theta(v)u$.

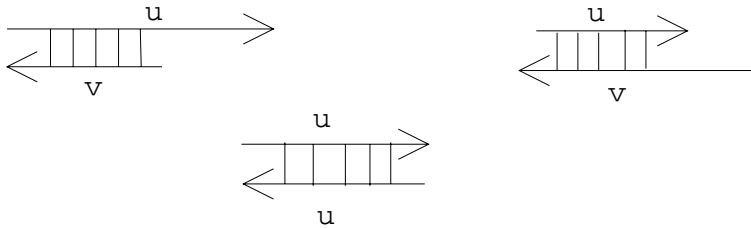


Fig. 3. If the DNA string u Watson-Crick commutes with v , then one of the intermolecular hybridizations (a) or (b) occurs and, in addition, u is a Watson-Crick palindrome (c).

If the word u Watson-Crick commutes with the word v , the characterization in Proposition 3 will show that u and v will form one of the hybridizations in Figure 3.

Observation 1 Let θ be either a morphic or an antimorphic involution on Σ^* .

1. For all $u \in \Sigma^+$, $u \in L_c^\theta(u)$, i.e., $u \leq_c^\theta u$.
2. $C_\theta(1) = \{u \in \Sigma^+ | v \leq_c^\theta u \Leftrightarrow v = u\}$.
3. For all $u \in \Sigma^+$ such that u is a θ -palindrome we have $\lambda \in L_c^\theta(u)$.
4. For all $a \in \Sigma$ such that $a \neq \theta(a)$, $a^+ \subseteq C_\theta(1)$.

Note that $C_\theta(1)$ is the set of all words that cannot be written as a catenation of two non-empty words x and y such that x θ -commutes with y . $C_\theta(1)$ is the set of all words u that have only one element in the set $L_c^\theta(u)$ namely u . In particular, θ -palindromes are not in $C_\theta(1)$. In the next lemma we show that for an antimorphic involution θ , the set $L_c^\theta(u)$ is a totally ordered set with respect to \leq_c^θ .

Lemma 3. *For an antimorphic involution θ and $u \in \Sigma^+$, $L_c^\theta(u)$ is a totally ordered set with \leq_c^θ .*

The proof technique of the following proposition is similar to that of Proposition 1 and hence we omit the proof.

Proposition 3. *Let $u, v \in \Sigma^+$ such that u θ -commutes with v , i.e., $uv = \theta(v)u$.*

1. *If θ is an antimorphic involution then $u = x(yx)^i$, $v = yx$ where $i \geq 0$ (Figure 3 (a), or (b)) and u (Figure 3, (c)) as well as x, y are θ -palindromes, where $x \in \Sigma^+$, $y \in \Sigma^*$.*
2. *If θ is a morphic involution then $u = x(yx)^i$ and $v = yx$ where $yx = \theta(x)\theta(y)$ and $i \geq 0$ with $y \in \Sigma^*$ and $x \in \Sigma^+$.*

It was shown in [15] that when $u = xy$ such that $x, y \in P_\theta$ and for an antimorphic involution θ , u can be written as $(\alpha\beta)^n$ with $x = (\alpha\beta)^i\alpha$ and $y = \beta(\alpha\beta)^{n-i-1}$. The authors also proved that for $u = xy = \theta(y)\theta(x)$ for a morphic involution θ , either $u = \alpha^m$ for $\alpha \in P_\theta$ or $u = [\alpha\theta(\alpha)]^n$ for some $\alpha \in \Sigma^+$. We use these results and Proposition 3 to deduce the following corollary.

Corollary 3. *Let $u, v \in \Sigma^+$ such that u θ -commutes with v .*

1. *If θ is a morphic involution then one of the following hold:*
 - (a) *$u = \alpha^m$ and $v = \alpha^n$ for some $m, n \geq 1$, $\alpha \in P_\theta$.*
 - (b) *$u = \theta(\alpha)[\alpha\theta(\alpha)]^m$ and $v = [\alpha\theta(\alpha)]^n$ for some $n \geq 1$ and $k \geq 0$ with $\alpha \in \Sigma^+$.*
2. *If θ is an antimorphic involution, $u = \beta(\alpha\beta)^n$ and $v = (\alpha\beta)^m$ for some $\alpha, \beta \in P_\theta$ with $m \geq 1$ and $n \geq 0$.*

Based on the definitions and the previous two results we have the following observation.

Lemma 4. *Let $w \in \Sigma^+$ and θ be an antimorphic involution. Then w is a θ -palindrome iff there exists $v \in \Sigma^*$ such that $v \neq w$ and $v \leq_c^\theta w$.*

Note that Lemma 4 states that, for an antimorphic involution θ , a word $w \in C_\theta(1)$ iff w is not a θ -palindrome, i.e., the set $L = \Sigma^* \setminus C_\theta(1)$ is the set of all θ -palindromes.

Note that for a word w which is not a θ -palindrome for an antimorphic involution θ , $L_c^\theta(w)$ may be an emptyset. For example, let $\Sigma = \{a, b\}$, and θ be an antimorphic involution that maps a to b and vice versa. Let $w = ababa$, then $\theta(w) = babab$. Clearly $w \neq \theta(w)$. Note that

- $w = abab \cdot a \neq \theta(a) \cdot abab = babab$.
- $w = aba \cdot ba \neq \theta(ba) \cdot aba = baaba$.

- $w = ab \cdot aba \neq \theta(aba) \cdot ab = babab.$
- $w = a \cdot baba \neq \theta(baba) \cdot a = babaa.$

Thus it is clear that for $w = ababa$ there does not exist a $v \in \Sigma^*$ such that $w = vx = \theta(x)v$ and thus $L_c^\theta(w) = \emptyset.$

Lemma 5. *Let θ be either a morphic or an antimorphic involution. For all $u \in \Sigma^+, \theta(L_c^\theta(u)) = L_c^\theta(\theta(u)).$*

Lemma 6. *Let θ be either a morphic or an antimorphic involution. Then for all $u \in C_\theta(1)$ we have $u^+ \subseteq C_\theta(1).$*

It is shown in [14], that if u and v are θ -unbordered for an antimorphic involution $\theta,$ then for $u = u_1u_2$ such that $u_1, u_2 \in \Sigma^+, u_1vu_2$ is also θ -unbordered. But it is not true for words in $C_\theta(1).$ We illustrate it in the following example.

Example 5. Let $\Sigma = \{a, b\}$ and let θ be an antimorphic involution that maps a to b and vice versa. Note that $u = abb \in C_\theta(1)$ since $u \neq \theta(u)$ and $ab \cdot b \neq a \cdot ab$ and $a \cdot bb \neq aa \cdot a.$ Let $v = a$ and $v \in C_\theta(1)$ since $v \neq \theta(v).$ But u_1vu_2 with $u_1, u_2 \in \Sigma^+$ is either $a \cdot a \cdot bb$ or $ab \cdot a \cdot b.$ Note that both $aabb, abab \notin C_\theta(1)$ since $aabb = \theta(aabb)$ and $abab = \theta(abab).$

Proposition 4. *Let $u, v \in C_\theta(1)$ and $\theta(\text{Pref}(u)) \cap \text{Suff}(v) = \emptyset.$*

1. *If θ is an antimorphic involution then $uv \in C_\theta(1).$*
2. *If θ is a morphic involution then uv is not necessarily in $C_\theta(1).$*

Proof. 1. Let θ be an antimorphic involution. Suppose for some $u, v \in C_\theta(1), uv \notin C_\theta(1)$ then there exists $\alpha \in \Sigma^+$ such that $uv = p\alpha = \theta(\alpha)p.$ Then we have the following cases. If $|\alpha| \leq |v|$ and $|\theta(\alpha)| \leq |u|,$ we have $v = r\alpha$ and $u = \theta(\alpha)s$ then $\alpha \in \theta(\text{Pref}(u)) \cap \text{Suff}(v)$ a contradiction. If $|\alpha| \leq |v|$ and $|\theta(\alpha)| \leq |uv|$ we have $v = r\alpha$ and $\theta(\alpha) = us$ then $\alpha = \theta(s)\theta(u)$ and $v = r\theta(s)\theta(u)$ for some $r, s \in \Sigma^*$ which implies $\theta(u) \in \theta(\text{Pref}(u)) \cap \text{Suff}(v)$ which is a contradiction. If $|\alpha| \leq |uv|$ and $|\theta(\alpha)| \leq |u|$ we have $\alpha = rv$ and $u = \theta(\alpha)s$ then $u = \theta(v)\theta(r)s$ for some $r, s \in \Sigma^*$ which implies $\theta(v) \in \text{Pref}(u)$ and hence $v \in \theta(\text{Pref}(u)) \cap \text{Suff}(v)$ which is a contradiction. If $|\alpha| \leq |uv|$ and $|\theta(\alpha)| \leq |uv|$ we have $\alpha = rv$ and $\theta(\alpha) = us$ then $\alpha = rv = \theta(s)\theta(u).$ Then we have the following subcases:

- If $|v| = |u|$ then $\theta(u) = v.$
- If $|u| < |v|$ then $v = \beta\theta(u)$ for some $\beta \in \Sigma^+$ and $\theta(u) \in \theta(\text{Pref}(u)) \cap \text{Suff}(v).$
- If $|v| < |u|$ then $\theta(u) = \beta v$ for some $c \in \Sigma^+$ and $\theta(u_2)\theta(u_1) = \beta v$ with $u = u_1u_2$ and $|u_1| = |v|$ which implies $\theta(u_1) \in \theta(\text{Pref}(u)) \cap \text{Suff}(v).$

All the above cases arrive at a contradiction. Hence $uv \in C_\theta(1).$

2. Let $u = ab$ and $v = a$ over the alphabet set $\Sigma = \{a, b\}$ and let θ be a morphic involution that maps a to b and vice versa. Then $\theta(u) = ba$ and $\theta(v) = b.$ Note that $u, v \in C_\theta(1).$ Also $\text{Pref}(u) = \{a, ab\}, \theta(\text{Pref}(u)) = \{b, ba\}$ and $\text{Suff}(v) = \{a\}.$ Note that $\theta(\text{Pref}(u)) \cap \text{Suff}(v) = \emptyset.$ But $uv = aba \notin C_\theta(1)$ since $a \cdot ba = \theta(ba) \cdot a = aba$ which implies that $a \in L_c^\theta(uv).$

□

Note that the converse of the statement 1 in Proposition 4 does not hold in general. Let $\Sigma = \{a, b\}$ and θ be an antimorphic involution that maps a to b and vice versa. Let $u = aba$ and $v = abb$. Then $\theta(u) = bab$ and $\theta(v) = aab$. Note that $u \neq \theta(u)$ and $v \neq \theta(v)$ and $u, v \in C_\theta(1)$. For $uv = abaabb$, $\theta(uv) = aabbab$ and

- $abaab \cdot b \neq \theta(b) \cdot abaab = aabaab.$
- $abaa \cdot bb \neq \theta(bb) \cdot abaa = aaabaa.$
- $aba \cdot abb \neq \theta(abb) \cdot aba = aababa.$
- $ab \cdot aabb \neq \theta(aabb) \cdot ab = aabbab.$
- $a \cdot baabb \neq \theta(baabb) \cdot a = aabbaa.$

Hence $uv \in C_\theta(1)$. But $\text{Pref}(u) = \{a, ab, aba\}$, $\text{Suff}(v) = \{b, bb, abb\}$ and $\theta(\text{Pref}(u)) = \{b, ab, bab\}$. Thus $b \in \theta(\text{Pref}(u)) \cap \text{Suff}(v) \neq \emptyset$.

Lemma 7. *Let θ be either a morphic involution or an antimorphic involution and let Σ be such that for all $a \in \Sigma$, $\theta(a) \neq a$. Then $D_\theta(1) \subseteq C_\theta(1)$.*

In [14], it was shown that for an antimorphic involution θ , the set of all θ -bordered words is regular. Note that from Lemma 4, $C_\theta(1)$ is the set of all non θ -palindromes for an antimorphic involution θ . We show using pumping lemma for regular languages that $\Sigma^* \setminus C_\theta(1)$ is not regular and hence $C_\theta(1)$ is not regular for an antimorphic involution θ .

Lemma 8. *When θ is an antimorphic involution, the set of all θ -palindrome words is not regular.*

Proof. Let $\Sigma = \{a, b\}$ and let θ be an antimorphic involution that maps $a \mapsto b$ and viceversa. Assume that the language L of all θ -palindromes is regular and let n be the constant given by the pumping lemma. Chose $w = a^n b^n$ and note that $w = \theta(w)$ and hence w is a θ -palindrome. Let $w = a^n b^n = xvy$ such that $|xv| \leq n$ and $|v| > 0$. Then $z = xv^i y$ contains more a 's than b 's for all i and hence z is not a θ -palindrome. Thus $L = \Sigma^* \setminus C_\theta(1)$ is not regular. \square

In our last proposition we construct a context-free grammar that generates the set of all θ -palindromes over a finite alphabet set for an antimorphic involution θ .

Proposition 5. *For an antimorphic involution θ , the set $L = \Sigma^* \setminus C_\theta(1)$ is context-free.*

Proof. Let Σ be a finite alphabet set and let $G = (\{X, Y\}, \Sigma, X, \mathcal{R})$ where $\mathcal{R} = \{X \rightarrow \lambda, Y \rightarrow \lambda, X \rightarrow a_i X \theta(a_i) \text{ for all } a_i \in \Sigma \text{ and } X \rightarrow b_i Y, Y \rightarrow b_i Y \text{ for all } b_i \in \Sigma \text{ such that } b_i = \theta(b_i)\}$. It is easy to check that G generates the set of all θ -palindromes over Σ and G is context-free. Hence $L(G) = \Sigma^* \setminus C_\theta(1)$. \square

It is shown in Proposition 5.4 in [14] that for a morphic involution θ , the set of all θ -bordered words is not context-free. It is also clear from Proposition 5.4 in [14] that $L = \Sigma^* \setminus C_\theta(1)$ is not context-free when θ is a morphic involution.

Acknowledgment. Research supported by NSERC and Canada Research Chair grants for Lila Kari.

References

1. Adleman, L.: Towards a mathematical theory of self-assembly. Technical Report 00-722, Department of Computer Science, University of Southern California (2000)
2. Berstel, J., Perrin, D.: *Theory of Codes*. Academic Press, Orlando Florida (1985)
3. Blanchet-Sadri, F., Luhman, D.: Conjugacy on partial words. *Theoretical Computer Science* 289, 297–312 (2002)
4. Boyer, R., Moore, J.: A fast string searching algorithm. *Communication of the ACM* 20, 762–772 (1977)
5. Crochemore, M., Perrin, D.: Two-way string matching. *Journal of Association of Computing Machinery* 38, 651–675 (1991)
6. Crochemore, M., Rytter, W.: *Jewels of Stringology*. World Scientific (2003)
7. Crochemore, M., Mignosi, F., Restivo, A., Salemi, S.: Text compression using antidictionaries. In: Wiedermann, J., van Emde Boas, P., Nielsen, M. (eds.) *ICALP 1999*. LNCS, vol. 1644, pp. 261–270. Springer, Heidelberg (1999)
8. Daley, M., McQuillan, I.: On computational properties of template-guided DNA recombination. In: Carbone, A., Pierce, N.A. (eds.) *DNA Computing*. LNCS, vol. 3892, pp. 27–37. Springer, Heidelberg (2006)
9. de Luca, A., de Luca, A.: Pseudopalindrome closure operators in free monoids. *Theoretical Computer Science* 362, 282–300 (2006)
10. Domaratzki, M.: Hairpin structures defined by DNA trajectories. In: Mao, C., Yokomori, T. (eds.) *DNA Computing*. LNCS, vol. 4287, pp. 182–194. Springer, Heidelberg (2006)
11. Garzon, M., Phan, V., Roy, S., Neel, A.: In search of optimal codes for DNA computing. In: Mao, C., Yokomori, T. (eds.) *DNA Computing*. LNCS, vol. 4287, pp. 143–156. Springer, Heidelberg (2006)
12. Kari, L., Konstantinidis, S., Losseva, E., Wozniak, G.: Sticky-free and overhang-free DNA languages. *Acta Informatica* 40, 119–157 (2003)
13. Kari, L., Konstantinidis, S., Losseva, E., Sosik, P., Thierrin, G.: Hairpin structures in DNA words. In: Carbone, A., Pierce, N.A. (eds.) *DNA Computing*. LNCS, vol. 3892, pp. 158–170. Springer, Heidelberg (2006)
14. Kari, L., Mahalingam, K.: Involution bordered words. *International Journal of Foundations of Computer Science* (accepted, 2007)
<http://www.csd.uwo.ca/~lila/invbor.pdf>
15. Kari, L., Mahalingam, K., Seki, S.: Language equations on Watson-Crick words, manuscript
16. Lothaire, M.: *Combinatorics of Words*. Cambridge University Press, Cambridge (1997)
17. Marathe, A., Condon, A., Corn, R.: On combinatorial DNA word design. In: Winfree, E., Gifford, D. (eds.) *Proc. of DNA Based Computers 5*, DIMACS Series in Discrete Math. and Theoretical Comp. Sci. pp. 75–89 (1999)
18. Margaritis, D., Skiena, S.: Reconstructing strings from substrings in rounds. In: *Proceedings of the 36th Annual Symposium on Foundations of Computer Science*, pp. 613–620 (1995)
19. Shyr, H.J.: *Free Monoids and Languages*. Hon Min Book Company (2001)
20. Soloveichik, D., Winfree, E.: Complexity of compact proofreading for self-assembled patterns. In: Carbone, A., Pierce, N.A. (eds.) *DNA Computing*. LNCS, vol. 3892, pp. 305–324. Springer, Heidelberg (2006)
21. Storer, J.A.: *Data Compression: Methods and Theory*. Computer Science Press, Rockville (1998)
22. Yu, S.S.: d-minimal languages. *Discrete Applied Mathematics* 89, 243–262 (1998)