On a Special Class of Primitive Words

Elena Czeizler, Lila Kari, and Shinnosuke Seki

Department of Computer Science, University of Western Ontario,
London, Ontario, Canada, N6A 5B7,
elenac, lila, sseki@csd.uwo.ca

Abstract. When representing DNA molecules as words, it is necessary to take into account the fact that a word u encodes basically the same information as its Watson-Crick complement $\theta(u)$, where θ denotes the Watson-Crick complementarity function. Thus, an expression which involves only a word u and its complement can be still considered as a repeating sequence. In this context, we define and investigate the properties of a special class of primitive words, called θ-primitive, which cannot be expressed as such repeating sequences. For instance, we prove the existence of a unique θ-primitive root of a given word, and we give some constraints forcing two distinct words to share their θ-primitive root. Also, we present an extension of the well-known Fine and Wilf Theorem, for which we give an optimal bound.

1 Introduction

Encoding information as DNA strands as in, e.g., DNA Computing, brings up for investigation new features based on the specific biochemical properties of DNA molecules. Recall that single-stranded DNA molecules can be viewed as words over the quaternary alphabet of bases $\{A, T, C, G\}$. Moreover, one of the main properties of DNA molecules is the Watson-Crick complementarity of the bases A and T and respectively G and C. Because of this property two Watson-Crick complementary single DNA strands with opposite orientation bind together to form a DNA double strand, in a process called base-pairing. Recently, there were several approaches to generalize notions from classical combinatorics on words in order to incorporate this major characteristic of DNA molecules, see, e.g., [6], [7], and [9]. Following these lines, in this paper, we generalize the concept of *primitivity* and define *θ-primitive words*.

The notion of primitivity plays an important role in various fields of theoretical computer science, such as algebraic coding theory, [11], and combinatorics on words, [8]. A word is called *primitive* if it cannot be decomposed as a power of another word. Thus, investigating the primitivity of a word is often the first step when analyzing its properties. Moreover, how a word can be decomposed and whether two words are powers of a common word are two questions which were widely investigated in language theory, see, e.g., [2], [8], and [12]. While, in classical combinatorics on words we look for repetitions of the form u^i for some word u and some $i \geq 2$, when dealing with DNA molecules (i.e., their abstract

representation as words) we should exploit the fact that a word u encodes the same information as its complement $\theta(u)$, where θ denotes the Watson-Crick complementarity function, or its mathematical formalization as an arbitrary antimorphic involution. In other words, we should look for expressions involving a word u and its complement $\theta(u)$. In this context, we define θ-primitive words as strings which cannot be decomposed using only some word u and its complement. Also, we define the *θ-primitive root* of a word w as the shortest word u such that w can be decomposed using only u and its complement. In classical combinatorics on words, there exist two equivalent definitions for the *primitive root* of a word w: the shortest word u such that $w = u^i$ for some $i \geq 1$, or the unique primitive word u such that $w = u^i$ for some $i \geq 1$. In our search for such equivalent definitions for the θ-primitive root of a word, we succeed to prove an extension of the well-known Fine and Wilf Theorem, one of the most widely used results on words. Although it was initially proved in connection with real functions, [5], the Fine and Wilf Theorem can be naturally interpreted also as a result on words, see, e.g., [2] and [8]. Moreover, several extensions of this theorem were proved so far, see, e.g., [1], [3], [4], and [10]. In this paper, we look at the case when a word w has two decompositions: one using a word u and its complement $\theta(u)$, and the other using some other word v and its complement $\theta(v)$. If w is longer than a given bound, then we prove that u and v share their θ-primitive root t and, thus w will have a refined decomposition depending on t and its complement. Moreover, we show that our bound is optimal, i.e., twice the length of the longer word (u or v) plus the length of the other word minus the greatest common divisor of the lengths of u and v.

The paper is organized as follows. In Section 2, we fix our terminology and recall some basic results. In Section 3 we investigate some basic properties of θ-primitive words. In particular, we give an extension of the Fine and Wilf Theorem which implies immediately that we can define the θ-primitive root of a word in two equivalent ways. In Section 4, we present several constraints forcing two words to share their θ-primitive root. In Section 5, we investigate some connections between the θ-primitive words that we introduced here and the θ-palindrome words, which were proposed and investigated in [7] and [9]. In Section 6, we present the optimal bound for our extension of the Fine and Wilf Theorem.

2 Preliminaries

Let Σ be a finite alphabet. We denote by Σ^* the set of all finite words over Σ, by ϵ the empty word, and $\Sigma^+ = \Sigma^* \setminus \{\epsilon\}$. The *length* of a word w, denoted by $|w|$, is the number of letters occurring in it, i.e., if $w = a_1 \dots a_n$ with $a_i \in \Sigma$, $1 \leq i \leq n$, then $|w| = n$. We say that u is a *prefix* (resp. a *suffix*) of v if $v = ut$ (resp. $v = tu$) for some $t \in \Sigma^*$. For any $0 \leq k \leq |v|$, we use the notation $pref_k(v)$ (resp. $suf_k(v)$) for the prefix (resp. suffix) of length k of a word v and Pref(v) (resp. Suff(v)) for the set of all prefixes (resp. all suffixes) of v. In particular

$pref_0(v) = \epsilon$ for any word $v \in \Sigma^*$. An integer $p \geq 1$ is a *period* of a word $w = a_1 \dots a_n$, with $a_i \in \Sigma$ for all $1 \leq i \leq n$, if $a_i = a_{i+p}$ for all $1 \leq i \leq n - p$.

A word $w \in \Sigma^+$ is called *primitive* if it cannot be written as a power of another word; that is, $w = u^n$ implies $n = 1$ and $w = u$. For a word $w \in \Sigma^+$, the shortest $u \in \Sigma^+$ such that $w = u^n$ for some $n \geq 1$ is called the *primitive root* of the word w and is denoted by $\rho(w)$. The following result gives an alternative, equivalent way for defining the primitive root of a word.

Theorem 1. *For a word $w \in \Sigma^*$, there exists a unique primitive word $t \in \Sigma^+$ such that $\rho(w) = t$, i.e., $w = t^n$ for some $n \geq 1$.*

The next result illustrates another useful property of primitive words.

Proposition 1. *Let $u \in \Sigma^+$ be a primitive word. Then, u cannot be a factor of u^2 in a nontrivial way, i.e., if $u^2 = xuy$, then necessarily either $x = \epsilon$ or $y = \epsilon$.*

We say that two words u and v *commute* if $uv = vu$. The following result characterizes the commutation of two words in terms of primitive roots.

Theorem 2. *For $u, v \in \Sigma^*$, the following conditions are equivalent: i) u and v commute; ii) u and v satisfy a non-trivial relation, i.e., an equation where the two sides are not graphically identical; iii) u and v have the same primitive root.*

Two words u and v are said to be *conjugate* if there exist words x and y such that $u = xy$ and $v = yx$. In other words, v can be obtained via a cyclic permutation of u. The next result characterizes the conjugacy of two words.

Theorem 3. *Let $u, v \in \Sigma^+$. Then, the following conditions are equivalent: i) u and v are conjugate; ii) there exists a word z such that $uz = zv$; moreover, this holds if and only if $u = pq$, $v = qp$, and $z = (pq)^i p$, for some $p, q \in \Sigma^*$ and $i \geq 0$; iii) the primitive roots of u and v are conjugate.*

Note that conjugacy is an equivalence relation, the *conjugacy class* of a word w consisting of all conjugates of w. The following is a well-known result.

Proposition 2. *If w is a primitive word, then its conjugacy class contains $|w|$ distinct primitive words.*

The following result, known as the Fine and Wilf theorem in its form for words, cf. [2] and [8], illustrates a fundamental periodicity property of words. As usual, $gcd(n, m)$ denotes the *greatest common divisor of n and m.*

Theorem 4. *Let $u, v \in \Sigma^*$, $n = |u|$, $m = |v|$, and $d = gcd(n, m)$. If two powers u^i and v^j of u and v have a common prefix of length at least $n + m - d$, then u and v are powers of a common word. Moreover, the bound $n + m - d$ is optimal.*

A mapping $\theta : \Sigma^* \to \Sigma^*$ is called a *morphism* (resp. an *antimorphism*) if for any words $u, v \in \Sigma^*$, $\theta(uv) = \theta(u)\theta(v)$ (resp. $\theta(uv) = \theta(v)\theta(u)$). Moreover, a mapping $\theta : \Sigma^* \to \Sigma^*$ is called an *involution* if, for all words $u \in \Sigma^*$, $\theta(\theta(u)) = u$.

For a mapping $\theta : \Sigma^* \to \Sigma^*$, a word $w \in \Sigma^*$ is called *θ-palindrome* if $w = \theta(w)$, see [7] and [9]. Now we say that a word $w \in \Sigma^+$ has a *θ-decomposition* if there exist a positive integer $k \geq 2$ and some words $t, w_1, \ldots, w_k \in \Sigma^+$ such that $w = w_1 \ldots w_k$ and $w_i \in \{t, \theta(t)\}$ for all $1 \leq i \leq k$. In this case, we say that w is *θ-periodic*, with *θ-period* $|t|$. We call a word $w \in \Sigma^+$ *θ-primitive* if it has no θ-decompositions, i.e., its least θ-period is $|w|$. We define *the θ-primitive root of* w, denoted by $\rho_\theta(w)$, as the shortest word t such that $w = w_1 \ldots w_k$ for some $k \geq 2$, $w_i \in \{t, \theta(t)\}$ for all $1 \leq i \leq k$, and $w_1 = t$. Note that if w is θ-primitive, then we can fix $\rho_\theta(w) = w$.

3 Properties of θ-Primitive Words

In this section, we consider $\theta : \Sigma^* \to \Sigma^*$ to be either a morphic or antimorphic involution, other than the identity function. We start by looking at some basic properties of θ-primitive words.

Proposition 3. *If a word $w \in \Sigma^+$ is θ-primitive, then it is also primitive. Moreover, the converse is not always true.*

Proof. Suppose that w is a θ-primitive word but not primitive. Then, there exists some $t \in \Sigma^+$ such that $w = t^n$ with $n \geq 2$. But then we can θ-decompose w as $w = w_1 \ldots w_n$, where $w_1 = \ldots = w_n = t$, which contradicts the θ-primitivity of w. For the converse, since θ is not the identity function, there exists a letter a such that $\theta(a) \neq a$. Then, if we take $w = a\theta(a)$, it is obvious that w is primitive, but not θ-primitive. □

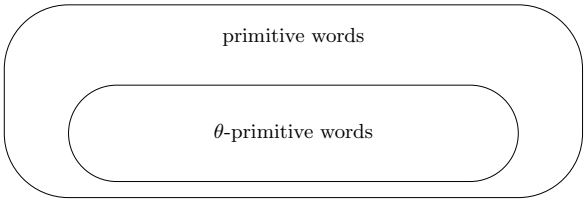

Fig. 1. The sets of primitive and θ-primitive words

Thus, the class of θ-primitive words is strictly included in the set of primitive ones, as illustrated in Fig. 1.

Proposition 4. *The θ-primitive root of a word is θ-primitive.*

Proof. Let $w \in \Sigma^+$ and $t = \rho_\theta(w)$ be its θ-primitive root. We can suppose, without loss of generality, that w is not θ-primitive; otherwise, $\rho_\theta(w) = w$ and thus the θ-primitive root is obviously θ-primitive. Then, we can write $w = w_1 \ldots w_n$, where $n \geq 2$ and $w_i \in \{t, \theta(t)\}$ for all $1 \leq i \leq n$. Suppose, now that t is not θ-primitive. Then, there exist a word $s \in \Sigma^+$ with $|s| < |t|$ and a positive integer $k \geq 2$, such that t has the θ-decomposition $t = t_1 \ldots t_k$, where for all $1 \leq i \leq k$, $t_i \in \{s, \theta(s)\}$. Thus, we obtain another θ-decomposition of w, i.e., $w = v_1 \ldots v_{kn}$, where all $v_i \in \{s, \theta(s)\}$ and $|s| < |t|$. But this contradicts the fact that t is the θ-primitive root of w. □

We also obtain the following result as an immediate consequence.

Corollary 1. *The θ-primitive root of a word is primitive.*

Contrary to the case of primitive words, a conjugate of a θ-primitive word need not be θ-primitive, as shown by the following two examples.

Example 1. Let $\theta : \{A, T, C, G\}^* \to \{A, T, C, G\}^*$ be the Watson-Crick antimorphic involution defined by $\theta(A) = T$, $\theta(T) = A$, $\theta(G) = C$, and $\theta(C) = G$. Then, the word $w = GCTA$ is θ-primitive, while its conjugate $w' = AGCT = AG\theta(AG)$ is not.

Example 2. Let $\theta : \{a, b, c, d\}^* \to \{a, b, c, d\}^*$ be a morphic involution defined by $\theta(a) = c$, $\theta(c) = a$, $\theta(b) = d$, and $\theta(d) = b$. Then, the word $w = abadcb$ is θ-primitive, while its conjugate $w' = babadc = (ba)^2\theta(ba)$ is not.

So, we can formulate the following result.

Proposition 5. *The class of θ-primitive words is not necessarily closed under circular permutations.*

Fine and Wilf's result on words, i.e., Theorem 4, constitutes one of the fundamental periodicity properties of words. Thus, a natural question is whether we can obtain an extension of this result when, instead of taking powers of two words u^n and v^m, we look at expressions over $\{u, \theta(u)\}$ and $\{v, \theta(v)\}$, respectively. In particular, since the mapping θ is an involution, we can suppose without loss of generality that the two expressions start with u and v, respectively. First, we analyze the case when θ is a morphic involution; it turns out that in this case we can obtain the same bound as in Theorem 4. However, since the proof of this result is analogous to the one for Theorem 4, see for instance [8], we will not include it here due to space limitations.

Theorem 5. *Let $\theta : \Sigma^* \to \Sigma^*$ be a morphic involution, $u, v \in \Sigma^+$ with $n = |u|$, $m = |v|$, and $d = gcd(n, m)$, $\alpha(u, \theta(u)) \in u\{u, \theta(u)\}^*$, and $\beta(v, \theta(v)) \in v\{v, \theta(v)\}^*$. If the two expressions $\alpha(u, \theta(u))$ and $\beta(v, \theta(v))$ have a common prefix of length at least $n + m - d$, then there exists a word $t \in \Sigma^+$ such that $u, v \in t\{t, \theta(t)\}^*$, i.e., $\rho_\theta(u) = \rho_\theta(v)$. Moreover, the bound $n + m - d$ is optimal.*

However, as illustrated by the following example, if the mapping θ is an antimorphic involution, then the bound given by Theorem 5 is not enough anymore.

Example 3. Let $\theta : \{a, b\}^* \to \{a, b\}^*$ be the mirror mapping defined as follows: $\theta(a) = a, \theta(b) = b$, and $\theta(w_1 \dots w_n) = w_n \dots w_1$, where $w_i \in \{a, b\}$ for all $1 \leq i \leq n$. Obviously, θ is an antimorphic involution on $\{a, b\}^*$. Let now $u = (ab)^k b$ and $v = ab$. Then, u^2 and $v^k\theta(v)^{k+1}$ have a common prefix of length $2|u| - 1 > |u| + |v| - gcd(|u|, |v|)$. However, $\rho_\theta(u) \neq \rho_\theta(v)$.

The next result gives a lower bound for the antimorphic case, for which we employ similar techniques as in [4], so we omit the proof here. As usual, $lcm(n, m)$ denotes the *least common multiple of n and m.*

Theorem 6. *Let $\theta : \Sigma^* \to \Sigma^*$ be an antimorphic involution, $u, v \in \Sigma^+$, and $\alpha(u, \theta(u)) \in u\{u, \theta(u)\}^*$, $\beta(v, \theta(v)) \in v\{v, \theta(v)\}^*$ be two expressions sharing a common prefix of length at least $lcm(|u|, |v|)$. Then, there exists a word $t \in \Sigma^+$ such that $u, v \in t\{t, \theta(t)\}^*$, i.e., $\rho_\theta(u) = \rho_\theta(v)$. In particular, if $\alpha(u, \theta(u)) = \beta(v, \theta(v))$, then $\rho_\theta(u) = \rho_\theta(v)$.*

Note that, in many cases there is a big gap between the bounds given in Theorems 5 and 6. Moreover, Theorem 6 does not give the optimal bound for the general case when θ is an antimorphic involution. In Section 6, we show that this optimal bound for the general case is $2|u|+|v|-gcd(|u|, |v|)$, where $|u| > |v|$, while for some particular cases we obtain bounds as low as $|u|+|v|-gcd(|u|, |v|)$. As an immediate consequence of Theorems 5 and 6, we obtain the following result.

Corollary 2. *For any word $w \in \Sigma^+$ there exists a unique θ-primitive word $t \in \Sigma^+$ such that $w \in t\{t, \theta(t)\}^*$, i.e., $\rho_\theta(w) = t$.*

Let us note now that, maybe even more importantly, just as in the case of primitive words, this result provides us with an alternative, equivalent way for defining the θ-primitive root of a word w, i.e., *the θ-primitive word t such that $w \in t\{t, \theta(t)\}^*$*. This proves to be a very useful tool in our future considerations.

Moreover, we also obtain the following two results as immediate consequences of Theorems 5 and 6.

Corollary 3. *Let $u, v \in \Sigma^+$ be two words such that $\rho(u) = \rho(v) = t$. Then, $\rho_\theta(u) = \rho_\theta(v) = \rho_\theta(t)$.*

Corollary 4. *If we have two words $u, v \in \Sigma^+$ such that $u \in v\{v, \theta(v)\}^*$, then $\rho_\theta(u) = \rho_\theta(v)$.*

4 Relations Imposing θ-Periodicity

It is well-known, due to Theorem 2, that any non-trivial equation over two distinct words forces them to be powers of a common word, i.e., to share a common period. Thus, a natural question is whether this would be the case also when we want two distinct words to have θ-decompositions depending on the same u and $\theta(u)$, i.e., to share a common θ-period. From [6], we already know that the equation $uv = \theta(v)u$ imposes $\rho_\theta(u) = \rho_\theta(v)$ only when θ is a morphic involution. In this section, we give several examples of equations over $\{u, \theta(u), v, \theta(v)\}$ forcing $\rho_\theta(u) = \rho_\theta(v)$ in the case when $\theta : \Sigma^* \to \Sigma^*$ is an antimorphic involution.

The first equation we look at is very similar to the commutation equation of two words, but it involves also the mapping θ.

Theorem 7. *Let $\theta : \Sigma^* \to \Sigma^*$ be an antimorphic involution over the alphabet Σ and $u, v \in \Sigma^+$. If $uv\theta(v) = v\theta(v)u$, then $\rho_\theta(u) = \rho_\theta(v)$.*

Proof. Since $uv\theta(v) = v\theta(v)u$, we already know, due to Theorem 2, that there exists a primitive word $t \in \Sigma^+$ such that $u = t^i$ and $v\theta(v) = t^j$, for some $i, j \geq 0$. If $j = 2k$ for some $k \geq 0$, then we obtain immediately that $v = \theta(v) = t^k$, i.e., $\rho(u) = \rho(v) = t$. Thus, $\rho_\theta(u) = \rho_\theta(t) = \rho_\theta(v)$. Otherwise, i.e., $j = 2k+1$, we can write $v = t^k t_1$ and $\theta(v) = t_2 t^k$, where $t = t_1 t_2$ and $|t_1| = |t_2| > 0$. Hence, $\theta(v) = \theta(t_1)\theta(t)^k = t_2 t^k$, which implies $t_2 = \theta(t_1)$. In conclusion, $u, v \in t_1\{t_1, \theta(t_1)\}^*$, for some word $t_1 \in \Sigma^+$, i.e., $\rho_\theta(u) = \rho_\theta(t_1) = \rho_\theta(v)$. □

Next, we modify a little bit the previous equation, such that on one side, instead of $v\theta(v)$, we take its conjugate $\theta(v)v$.

Theorem 8. *Let $\theta : \Sigma^* \to \Sigma^*$ be an antimorphic involution over the alphabet Σ and $u, v \in \Sigma^+$. If $v\theta(v)u = u\theta(v)v$, then $\rho_\theta(u) = \rho_\theta(v)$.*

Proof. If we concatenate the word $\theta(v)$ to the right on both sides of the equation $v\theta(v)u = u\theta(v)v$, then we obtain $(v\theta(v))(u\theta(v)) = (u\theta(v))(v\theta(v))$. Due to Theorem 2, this means that there exists a primitive word $t \in \Sigma^+$ such that $v\theta(v) = t^i$ and $u\theta(v) = t^j$, for some $i, j \geq 0$, $j \geq \lceil i/2 \rceil$. If $i = 2k$ for some $k \geq 0$, then $\theta(v) = v = t^k$ and thus also $u = t^{j-k}$, i.e., $\rho(u) = \rho(v) = t$. Henceforth, $\rho_\theta(u) = \rho_\theta(t) = \rho_\theta(v)$. Otherwise, i.e., $j = 2k + 1$, we can write $v = t^k t_1$ and $\theta(v) = t_2 t^k$, where $t = t_1 t_2$ and $|t_1| = |t_2| > 0$. Hence, we achieve again $t_2 = \theta(t_1)$, which implies that $v \in t_1\{t_1, \theta(t_1)\}^*$. Moreover, since $u\theta(v) = t^j$, we also obtain $u = t^{j-k-1}t_1 \in t_1\{t_1, \theta(t_1)\}^*$. Thus, $\rho_\theta(u) = \rho_\theta(t_1) = \rho_\theta(v)$. □

Next, we look at the case when both uv and vu are θ-palindrome words, which also proves to be enough to impose that $u, v \in \{t, \theta(t)\}^*$ for some $t \in \Sigma^+$.

Theorem 9. *Let $u, v \in \Sigma^*$ be two words such that both uv and vu are θ-palindrome words and let $t = \rho(uv)$. Then, $t = \theta(t)$ and either $\rho(u) = \rho(v) = t$ or $u = (t_1\theta(t_1))^i t_1$ and $v = \theta(t_1)(t_1\theta(t_1))^j$, where $t = t_1\theta(t_1)$ and $i, j \geq 0$.*

Proof. The equality $uv = \theta(uv)$ immediately implies that $t = \theta(t)$. Moreover, if u and v commute, then $\rho(u) = \rho(v) = \rho(uv) = t$. Assume now that u and v do not commute. Since $\rho(u) \neq \rho(v)$ and $uv = t^n$ for some $n \geq 1$, we can write $u = t^i t_1$ and $v = t_2 t^{n-i-1}$ for some $i \geq 0$ and $t_1, t_2 \in \Sigma^+$ such that $t = t_1 t_2$. Thus, $vu = t_2 t^{n-1} t_1 = (t_2 t_1)^n$ and since $vu = \theta(vu)$ we obtain that also $t_2 t_1$ is θ-palindrome, i.e., $t_2 t_1 = \theta(t_2 t_1) = \theta(t_1)\theta(t_2)$. Now, if $|t_1| = |t_2|$, then $t_2 = \theta(t_1)$ and thus $t = t_1\theta(t_1)$, $u = t^i t_1$, and $v = \theta(t_1)t^{n-i-1}$. Otherwise, either $|t_1| > |t_2|$ or $|t_1| < |t_2|$. We consider next only the case $|t_1| > |t_2|$, the other one being similar. Since $t_2 t_1 = \theta(t_1)\theta(t_2)$, we can write $\theta(t_1) = t_2 x$ and $t_1 = x\theta(t_2)$ for some word $x \in \Sigma^+$ with $x = \theta(x)$. Then, since $t = \theta(t)$ we have that $t = t_1 t_2 = x\theta(t_2)t_2 = \theta(x\theta(t_2)t_2) = \theta(t_2)t_2 x$. Hence, x and $\theta(t_2)t_2$ commute, which contradicts the primitivity of t. □

As an immediate consequence we obtain the following result.

Corollary 5. *For $u, v \in \Sigma^*$, if $uv = \theta(uv)$ and $vu = \theta(vu)$, then $\rho_\theta(u) = \rho_\theta(\theta(v))$. In particular, there exists some $t \in \Sigma^+$ such that $u, v \in \{t, \theta(t)\}^*$.*

5 On θ-Primitive and θ-Palindrome Words

In this section, we investigate two word equations under which a θ-primitive word must be θ-palindrome. Throughout this section we consider $\theta : \Sigma^* \to \Sigma^*$ to be an antimorphic involution over the alphabet Σ.

Theorem 10. *Let $\theta : \Sigma^* \to \Sigma^*$ be an antimorphic involution over the alphabet Σ and $v \in \Sigma^+$ be a θ-primitive word. If $\theta(v)vx = yv\theta(v)$ for some words $x, y \in \Sigma^*$ with $|x|, |y| < |v|$, then v is θ-palindrome and $x = y = \epsilon$.*

Proof. Assume there exist some words $x, y \in \Sigma^*$ with $|x|, |y| < |v|$, such that $\theta(v)vx = yv\theta(v)$, as illustrated in Fig. 2.

Then, we can write $v = v_1v_2 = v_2v_3$, with $v_1, v_2, v_3 \in \Sigma^*$, $y = \theta(v_2) = x$, $v_1 = \theta(v_1)$, $v_3 = \theta(v_3)$. Since $v_1v_2 = v_2v_3$, we can write $v_1 = pq$, $v_3 = qp$, $v_2 = (pq)^ip$, and $v = (pq)^{i+1}p$ for some words $p, q \in \Sigma^*$ and some $i \geq 0$. Thus, $pq = \theta(pq)$ and $qp = \theta(qp)$, which, due to Theorem 9, leads to one of the following two cases. First, if $p = t^kt_1$ and $q = \theta(t_1)t^j$, where $k, j \geq 0$ and $t = t_1\theta(t_1)$ is the primitive root of pq, then we obtain that $v = t^{(k+j+1)(i+1)+k}t_1$ with $(k+j+1)(i+1)+k \geq 1$, which contradicts the θ-primitivity of v. Second,

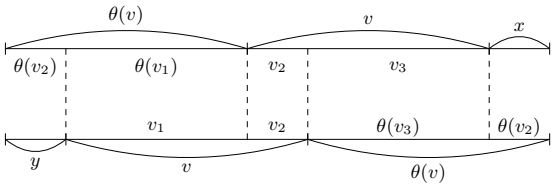

Fig. 2. The equation $\theta(v)vx = yv\theta(v)$

if $\rho(p) = \rho(q) = t$, then also $v \in \{t\}^*$ where $t = \theta(t)$. Thus, $v = \theta(v)$, and the initial identity becomes $v^2x = yv^2$. But, since v is θ-primitive and thus also primitive, we immediately obtain, due to Proposition 1, that $x = y = \epsilon$. □

In other words, the previous result states that if v is a θ-primitive word, then $\theta(v)v$ cannot overlap with $v\theta(v)$ in a nontrivial way. However, the following example shows that this is not the case anymore if we look at the overlaps between $\theta(v)v$ and v^2, or between $v\theta(v)$ and v^2, respectively, even if we consider the larger class of primitive words.

Example 4. Let $\theta : \Sigma^* \to \Sigma^*$ be an antimorphic involution over the alphabet Σ, $p, q \in \Sigma^+$ such that $\rho(p) \neq \rho(q)$, $p = \theta(p)$, and $q = \theta(q)$, and let $v = p^2q^2p$ and $u = pq^2p^2$. It is easy to see that u and v are primitive words. In addition, if we take $\Sigma = \{a, b\}$, the mapping θ to be the mirror image, $p = a$, and $q = b$, then u and v are actually θ-primitive words. Since $\theta(v) = pq^2p^2$ and $\theta(u) = p^2q^2p$, we can write $xv^2 = v\theta(v)y$ and $y\theta(u)u = u^2z$ where $x = p^2q^2$, $y = pq^2p$, and $z = q^2p^2$. Thus, for primitive (resp. θ-primitive) words u and v, $v\theta(v)$ can overlap with v^2 and $\theta(u)u$ with u^2 in a nontrivial way.

Maybe even more surprisingly, the situation changes again if we try to fit v^2 inside $v\theta(v)v$, as shown by the following result.

Theorem 11. *Let $\theta : \Sigma^* \to \Sigma^*$ be an antimorphic involution over the alphabet Σ and $v \in \Sigma^+$ be a primitive word. If $v\theta(v)v = xv^2y$ for some words $x, y \in \Sigma^*$, then v is θ-palindrome and either $x = \epsilon$ and $y = v$ or $x = v$ and $y = \epsilon$.*

Proof. Suppose that $v\theta(v)v = xv^2y$ for some words $x, y \in \Sigma^*$, as illustrated in Fig. 3. If we look at this identity from left to right, then we can write $v = xv_1 = v_1v_2$, with $v_1, v_2 \in \Sigma^*$ such that $|x| = |v_2|$ and $\theta(v) = \theta(v_2)\theta(v_1)$. Now, if we look at the right sides of this identity, then we immediately obtain that $x = v_2$ and $v_1 = y$. Thus, $v = xy = yx$, implying that $x, y \in \{t\}^*$, for some primitive word t. However, since v is primitive, this means that either $x = \epsilon$ and $y = v$

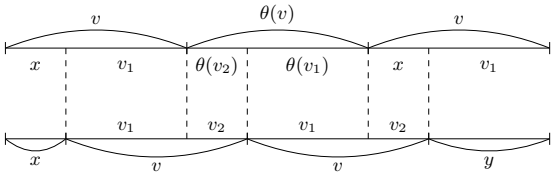

Fig. 3. The equation $v\theta(v)v = xv^2y$

or $x = v$ and $y = \epsilon$. Moreover, in both cases we also obtain $v = \theta(v)$. □

6 An Optimal Bound for the Antimorphic Extension of the Fine and Wilf Theorem

Throughout this section we take $\theta : \Sigma^* \to \Sigma^*$ to be an antimorphic involution, $u, v \in \Sigma^+$ with $|u| > |v|$, $\alpha(u, \theta(u)) \in \{u, \theta(u)\}^+$, and $\beta(v, \theta(v)) \in \{v, \theta(v)\}^+$. Since θ is an involution, we can suppose, without loss of generality, that $\alpha(u, \theta(u))$ and $\beta(v, \theta(v))$ start with u and v, respectively. We start our analysis with the case when v is θ-palindrome.

Theorem 12. *Let u and v be two words with $|u| > |v|$ and $v = \theta(v)$. If there exist two expressions $\alpha(u, \theta(u)) \in u\{u, \theta(u)\}^*$ and $\beta(v, \theta(v)) \in v\{v, \theta(v)\}^*$ having a common prefix of length at least $|u| + |v| - gcd(|u|, |v|)$, then $\rho_\theta(u) = \rho_\theta(v)$.*

Proof. First, we can suppose, without loss of generality that $gcd(|u|, |v|) = 1$. Otherwise, i.e., $gcd(|u|, |v|) = d \geq 2$, we consider a new alphabet $\Sigma' = \Sigma^d$, where the new letters are words of length d in the original alphabet, and we look at the words u and v as elements of $(\Sigma')^+$. In the larger alphabet $gcd(|u|, |v|) = 1$, and if we can prove the theorem there it immediately gives the general proof.

Since $v = \theta(v)$, $\beta(v, \theta(v)) = v^n$ for some $n \geq 2$. Moreover, if $v \in \Sigma$, then trivially $u \in v\{v, \theta(v)\}^*$, i.e., $\rho_\theta(u) = \rho_\theta(v)$. So, suppose next that $|v| \geq 2$ and, since $gcd(|u|, |v|) = 1$, $u = v^i v_1$, where $i \geq 1$ and $v = v_1v_2$ with $v_1, v_2 \in \Sigma^+$.

If $\alpha(u, \theta(u)) = u^2\alpha'(u, \theta(u))$, then u^2 and v^n have a common prefix of length at least $|u| + |v| - gcd(|u|, |v|)$, which, due to Theorem 4, implies that $\rho(u) = \rho(v) = t$, for some primitive word $t \in \Sigma^+$, and thus $\rho_\theta(u) = \rho_\theta(t) = \rho_\theta(v)$.

Otherwise, $\alpha(u, \theta(u)) = u\theta(u)\alpha'(u, \theta(u))$ for some $\alpha'(u, \theta(u)) \in \{u, \theta(u)\}^*$. Now, we have two cases depending on $|v_1|$ and $|v_2|$. We present here only the

case when $|v_1| \leq |v_2|$, see Fig. 4, the other one being symmetric. Now, since θ is an antimorphism, $\theta(suf_{|v|-1}(u)) = pref_{|v|-1}(\theta(u))$. So, we can write $v_2 = \theta(v_1)z$ for some $z \in \Sigma^*$, since $|v_1| \leq |v_2| \leq |v| - 1 = |v| - gcd(|u|, |v|)$. Now, to the

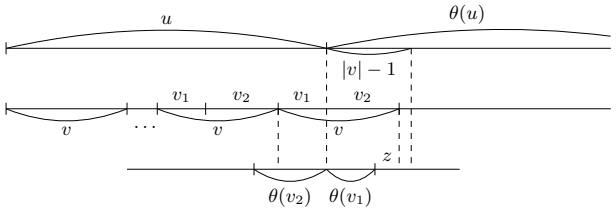

Fig. 4. The common prefix of $u\theta(u)$ and v^n of length $|u| + |v| - 1$

left of the border-crossing v there is at least one occurrence of another v, so we immediately obtain $z = \theta(z)$, as $v_2 = \theta(v_1)z$ and $\theta(v_2) = \theta(z)v_1$. Then, $v = v_1\theta(v_1)z = zv_1\theta(v_1) = \theta(v)$ implying, due to Theorem 7, $\rho_\theta(v_1) = \rho_\theta(z)$. So, since $v = v_1\theta(v_1)z$ and $u = v^i v_1 = (v_1\theta(v_1)z)^i v_1$, we obtain $\rho_\theta(u) = \rho_\theta(v)$. □

Let us look next at the case when u is θ-palindrome.

Theorem 13. *Let u and v be two words with $|u| > |v|$ and $u = \theta(u)$. If there exist two expressions $\alpha(u, \theta(u)) \in u\{u, \theta(u)\}^*$ and $\beta(v, \theta(v)) \in v\{v, \theta(v)\}^*$ having a common prefix of length at least $|u| + |v| - gcd(|u|, |v|)$, then $\rho_\theta(u) = \rho_\theta(v)$.*

Proof. As before, we can suppose without loss of generality that $gcd(|u|, |v|) = 1$. Also, since $u = \theta(u)$, we actually have $\alpha(u, \theta(u)) = u^n$ for some $n \geq 2$. Moreover, since u starts with v and $u = \theta(u)$, we also know that u ends with $\theta(v)$. Now, if $v \in \Sigma$, then trivially $u \in v\{v, \theta(v)\}^*$, i.e., $\rho_\theta(u) = \rho_\theta(v)$. So, we can suppose next that $|v| \geq 2$ and thus, since $gcd(|u|, |v|) = 1$, we have $u = \beta'(v, \theta(v))v'$, where $\beta'(v, \theta(v))$ is a prefix of $\beta(v, \theta(v))$ and $v' \in \Sigma^+$, $v' \in Pref(v) \cup Pref(\theta(v))$.

Case 1: We begin our analysis with the case when the border between the first two u's falls inside a v, as illustrated in Fig. 5. Then, we can write $v =$

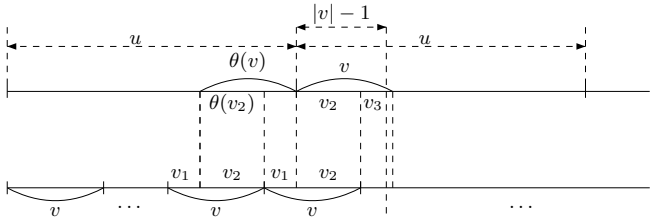

Fig. 5. The common prefix of u^2 and $\beta(v, \theta(v))$ of length $|u| + |v| - 1$

$v_1v_2 = v_2v_3$ where $v_1, v_2, v_3 \in \Sigma^+$, implying that $v_1 = xy$, $v_3 = yx$, and $v_2 = (xy)^j x$ for some $j \geq 0$ and $x, y \in \Sigma^*$. Moreover, since u ends with $\theta(v)$, we also have $v_1 = \theta(v_1)$, i.e., $xy = \theta(y)\theta(x)$. If $x = \epsilon$, then $v_1, v_2, v_3, v \in \{y\}^*$, which implies that also $u \in y\{y, \theta(y)\}^*$, i.e., $\rho_\theta(u) = \rho_\theta(v) = \rho_\theta(y)$; moreover, since $gcd(|u|, |v|) = 1$ we actually must have $y \in \Sigma$. Similarly, we also obtain $\rho_\theta(u) = \rho_\theta(v)$ when $y = \epsilon$. So, from now on we can suppose that $x, y \in \Sigma^+$.

Let us consider next the case when, before the border-crossing v we have an occurrence of another v, as illustrated in Fig. 5. Then, we have that $v_2 = \theta(v_2)$, i.e., $(xy)^j x = (\theta(x)\theta(y))^j\theta(x)$. If $j \geq 1$, then this means that $x = \theta(x)$ and $y = \theta(y)$. Then, the equality $xy = \theta(y)\theta(x)$ becomes $xy = yx$. So, there exists a word $t \in \Sigma^+$ such that $x, y \in \{t\}^*$, and thus also $v \in \{t\}^+$ and $u \in t\{t, \theta(t)\}^*$, i.e., $\rho_\theta(u) = \rho_\theta(v)$. Otherwise, $j = 0$ and we have $x = \theta(x)$. But then, the equality $xy = \theta(y)\theta(x)$ becomes $xy = \theta(y)x$, implying that $x = p(qp)^n$ and $y = (qp)^m$ for some $m \geq 1$, $n \geq 0$, and some words p and q with $p = \theta(p)$ and $q = \theta(q)$, see [6]. Since u^2 and $\beta(v, \theta(v))$ share a common prefix of length at least $|u|+|v|-gcd(|u|, |v|) = |u|+|v|-1$, v_3 and some $\beta'(v, \theta(v))$ share a prefix of length $|v_3| - 1$. Furthermore, as $v_3 = yx = (qp)^m p(qp)^n$, $v = v_1v_2 = p(qp)^{m+n}p(qp)^n$, and $\theta(v) = (pq)^n p(pq)^{m+n}p$, this means that independently of what follows to the right the border-crossing v, either v or $\theta(v)$, we have two expressions over p and q sharing a common prefix of length at least $|p| + |q|$. Thus, from [2], we can conclude that $p, q \in \{t\}^*$ for some $t \in \Sigma^+$, which implies that also $x, y, v \in \{t\}^+$ and $u \in t\{t, \theta(t)\}^*$, i.e., $\rho_\theta(u) = \rho_\theta(v)$.

Now, suppose that before the border-crossing v we have an occurrence of $\theta(v)$. If $|u| < 2|v| + |v_1|$, then, since $\beta(v, \theta(v))$ starts with v, we must have $v = \theta(v)$, in which case we can use Theorem 12 to conclude that $\rho_\theta(u) = \rho_\theta(v)$. Otherwise, $|u| \geq 2|v| + |v_1|$ and since $u = \theta(u)$, u ends either with $v\theta(v)$ or with $\theta(v)\theta(v)$. In the first case, we obtain $v_3 = \theta(v_3)$, i.e., $yx = \theta(yx)$, which together with $xy = \theta(xy)$ imply, due to Corollary 5, that $x, y \in \{t, \theta(t)\}^*$, for some $t \in \Sigma^+$ and thus, $\rho_\theta(u) = \rho_\theta(v)$. In the second case, we obtain $v_1 = v_3$, i.e., $xy = yx$. So, $x, y \in \{t\}^*$, and thus also $v \in \{t\}^+$ and $u \in t\{t, \theta(t)\}^*$, i.e., $\rho_\theta(u) = \rho_\theta(v)$.

Case 2: The case when the border between the first two u's falls inside $\theta(v)$ is similar to the one above. So, due to page limitations, we omit it here. □

Although the previous two results give a very short bound, i.e., $|u| + |v| - gcd(|u|, |v|)$, this is not enough in the general case, as illustrated also in Example 3. However, we can prove that, independently of how the expression $\alpha(u, \theta(u))$ starts, $2|u| + |v| - gcd(|u|, |v|)$ is enough to impose θ-periodicity of u and v. The first case we consider is when $\alpha(u, \theta(u))$ starts with u^2. The proofs of Theorems 14 and 16 are rather complex and necessitate the analysis of many cases. Their inclusion would double the length of this paper and we therefore omit them here.

Theorem 14. *Given two distinct words $u, v \in \Sigma^+$ with $|u| > |v|$, if there exist two expressions $\alpha(u, \theta(u)) \in u\{u, \theta(u)\}^*$ and $\beta(v, \theta(v)) \in v\{v, \theta(v)\}^*$ having a common prefix of length at least $2|u| + |v| - gcd(|u|, |v|)$ and, moreover, $\alpha(u, \theta(u)) = u^2\alpha'(u, \theta(u))$ for some $\alpha'(u, \theta(u)) \in \{u, \theta(u)\}^+$, then $\rho_\theta(u) = \rho_\theta(v)$.*

The next result considers the case when $\alpha(u, \theta(u))$ starts with $u\theta(u)u$, which is an immediate consequence of Theorem 13.

Theorem 15. *Given two distinct words $u, v \in \Sigma^+$ with $|u| > |v|$, if there exist two expressions $\alpha(u, \theta(u)) \in u\{u, \theta(u)\}^*$ and $\beta(v, \theta(v)) \in v\{v, \theta(v)\}^*$ having a common prefix of length at least $2|u| + |v| - gcd(|u|, |v|)$ and, moreover, $\alpha(u, \theta(u)) = u\theta(u)u\alpha'(u, \theta(u))$ with $\alpha'(u, \theta(u)) \in \{u, \theta(u)\}^*$, then $\rho_\theta(u) = \rho_\theta(v)$.*

The only case left is when $\alpha(u, \theta(u))$ starts with $u\theta(u)\theta(u)$.

Theorem 16. *Let $u, v \in \Sigma^+$ be two words with $|u| > |v|$. If there exist two expressions $\alpha(u, \theta(u)) \in u\{u, \theta(u)\}^*$ and $\beta(v, \theta(v)) \in v\{v, \theta(v)\}^*$ having a common prefix of length at least $2|u| + |v| - gcd(|u|, |v|)$, and, moreover, $\alpha(u, \theta(u)) = u\theta(u)\theta(u)\alpha'(u, \theta(u))$ for some $\alpha'(u, \theta(u)) \in \{u, \theta(u)\}^*$, then $\rho_\theta(u) = \rho_\theta(v)$.*

To conclude, in this section we proved that if θ is an antimorphic involution, then we only need two expressions $\alpha(u, \theta(u))$ and $\beta(v, \theta(v))$ to share a common prefix of length $2|u| + |v| - gcd(|u|, |v|)$, where $|u| > |v|$, in order to impose $\rho_\theta(u) = \rho_\theta(v)$. Moreover, the following examples show that this bound is optimal.

Example 5. Let $\theta : \{a, b\}^* \to \{a, b\}^*$ be the mirror involution, $u_1 = a^2ba^3b$, $v_1 = a^2ba$, with $gcd(|u_1|, |v_1|) = 1$, and $u_2 = ba^2baba$, $v_2 = ba^2ba$, with $gcd(|u_2|, |v_2|) = 1$. Then, u_1^3 and $v_1^2\theta(v_1)^2v_1$ have a common prefix of length $2|u_1| + |v_1| - 2$, but $\rho_\theta(u_1) \neq \rho_\theta(v_1)$. Also, $u_2\theta(u_2)^2$ and v_2^4 have a common prefix of length $2|u_2| + |v_2| - 2$, but $\rho_\theta(u_2) \neq \rho_\theta(v_2)$.

Acknowledgments. This research was supported by Natural Sciences and Engineering Research Council of Canada Discovery Grant and Canada Research Chair Award of Lila Kari.

References

1. Berstel, J., Boasson, L., *Partial words and a theorem of Fine and Wilf*, WORDS 1997 (Rouen), Theoret. Comput. Sci. 218, no. 1, 135-141, (1999).
2. Choffrut, C., Karhumäki, J., Combinatorics of words. In: G. Rozenberg, A. Salomaa (eds), *Handbook of Formal Languages*, Vol. 1, Springer-Verlag, Berlin, 329–438, (1997).
3. Constantinescu, S., Ilie, L., *Generalized Fine and Wilfs theorem for arbitrary number of periods*, Theoret. Comput. Sci. 339(1), 49–60, (2005).
4. Constantinescu, S., Ilie, L., *Fine and Wilf's theorem for abelian periods*, Bulletin of EATCS 89, 167–170, (2006).
5. Fine, N.J., Wilf, H.S.,, *Uniqueness theorem for periodic functions*, Proc. Amer. Math. Soc., **16**, 109–114, (1965).
6. Kari, L., Mahalingam, K., *Watson-Crick Conjugate and Commutative Words*, Preproceedings of the International Conference on DNA 13, (Max Garzon and Hao Yan editors), 75–87, (2007).
7. Kari, L., Mahalingam, K., *Watson-Crick Palindromes*, Submitted.
8. Lothaire, M., *Combinatorics on words*, Encyclopedia of Mathematics and its applications **17**, Addison-Wesley Publishing Co., (1983).
9. de Luca, A., De Luca, A., *Pseudopalindrome closure operators in free monoids*, Theoret. Comput. Sci. 362, 282–300, (2006).
10. Mignosi, F., Restivo, A., Silva, P.S., *On Fine and Wilfs theorem for bidimensional words*, Selected papers in honor of Jean Berstel, Theoret. Comput. Sci. 292, no. 1, 245-262, (2003).
11. Shyr, H.J., Thierrin, G., *Disjunctive languages and codes*, Proc. FCT77, LNCS 56, Springer-Verlag, Berlin, Heidelberg, New York, 171–176, (1977).
12. Yu, S.S., *Languages and Codes*, Lecture Notes, Department of Computer Science, National Chung-Hsing University, Taichung, Taiwan 402, (2005).