Periodicity properties on partial words

By: F. Blanchet-Sadri, Kevin Corcoran, Jenell Nyberg

F. Blanchet-Sadri, K. Corcoran, and J. Nyberg, "Periodicity Properties on Partial Words." *Information and Computation*, Vol. 206, 2008, pp 1057-1064. doi:10.1016/j.ic.2008.03.007

Made available courtesy of Elsevier: http://dx.doi.org/10.1016/j.ic.2008.03.007

Abstract:
The concept of periodicity has played over the years a centra1 role in the development of combinatorics on words and has been a highly valuable too1 for the design and analysis of algorithms. Fine and Wilf's famous periodicity result, which is one of the most used and known results on words, has extensions to partia1 words, or sequences that may have a number of "do not know" symbols. These extensions fal1 into two categories: the ones that relate to strong periodicity and the ones that relate to weak periodicity. In this paper, we obtain consequences by generalizing, in particular, the combinatoria1 property that "for any word u over $\{a, b\}$, ua or ub is primitive," which proves in some sense that there exist very many primitive partia1 words.

Keywords: Formal languages, Combinatorics on words, Fine and Wilf's periodicity result, Partial words, Primitive partial words, Periods, Weak periods

1. Introduction
The problem of computing patterns in *words*, or finite sequences of symbols from a finite alphabet, has important applications in data compression, string searching and pattern matching algorithms. The notion of *period* of a word is central in combinatorics on words. There are many fundamental results on periods of words. Among them is the well known periodicity result of Fine and Wilf [12] which intuitively determines how far two periodic events have to match in order to guarantee a common period. More precisely, any word u having periods p,q and length at least $p + q - \gcd(p,q)$ has also period $\gcd(p,q)$. Extensions to more than two periods are given in [10,12,17,22]. Other generalizations of Fine and Wilf's theorem have been made, including a generalization for abelian periods [14].

Partial words, or finite sequences that may contain a number of "do not know" symbols or *holes*, appear in natural ways in several areas such as molecular biology, data communication, DNA computing, etc. In this case there are two notions of periodicity: one is that of *(strong) period* and the other is that of *weak period* (see Section 2.2). The original Fine and Wilf's result has been generalized to partial words in two ways:

First, any partial word u with h holes and having weak periods p,q and length at least $l_{(h,p,q)}$ has also period $\gcd(p,q)$ provided u is not (h,p,q)-special. This extension was done for one hole by Berstel and Boasson in their seminal paper [1] where the class of $(1,p,q)$-special partial words is empty; for two-three holes by Blanchet-Sadri and Hegstrom [8]; and for an arbitrary number of holes by Blanchet-Sadri [2]. Closed formulas for the bounds $l_{(h,p,q)}$ were given and shown to be optimal. Extensions to more than two weak periods are given in [9].

Second, any partial word u with h holes and having (strong) periods p,q and length at least $L_{(h,p,q)}$ has also period $\gcd(p,q)$. The study of the bounds $L_{(h,p,q)}$ was initiated by Shur and Gamzova [19]. In particular, they gave a closed formula for $L_{(h,p,q)}$ in the case where $h = 2$ (the cases where $h = 0$ or $h = 1$ are implied by the above mentioned results). In [5], Blanchet-Sadri et al. give closed formulas for the optimal bounds $L_{(h,p,q)}$ in the case where $p = 2$ and also in the case where q is large. In addition, they give upper bounds when q is small and $h =$ 3,4,5,6 or 7. Their proofs are based on connectivity in graphs associated with partial words.

In this paper, we obtain consequences of the generalizations of Fine and Wilf's periodicity result to partial words. In particular, we generalize the following combinatorial property: "For any word u over $\{a,b\}$, ua or ub is primitive." This property proves in some sense that there exist very many primitive words. The study of primitive partial words was initiated in [3]. Testing primitivity of partial words can be done in linear time in the length of the word [4].

2. Preliminaries

This section is devoted to reviewing basic concepts on words and partial words.

2.1. Words

Let A be a nonempty finite set of symbols called an *alphabet*. Symbols in A are called *letters* and any finite sequence over A is called a *word* over A. The *empty word*, that is the word containing no letter, is denoted by ε ε. The set of all words over A is denoted by A^*.

For any word u over A, $|u|$ denotes the length of u. In particular, $|\varepsilon| = 0$. The set of symbols occurring in a word u is denoted by $\alpha(u)$. A word of length n over A can be defined by a total function $u : \{0,\ldots,n-1\} \to A$ and is usually represented as $u = a_0a_1,\ldots,a_{n-1}$ with $a_i \in A$. A word v is a *factor of* u if there exist words x,y such that $u = xvy$. The word v is a *prefix of* (respectively, *suffix of*) u if $x = e$ (respectively, $y = e$). If $u = a_0,\ldots,a_{n-1}$ with $a_i \in A$, then *a period of* u is a positive integer p such that $a_i = a_{i+p}$ for $0 \leq i < n - p$.

A nonempty word u is *primitive* if there exists no word v such that $u = v^n$ with $n \geq 2$. Note the fact that the empty word is not primitive. If u is a nonempty word, then there exist a unique primitive word v and a unique positive integer n such that $u = v^n$.

2.2. Partial words

A *partial word* u of length n over A is a partial function $u : \{0, \ldots ,n-1\} \to A$. For $0 \leq i < n$, if $u(i)$ is defined, then we say that i belongs to the *domain* of u (denoted by $i \in D(u)$), otherwise we say that i belongs to the *set of holes* of u (denoted by $i \in H(u)$). A word over A is a partial word over A with an empty set of holes (we sometimes refer to words as *full* words).

If u is a partial word of length n over A, then the *companion* of u (denoted by $u_\Diamond$) is the total function $u_\Diamond : \{0,\ldots,n-1\} \to A \cup \{\Diamond\}$ defined by

$$u_\Diamond(i) = \begin{cases} u(i) \; if \; i \in D(u) \\ \Diamond \text{ otherwise} \end{cases}$$

The bijectivity of the mapping $u \mapsto u_\Diamond$ allows us to define for partial words concepts such as concatenation, powers and factors in a trivial way. The symbol $\Diamond$ is viewed as a "do not know" symbol. The word $u_\Diamond = abb\Diamond bbcb$ is the companion of the partial word u of length 8 where $D(u) = \{0,1,2,4,5,6,7\}$ and $H(u) = \{3\}$. For convenience, we will refer to a partial word over A as a word over the enlarged alphabet $A \cup \{\Diamond\}$, where the additional symbol $\Diamond$ plays a special role. This allows us to say for example "the partial word $aba\Diamond aa\Diamond$" instead of "the partial word with companion $aba\Diamond aa\Diamond$".

A *(strong) period* of a partial word u over A is a positive integer p such that $u(i) = u(j)$ whenever $i ,j \in D(u)$ and $i \equiv j \bmod p$. In such a case, we call u *p-periodic*. Similarly, a *weak period* of u is a positive integer p such that $u(i) = u(i + p)$ whenever $i,i + p \in D(u)$. In such a case, we call u *weakly p-periodic*. The partial word with companion *abbobbcbb* is weakly 3-periodic but is not 3-periodic. The latter shows a difference between partial words and full words since every weakly p-periodic full word is p-periodic. Another difference worth noting is the fact that even if the length of a partial word u is a multiple of a weak period of u, u is not necessarily a power of a shorter partial word. The minimum period of u is denoted by $p(u)$, and the minimum weak period by $p'(u)$. The set of all periods (respectively, weak periods) of u is denoted by $P(u)$ (respectively, $P'(u)$).

For a partial word u, positive integer p and integer $0 \leq i < p$, define

$$u_{i,p} = u(i)u(i+p)u(i+2p) \ldots u(i+jp)$$

where j is the largest nonnegative integer such that $i + jp < |u|$. Then u is (strongly) p-periodic if and only if $u_{i,p}$ is (strongly) 1-periodic for all $0 \leq i < p$, and u is weakly p-periodic if and only if $u_{i,p}$ is weakly 1-periodic for all $0 \leq i < p$. Strongly 1-periodic partial words as well as the full factors, that is factors that are full words, of weakly 1-periodic partial words are over a singleton alphabet.

If u and v are two partial words of equal length, then u is said to be contained in v, denoted by $u \subset v$, if $D(u) \subset D(v)$ and $u(i) = v(i)$ for all $i \in D(u)$. The order $u \subset v$ on partial words is obtained when we let $\lozenge < a$ and $a \leq a$ for all $a \in A$. A partial word u is *primitive* if there exists no word v such that $u \subset v^n$ with $n \geq 2$. In this definition of primitivity, v is (or can be) assumed to be a "full" word in $u \subset v^n$ (not just a partial word). Note that if v is primitive and $v \subset u$, then u is primitive as well. It was shown in [3] that if u is a nonempty partial word, then there exist a primitive word v and a positive integer n such that $u \subset v^n$. However uniqueness does not hold as seen with the partial word $u = \lozenge a$ (here $u \subset a^2$ and $u \subset ba$ for distinct letters a,b).

Partial words u and v of equal length are *compatible*, denoted by $u \uparrow v$, if there exists a partial word w such that $u \subset w$ and $v \subset w$. In other words, $u(i) = v(i)$ for every $i \in D(u) \cap D(v)$. Note that for full words, the notion of compatibility is simply that of equality.

3. Fine and Wilf's periodicity result and generalizations to partial words

In this section, we discuss in details the two ways Fine and Wilf's periodicity result has been extended to partial words. For easy reference, we recall Fine and Wilf' s result (the bound $p + q - \gcd(p,q)$ was shown to be optimal [11]).

Theorem 1 ([14]).
Let p and q be positive integers satisfying $p < q$. Let u be a full word. If u is p-periodic and q-periodic and $|u| \geq p + q - \gcd(p,q)$, then u is $\gcd(p,q)$-periodic.

First, we review the generalization related to weak periodicity [1,2,8].

We first recall Berstel and Boasson's result for partial words with exactly one hole where the bound $p + q$ is optimal.

Theorem 2 ([1]).
Let p and q be positive integers satisfying $p < q$. Let u be a partial word with one hole. If u is weakly p-periodic and weakly q-periodic and $|u| \geq l_{(1,p,q)} = p + q$, then u is $\gcd(p,q)$-periodic.

When we discuss partial words with $h \geq 2$ holes, we need the extra assumption of u not being (h,p,q)-*special* for a similar result to hold true. Indeed, if p and q are positive integers satisfying $p < q$ and $\gcd(p,q) = 1$, then the infinite sequence $(ab^{p-1}\lozenge b^{q-p-1}\lozenge b^n)_{n>0}$ consists of $(2,p,q)$-special partial words with two holes that are weakly p-periodic and weakly q- periodic but not $\gcd(p,q)$-periodic.

In order to define the concept of (h,p,q)-speciality, note that a partial word u that is weakly p-periodic and weakly q-periodic can be represented as a two-dimensional structure. Consider for example the partial word

$$u = ababa\lozenge\lozenge\lozenge bab\lozenge bb\lozenge bbbbbbbb \qquad (1)$$

where $p = 2$ and $q = 5$. The array looks like:

	$u(0)$	$u(5)$	$u(10)$	$u(15)$	$u(20)$
	$u(2)$	$u(7)$	$u(12)$	$u(17)$	$u(22)$
	$u(4)$	$u(9)$	$u(14)$	$u(19)$	
$u(1)$	$u(6)$	$u(11)$	$u(16)$	$u(21)$	
$u(3)$	$u(8)$	$u(13)$	$u(18)$	$u(23)$	

A World Wide Web server interface has been established at

www.uncg.edu/mat/research/finewilf

for automated use of a program that builds two- (and three-)dimensional representations out of a partial word based on two of its weak periods.

In general, if $\gcd(p,q) = d$, we get d arrays. Each of these arrays is associated with a subgraph $G = (V,E)$ of $G(p,q)$ (u) as follows: V is the subset of $D(u)$ comprising the defined positions of u within the array, and $E = E_1 \cup E_2$ where $E_1 = \{\{i,i-q\} \mid i,i-q \in V\}$ and $E_2 = \{\{i,i-p\} \mid i,i-p \in V\}$. For $0 \leq j < \gcd(p,q)$, the subgraph of $G(p,q)$ (u) corresponding to

$$D(u) \cap \{i \mid i \geq \geq 0 \text{and} i \equiv j \bmod \gcd(p,q)\}$$

will be denoted by $G^j_{(p,q)}(u)$. Whenever $\gcd(p,q) = 1$, $G^0_{(p,q)}(u)$ is just $G_{(p,q)}(u)$. Referring to the partial word u in (1) above, the graph $G_{(2,5)}$ (u) is disconnected (u is (5,2,5)-special).

We now define the concept of speciality.

Definition 1 *([2])***.**
Let p and q be positive integers satisfying $p < q$, and let h be a nonnegative integer. Let

$$l_{(h,p,q)} = \begin{cases} \left(\frac{h}{2} + 1\right)(p+q) - \gcd(p,q) & \text{if } h \text{ is even} \\ \left(\left\lfloor\frac{h}{2}\right\rfloor + 1\right)(p+q) & \text{otherwise} \end{cases}$$

Let u be a partial word with h holes of length at least $l_{(h,p,q)}$. Then u is (h,p,q)-special if $G^j_{(p,q)}(u)$ is disconnected for some $0 \leq j < \gcd(p,q)$.

It turns out that the bound $l_{(h,p,q)}$ is optimal for a number of holes h.

Theorem 3 ([2])**.**
Let p and q be positive integers satisfying $p < q$, and let u be a non (h,p,q)-special partial word with h holes. If u is weakly p-periodic and weakly q-periodic and $|u| \geq l_{(h,p,q)}$, then u is $\gcd(p,q)$-periodic.

Now, we review the generalizations related to strong periodicity [5,18,19,20]. Note that there exists an integer L such that if a partial word u with h holes has periods p and q satisfying $p < q$ and $|u| \geq L$, then u has period $\gcd(p,q)$ [20]. Let $L_{(h,p,q)}$ be the smallest such integer L.

The following result is a direct consequence of Berstel and Boasson's result.

Theorem 4 ([1])**.**
The equality $L(1,p,q) = p + q$ holds.

For at least two holes, we have the following results.

Theorem5 ([19,20]).
The equality $L_{(2,p,q)} = 2p + q - \gcd(p,q)$ *holds.*

Theorem 6 ([5,18–20])**.**
The equality $L_{(h,2,q)} = (2\left\lfloor\frac{h}{q}\right\rfloor + 1)q + h \bmod q + 1$ *holds.*

Setting $h = nq + m$ where $0 \le m < q, L_{(h,2,q)} = (2n + 1)q + m + 1$. Now let $W_{h,p,q} = \{w \mid w$ has periods p and q, $\|H(w)\| = h$ and $|w| = L_{(h,p,q)} - 1\}$ and let $V_{h,p,q} = \{v \mid v \in W_{h,p,q}$ and v does not have period gcd(p,q)$\}$. Consider the word $u = \Diamond^m w(\Diamond^q w)^n$ where w is the unique element in $V_{0,2,q}$ of length q. Note that u is an optimal word. Indeed, $|u| = (2n + 1)q + m$, u has h holes, and since w is not 1-periodic, we also have that u is not 1-periodic. It is easy to show that u is 2- and q-periodic.

In [19], the authors proved that if $q > p \geq 3$ and $\gcd(p,q) = 1$ and h is large enough, then

$$\frac{pq}{p+q-2}\,(h+1) \leq L_{(h,p,q)} < \frac{pqh}{p+q-2} + 4(q-1)$$

4. Consequences of Fine and Wilf's generalized periodicity results

In this section, we consider some combinatorial properties of words and extend them to partial words. In particular, Propositions 1 and 2, Theorem 8, Lemma 1 and Theorem 9 are consequences of the generalizations of Fine and Wilf's periodicity result of Section 3.

To motivate this section, an unexpected result of Guibas and Odlyzko [15] states that for every word u over an arbitrary alphabet A, there exists a word v of length $|u|$ over the alphabet $\{0,1\}$ such that the set of all periods of u coincides with the set of all periods of v. The proof of this result is somewhat complicated and uses properties of correlations. In [16], Halava et al. gave a simple constructive proof which computes v in linear time. This result was later proved for partial words with one hole by extending Halava et al.'s approach which is based on properties mentioned in this section. More specifically, Blanchet-Sadri and Chriscoe [6] showed that for every partial word u with one hole over A, a partial word v over $\{0,1\}$ satisfying (1) $|v| = |u|$; (2) $P(v) = P(u)$; (3) $P'(v) = {}_{P}\sim(u)$; and (4) $H(v) \subset H(u)$ can be computed in linear time. More recently, Blanchet-Sadri et al. [7] showed that Conditions (1)–(3) can be satisfied simultaneously for *any* partial word u. However all the four conditions *cannot* be satisfied simultaneously in the case of two holes or more. For the partial word *abaca*◊◊*acaba* can be checked by brute force to have no such binary reduction. Can we characterize the partial words that have such a binary reduction? We believe that some of the results in this section may help answering this open question.

First, we characterize the set of periods and weak periods of partial words. We consider the following combinatorial property of words [16]: "For any word u over an alphabet A, if q is a period of u satisfying $|u| \geq p(u) + q$, then q is a multiple of $p(u)$."

The following proposition gives the structure of the set of weak periods of a partial word u with h holes.

Proposition 1. *Let u be anon* $(h,p'(u),q)$*-special partial word with h holes over an alphabet A. If q is a weak period of u satisfying* $|u| \geq l_{(h,p'(u),q)}$, *then q is a multiple of* $p'(u)$.

Proof. By Theorem 3, $\gcd(p'(u),q)$ is a period of u since $|u| \geq l_{(h,p'(u),q)}$. Since $p(u)$ is the minimum period of u and $p'(u)$ is the minimum weak period of u, we get $p'(u) \leq p(u) \leq \gcd(p'(u),q)$. We conclude that $p'(u) = \gcd(p'(u),q)$ and so $p'(u)$ divides q. ■

Another version of the property and a similar consequence follow in the case of strong periodicity.

Proposition 2. *Let u be a partial word with h holes over an alphabet A. If q is a (strong) period of u satisfying* $|u| \geq L_{(h,p(u),q)}$, *then q is a multiple of p(u).*

Second, we consider the following combinatorial property of words: "For any non empty word u over an alphabet A with minimum period $p(u)$, there exist a positive integer k, a (possibly empty) word v, and a nonempty word w such that $u = (vw)^k v$ and $p(u) = |vw|$."

Proposition 3.
Let u be a nonempty partial word over an alphabet A with minimum weak period p'(u). Then there exist a positive integer k, (possibly empty) partial words v1,v2, ...,vk+1, and nonempty partial words w1,w2, ... ,wk such that

$$u = v_1 w_1 v_2 w_2 \ldots v_k w_k v_{k+1}$$

where $p'(u) = |v_1 w_1| = |v_2 w_2| = \cdots = |v_k w_k|$, *where* $|v_1| = |v_2| = \cdots = |v_k| = |v_{k+1}|$, *and where* $v_i \uparrow v_{i+1}$ *for all* $1 \leq i \leq k$, *and* $w_i \uparrow w_{i+1}$ *for all* $1 \leq i < k$.

Proof. Let u be a nonempty partial word over A with minimum weak period $p'(u)$. Then $|u| = kp'(u) + r$ where $0 \leq r < p'(u)$. Put $u = v_1 w_1 v_2 w_2 \ldots v_k w_k v_{k+1}$ where $|v_1 w_1| = |v_2 w_2| = \cdots = |v_k w_k| = p'(u)$ and $|v1| = |v2| = \cdots = |v_k| = |v_{k+1}| = r$. If w_i is empty, then $r = |v_{k+1}| = |v_k| = p'(u)$, a contradiction. If $k = 0$, then $u = v_{k+1}$ and u has weak period $|v_{k+1}| < p'(u)$ contradicting the fact that $p'(u)$ is the minimum weak period of u. Since $p'(u)$ is the minimum weak period of u, we get $v_i w_i \uparrow v_{i+1} w_{i+1}$ for all $1 \leq i < k$ and $v_k \uparrow v_{k+1}$. The result follows. ■

Third, we consider the following combinatorial property of words [21]: "For any word u over an alphabet A, if a and b are distinct letters of A, then ua or ub is primitive." This property has been generalized to partial words with one hole by Blanchet-Sadri and Chriscoe [6].

The following result treats the one-hole case.

Theorem 7 ([3]).
Let u be a partial word with one hole over an alphabet A which is not of the form x◊x for any word x. If a and b are distinct letters of A, then ua or ub is primitive.

Note that Theorem 7 does not hold for partial words with at least two holes. Consider for example $u = a\Diamond a\Diamond a$. Neither ua nor ub is primitive since $ua \subset a^6$ and $ub \subset (ab)^3$.

We now characterize all partial words u with at least two holes over an alphabet A such that if a and b are distinct letters of A, then ua or ub is primitive. Let u be a partial word with at least two holes, and let H denote $\|H(u)\|$, the cardinality of $H(u)$. Set $u = u_1\Diamond u_2\Diamond \ldots u_H\Diamond u_{H+1}$ where the u_j's do not contain any holes. We define a set S_H as follows: For all $1 \leq m \leq H$, if there exist a word x and integers $0 = i_0 < i_1 < i_2 < \cdots < i_m \leq H$ such that

$u_1\Diamond u_2\Diamond \ldots \Diamond u_{i_1} \subset x$,
$u_{i_1+1}\Diamond u_{i_1+2}\Diamond \ldots \Diamond u_{i_2} \subset x$,
$\vdots$
$u_{i_{m-1}+1}\Diamond u_{i_{m-1}+2}\Diamond \ldots \Diamond u_{i_m} \subset x$,
$u_{i_{m+1}}\Diamond u_{i_{m+2}}\Diamond \ldots \Diamond u_{H+1} \subset x$,

then put u in the set S_H. Otherwise, do not put u in S_H. For example, S_2 consists of the partial words of the form $x\Diamond x\Diamond x$ for a word x, or $x_1 \Diamond x_2 \Diamond x_1\, a x_2$ or $x_1\, a x_2 \Diamond x_1 \Diamond x_2$ for words x_1, x_2 and letter a.

Theorem 8.
Let u be a partial word with at least two holes over an alphabet A which is not in $S_{\|H(u)\|}$. *If a and b are distinct letters of A, then ua or ub is primitive* (*or there exists at most one letter* λ *such that* $u\lambda$ *is not primitive*).

Proof.
Set $\|H(u)\| = H$. Assume that $ua \subset v^k$, $ub \subset w^l$ for some primitive full words v,w and integers $k,l \geq 2$. Both $|v|$ and $|w|$ are periods of u, and, since $k,l \geq 2$, $|u| = k|v| - 1 = l|w| - 1 \geq 2 \max\{|v|,|w|\} - 1 \geq |v| + |w| - 1$. Without loss of generality, we can assume that $k \geq l$ or $|v| \leq |w|$. Set $u = u_1\Diamond u_2\Diamond \ldots u_H\Diamond u_{H+1}$ where the u_j's do not contain any holes. Since v ends with a and w with b, write $v = xa$ and $w = yb$. We have $u \subset (xa)^{k-1}x$ and $u \subset (yb)^{l-1}y$.

Case 1. k = l
Here $|v| = |w|$ and $|x| = |y|$. Note that $2 \leq k = l \leq H + 1$. First, assume that $k = l = H + 1$. In this case, it is clear that $u_1 = u_2 = \cdots = u_{H+1} = x$, a contradiction since $u \notin S_H$. Now, assume that $k = l \leq H$. There exist integers $0 = i_0 < i_1 < i_2 < \cdots < i_{k-1} \leq H$ such that

$u_{1_0+1}\Diamond u_{i_0+2}\Diamond \ldots \Diamond\, u_{i_1}\Diamond \subset xa$ and $u_{i_0+2}\Diamond \ldots \Diamond u_{i_1}\Diamond \subset yb$,
$u_{1_1+1}\Diamond u_{i_1+2}\Diamond \ldots \Diamond\, u_{i_2}\Diamond \subset xa$ and $u_{i_1+2}\Diamond \ldots \Diamond u_{i_2}\Diamond \subset yb$,
$\vdots$
$u_{i_{k-2}+1}\Diamond u_{i_{k-2}+2}\Diamond \ldots \Diamond u_{i_{k-1}}\Diamond \subset xa$ and $u_{i_{k-2}+1}\ \Diamond u_{i_{k-2}+1}\Diamond \ldots \Diamond u_{i_{k-1}}\Diamond \subset yb$,
$u_{i_{k-1}+1}\Diamond u_{i_{k-1}+2}\Diamond \ldots \Diamond u_{H+1} \subset x$ and $u_{i_{k-1}+1}\Diamond u_{i_{k-1}+2}\Diamond \ldots \Diamond u_{H+1} \subset y$.

We get

$u_{i_0+1}\Diamond u_{i_0+2}\Diamond \ldots u_{i_1} \subset x$,
$u_{i_1+1}\Diamond u_{i_1+2}\Diamond \ldots u_{i_2} \subset x$,
$\vdots$
$u_{i_{k-2}+1}\Diamond u_{i_{k-2}+2}\Diamond \ldots \Diamond u_{i_{k-1}} \subset x$,
$u_{i_{k-1}+1}\Diamond u_{i_{k-1}+2}\Diamond \ldots \Diamond u_{H+1} \subset x$,

a contradiction with the fact that $u \notin SH$.

Case 2. k > l
Here $|v| < |w|$ and $|u| \geq |v| + |w|$ (otherwise, $|u| = |v| + |w| - 1$ and $k = l = 2$).

First, assume that $|u| \geq L_{(H,|v|,|w|)}$. Referring to Section 3, u is also $\gcd(|v|,|w|)$-periodic. However, $\gcd(|v|,|w|)$ divides $|v|$ and $|w|$, and so $u \subset z^m$ with $|z| = \gcd(|v|,|w|)$. Since v ends with a and w with b, we get that z ends with a and b, a contradiction.

Now, assume that $|u| < L_{(H,|v|,|w|)}$. Set $k = lp + r$ where $0 \leq r < l$. We consider the case where $r = 0$ (the case where $r > 0$ is similar). We have that $k = lp$. The latter and the fact that $k > l$ imply that $p > 1$. Since $ua \subset (xa)^{lp}$ and $ub \subset (yb)^l$, we can write $y = x_1b_1x_2b_2 \ldots x_{p-1}b_{p-1}x_p$ where $|x_1| = \cdots = |x_p| = |x|$ and $b_1,\ldots,b_{p-1} \in A$. The containments $u \subset (xa)^{lp-1}x$ and $u \subset (x_1b_1x_2b_2 \ldots x_{p-1}b_{p-1}x_pb)^{l-1}x_1b_1x_2b_2 \ldots x_{p-1}b_{p-1}x_p$ allow us to write

$$u = v_1\Diamond v_2\Diamond \ldots v_{l-1}\Diamond v_l$$

where

$$v_j \subset x\ a\ x\ a \ldots x\ a\ x$$
$$v_j \subset x_1\ b_1\ x_2\ b_2 \ldots x_{p-1}\ b_{p-1}\ x_p$$

for all $1 \leq j \leq l$. If $l - 1 = H$, then $vj = uj = (xa)^{p-1}x$ for all j, and we obtain a contradiction with the fact that $u \notin S_H$. If $l - 1 < H$, then there exist integers $0 = i_0 < i_1 < i_2 < \cdots < i_{l-1} \leq H$ such that

$$u_{i_0+1}\lozenge u_{i_0+2}\lozenge\ldots\lozenge u_{i_1}\lozenge = v_1,$$
$$u_{i_1+1}\lozenge u_{i_1+2}\lozenge\ldots\lozenge u_{i_2}\lozenge = v_2,$$
$$\vdots$$
$$u_{i_{l-2}+1}\lozenge u_{i_{l-2}+2}\lozenge\ldots\lozenge u_{i_{l-1}}\lozenge = v_{l-1},$$
$$u_{i_{l-1}+1}\lozenge u_{i_{l-1}+2}\lozenge\ldots\lozenge u_{H+1} = v_l.$$

We get

$$u_{i_0+1}\lozenge u_{i_0+2}\lozenge\ldots\lozenge u_{i_1} \subset (xa)^{p-1}x,$$
$$u_{i_1+1}\lozenge u_{i_1+2}\lozenge\ldots\lozenge u_{i_2} \subset (xa)^{p-1}x,$$
$$\vdots$$
$$u_{i_{l-2}+1}\lozenge u_{i_{l-2}+2}\lozenge\ldots\lozenge u_{i_{l-1}} \subset (xa)^{p-1}x,$$
$$u_{i_{l-1}+1}\lozenge u_{i_{l-1}+2}\lozenge\ldots\lozenge u_{H+1} \subset (xa)^{p-1}x,$$

a contradiction with the fact that $u \notin S_H$. ■

Finally, we characterize a class of special partial words u with two holes over the binary alphabet $\{a,b\}$ where both ua and ub are non-primitive.

The concept of $(2,p,q)$-speciality can be rephrased as follows.

Definition 2 *([8])*.
Let p and q be positive integers satisfying $p < q$. A partial word u with two holes is called $(2,p,q)$-special if at least one of the following holds:

(1) There exists $0 \leq i < p$ such that $i + p, i + q \in H(u)$ (the position i is said to be 1-isolated),
(2) $q = 2p$ and there exists $p \leq i < |u| - 4p$ such that $i + p, i + 2p \in H(u)$ (the position i is said to be 2-isolated),
(3) There exists $|u| - p \leq i < |u|$ such that $i - p, i - q \in H(u)$ (the position i is said to be 3-isolated).

Lemma 1.
Let u be a partial word with two holes over the alphabet $\{a,b\}$ which is $(2,p,q)$-special according to Definition 2(1) for some integers $p < q$. Let i be the only position with letter a and assume that i is 1-isolated by $i + p$ and $i + q$ where $0 \leq i < p$. Then the following hold:

(1) *$ua \subset v^2$ for some word v imply $|ua| = 2q$ and $i = q - p - 1$,*
(2) *$ua \subset v^3$ for some word v imply $|ua| = 3p$, $q < 2p$ and $i = p - 1$,*
(3) *$ua \subset v^4$ for some word v imply $|ua| = 4p$, $q = 2p$, and $i = p - 1$,*
(4) *$ua \subset v^k$ and $ub \subset w^k$ for some words v,w and integer $k > 1$ imply $k = 2$,*
(5) *$ub \subset w^l$ for some word w and integer $l > 2$ imply $|ub| = 3|w| = 3p$ and $q = 2p$.*

Proof. We prove Statement 2 (the other statements are similar). So suppose $ua \subset v^3$ for some word v. Partitioning ua into segments of size $|v|$

u(0) u(1) … (|v| - 2) u(|v| - 1)
u(|v|) u(|v| + 1)… u(|v| - 2) u(2|v| - 1)
u(2|v|) u(2|v| + 1)… u(3|v| - 2) a

all the elements in each column must be contained within the same letter c for some $c \in \{a,b\}$. Thus $|v| - 1, 2|v| - 1$ have symbol $\Diamond$ or a. There are three cases to consider:

Case 1. $|v| - 1 = i$ and $2|v| - 1 = i + p$

We have $|v| - 1 = i$ and $2|v| - 1 = i + p$ imply $|v| = p$ and $i = p - 1$. Also, $i + q \leq 3p - 2$, so $q < 2p$.

Case 2. $|v| - 1 = i$ and $2|v| - 1 = i + q$

Then $|v| = q$ where $|v| - 1 = i$. But $i < p$, so we have $q - 1 < p$, which implies $q \leq p$, a contradiction.

Case 3. $|v| - 1 = i + p$ and $2|v| - 1 = i + q$

If $|v| - 1 = i + p$ and $2|v| - 1 = i + q$, then the column containing i also contains $i + |v|$, which has letter b. Then $ua \notin v^3$, a contradiction. ■

Theorem 9. *Let u be a $(2,p,q)$-special partial word according to Definition 2(1) with two holes over the binary alphabet $\{a,b\}$ where i is the only position with letter a and assume that i is 1-isolated by $i + p$ and $i + q$ where $0 \leq i < p$. Then $ua \subset v^k$ and $ub \subset w^l$ for some words v,w and integers $k,l \geq 2$ if and only if one of the following holds:*

(1) *$ua \subset v^2$ and $ub \subset w^2$ where $|ua| = |ub| = 2q$ and $i = q - p - 1$,*
(2) *$ua \subset v^3$ and $ub \subset w^2$ where $|v| = p$, $|w| = q$, $i = p - 1$ and $p = 2m$, $q = 3m$ for some integer $m \geq 1$,*
(3) *$ua \subset v^4$ and $ub \subset w^2$ where $|ua| = 4p$, $|ub| = 2q$, $i = p - 1$ and $q = 2p$.*

Proof.
First, we claim that if $uc \subset v^k$ for some word v, integer $k \geq 2$, and letter $c \in A$, then $k \leq \|H(u)\| + \|I(u)\| + 1$ where $I(u) = \{i \in D(u) \mid i\text{'s letter is } c\}$. To see this, partitioning uc into segments of size $|v|$, consider the column containing position $k|v| - 1$, which has letter c. Every other element in the column has symbol c or $\Diamond$. The elements in u which satisfy this requirement are the elements in $H(u)$ or $I(u)$. Thus, the number of rows in our array is less than or equal to $\|H(u)\| + \|I(u)\| + 1$.

Now, suppose $ua \subset v^k$, $ub \subset w^l$ for some words v and w and $k,l \geq 2$. It must be the case that $k \leq 4$ and $l \leq 3$ by the above claim and Lemma 1(5). Let us consider our possibilities:

Case 1. $k = 2$, $l = 2$

By Lemma 1(1), $|v| = q$ and $i = q - p - 1$.

Case 2. $k = 2$, $l = 3$

By Lemma 1(1), we have that $|ua| = 2q$ and $i = q - p - 1$. Now $ub \subset w^3$ for some word w. So it must be the case that $i + |w| = i + p$, $i + 2|w| = i + q$, which implies $|w| = p$, $q = 2p$, and $i \neq p - 1$. Observe that $i = q - p - 1 = p - 1$. We therefore have a contradiction and conclude that $k = 2$, $l = 3$ can never happen.

Case 3. $k = 3$, $l = 2$

By Lemma 1(2), we have $|v| = p$, $q < 2p$, and $p - 1 = i$. Since $ub \subset w^2$, we also have $i + |w| = i + p$ or $i + |w| = i + q$, and $2|w| - 1 > i + q$. This implies that $p + q < 2|w|$ and so $|w| = q$. In sum, we have that $ua \subset v^3$ and $ub \subset w^2$ where $|v| = p$ and $|w| = q$. Here $|ua|$ is a positive multiple of lcm(k,l), so we have that $p = 2m$ and $q = 3m$ for some integer $m \geq 1$.

Case 4. $k = 3$, $l = 3$

By Lemma 1(4), this can never happen.

Case 5. $k = 4$, $l = 2$

By Lemma 1(3), we have $|ua| = 4p$, $q = 2p$, and $i = p - 1$. Therefore, $|ub| = 4p = 2q$.

Case 6. $k = 4$, $l = 3$

By Lemma 1(3) and Lemma 1(5), we have $|v| = p$, $|w| = p$, which is a clear contradiction. ■

The converse is trivial.

5. Conclusion

In this paper, we presented some new periodicity properties on partial words which are built on Fine and Wilf' s periodicity result generalized to partial words. As was discussed in Section 3, these generalizations fall into two categories: the ones that relate to strong periodicity and the ones that relate to weak periodicity. Our main result shows that there exist very many primitive partial words. We believe that some of the results in Section 4 can have interesting applications, such as answering the open question that was discussed at the beginning of Section 3.

References:

[1] J. Berstel, L. Boasson, Partia1 words and a theorem of Fine and Wilf, Theor. Comput. Sci., 218 (1999)135–141.

[2] F. Blanchet-Sadri, Periodicity on partia1 words, Comput. Math. Appl. 47 (2004) 71–82.

[3] F. Blanchet-Sadri, Primitive partia1 words, Discrete Appl. Math. 148 (2005)195–213.

[4] F. Blanchet-Sadri, A.R. Anavekar, Testing primitivity on partia1 words, Discrete Appl. Math.155 (2007) 279–287. <www.uncg.edu/mat/primitive/>

[5] F. Blanchet-Sadri, D. Bal, G. Sisodia, Graph connectivity, partia1 words, and a theorem of Fine and Wilf, Inform. Comput. 206 (2008) 676–693. <www.uncg.edu/mat/research/finewilf3/>

[6] F. Blanchet-Sadri, Ajay Chriscoe, Loca1 periods and binary partia1 words: an algorithm, Theor. Comput. Sci. 314 (2004) 189–216. <www.uncg.edu/mat/AlgBin/>.

[7] F. Blanchet-Sadri, J. Gafni, K. Wilson, Correlations of partia1 words, in: W. Thomas, P. Wei1 (Eds.), STACS 2007, LNCS 4393, Springer-Verlag, Berlin, 2007, pp. 97–108. <www.uncg.edu/mat/research/correlations/>.

[8] F. Blanchet-Sadri, Robert A. Hegstrom, Partia1 words and a theorem of Fine and Wilf revisited, Theor. Comput. Sci. 270 (2002) 401–419.

[9]F. Blanchet-Sadri, T. Oey, T. Rankin, Partia1 words and generalized Fine and Wilf's theorem for an arbitrary number of weak periods. <www.uncg.edu/mat/research/finewilf2/>.

[10] M.G. Castelli, F. Mignosi, A. Restivo, Fine and Wilf's theorem for three periods and a generalization of Sturmian words, Theor. Comput. Sci. 218 (1999) 83–94.

[11] C. Choffrut, J. Karhumäki, Combinatorics of words, in: G. Rozenberg, A. Salomaa (Eds.), Handbook of Forma1 Languages, vol. 1, Springer-Verlag, Berlin, 1997, pp. 329–438.

[12] S. Constantinescu, L. Ilie, Generalised Fine and Wilf's theorem for arbitrary number of periods, Theor. Comput. Sci. 339 (2005) 49–60.

[13] S. Constantinescu, L. Ilie, Fine and Wilf's theorem for abelian periods, Bull. Eur. Assoc. Theor. Comput. Sci. EATCS 89 (2006)167–170.

[14] N.J. Fine, H.S. Wilf, Uniqueness theorems for periodic functions, Proc. Am. Math. Soc.16 (1965)109–114.

[15] L.J. Guibas, A.M. Odlyzko, Periods in strings, J. Comb. Theory A 30 (1981)19–42.

[16] V. Halava, T. Harju, L. Ilie, Periods and binary words, J. Comb. Theory A 89 (2000) 298–303.

[17] J. Justin, On a paper by Castelli, Mignosi, Restivo, Theor. Inform. Appl. 34 (2000) 373–377.

[18] A.M. Shur, Y.V. Gamzova, Periods' interaction property for partia1 words, in: T. Harju, J. Karhumäki (Eds.), Words 2003, TUCS 27, Turku, 2003, pp. 75–82.

[19] A.M. Shur, Y.V. Gamzova, Partia1 words and the periods' interaction property, Izv. RAN 68 (2004)199–222.

[20] A.M. Shur, Y.V. Konovalova, On the periods of partia1 words, Lecture Notes Comput. Sci. 2136 (2001) 657–665.

[21] H .J. Shyr, Free Monoids and Languages, Hon Min Book Company, Taichung, Taiwan, 1991

[22] R. Tijdeman, L. Zamboni, Fine and Wilf words for any periods, Indag. Math. 14 (2003)135–147.