Tiling Periodicity

Juhani Karhumäki, Turku University, Finland
http://users.utu.fi/karhumak/

Yury Lifshits, Steklov Institute of Mathematics St.Petersburg, Russia
http://logic.pdmi.ras.ru/~yura/

Wojciech Rytter, Warsaw University
http://www.mimuw.edu.pl/ rytter/

Abstract

A word is called tiling periodic if it can be split in a finite number of copies of a partially defined word. For example the word XXYY is periodic, since it consists of two copies of the partially defined word X_Y. We investigate this notion, study its properties and present an almost linear time (within polylogarithmic factor)c algorithm for finding minimal tiling periods. As a side product of the paper we solve two open problems posted by T. Harju and N.J.A. Sloane, respectively.

1 Introduction

The word S is called *purely periodic* if $S = W^k = W \dots W$. Equivalently, pure periodicity means $p|n$ and the equality $s_i = s_{i+p}$ for all $1 \le i < i + p \le n$. In this paper we deal with pure periodicity only. In this paper we are going to introduce a more general notion of periodicity, according to which these words become periodic. Informally, a word T is called to be *tiling periodic* if there exists a partially defined word S (we call it tiler) such that T can be split in several parallel copies of S. A **tiler** (or partial word) is a word over $\Sigma \cup \{\diamond\}$ alphabet, where $\diamond$ is a special transparent (or undefined) letter. In other words, a tiler is a sequence of words (blocks) with gaps of fixed sizes between the blocks. A tiler S is called the *tiling period* of an (ordinary) word T if we can split T in shifted copies of S satisfying the following: **(a)** All defined (visible) letters of S-copies match the text letters; **(b)** Every text letter is covered by *exactly one* defined (visible) letter.
For example $a\,a\,\diamond\diamond\,b\,b$, $a\,\diamond\diamond\diamond\,b$ are tiling periods of $aaaabbbbaaaabbbb$).

Problem. This paper is devoted to understanding the basic properties of tiling periodicity and comparing it to the classical notion. We address the following questions: (1) How to enlist all possible tiling periods? (2) Does any word have a unique *minimal* tiling period? (3) How to find all minimal tiling periods? (4) How many periods can have a word of length n? (5) What is the relation between the minimal classical and the minimal tiling periods?

We have several reasons to be interested in tiling periodicity. First of all, it is a natural generalization of classical notion, i.e. any classical period is also a tiling period, but not necessary other way around. Second, in the case when a tiling period is relatively small, we can describe a long word by just its tiling period and the length. Hence, we get a new class of words with low Kolmogorov complexity. Tiling periodicity might be useful for the new text compression methods (especially for generalizing run-length encoding). Note here, that the ratio between the "size" of tiling period and the length of minimal classical period may be arbitrarily small. Next, the notion of tiling periodicity provides a geometrical intuition about structure of the text. We have a conjecture (look at Section **??**), that tiling periodicity is not expressible by word equations. Yet another motivation for studying tiling periodicity is a hope for applications in pattern discovery in some real data. We suggest to research community to perform experimental projects in this direction. We want to show one natural source of tiling periodicity. Consider multidimensional rectangle $n_1 \times \cdots \times n_d$. Let every integer point in it be colored. We sort all points in lexicographical order of their coordinates (from $(0, \ldots, 0)$ to $(n_1 - 1, \ldots, n_d - 1)$) and write down all their colors in a single sequence. Assume that the initial rectangle was tiled by a smaller one $m_1 \times \cdots \times m_d$ where $m_1 | n_1, \ldots, m_d | n_d$ and every copy of the smaller rectangle was colored in the same way. Then the color sequence for the bigger rectangle has a tiling period. We comment this example later after introducing some useful terminology. Automatically generated texts and XML-files might be another possible sources of tiling periodicity.

Our results. The main result of the paper is definitely the notion of tilling periodicity itself. It looks very simple and natural, but up to our knowledge it was never formulated before. Next, we introduce a partial order on tiling periods and discover that contrary to the classical case there might be several incomparable minimal tiling periods. We present the shortest known example (24 letters) with that property in Section **??**. This example helps to disprove a *common subtiler* conjecture by Tero Harju.

However, in Subsection **??** we prove that every minimal tiling period of a word T is also a tiling period of the minimal classical period of T. This property tells us that tiling periodicity lives "inside" classical one. The proof is based on the extended Euclid algorithm.

In Subsection **??** we present a classification of all possible tiling periods. The most difficult and tricky part was to prove that our construction is complete, i.e. it contains all possible tiling periods. We get a recursive formula for computing function $L(n)$ that gives the maximal number of tiling periods for any word of length n. The value $L(n)$ might be even more than the text length. For example, a word of length 36 can have up to $L(36) = 52$ tiling periods. As a side product we prove that sequences A067824 and A107736 from the On-line Encyclopedia of Integer Sequences [15] coincide.

Finally, we present an algorithm for finding (all) minimal tiling periods of a given text in Subsection 3. We can compute the answer in time $n^{1+o(1)}$.

Related results One of the most important results for classical periodicity is a theorem by Fine and Wilf [7]. They studied the necessary and sufficient condition under which from p-periodicity and q-periodicity we can derive $\gcd(p, q)$-periodicity (recall that gcd is the greatest common divisor). We tried to prove the similar property for tiling periods. Surprisingly, it does not hold. Even two *pure* tiling periods of the same text may have no common subperiod. Our attempt to generalize the notion of periodicity is not the first one. Recently, Simpson and Tijdeman generalized Fine and Wilf theorem for multidimensional periodicity [14]. They say that some vector $\bar{v}$ is a period for multidimensional colored figure if for any point x it is colored in the same way as $x + \bar{v}$. Another generalization of theorem by Fine and Wilf is presented in [5]. Berstel and Boasson [10] followed by Shur and Konovalova [13] and Blanchet-Sadri [2, 3, 4] extensively studies partial words and their periodicity properties. However, overlapping of periods (where every letter of the text is covered exactly once) were not considered and therefore in their terms only partial words may have partial periods (i.e. periods with gaps). In the paper [1] the notion of cover was introduced by Apostolico, Farach and Iliopoulos. A word C is a cover for a word T if any letter of T belongs to some occurrence of C in T. Hence, there are two differences with our notion: covers are connected (no gaps) and covers can overlap. Katona and Szàsz introduced in [11] a sort of tiling periodicity in two dimensions. Unfortunately, they consider tilers consisting only of two letters! Finally, the notion of periodicity for continuous functions was generalized under the name *almost periodic functions* [6].

2 Properties of Tiling Periodicity

Partial Order on Tilers. We say that one tiler S is *smaller* than another tiler Q, if Q can be split in several parallel copies of S satisfying the following: **(a)** All defined (visible) letters of S-copies match the visible Q letters; **(b)** letter.
In particular, any tiling period is smaller than the original text. Actually, this definition just extend tiling periodicity to partial words.

In the classical periodicity any text has a single minimal pure period. Here by "minimal" we mean that any other period is just formed by several repetitions of the minimal one. For finding tiling periodicity it would be very important to have the same result for tiling periodicity. We formulate it below as **Minimal Tiling Period** Question: is it true that for every ordinary word there exists a unique minimal tiling period (i.e. it is smaller than any other tiling period)? Reformulation: any two tiling periods have a common tiling "subperiod". Surprisingly, this naturally looking conjecture is wrong. Figure 1 presents the shortest known example (24 letters) with two incomparable periods:

a a a	a a a	a a b	a a b	b a a	b a a	a a a	a a a

a a a a	a a a a	b a a b	b a a b	a a a a	a a a a

Figure 1: Two minimal tiling periods of an example word, both of size 12.

Let us give a short sketch why they have no common subperiod. Assume that they have. Then this tiler has only one letter B. Since the distance between the first B and the second B in the text is equal to three, the shift between the first and the second copy of our tiler is also equal to three. Hence, the tiler begins with three consecutive A. But the distance between the second and the third copy of our tiler is equal to one. And therefore, starting blocks of the second and the third copy overlap.

How to get more than two incomparable tiling periods? We can use a "nested construction". Let T be the text presented above and let $T_{A_1B_1}$ and $T_{A_2B_2}$ be obtained from T by just using different letters. We can construct the text T_2 by replacing every A in T by $T_{A_1B_1}$ and every B in T by $T_{A_2B_2}$. Now we can describe four incomparable tiling periods for T_2. The text T_2 has length 24^2. Group all letters to 24 blocks of 24 letters. Choose 12 of blocks using one of two partial words above. Then inside each block also keep only 12 letters using either first or second tiling period (the same in every block). We can repeat the construction several times. The text T_k has size 24^k and has 2^k incomparable periods. Asymptotically it has $n^{\frac{\log 2}{\log 24}} > \sqrt[5]{n}$ incomparable minimal tiling periods.

In 2003 in his lecture course [8] Tero Harju asked the following question. Assume that some colored cellular figure has tiling by one pattern and by another one. Is it always true that there exists the third one such that both first two can be tiled by that third one? This question is motivated by studying defect effect in combinatorics of words [9]. Our example shows that the answer is negative even for one-dimensional (but disconnected!) figures.

We consider now the situation when the number of tilers in a word is maximal, this happens for a unary word of length n (i.e only one character is used). In unary ase we reformulate tiling periodicity in the algebraic terms (like $s_i = s_{i+p}$ for classical periodicity). This reformulation helps us to prove that any minimal tiling period is smaller than the minimal classical period and to compute all minimal tiling periods.

Theorem 1. *There is a one-to-one correspondence between tiling periods of a unary word and decompositions $n = n_1 \cdot \ldots \cdot n_k$, where $n_2, \ldots, n_k \geq 2$.*

Proof. At first, we describe a set of tiling periods (hierarchial construction) for the word of length n over the unary alphabet. Then we prove that this set is complete, i.e. contains all possible tiling periods.

We divide the set of all periods in levels. Every period in our system is associated with a special *code*. The whole text itself is the only period from 0-level and has the code n. For every decomposition $n = n_1 \cdot n_2$ with $n_2 \geq 2$ the block of size n_1 is a period from 1-level with a code $n_1 \cdot n_2$. The 1-level actually contains all classical (i.e. connected) periods.

We now explain how to construct a period Q of $k+2$-level from a period P of k-level with the code $n_1 \cdot \ldots \cdot n_k$. Take any decomposition $n_1 = m_1 \cdot m_2 \cdot m_3$ with $m_2, m_3 \geq 2$. Through all our construction the first number of a code is equal to block size of a tiling period (all continuous blocks have equal sizes in our constructions). To construct the new tiler Q we take every block of P, divide it into m_3 groups of size $m_1 \cdot m_2$, and in every such group keep only first m_1 letters. We can notice that P can be tiled by m_2 copies of Q (with shifts $0, m_1, \ldots, m_1 \cdot (m_2 - 1)$). Hence, Q is also a tiling period for the text with the code $m_1 \cdot m_2 \cdot m_3 \cdot n_2 \cdot \ldots \cdot n_k$. Note, that our construction maintains the inequality $n_2, \ldots, n_k \geq 2$ for all codes.

We now prove that any tiling period is included in our construction. Take any tiling period P. It must have all continuous blocks of the same size s and the length of any gap in P must be a multiple of s. Indeed, in the text tiling the second copy of P is shifted by the size of the first block s_1. Hence, for avoiding overlapping all other blocks are smaller or equal than s_1. Assume now, that the block b is the first one which is *strictly* smaller than s_1. Then the gap between b in the first copy of P and b in the second copy of P is smaller than s_1. Hence, that gap cannot be filled by anything. Every gap in P is filled by several blocks of some copies of P. Hence it must be a multiple of s.

We now define a *plain power* for every tiler P. Let s be the block size and $g \cdot s$ be the length of the first gap. The plain power of P is the union of P-copies shifted by $0, s, 2s, \ldots, g \cdot s$.

Claim: The plain power of any tiling period P is also a tiling period.

Proof: take a text splitting in P-copies. Let s be the block size and $g \cdot s$ be the length of the first gap. We divide all copies of P into groups of $g+1$ copies in the following way. Consider all P-copies from the left to the right. The gap between first two blocks of the first copy can be filled *only by first* blocks of other P-copies. These copies together with the first one form the first group. Now assume that we already formed several such groups. Consider the first copy of P unused so far. Look at its first gap. All P-copies from the previous groups has long (at least $s\cdot(g+1)$) continuous blocks. Hence, this gap is also filled by new, still unused P-copies. Since we process all copies from the left one to the right one, all of them contribute to this gap filling by their first block. Therefore these copies form the next group we need. Every group itself is exactly a plain power of P. Hence, the initial text has also splitting in the plain power copies. The claim is proved.

If there exists some P outside of our construction, then there also exists some tiling period P' such that (1) it is outside of our hierarchy and (2) its plain power Q generated by our hierarchy. Let us derive a contradiction from that. Indeed, let $n_1\cdot\ldots\cdot n_k$ be the code of Q, let s be the block size of P' and $s\cdot g$ be the length of its first gap. Then n_1 is the block size of Q and it is a multiple of $s\cdot(g+1)$. Let $m_1=s, m_2=g+1, m_3=n_1/(s\cdot(g+1))$. Now we see that P' is in fact a tiling period with the code $m_1\cdot m_2\cdot m_3\cdot n_2\cdot\ \ldots\ \cdot n_k$. □

Corollary 1. Let $L(n)$ be the number of periods for the word of length n over a unary alphabet. The theorem above states that $L(n)$ is equal to the number of factorizations $n=n_1\cdot\ldots n_k$, where $n_2,\ldots,n_k\geq 2$. By grouping all decompositions by the rightmost factor we obtain the recurrence: $L(1)=1;\ L(n)=1+\sum_{d|n,d\neq n}L(d)$.

Remark. Two related sequences are included in the On-line Encyclopedia of Integer Sequences [15] maintained by N.J.A. Sloane. The sequence A067824 is defined by the formula in Corollary 1. The sequence A107736 is the number of polynomials p with coefficients in $\{0,1\}$ that divide x^n-1 and such that $(x^n-1)/((x-1)p(x))$ has all coefficients in $\{0,1\}$. But this is exactly the number of tiling periods for the unary text of the length n. Indeed, multiplying p by some polynomial with coefficients in $\{0,1\}$ we are trying to split n "ones" in several parallel copies of p without overlapping. In the Encyclopedia it was indicated that A067824 and A107736 agree at least on first 300 terms, but no proof of equivalence was known so far. Theorem 1 and Corollary 1 give such a proof. Indeed, the number of polynomials is equal to the number of tiling periods, the number of tiling periods is equal to the number of factorizations (Theorem 1), the number of factorizations satisfies the recursive formula (Corollary 1).

The function $L(n)$ and the number of factorizations also appear in Knuth's book "The Art of Computer Programming" [12]. However, no closed formula for $L(n)$ is known so far.

Example. We have for example $L(36)=52$. Hence the word of length 36 can have up to 52 tiling periods. Therefore, there exist two different tilers with the same wing-size (the distance between the first and the last letter).

Definition. We say that a word $T=T_0\ldots T_{n-1}$ has a *pure ranged period* (a,b) (or a period a ranged by b) if $b|n$, $a|b$, $a<b$ and the following statement holds:

$$\left\lfloor\frac{i}{b}\right\rfloor=\left\lfloor\frac{i+a}{b}\right\rfloor\Rightarrow T_i=T_{i+a}.$$

In other words, if we divide the whole text into blocks of length b, then each of them has the pure period a. Classical pure period p coincides with ranged period (p,n).

Lemma 1. *A word T has a tiling period with the code $n_1\cdot\ldots\cdot n_{2k}$ or $n_1\cdot\ldots\cdot n_{2k+1}$ iff it has ranged periods respecively*

$$(n_1,n_1n_2),\ldots,(\prod_{i=1}^{2k-1}n_i,\prod_{i=1}^{2k}n_i)\quad or\quad (n_1,n_1n_2),\ldots,(\prod_{i=1}^{2k-1}n_i,\prod_{i=1}^{2k}n_i).$$

Proof. Step 1: from tiling period P to ranged periodicity. Let us use induction over tiler's level. Consider the corresponding text tiling. Recall that all copies of P can be divided into groups of size n_2 with internal shifts $0, n_1, \dots, n_1(n_2-1)$. Hence, if we divide the whole text into the blocks of size $n_1 n_2$, every block is covered by P-copies from the same groups, and therefore it is n_1-periodic (inside the block). We proved $(n_1, n_1 n_2)$-ranged periodicity. All others follow from the induction hypothesis for the plain power of P.

Step 2: from ranged periods to a tiling periodicity. Consider the *top* ranged period $(\prod_{i=1}^{2k-1} n_i, \prod_{i=1}^{2k} n_i)$. Let us consider the text tiling by tiler Q with the code $(\prod_{i=1}^{2k-1} n_i) \cdot n_{2k}$ for the first statement of lemma and Q with the code $(\prod_{i=1}^{2k-1} n_i) \cdot n_{2k} \cdot n_{2k+1}$ for the second case. Directly from this top ranged periodicity every copy of Q has the same letters on the same places. Hence, Q is a tiling period for the text. Continuing the reasoning, with the help of the second ranged periodicity we find another tiling period R inside Q. Finally, the last ranged periodicity gives us the tiling period with the code we promised in the lemma's statement. □

Lemma 2. *If the text T has a classical period p and a ranged period (a, b), then either $b|p$ or the text has also classical period $gcd(a, p)$.*

Proof. Take any letter T_i. We are going to make several moves of size $\pm p$ and $+a$ for reaching position $i + \gcd(a, p)$. We want to be on the same character every time and hence we can not make a move of size $+a$ from the last a-blocks in every b-block. As we know from extended Euclid algorithm, there exist integers k and l such that $\gcd(a, p) = ka - lp$. We will use the following *greedy* strategy. If we are able to make a move of size $+a$ we do this. Otherwise we try to make several moves of size $\pm p$. After making exactly k moves of size $+a$ we just make all the remaining moves of size $\pm p$ and we are done. We cannot follow greedy strategy only if for some position $j < p$ from all points $j, j+p, \dots j + (\frac{n}{p} - 1)p$ we cannot jump $+a$. This means that all that points belong to the last a-groups in b-blocks. Suppose now that p is not divided by b. Then $u = \gcd(p, b) \leq \frac{b}{2}$. Since $p|n$ and $b|n$, among the numbers $j \bmod b, \dots, j + (\frac{n}{p} - 1)p \bmod b$, there exists all residues h modulo b such that $h \equiv j (\bmod u)$. Hence, one of these values is smaller than u which does not exceed $b/2$. But this means that the corresponding point does not belong to the last a-group in the b-block. □

Theorem 2. *Any minimal tiling period Q of word T is also a tiling period for the minimal classical period of T.*

Proof. We use induction over the text length. In the case of the one-letter-text theorem is true. Let now p be the length of the classical period and (a, b) be the top ranged periodicity from the code of Q (here we use Lemma 1). From $p = n$ theorem follows immediately. Assume now that $p < n$. By hierarchial construction Q consists of some letters from the first a-block in every b-blocks in the text. Let us apply Lemma 2. If $b|p$, then we can produce a new tiling period Q' restricting Q to the first p symbols in the text. By p-periodicity Q can be split in several copies of Q' with shifts $0, p, 2p, \dots, (\frac{n}{p} - 1)p$. We get the contradiction against the minimality of Q. Therefore the text has classical period $\gcd(a, p)$. But p is the minimal classical period. Hence, $p|a$. Since the text is p-periodic, it is also a-periodic and b-periodic. That means that we can again restrict Q only for the first a-block. Either we get a new smaller tiling period Q' (and that gives us contradiction) or $b = n$ and all letters of Q belong to the first a-block. We now see that both classical period p and the tiling period Q are periods for the first a-block. Since $a \leq b/2 \leq n/2$, we can apply induction hypothesis. □

Corollary 2. Take any tiling period Q and any classical period p. Then they have a common "tiling subperiod".

Proof. Consider a minimal tiling period Q' that is smaller than Q. Consider the minimal classical period p'. From folklore we know that $p'|p$. By Theorem 2 Q' is a tiling period for the first p'-block of the text. Hence, Q' is the required tiling subperiod for Q and p. □

Remark. Using the technique from Lemma 1, Lemma 2 and Theorem 2 it is possible to prove that the minimal tiling period is unique for $n = 2^k$. We just suggest the proof scheme here. Take two minimal incomparable tiling periods. Consider their top ranged periodicities. Apply the reasoning of Lemma 2. Either one of the periods is not minimal or these range periodicities are the same. Going down we prove either their equivalence or their nonminimality.

3 Algorithm for Finding Minimal Tiling Periods

We define the size of tiling period as the number of defined letters in it. In the algorithm we ue the fact that for every tiling period there is a corresponding chain of *embedded* ranged periodicities $(a_1, b_1), \ldots, (a_k, b_k)$. By "embedded" we mean that $b_i | a_{i+1}$ holds for every i. Notice that the size of the tiling period is equal to $n \prod_{i=1}^{k} \frac{a_i}{b_i}$

ALGORITHM *Compute-Minimal-Tilers*

1. Construct the acyclic tiling graph $\mathcal{G}$ of ranged periods.
   (a) $V = \{(a, b) : 1 \leq a < b \leq n, \ \& \ a|b \ \& \ b|n \ \& \ RangedPeriod(a, b) \ )$
   (b) $E = \{(a, b) \to (c, d) : b|c\}$.
   (c) Assign the weight b/a to every node $(a, b) \in V$.
2. Find in $\mathcal{G}$ a path π having maximal product $val(\pi)$ of its node-weights;
3. **Output** $tiler(\pi)$ (of size $\frac{n}{val(\pi)}$).

For a string from Figure 1 $\mathcal{G}$ consists only of two nodes $(4, 8)$, $(3, 6)$. There are two maximal paths, each consists of a single node. Both produce minimal size tilers.

Remark. The following $O(n)$ size *max-paths graph* representation $\mathcal{G}' \subseteq \mathcal{G}$ of all minimal tiling periods can be constructed in an obvious way with additional linear work (sice the graph is extremely small). For each vertex v of $\mathcal{G}$ compute all outgoing edges which are on a maximal path starting from v. The graph of all edges on maximal-product paths represents all minimal tiling periods. In particular we can compute easily the number of such periods.

Lemma 3. *Assume we can preprocess the string in time $O(D(n))$ to compute each query RangedPeriod in logarithmic time. Then the algorithm* Compute-Minimal-Tilers*works in time $O(n + D(n))$,* *Proof.* By $d(n)$ we denote the number of divisors of n. It is known that $d(n) = O(n^{\epsilon})$, for any constant $\epsilon > 0$. Hence the number of nodes and edges is $O(n)$. The constrution of the graph can be done in linear time after preprocessing. Computation of a maximal path in time $O(n)$ is very easy, since it is an acyclic graph with linear number of edges. □

Let x be a string of length n and let $Divisors(n)$ denotes the set of divisors of n. For a divisor p of n, by p-block we mean a subword of the form $x[i \cdot p..(i+1) \cdot p - 1]$. The subwords are given by their starting-endding positions in x. Denote by $Blocks(x, p)$ the set of all p-blocks of x. Assume we count positions starting from 0. Denote by $period(w)$ the size of the smallest full period of a string w.
The **main operation** in the algorithm *Compute-Minimal-Tilers*is the test $RangedPeriod$ which can be expressed in terminology of the blocks as follows:

RangedPeriod(p, q):
for given $p, q \in Divisors(n)$ check if $\forall\ w \in Blocks(x, p)$ w has a full period q.

We call this operation the *main query*. We show.

Theorem 3. [Fast-Preprocessing]
We can preprocess the string in $O(n \log n \log\log n)$ time in such a way that each main query can be answered in constant time.

The *Compute-Minimal-Tilers*algorithm is doing a sublinear number of $RangedPeriod$ quieris, hence the theorem implies immediately the following speed-up result.

Theorem 4. *The minimal tiler of a string can be found in $O(n \log n \log\log n)$ time.*

We now show the proof of the *Fast-Preprocessing theorem.* Firstly we introduce and concentrate on *small queries.*

A **small query** operation is to compute for a subword w of x the value $period(w)$.

Lemma 4. [Small Queries]
We can preprocess the string x in $O(n \log n)$ time in a way that each small query can be answered in logarithmic time.

Proof. Denote by $powers(n)$ the set of all powers of two, not exceeding n. Our basic data structure is the dictionary of basic factors, we refer to [**?**] for detailed definition. Denote this data structure by DBF(x). For each subword w of x of length $k \in powers(n)$ we have a unique label $NAME(k, w) \in [1..n]$ such that two subwords (given by their intervals in x) are the same as words if their labels are the same. The subwords which lengath are in powers(n) are called basic factors.

The DBF data structure allows to check in constant time subword which lengths are not necessarily in $powers(n)$ (decomposing them into ones which are, possibly overlapping each other).

We augment DBF(x) with additional structure. For each $k \in powers(n)$ we keep a sorted list of all subwords of size k, sorted lexicographically with respect to their names and starting positions in x. Let us called the resulted structure the Extended DBF and denote it by

EDBF(x). Using the techniques from [**?**], or from the algorithm of Karp-Miller-Rosenberg one can easily prove.

Claim. EDBF(x) can be computed in time $O(n \log n)$.

We show now how we compute $period(w)$ for $w = x[p..q]$. We can check if w has a full period of exponent 2 or 3 in constant time, since we can check equality of words in constant time.

Now let us go to larger exponents, assume $period(w) \leq |w|/4$. Let us take the prefix u of w which size k is in $powers(n)$ and such that $|w|/4 \leq k \leq |w|/2$. We can find the first *internal* position in w where an occurrence of u starts. Assume it is a position j in w, counting from 0.

Claim. $period(w) \leq |w|/4$ implies $j = period(w)$.

The claim follows from the fact that period(w) in this case is a size of a primitive word v such that w is a full power of v. This primitive word can start an occurrence only at positions which are multiples of w, due to primitivity. Hence the first such internal position should be equal to $|v| = period(w)$. This completes the proof of the lemma. □

We define now the following data structure. For each $p \in Divisors(n)$ define $GCDP(p)$ as the greatest common divisors of the smallest full periods of all p-blocks of x.

Lemma 5. [GCDP-Computation]
We can compute GCDP(p) for all divisors p of n in $O(n \log n \log\log n)$ time.

Proof. Assume we constructed EDBF(x). For each $p \in Divisors(n)$ we compute the set of periods of p-blocks in time $O(p)$. Then we compute the greatest common divisor of all these periods in time $O(p) \log n$. Now the thesis follows from the well known number-theory fact that

$$\sum_{p \in Divisors(n)} p \;=\; O(n \log\log n)$$

This completes the proof of the lemma. □

We can finish now the proof of the Fast-Preprocessing theorem. The main query $RangedPeriod(p, q)$ is equivalent to the property:

$$GCDP(p) \text{ is a divisor of } q.$$

This property can be checked in logarithmic time using the precomputed values of $GCDP(p)$. Consequently, this completes the proof of the Fast-Preprocessing theorem.

References

[1] A. Apostolico, M. Farach, and C. S. Iliopoulos. Optimal superprimitivity testing for strings. *Inf. Process. Lett.*, 39(1):17–20, 1991.

[2] F. Blanchet-Sadri. Periodicity on partial words. *Computers and Mathematics with Applications*, 47(1):71–82, 2004.

[3] F. Blanchet-Sadri. Primitive partial words. *Discrete Applied Mathematics*, 148:195–213, 2005.

[4] F. Blanchet-Sadri and R.A. Hegstrom. Partial words and a theorem of Fine and Wilf revisited. *Theoretical Computer Science*, 270(1/2):401–419, 2002.

[5] S. Constantinescu and L. Ilie. Generalised Fine and Wilf's theorem for arbitrary number of periods. *Theor. Comput. Sci.*, 339(1):49–60, 2005.

[6] C. Corduneanu and H. Bohr. *Almost Periodic Functions.* Chelsea Publishing Company; 2nd edition, 1989.

[7] N. Fine and H. Wilf. Uniqueness theorems for periodic functions. *Proc. Amer. Math. Soc.*, 16:109–114, 1965.

[8] T. Harju. Defect theorem, lecture notes of "Combinatorics of words" Tarragona course, 2002/2003.

[9] T. Harju and J. Karhumäki. Many aspects of defect theorems. *Theor. Comput. Sci.*, 324(1):35–54, 2004.

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

[11] G.O.H. Katona and D.O.H. Szász. Matching problems. *J. of Combinatorial Theory Ser B*, 10(1):60–92, 1971.

[12] Donald Knuth. *The Art of Computer Programming, Volume 4, Fascicle 3: Generating All Combinations and Partitions.* Addison-Wesley, 2005.

[13] A.M. Shur and Y.V. Konovalova. On the periods of partial words. In *MFCS'01*, number 2136 in LNCS, pages 657–665. Springer-Verlag, 2001.

[14] R.J. Simpson and R. Tijdeman. Multi-dimensional versions of a theorem of Fine and Wilf and a formula of Sylvester. *Proc. Amer. Math. Soc.*, 131:1661–1667, 2003.

[15] N.J.A. Sloane. Sequence A067824. The On-Line Encyclopedia of Integer Sequences, http://www.research.att.com/~njas/sequences/A067824.