Tiling Periodicity

Juhani Karhumäki[1], **Yury Lifshits**[2] and Wojciech Rytter[3]

[1]University of Turku,
[2]Steklov Institute of Mathematics at St.Petersburg,
[3]Warsaw University

Combinatorial Pattern Matching 2007

The string S has a **full period** if

$$S = W^k = W \ldots W$$

Equivalently,

$$\forall \quad 1 \leq i < i + p \leq n : \quad s_i = s_{i+p}$$

The string S has a **full period** if

$$S = W^k = W \dots W$$

Equivalently,

$$\forall \quad 1 \le i < i + p \le n : \qquad s_i = s_{i+p}$$

Does the following string have full period?

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

The string S has a **full period** if

$$S = W^k = W \dots W$$

Equivalently,

$$\forall \quad 1 \leq i < i + p \leq n : \quad s_i = s_{i+p}$$

Does the following string have full period?

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

Not in the classical sense. But...

Outline of the Talk

1. Notion of Tiling Periodicity
2. Properties of Tiling Periodicity
3. Finding Tiling Periods of Minimal Size
4. Conclusions and Future Work

Notion of Tiling Periodicity

Motivating Examples

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

The string above is not periodic, but the red **structure**

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

is a kind of period, since we can cover initial string by **four parallel copies** of it:

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

Motivating Examples

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

The string above is not periodic, but the red **structure**

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

is a kind of period, since we can cover initial string by **four parallel copies** of it:

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

The simplest example:

A	A	B	B

Formal Definition

A **partially defined word** (tiler) is a string over $\Sigma \cup \diamond$ alphabet, where $\diamond$ is a special **transparent** (undefined) letter.

Formal Definition

A **partially defined word** (tiler) is a string over $\Sigma \cup \diamond$ alphabet, where $\diamond$ is a special **transparent** (undefined) letter.

A tiling string S is called the **tiling period** of (ordinary) string T if we can cover T by parallel copies of S satisfying the following:

- All defined (visible) letters of S-copies match the text letters
- Every text letter covered by **exactly one** defined (visible) letter

Why Tiling Periodicity

- Natural generalization of the classical notion

Why Tiling Periodicity

- Natural generalization of the classical notion
- New tool for text compression

Why Tiling Periodicity

- Natural generalization of the classical notion
- New tool for text compression
- New structural properties of texts

  Conjecture: tiling periodicity is not expressible in word equations

Why Tiling Periodicity

- Natural generalization of the classical notion
- New tool for text compression
- New structural properties of texts

  Conjecture: tiling periodicity is not expressible in word equations
- Relations to multidimensional periodicity

Why Tiling Periodicity

- Natural generalization of the classical notion
- New tool for text compression
- New structural properties of texts

  Conjecture: tiling periodicity is not expressible in word equations
- Relations to multidimensional periodicity
- Pattern discovery

Why Tiling Periodicity

- Natural generalization of the classical notion
- New tool for text compression
- New structural properties of texts

  Conjecture: tiling periodicity is not expressible in word equations
- Relations to multidimensional periodicity
- Pattern discovery. At least my talk is in "pattern discovery" session :-)

Partial Order on Tilers

Definition: a tiler S is **smaller** than a tiler Q iff Q can be splitted into several parallel copies of S satisfying the following:

- All defined (visible) letters of S-copies match the visible Q letters
- Every Q letter covered by **exactly one** defined (visible) letter

Partial Order on Tilers

Definition: a tiler S is **smaller** than a tiler Q iff Q can be splitted into several parallel copies of S satisfying the following:

- All defined (visible) letters of S-copies match the visible Q letters
- Every Q letter covered by **exactly one** defined (visible) letter

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

is less than

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

Primitive Tiling Period Conjecture

Conjecture: For every ordinary string there exists a unique primitive tiling period (it is less than any other tiling period).

Reformulation: Any two tiling periods have a common tiling "subperiod"

Primitive Tiling Period Conjecture

Conjecture: For every ordinary string there exists a unique primitive tiling period (it is less than any other tiling period).

Reformulation: Any two tiling periods have a common tiling “subperiod”

Surprisingly, the conjecture is wrong! Look at the (minimal known) counterexample:

A	A	A	A	A	A	A	A	B	A	A	B	B	A	A	B	A	A	A	A	A	A	A	A

and

A	A	A	A	A	A	A	A	B	A	A	B	B	A	A	B	A	A	A	A	A	A	A	A

Properties of Tiling Periodicity

How Many Tiling Periods? (1/2)

Bodini & Rivals (CPM'06) studied number of tilings $L(n)$ of **unary** word of length n:

- $L(1) = 1$, for every $n > 1$
  $L(n) = 1 + \sum_{d|n, d \neq n} L(d)$
- $L(36) = 52$

How Many Tiling Periods? (2/2)

Our result:

Theorem

There is one-to-one correspondence between tiling of unary word of length n and factorizations $n = n_1 \cdot \cdots \cdot n_k$ where $n_2, \ldots, n_k \geq 2$

Tiling Periods Live Inside

Theorem

Take any pair of tiling period and classical period. Then they have a common "tiling subperiod"

Tiling Periods Live Inside

Theorem

Take any pair of tiling period and classical period. Then they have a common "tiling subperiod"

Reformulation

Any primitive tiling period of string T is also a tiling period of any classical period of T

Finding Tiling Periods of Minimal Size

Auxiliary Definition: Multi-Period

A word has **multi-period** (a, b) iff $a|b$ and for every k a $[kb+1, (k+1)b]$ block has the full period a

Auxiliary Definition: Multi-Period

A word has **multi-period** (a, b) iff $a|b$ and for every k a $[kb + 1, (k + 1)b]$ block has the full period a

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

The text above has multi-periods $(1, 2)$ and $(4, 8)$

Auxiliary Definition: Multi-Period

A word has **multi-period** (a, b) iff $a|b$ and for every k a $[kb + 1, (k + 1)b]$ block has the full period a

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

The text above has multi-periods $(1, 2)$ and $(4, 8)$

Definition: multi-period (a, b) is **embedded** into another one a', b' iff $b|a'$

Tiling Period and “his” Multi-Periods

Multi-Period Lemma

Every tiling period corresponds to some sequence of embedded multi-periods $(a_1, b_1) \ldots (a_k, b_k)$. The size of period is equal $n \prod_{i=1}^{k} \frac{a_i}{b_i}$

Preprocessing Step

Preprocessing Lemma

There is $\mathcal{O}(n \log n \log \log n)$ preprocessing of the text such every query "is (a, b) a multi-period" can be answered in $\mathcal{O}(\log n)$ time

Preprocessing Step

Preprocessing Lemma

There is $\mathcal{O}(n \log n \log \log n)$ preprocessing of the text such every query "is (a, b) a multi-period" can be answered in $\mathcal{O}(\log n)$ time

Trick: Karp-Miller-Rosenberg algorithm

Finding Tiling Periods of Minimal Size

Theorem

There is $\mathcal{O}(n \log n \log \log n)$ algorithm for finding a tiling period of minimal size

Conclusions and Future Work

Directions for Further Research

- Whether all primitive tiling periods have the same number of visible letters?
- How often strings are tiling periodic?
- Introduce **not full** tiling periods. How to find the one of minimal size?
- Find natural sources of tiling periodicity

Summary

- New notion: tiling periodicity

A A B B A A B B C C D D C C D D

Summary

- New notion: tiling periodicity

A	A	B	B	A	A	B	B	C	C	D	D	C	C	D	D

- **Result 1:** the primitive tiling period is not necessary unique

Summary

- New notion: tiling periodicity

- **Result 1:** the primitive tiling period is not necessary unique
- **Result 2:** tiling periods live “inside” classical

Summary

- New notion: tiling periodicity

  A A B B A A B B C C D D C C D D

- **Result 1:** the primitive tiling period is not necessary unique
- **Result 2:** tiling periods live “inside” classical
- **Result 3:** there is bijection between tiling periods of unary words and length factorizations

Summary

- New notion: tiling periodicity

A A B B A A B B C C D D C C D D

- **Result 1:** the primitive tiling period is not necessary unique
- **Result 2:** tiling periods live “inside” classical
- **Result 3:** there is bijection between tiling periods of unary words and length factorizations
- **Result 4:** $\mathcal{O}(n \log n \log \log n)$ algorithm for tiling periods of minimal size

Last Slide

Search "**Lifshits**" or visit http://logic.pdmi.ras.ru/~yura/

Juhani Karhumäki, Yury Lifshits and Wojciech Rytter

Tiling Periodicity

CPM'07 // on-line version

Last Slide

Search "**Lifshits**" or visit http://logic.pdmi.ras.ru/~yura/

Juhani Karhumäki, Yury Lifshits and Wojciech Rytter

Tiling Periodicity

CPM'07 // on-line version

Thank you for your attention! Questions?