TUCS

Jetro Vesti

Rich Words and Balanced Words

TURKU CENTRE *for* COMPUTER SCIENCE

TUCS Dissertations
No 213, September 2016

Rich Words and Balanced Words

Jetro Vesti

To be presented, with the permission of the Faculty of Mathematics and Natural Sciences of the University of Turku, for public criticism in Auditorium I/Tauno Nurmela-sali on September 23, 2016, at 12 noon.

University of Turku
Department of Mathematics and Statistics
FI-20014 Turku, Finland

2016

Supervisors

Professor Luca Q. Zamboni
CNRS UMR 5208, Institut Camille Jordan
Université de Lyon Université Lyon 1
43 Boulevard du 11 novembre 1918, 69622 Villeurbanne Cedex
France

Professor Tero Harju
Department of Mathematics and Statistics
University of Turku
FI-20014 Turku
Finland

Reviewers

Doctor Amy Glen
Mathematics & Statistics
School of Engineering & Information Technology
Murdoch University
Murdoch, WA, 6150
Australia

Doctor Sébastien Labbé
Département de Mathématique (B 37)
Université de Liège
Allée de la découverte 12, B-4000 Liège
Belgium

Opponent

Professor Edita Pelantová
Dept. of Mathematics FNSPE
Czech Technical University in Prague
Trojanova 13, Prague 12000
Czech Republic

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

ISBN 978-952-12-3404-0
ISSN 1239-1883

Abstract

This thesis is mostly focused on palindromes. Palindromes have been studied extensively, in recent years, in the field of combinatorics on words. Our main focus is on rich words, also known as full words. These are words which have maximum number of distinct palindromes as factors. We shed some more light on these words and investigate certain restricted problems.

Finite rich words are known to be extendable to infinite rich words. We study more closely how many different ways, and in which situations, rich words can be extended so that they remain rich. The defect of a word is defined to be the number of palindromes the word is lacking. We will generalize the definition of defect with respect to extending the word to be infinite. The number of rich words, on an alphabet of size n, is given an upper and a lower bound.

Hof, Knill and Simon presented (Commun. Math. Phys. 174, 1995) a well-known question whether all palindromic subshifts which are generated by primitive substitutions arise from substitutions which are in class P. Over the years, this question has transformed a bit and is nowadays called the class P conjecture. The main point of the conjecture is to attempt to explain how an infinite word can contain infinitely many palindromes. We will prove a partial result of the conjecture.

Rich square-free words are known to be finite (Pelantová and Starosta, Discrete Math. 313, 2013). We will give another proof for that result. Since they are finite, there exists a longest such word on an n-ary alphabet. We give an upper and a lower bound for the length of that word.

We study also balanced words. Oliver Jenkinson proved (Discrete Math., Alg. and Appl. 1(4), 2009) that if we take the partial sum of the lexicographically ordered orbit of a binary word, then the balanced word gives the least partial sum. The balanced word also gives the largest product. We will show that, at the other extreme, there are the words of the form $0^{q-p}1^p$ (p and q are integers with $1 \leq p < q$), which we call the most unbalanced words. They give the greatest partial sum and the smallest product.

Tiivistelmä

Tässä väitöskirjassa käsitellään pääasiassa palindromeja. Palindromeja on tutkittu viime vuosina runsaasti sanojen kombinatoriikassa. Suurin kiinnostuksen kohde tässä tutkielmassa on rikkaissa sanoissa. Nämä ovat sanoja joissa on maksimaalinen määrä erilaisia palindromeja tekijöinä. Näitä sanoja tutkitaan monesta eri näkökulmasta.

Äärellisiä rikkaita sanoja voidaan tunnetusti jatkaa äärettömiksi rikkaiksi sanoiksi. Työssä tutkitaan tarkemmin sitä, miten monella tavalla ja missä eri tilanteissa rikkaita sanoja voidaan jatkaa siten, että ne pysyvät rikkaina. Sanan vajauksella tarkoitetaan puuttuvien palindromien lukumäärää. Vajauksen käsite yleistetään tapaukseen, jossa sanaa on jatkettava äärettömäksi sanaksi. Rikkaiden sanojen lukumäärälle annetaan myös ylä- ja alaraja.

Hof, Knill ja Simon esittivät kysymyksen (Commun. Math. Phys. 174, 1995), saadaanko kaikki äärettömät sanat joissa on ääretön määrä palindromeja tekijöinä ja jotka ovat primitiivisen morfismin generoimia, morfismeista jotka kuuluvat luokkaan P. Nykyään tätä ongelmaa kutsutaan luokan P konjektuuriksi ja sen tarkoitus on saada selitys sille, millä tavalla äärettömässä sanassa voi olla tekijöinä äärettömän monta palindromia. Osittainen tulos tästä konjektuurista todistetaan.

Rikkaiden neliövapaiden sanojen tiedetään olevan äärellisiä (Pelantová ja Starosta, Discrete Math. 313, 2013). Tälle tulokselle annetaan uudenlainen todistus. Koska kyseiset sanat ovat äärellisiä, voidaan selvittää mikä niistä on pisin. Ylä- ja alaraja annetaan tällaisen pisimmän sanan pituudelle.

Työssä tutkitaan myös tasapainotettuja sanoja. Tasapainotetut sanat antavat pienimmän osittaissumman binäärisille sanoille (Jenkinson, Discrete Math., Alg. and Appl. 1(4), 2009). Lisäksi ne antavat suurimman tulon. Muotoa $0^{q-p}1^p$ (p ja q ovat kokonaislukuja joille $1 \leq p < q$) olevien sanojen todistetaan vastaavasti antavan suurimman osittaissumman ja pienimmän tulon. Ne muodostavat täten toisen ääripään tasapainotetuille sanoille, ja asettavat kaikki muut sanat näiden väliin.

Acknowledgements

First of all, I want to thank Professor Juhani Karhumäki, for taking me in and providing me funding for my doctoral studies. Without him, I may not have had courage to take this journey. I want to express huge gratitude for my supervisors, Professor Luca Zamboni and Professor Tero Harju, who gave me introduction to combinatorics on words, offered many interesting problems to think, gave support and provided me with good insights. Especially, I want to thank Luca for introducing me to rich words, which my thesis mostly consists of. I want to thank my roommates and colleagues, Jarkko Peltomäki, Markus Whiteland, Mikhail Barash and Michal Szabados, for several interesting and fun discussions and helping me with many things I was sometimes struggling with.

I want to thank all my other thesis supervisors also, Iiro Honkala for my master's thesis, Juha Honkala for my bachelor's thesis and Sanna Ranto for my essay in mathematics. I want to thank all my fellow students from the undergraduate phase, especially Toni Ernvall, for taking the first steps in mathematics with me. I want to thank all the lecturers who gave interesting and insightful lectures during my studies, especially Tero Harju, Iiro Honkala, Jarkko Kari, Kari Ylinen, Juhani Karhumäki and Hannu Laakso. I also want to thank my high school mathematics teachers, Jorma Sorvoja and Jarno Kukonlehto.

I want to thank Professor Matti Vuorinen, with whom I wrote my first scientific article as a young summer assistant. Only then I realized for the first time that it is not impossible to do science. This had a huge effect on me.

I also want to give a short moment for all the previous mathematicians who have contributed to this beautiful science and who have brought us this far. And for all those unsung heroes who have been forgotten or have not been appreciated at their time. Especially Galois, Abel and Hypatia, who all died too early and tragically.

I want to thank the reviewers and the opponent of my thesis, for taking the time to do this. I want to thank Turku Centre for Computer Science and Academy of Finland for funding my doctoral studies, and therefore all the hardworking taxpayers also. I want to thank the Department of Mathematics and Statistics, for a great atmosphere and for all the administrative people for taking care of all the practical matters.

I want to thank University of Turku Metal Club TYRMY and all the lovely TYRMY-people for supporting me and for being awesome. You know who you are, since you have been a crucial part of my social life for a long time. I also want to thank all my other close friends over the years, you have been extremely important. I want to thank Proffan Kellari & Teppo and Three Beers, in which parts of this thesis were written.

I want to give special thanks to my whole family. I couldn't have got it any better. You have always been there for me and you gave me the means to strive for life and happiness. This has been a great journey already now.

Finally, I want to thank my parents from everything.

- Turku, March 2016 Jetro Vesti

List of original publications

1. Jetro Vesti, *Extensions of rich words*, Theoretical Computer Science, Vol. 548, pp. 14-24, 2014, DOI 10.1016/j.tcs.2014.06.033.

2. Tero Harju, Jetro Vesti, Luca Q. Zamboni, *On a question of Hof, Knill and Simon on palindromic substitutive systems*, Monatshefte für Mathematik, Vol. 179, pp. 379-388, 2016, DOI 10.1007/s00605-015-0828-2.

3. Jetro Vesti, *Rich square-free words*, preprint, submitted to Theoretical Computer Science 03/2016, available in arXiv.

4. Jetro Vesti, *The most unbalanced words* $0^{q-p}1^p$ *and majorization*, Discrete Mathematics, Algorithms and Applications, Vol. 7, No. 3, 1550028, 22 pages, 2015, DOI 10.1142/S1793830915500287. Revised version of the preprint.

Contents

I Summary **1**
1 Introduction . . . 3
2 Preliminaries . . . 4
3 The structure of the thesis . . . 6
3.1 Extensions of rich words . . . 6
3.2 Class P conjecture . . . 7
3.3 Rich square-free words . . . 8
3.4 The most unbalanced words . . . 9

II Original publications **15**

Part I

Summary

1 Introduction

The subjects discussed in this thesis belong to the area of mathematics called *combinatorics on words*. Combinatorics on words, as an area of its own, can be traced back to the works of Axel Thue on repetition-free words in the early 1900s (see [Th1],[Th2]). This was the first time when words were studied as their own interest. Unfortunately his work was forgotten for a long time, only to be rediscovered later. Combinatorics on words started to become a systematic and more clearly shaped field after the 1950s. The theory evolved and it culminated to the famous book of Lothaire in 1983, see [Lot1]. After this, it has been widely studied and is nowadays recognized as an own area of discrete mathematics in relation to computer science. Much of this research is gathered to the second and third book of Lothaire, see [Lot2] and [Lot3].

This thesis is mostly focused on *palindromes*. Palindrome is a word which is equal to its reversal, for example *aababaa*. In my native language, Finnish, there are a lot of palindromic words and sentences, for example *saippuakauppias* and *neulo taas niin saat oluen*, which mean *soap vendor* and *if you knit again then I will give you a beer*. In combinatorics on words, however, we study combinatorial properties of abstract words. This means the words do not have any semantic meaning in any natural language.

Palindromes have been a topic of wide interest in combinatorics on words since the articles of de Luca [deL] and Droubay and Pirillo [DP], where they studied palindromic factors inside *Sturmian words*. Sturmian words are another important topic in combinatorics on words. They have been known for a long time, in one form or another, but the systematic study of these words can be traced back to Morse and Hedlund [MH] and the year 1940, although Johann III Bernoulli was interested about them already in the 1700s, see [Ber]. They have many important applications outside mathematics, for example the Nobel Prize in Chemistry for 2011 was awarded to Dan Shechtman [SBGC] for the discovery of quasicrystals, which are closely linked to Sturmian words.

When it comes to palindromes, in this thesis we are especially interested in words which have maximum number of palindromes as factors. Droubay, Justin and Pirillo proved in [DJP] that every word w has at most $|w|+1$ many distinct palindromic factors, where $|w|$ is the length of the word. This paper is the foundation for the study of *rich* words, or equivalently *full* words. A word is rich if it has exactly this maximum number of palindromes as factors. This new class of words was defined first by Brlek, Hamel, Nivat and Reutenauer in [BHNR]. The first unified approach to the study of these words as a whole was done by Glen, Justin, Widmer and Zamboni in [GJWZ]. We will prove several new results considering how rich words can be extended so that they remain rich.

Another important article considering palindromes, is a paper [HKS] by Hof, Knill and Simon, where the authors connected infinite words which have infinitely many palindromes as factors to one-dimensional quasicrystals in theoretical physics. From this paper arose a general question and a conjecture how an infinite word can contain infinitely many palindromes. This conjecture has been investigated in several papers, including one in this thesis.

Avoidability of patterns in words has been a core topic in combinatorics on words since Thue, who constructed an infinite overlap-free word on binary alphabet and an infinite square-free word on ternary alphabet. Many problems in mathematics can be stated in the terms of avoiding some certain patterns inside a word, which means pattern avoidance has many natural applications in other areas, for example in Ramsey theory and Burnside problems in algebra. In this thesis, we will discuss square-free words which are also rich.

Besides palindromes, we study a certain problem related to *finite balanced words*. They are closely linked to Sturmian words, since every Sturmian word is balanced and every finite balanced word is a factor of some Sturmian word. Finite balanced words were studied closely by Jenkinson and Zamboni in [JZ]. Jenkinson gave in [Jen] new properties for finite balanced words with respect to majorization. We will prove that the words of the form $0^{q-p}1^p$, where p and q are integers with $1 \leq p < q$, have opposite extremal properties to the finite balanced words.

2 Preliminaries

Next we define some basic terminology used in this thesis. For more comprehensive presentation about the definitions and notation in combinatorics on words, one can look [Lot1] and [Lot2].

An *alphabet* A is a non-empty finite set of symbols, which we call *letters*. A *word* is a finite sequence of letters from A. The *empty* word ϵ is the empty sequence. We denote by A^* the set of all finite words.

An *infinite word* is a sequence indexed by $\mathbb{N}$ with values in A. We denote the set of all infinite words by A^ω and define $A^\infty = A^* \cup A^\omega$. An infinite word is *ultimately periodic* if it is of form $uv^\infty = uvvv\cdots$, where $v \neq \epsilon$. If $u = \epsilon$, then we say the infinite word is *periodic*. An infinite word that is not ultimately periodic is *aperiodic*.

The *length* of a word $w = a_1a_2\dots a_n$ is denoted by $|w| = n$. The empty word ϵ has length 0. A word x is a *factor* of a word w if $w = uxv$. If $u = \epsilon$ (resp. $v = \epsilon$) then we say that x is a *prefix* (resp. *suffix*) of w. We denote by $|w|_a$ the number of occurrences of letter a in w. If there occurs n distinct letters in w then we say that w is *n-ary*.

A factor x of a word w is said to be *unioccurrent* in w if x has exactly one occurrence in w. Two occurrences of factor x are said to be *consecutive* if there is no occurrence of x between them. A factor of w having exactly two occurrences of a non-empty factor u, one as a prefix and the other as a suffix, is called a *complete return* to u in w. An infinite word w is *recurrent* if each factor x of w occurs infinitely many times in w. To each infinite recurrent word w we associate the *subshift* $\Omega(w)$ of all infinite words having the same factors as w.

An *overlapping* word is a word of form uuv, where v is a non-empty prefix of u. A word of form uu, where $u \neq \epsilon$, is called a *square*. A word which does not contain a square or an overlap, is called *square-free* or *overlap-free*, respectively. Generally, a word of form u^r is called r-power if $r \in \mathbb{Q}$ and $|u| \cdot r \in \mathbb{N}$.

The *reversal* of $w = a_1 a_2 \dots a_n$ is defined as $\widetilde{w} = a_n \dots a_2 a_1$. A word w is called a *palindrome* if $w = \widetilde{w}$. The empty word ϵ is assumed to be a palindrome. Let $w = vu$ be a word and u its longest palindromic suffix. The *palindromic closure* of w is defined as $w^{(+)} = vu\widetilde{v}$. An infinite word is *palindromic* if it contains infinitely many distinct palindromes as factors.

We already mentioned that from [DJP] we get that every word w has at most $|w| + 1$ many palindromic factors. Rich words are defined to be the words which achieve this limit.

Definition. *A word w is* rich *if it has exactly $|w| + 1$ distinct palindromic factors, including the empty word. An infinite word is* rich *if all of its factors are rich.*

An important characterization of rich words is the following result from [GJWZ] (Thm. 2.14).

Proposition. *A finite or infinite word w is rich if and only if all complete returns to any palindromic factor in w are themselves palindromes.*

The *defect* of a finite word w, denoted $D(w)$, is defined as $D(w) = |w| + 1 - |\mathrm{Pal}(w)|$, where $\mathrm{Pal}(w)$ is the set of palindromic factors in w. The defect of an infinite word w is defined as $D(w) = \sup\{D(u)|\ u \text{ is a factor of } w\}$. In other words, the defect tells how many palindromes the word is lacking. Rich words are exactly those whose defect is equal to 0.

An infinite word w is *balanced* if for every two factors x, y of w of the same length and for every letter a of w we have $||x|_a - |y|_a| \leq 1$. An infinite binary word is *Sturmian* if it is balanced and aperiodic.

Let w be an infinite word. It is closed under reversal if for every factor u of w we have that $\widetilde{u}$ is also a factor of w. A factor u of w is *right special* in w if there exist two distinct letters a, b of w such that both ua and ub are factors of w. The word w is *episturmian* if it is closed under reversal and has at most one right special factor of each length.

A function $\tau : A \to A^+$ is called a *substitution*. The definition of substitution extends by concatenation to morphisms $A^* \to A^*$ and $A^\omega \to A^\omega$. A substitution $\tau : A \to A^+$ is *primitive* if there exists $n \in \mathbb{N}$ such that $|\tau^n(a)|_b > 0$ for all $a, b \in A$. An infinite word w is a called a *fixed point* of a substitution τ if $\tau(w) = w$, and *pure primitive morphic* if it is a fixed point of some primitive substitution. An infinite word $w \in A^\omega$ is called *primitive morphic* if there exists a pure primitive morphic word $u \in B^\omega$ and a substitution $\tau : B^* \to A^*$ such that $w = \tau(u)$.

Let us set $A = \{0, 1\}$ and define $0 < 1$. The *lexicographic order* on words $u = u_1 \dots u_n$ and $v = v_1 \dots v_n$ in A^n is defined by: $u < v$ if there exists $j \in \{1, \dots, n\}$ such that $u_k = v_k$ for all $k = 1, \dots, j-1$ and $u_j < v_j$. We denote $u \leq v$ if either $u < v$ or $u = v$. The *cyclic shift* $\sigma : A^n \to A^n$ is defined by $\sigma(w_1 \dots w_n) = w_2 \dots w_n w_1$. The *orbit* $\mathcal{O}(w)$ of a word $w \in A^n$ is the vector $\mathcal{O}(w) = (\mathcal{O}_1(w), \dots, \mathcal{O}_n(w))$, where the words $\mathcal{O}_i(w)$ are the iterated cyclic shifts $w, \sigma(w), \dots, \sigma^{n-1}(w)$ arranged in lexicographic order from the smallest to the largest. We will set $(w)_2 = \sum_{i=1}^{n} w_i 2^{n-i}$ for a word $w = w_1 w_2 \dots w_n$ and define the *base-2 orbit* of w by $\mathcal{I}(w) = (\mathcal{I}_1(w), \dots, \mathcal{I}_n(w)) = ((\mathcal{O}_1(w))_2, \dots, (\mathcal{O}_n(w))_2)$.

Let p and q be coprime integers such that $1 \leq p < q$. $\mathcal{W}_{p,q}$ will denote the set of binary words $w \in \{0, 1\}^q$ such that $|w|_1 = p$ and $|w|_0 = q - p$. A finite word $w \in \{0, 1\}^*$ is (cyclically) *balanced* if for every equal-length factors u and v of every cyclic shift of w we have $||u|_1 - |v|_1| \leq 1$. From [BS] we know that there are q balanced words in $\mathcal{W}_{p,q}$ and they are all in the same orbit. We define $\mathbb{W}_{p,q}$ to be the set of all orbits in $\mathcal{W}_{p,q}$. This means there is a unique balanced orbit in each $\mathbb{W}_{p,q}$.

3 The structure of the thesis

The thesis consists of four journal articles and manuscripts. The first three are focused on rich words and palindromes, and the fourth one is related to finite balanced words. In this section, we will introduce the articles and present some results.

3.1 Extensions of rich words

The authors of [GJWZ] initiated the systematic study of rich words, which we desire to continue. They proved several basic results on these words. One of them was that, if w is a rich word then there exist letters x and z, which occur in w, such that wx and zw are rich. This means that every rich word can be extended to be an infinite rich word. However, it does not say whether it can be extended with two or more distinct letters to the right, or to the left.

In [V1], we will prove that if w is a rich word then there exists a word u (over the same alphabet as w) such that wu is rich, $|u| < 2|w|$ and wu can be extended richly with at least two distinct letters. The length of our word u is most definitely not optimal, but this means that eventually we can extend a rich word with two letters. We leave as an open question the shortest length of necessary u. Solving it would give us much insight about the structure of the tree of rich words. We will study also many other problems related to extending a rich word.

In [GJWZ], it was proved that the palindromic closure preserves richness. We will prove that it also preserves the number of distinct rich extensions. Periodic rich infinite words were studied in [BHNR] and [GJWZ]. We will prove that every rich word can be extended to a periodic rich infinite word.

In [V1], we will also study the number of rich words, on an alphabet of size n. For the lower bound, we will use the fact that every rich word can be eventually extended in at least two ways.

The concept of defect has been studied from many points of views. We will generalize this with respect to how much the defect must increase if a word has to be extended to an infinite word. We call this the *infinite defect*. We give several upper bounds for it and point out few properties.

There are several open questions left out in [V1]. Most interesting is the length of the shortest u such that wu can be extended with at least two letters. This would help us to understand the structure of rich words better and also to count how many there are.

3.2 Class P conjecture

Hof, Knill and Simon define in [HKS] the class $\mathcal{P}$ of morphisms $f : A^* \to B^*$ of the form $a \mapsto pq_a$, where p and q_a are palindromes. They present a remark, where they ask if every palindromic subshift generated by a primitive substitution is generated by a substitution in $\mathcal{P}$.

It turns out it is meaningful to replace the class $\mathcal{P}$ of morphisms in the question with a class of morphisms which are conjugate to some morphism in $\mathcal{P}$. We refer to this class of morphisms with $\mathcal{P}'$ and denote by $\mathcal{FP}'$ the set of all infinite words which are fixed by some primitive morphism in class $\mathcal{P}'$.

The original problem of Hof, Knill and Simon has transformed slightly over time and is nowadays known as the class P conjecture.

Conjecture. *If x is a palindromic word fixed by a primitive morphism, then $x \in \mathcal{FP}'$.*

Some partial results of this conjecture have been proved, for example for periodic words in [ABCD] and for binary words in [Tan]. In any case, the conjecture is not generally true. Labbé gave a counterexample to it on a ternary alphabet in [Lab2]. Note that this counterexample is still not a counterexample to the original problem, which was stated for subshifts.

The natural way to see the class P conjecture is that it tries to tell what it takes for a fixed point of primitive substitution to contain infinitely many palindromes. A random primitive substitution does not generally preserve palindromes, which means that a substitution that generates infinitely many palindromes has probably some kind of special structure.

In [HVZ], we make some remarks about the conjecture and prove a partial result of it:

Theorem. *Let y be a primitive morphic word with finite defect. Then there exists a morphism $f \in \mathcal{P}'$ and $x \in \mathcal{FP}'$ such that $y = f(x)$.*

3.3 Rich square-free words

Pelantová and Starosta proved in [PS] that every recurrent word with finite Θ-defect contains infinitely many overlapping factors. Here, Θ is any involutive antimorphism. If Θ is equal to the reversal mapping, which is also an involutive antimorphism, then Θ-defect is equal to the defect. This means that a corollary of the mentioned result in [PS] is that every infinite rich word contains a square.

Since every infinite rich word contains a square, we get that all rich square-free words are finite. This means we can look for a longest one. The length of a longest rich square-free word, on an alphabet of size n, was denoted by $r(n)$ in [PS]. The exact formula for $r(n)$ was left as an open problem.

In [V3], we will give a recursive construction for rich square-free words for every size of the alphabet. The lengths of these words trivially give us a lower bound for $r(n)$. We make a conjecture that the exact formula of $r(n)$ can be achieved using these words. We will also give an upper bound for $r(n)$. Totally, we prove:

Proposition. $2,008^n \approx 1068^{\frac{n}{10}} \leq r(n) \leq \sqrt{5}^n \approx 2,237^n$, *for* $n \geq 5$.

This square-freeness is related to a more general problem about repetitions in words. The *repetition threshold*, on an alphabet of size n, is the smallest number r such that there exists an infinite word which avoids greater than r-powers. We denote this number by $RT(n)$. Dejean gave in [Dej] a famous conjecture about this number, which have now been proven (see [Rao]).

We note that the repetition threshold can be studied also for a limited class of infinite words. For example, the *episturmian repetition threshold* $ERT(n)$ is the smallest number r such that there exists an episturmian word which avoids greater than r-powers. Similarly, we will define the *rich* *repetition threshold* $RRT(n)$. Since episturmian words are known to be rich by [DJP], we get from [PS] that $RRT(n), ERT(n) \geq 2$. The exact values of these numbers are left as an open question.

Open problem. *Determine the repetition threshold for episturmian words and for rich words, on an alphabet of size* n*.*

3.4 The most unbalanced words

Jenkinson studied finite balanced words in [Jen] with respect to majorization. Majorization has several applications and will come up in several areas of mathematics, especially in probability, statistics and graph theory. He proved that these words have the least partial sum, with respect to majorization, and the largest product, among all words in the same set of orbits.

For $w, w' \in \mathbb{W}_{p,q}$ the base-2 orbit $\mathcal{I}(w)$ of w is said to *majorize* the base-2 orbit $\mathcal{I}(w')$ of w', denoted $w' \prec w$, if

$$\sum_{k=1}^{i} \mathcal{I}_k(w') \geq \sum_{k=1}^{i} \mathcal{I}_k(w) \quad \text{for } 1 \leq i \leq q.$$

We denote the partial sums of the orbit of w by $\mathcal{S}_i(w) = \sum_{k=1}^{i} \mathcal{I}_k(w)$. To be exact, he proved:

Theorem. *Let* b *be the balanced orbit in* $\mathbb{W}_{p,q}$*. For any* $w \in \mathbb{W}_{p,q}$ *we have*

$$\mathcal{S}_i(b) \geq \mathcal{S}_i(w) \quad \text{for all } 1 \leq i \leq q.$$

Theorem. *For* $w \in \mathbb{W}_{p,q}$ *the product* $P(w) = \prod_{i=1}^{q} \mathcal{I}_i(w)$ *is maximized precisely when* w *is balanced.*

In [V2], we will study words of the form $0^{q-p}1^p$, where p and q are integers with $1 \leq p < q$. We will notice that these words have exactly the opposite extremal properties to the finite balanced words. This is the reason we call them *the most unbalanced words*.

We prove that the most unbalanced word is the greatest element in $\mathbb{W}_{p,q}$, with respect to partial sum. We also prove that the product of the most unbalanced word is the smallest among words in $\mathbb{W}_{p,q}$, if $p < q - p$. These results places every other word in $\mathbb{W}_{p,q}$ between these two extremal words, with respect to these properties.

Bibliography

[ABCD] J.-P. Allouche, M. Baake, J. Cassaigne, D. Damanik: *Palindrome complexity*, Theoret. Comput. Sci., 292(1) (2003), pp. 9-31.

[AFMP] P. Ambrož, C. Frougny, Z. Masáková, E. Pelantová: *Palindromic complexity of infinite words associated with simple Parry numbers*, Ann. Inst. Fourier (Grenoble) 56 (2006), pp. 2131-2160.

[BPS1] L. Balková, E. Pelantová, Š. Starosta: *Infinite words with finite defect*, Adv. in Appl. Math., 47(3) (2011), pp. 562-574.

[BPS2] L. Balková, E. Pelantová, Š. Starosta: *Proof of the Brlek-Reutenauer conjecture*, Theoret. Comput. Sci. 475 (2013), pp. 120-125.

[BPS3] L. Balková, E. Pelantová, Š. Starosta: Private communication (2013).

[BGT] J. M. Barbaroux, F. Germinet, S. Tcheremchantsev: *Fractal dimensions and the phenomenon of intermittency in quantum dynamics*, Duke Math. J., 110 (2001), pp. 161-193.

[Ber] J. Bernoulli III: *Sur une nouvelle espece de calcul*, Recueil pour les Astronomes, vol. 1, Berlin (1772), pp. 255-284.

[BS] J. Berstel, P. Séébold: *Sturmian words,* in: M. Lothaire (Ed.), Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, UK, 2002.

[BV] V. Berthe, L. Vuillon: *Tilings and rotations: A two-dimensional generalization of Sturmian sequences*, Discrete Math. 223 (2000), pp. 27-53.

[BG] A. Bovier, J.-M. Ghez: *Spectral properties of one-dimensional Schrödinger operators with potentials generated by substitutions,* Commun. Math. Phys. 158 (1993), pp. 45-66.

[BHNR] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer: *On the palindromic complexity of infinite words*, Internat. J. Found. Comput. 15 (2004), pp. 293-306.

[BR] S. Brlek, C. Reutenauer: *Complexity and palindromic defect of infinite words*, Theoret. Comput. Sci. 412 (2011), pp. 493-497.

[BDGZ1] M. Bucci, A. De Luca, A. Glen, L. Q. Zamboni: *A connection between palindromic and factor complexity using return words*, Advances in Applied Mathematics 42 (2009), pp. 60-74.

[BDGZ2] M. Bucci, A. De Luca, A. Glen, L. Q. Zamboni: *A new characteristic property of rich words*, Theoret. Comput. Sci. 410 (2009), pp. 2860-2863.

[CD] A. Carpi, A. de Luca: *Special factors, periodicity, and an application to Sturmian words*, Acta Informatica 36 (2000), pp. 983-1006.

[CW] O. Carton, T. Wolfgang: *The monadic theory of morphic infinite words and generalizations,* Inf. Comput. 176(1) (2002), pp. 51-65.

[CR] J. D. Currie, N. Rampersad: *A proof of Dejeans conjecture*, Math. Comput. 80(274) (2011), pp. 1063-1070.

[Da] D. Damanik: *Singular continuous spectrum for a class of substitution Hamiltonians,* Lett. Math. Phys. 46 (1998), pp. 303-311.

[DKL] D. Damanik, R. Killip, D. Lenz: *Uniform spectral properties of one-dimensional quasicrystals, III. α-Continuity,* Commun. Math. Phys. 212 (2000), pp. 191-204.

[DZ] D. Damanik, D. Zare: *Palindrome complexity bounds for primitive substitution sequences,* Dicret. Math. 222 (2000), pp. 259-267.

[Dej] F. Dejean: *Sur un théorème de Thue*, J. Combin. Theory Ser. A 13 (1972), pp. 90-99.

[deL] A. de Luca: *Sturmian words: structure, combinatorics and their arithmetics,* Theoret. Comput. Sci. 183 (1997), pp. 45-82.

[DGZ] A. de Luca, A. Glen, L. Q. Zamboni: *Rich, Sturmian, and trapezoidal words*, Theoret. Comput. Sci. 407 (2008), pp. 569-573.

[DJP] X. Droubay, J. Justin, G. Pirillo: *Episturmian words and some constructions of de Luca and Rauzy*, Theoret. Comput. Sci. 255 (2001), pp. 539-553.

[DP] X. Droubay, G. Pirillo: *Palindromes and Sturmian words,* Theoret. Comput. Sci. 223 (1999), pp. 73-85.

[Du] F. Durand: *A characterization of substitutive sequences using return words*, Discrete Math. 179 (1998), pp. 89-101.

[EKM] C. Epifanio, M. Koskas, F. Mignosi: *On a conjecture on bidimensional words*, Theoret. Comput. Sci. 299 (1-3) (2003), pp. 123-150.

[GJ] A. Glen, J. Justin: *Episturmian words: A survey*, Theor. Inform. Appl. 43(3) (2009), pp. 403-442.

[GJWZ] A. Glen, J. Justin, S. Widmer, L. Q. Zamboni: *Palindromic richness*, European Journal of Combinatorics 30 (2009), pp. 510-531.

[Ha] M. Hall, Jr.: *Generators and Relations in Groups - The Burnside Problem,* Lectures on Modern Mathematics, Vol. ii (1964), pp. 42-92.

[HVZ] T. Harju, J. Vesti, L. Q. Zamboni: *On a question of Hof, Knill and Simon on palindromic substitutive systems,* Monatshefte für Mathematik, Vol. 179 (2016), pp. 379-388.

[HPS] K. Hare, H. Prodinger, J. Shallit: *Three Series for the Generalized Golden Mean*, Fibonacci Quart. 52 (2014), pp. 307-313.

[HKS] A. Hof, O. Knill, B. Simon: *Singular continuous spectrum for palindromic Schrödinger operators*, Commun. Math. Phys. 174 (1995), pp. 149-159.

[Jen] O. Jenkinson: *Balanced words and majorization,* Discrete Math., Alg. and Appl. 1(4) (2009), pp. 463-484.

[JZ] O. Jenkinson, L. Q. Zamboni: *Characterisations of balanced words via orderings,* Theoret. Comput. Sci. 310 (1-3) (2004), pp. 247-271.

[JS] S. Jitomirskaya, B. Simon: *Operators with singular continuous spectrum: III. Almost periodic Schrödinger operators,* Commun. Math. Phys. 165 (1994), pp. 201-205.

[Lab1] S. Labbé: *Propriétés combinatoires des f-palindromes,* Master's thesis, Université du Québec Montrèal, Montrèal, 2008, M10615.

[Lab2] S. Labbé: *A counterexample to a question of Hof, Knill and Simon,* Electron. J. Comb. 21(3) (2014), paper 3.11, 12 pages.

[Lot1] M. Lothaire: *Combinatorics on Words,* in: Encyclopedia of Mathematics and its Applications, vol 17, Addison-Wesley, Reading, MA, 1983.

[Lot2] M. Lothaire: *Algebraic Combinatorics on Words,* in: Encyclopedia of Mathematics and its Applications, vol 90, Cambridge University Press, UK, 2002.

[Lot3] M. Lothaire: *Applied Combinatorics on Words,* Cambridge University Press, Cambridge 2005.

[MP] F. Mignosi, G. Pirillo: *Repetitions in the Fibonacci infinite word*, Informatique théorique et applications, tome 26, n. 3 (1992), pp. 199-204.

[MH] M. Morse, G. A. Hedlund: *Symbolic Dynamics II. Sturmian Trajectories,* American Journal of Mathematics 62(1) (1940), pp. 1-42.

[PS] E. Pelantová, Š. Starosta: *Languages invariant under more symmetries: overlapping factors versus palindromic richness*, Discrete Math. 313 (2013), pp. 2432-2445.

[PA] S. A. Puzynina, S. V. Avgustinovich: *On periodicity of two-dimensional words*, Theoret. Comput. Sci. 391 (2008), pp. 178-187.

[QZ] A. Quas, L. Q. Zamboni: *Periodicity and local complexity*, Theoret. Comput. Sci. 319 (1-3) (2004), pp. 229-240.

[Que] M. Queffélec: *Substitution Dynamical Systems - Spectral Analysis,* Lecture Notes in Mathematics. 2nd ed. 2010. Springer-Verlag, Berlin.

[Rao] M. Rao: *Last cases of Dejeans conjecture*, Theoret. Comput. Sci. 412(27) (2011), pp. 3010-3018.

[RR] A. Restivo, G. Rosone: *Burrows-Wheeler transform and palindromic richness*, Theoret. Comput. Sci. 410 (2009), pp. 3018-3026.

[SBGC] D. Shechtman, I. Blech, D. Gratias, J. W. Cahn: *Metallic phase with long range orientational order and no translation symmetry,* Physical Review Letters 53(20) (1984), pp. 1951-1954.

[Tan] B. Tan: *Mirror substitutions and palindromic sequences,* Theoret. Comput. Sci. 389(1-2) (2007), pp. 118-124.

[Th1] A. Thue: *Uber unendliche Zeichenvihen*, Kra. Vidensk. Selsk. Skrifter. I Mat.-Nat.Kl., Christiana, Nr. 7, 1906.

[Th2] A. Thue: *Uber die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*, Kra. Vidensk. Selsk. Skrifter. I Mat.-Nat.Kl., Christiana, Nr. 12, 1912.

[V1] J. Vesti: *Extensions of rich words,* Theoret. Comput. Sci. 548 (2014), pp. 14-24.

[V2] J. Vesti: *The most unbalanced words* $0^{q-p}1^p$ *and majorization,* Discrete Mathematics, Algorithms and Applications, Vol. 7, No. 3, 1550028 (2015), 22 pages.

[V3] J. Vesti: *Rich square-free words,* preprint, submitted to Theoretical Computer Science 03/2016, available in arXiv.

Part II

Original publications

Publication I

Jetro Vesti, *Extensions of rich words*, Theoretical Computer Science, Vol. 548, pp. 14-24, 2014, DOI 10.1016/j.tcs.2014.06.033.

The webpage of the article:
http://www.sciencedirect.com/science/article/pii/S0304397514004800

Theoretical Computer Science 548 (2014) 14–24

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

Extensions of rich words

Jetro Vesti

University of Turku, Department of Mathematics and Statistics, 20014 Turku, Finland

ARTICLE INFO

Article history:
Received 18 December 2013
Received in revised form 11 June 2014
Accepted 25 June 2014
Available online 1 July 2014
Communicated by M. Crochemore

Keywords:
Combinatorics on words
Palindromes
Rich words
Sturmian words
Defect
Two-dimensional words

ABSTRACT

A word w is *rich* if it has $|w|+1$ many distinct palindromic factors, including the empty word. This article contains several results about rich words, particularly related to extending them. A word w can be *eventually extended richly* in n ways if there exists a finite word u and n distinct letters $a \in \mathrm{Alph}(w)$ such that wua is rich. We will prove that every (non-unary) rich word can be eventually extended richly in at least two different ways, but not always in three or more ways. We will also prove that every rich word can be extended to both periodic and aperiodic infinite rich words.

The *defect* of a finite word w is defined by $\mathrm{D}(w) = |w|+1-|\mathrm{Pal}(w)|$. This concept has been studied in various papers. Here, we will define a new concept, *infinite defect*. For a finite word w the definition is $\mathrm{D}_\infty(w) = \min\{\mathrm{D}(z) \mid z$ is an infinite word which has factor $w\}$. We will show that the infinite defect of a finite word is always finite and give some upper bounds for it. The difference between defect and infinite defect is also investigated.

We will also give an upper and a lower bound for the number of rich words. A new class of words, two-dimensional rich words, is also introduced.

1. Introduction

In [11], it was proved that every word w has at most $|w|+1$ many distinct palindromic factors, including the empty word. The class of words which achieve this limit was introduced in [6] with the term *full* words. The authors of [13] studied these words thoroughly and named them *rich* (in palindromes). This class of words has been studied in several other papers from various points of view, for example in [1,8–10] and [18].

In Section 2 we will prove several results about extending rich words. A rich word w can be *extended richly* with a word $u \in \mathrm{Alph}(w)^+$ if wu is rich. In [13] it was proved that every rich word can be extended with at least one letter. We will prove that every rich word w can be extended richly with at least two different letters, after it has been extended with a word of length at most $2|w|$. This fact will be used in several places. We will also show that every rich word can be extended to both an infinite aperiodic and infinite periodic rich word. Also, all Sturmian words can be extended richly in two ways.

In Section 3 we will define a new concept, the *infinite defect* of a finite word. The *defect* of a finite word w is defined by $\mathrm{D}(w) = |w|+1-|\mathrm{Pal}(w)|$. We can also study how many defects a finite word must have if it has to be extended to an infinite word. Hence, we define the *infinite defect* of a finite word w with $\mathrm{D}_\infty(w) = \min\{\mathrm{D}(z) \mid z$ is an infinite word which has factor $w\}$, where we suppose $\mathrm{Alph}(z) \subseteq \mathrm{Alph}(w)$. We will show that this number is always finite and give some upper bounds for it. We will also study how the defect and the infinite defect can differ.

E-mail address: jejove@utu.fi.

http://dx.doi.org/10.1016/j.tcs.2014.06.033

In Section 4 we will give upper and lower bounds for the number of rich words of length n. For the lower bound we will use the fact that every rich word can be extended in at least two different ways after a limited extension. There have been no previous studies investigating the number of these words.

In Section 5 we will shortly introduce and study two-dimensional rich words and their extensions.

In Section 6 we will give some open problems from the previous sections.

1.1. Definitions and notation

An *alphabet* A is a non-empty finite set of symbols, called *letters*. A *word* is a finite sequence of letters from A. The *empty* word ϵ is the empty sequence. The set A^* of all finite words over A is a *free monoid* under the operation of concatenation. The *free semigroup* $A^+ = A^* \setminus \{\epsilon\}$ is the set of non-empty words over A.

A *right* (resp. *left*) *infinite word* is a sequence indexed by $\mathbb{Z}_+$ (resp. $\mathbb{Z}_-$) with values in A. A *two-way infinite word* is a sequence indexed by $\mathbb{Z}$. We denote the set of all infinite words over A by A^ω and define $A^\infty = A^* \cup A^\omega$. A right infinite word is *ultimately periodic* if it can be written as $uv^\infty = uvvv\cdots$, for some $u, v \in A^*$, $v \neq \epsilon$. If $u = \epsilon$, then we say the infinite word is *periodic*. An infinite word that is not ultimately periodic is *aperiodic*.

The *length* of a word $w = a_1a_2\ldots a_n \in A^+$, with each $a_i \in A$, is denoted by $|w| = n$. The empty word ϵ is the unique word of length 0. By $|w|_a$ we denote the number of occurrences of a letter a in w. The *reversal* of w is denoted by $\tilde{w} = a_n \ldots a_2a_1$. A word w is called a *palindrome* if $w = \tilde{w}$. The empty word ϵ is assumed to be a palindrome.

A word x is a *factor* of a word $w \in A^\infty$ if $w = uxv$, for some $u, v \in A^\infty$. If $u = \epsilon$ $(v = \epsilon)$ then we say that x is a *prefix* (resp. *suffix*) of w. A factor x of a word w is said to be *unioccurrent* in w if x has exactly one occurrence in w. Two occurrences of factor x are said to be *consecutive* if there is no occurrence of x between them. A factor of w having exactly two occurrences of a non-empty factor u, one as a prefix and the other as a suffix, is called a *complete return* to u in w.

If $w = uv \in A^+$, we use the notation $u^{-1}w = v$ or $wv^{-1} = u$ to mean the removal of a prefix or a suffix of w. The *right* (resp. *left*) *palindromic closure* of a word w is the unique shortest palindrome $w^{(+)}$ (resp. ${}^{(+)}w$) having w as a prefix (resp. suffix). If u is the (unique) longest palindromic suffix of $w = vu$ then $w^{(+)} = vu\tilde{v}$.

Let w be a finite or infinite word. The set $\mathrm{F}(w)$ is the set of all factors of w, the set $\mathrm{Alph}(w)$ is the set of all letters that occur in w and the set $\mathrm{Pal}(w)$ is the set of all palindromic factors of w. We say that a word w is *unary* if $|\mathrm{Alph}(w)| = 1$, *binary* if $|\mathrm{Alph}(w)| = 2$, *ternary* if $|\mathrm{Alph}(w)| = 3$ and *n-ary* if $|\mathrm{Alph}(w)| = n$

Other basic definitions and notation in combinatorics on words can be found from Lothaire's books [14] and [15].

1.2. Basic properties of rich words

In this subsection we provide some basic definitions and state some already known properties and characterizations of rich words.

Proposition 1.1. *(See [11, Proposition 2].) A word w has at most $|w| + 1$ distinct palindromic factors, including the empty word.*

Definition 1.2. A word w is *rich* if it has exactly $|w| + 1$ distinct palindromic factors.

Definition 1.3. An infinite word is *rich* if all of its factors are rich.

Proposition 1.4. *(See [13, Corollary 2.5].) A word w is rich if and only if all of its factors are rich.*

Proposition 1.5. *(See [13, Corollary 2.5].) If w is rich, then it has exactly one unioccurrent longest palindromic suffix (referred to later as* lps *or* $\mathrm{lps}(w)$*).*

From Corollary 2.5 in [13] we also get that if w is rich then $\tilde{w}$ is rich. From this we see that the above proposition holds for prefixes also and we refer to the unioccurrent longest palindromic prefix of w as lpp or $\mathrm{lpp}(w)$.

Proposition 1.6. *(See [13, Theorem 2.14].) A finite or infinite word w is rich if and only if all complete returns to any palindromic factor in w are themselves palindromes.*

Proposition 1.7. *(See [13, Proposition 2.8].) Suppose w is a rich word. Then there exist letters $x, z \in \mathrm{Alph}(w)$ such that wx and zw are rich.*

Proposition 1.8. *(See [13, Proposition 2.6].) Palindromic closure preserves richness.*

Let w be a word and $u \neq w$ its longest proper palindromic suffix. The *proper palindromic closure* of $w = vu$ is defined as $w^{(++)} = vu\tilde{v}$. From the proof of Proposition 2.8 in [13] we get that also the proper palindromic closure preserves richness using the fact that the longest proper palindromic suffix (referred to later as $\mathrm{lpps}(w)$ or lpps) can occur only in the beginning and the end of the word. Hence, we create a new palindrome in every step when we are taking the proper palindromic closure and the word stays rich.

Proposition 1.9. *(See [13, proof of Proposition 2.8].) The proper palindromic closure preserves richness.*

2. Extensions of rich words

We say that a finite rich word w can be *extended richly* with a word $u \in \mathrm{Alph}(w)^+$ if wu is rich. The word wu is called a *rich extension* of w with the word u. We also say that w can be extended richly in n ways if there exists n distinct letters $a \in \mathrm{Alph}(w)$ such that wa is rich. The word w can be *eventually extended richly* in n ways if there exists a finite word u such that the word wu is rich and can be extended richly in n ways.

Theorem 2.1. *Let w be a non-unary rich word. There exists a word u such that wu is rich, $|u| < 2|w|$ and wu can be extended richly in at least two ways.*

Proof. The idea of the proof is to take the largest power a^n of any letter a in w and then extend w richly with a word u such that a^n is a suffix of wu. We choose u so that wu does not have a factor a^{n+1}. After this, we can extend wu richly with the letter a and with the letter that is before the $\mathrm{lpps}(wu)$, which we denote by b. The extension with a gives us a new palindrome a^{n+1} and the extension with b gives us a new palindrome $b\,\mathrm{lpps}(wu)b$. The letter b is different than a because otherwise there would be a power a^{n+1} in wu.

Now we only have to show that we actually can extend w richly such that a^n is a suffix. Suppose $w = zv$, where $v = \mathrm{lpps}(w)$. Now $w^{(++)} = zv\tilde{z}$ is rich because of Proposition 1.9. If a^n is a factor of z then we can cut the word $w^{(++)} = w\tilde{z}$ inside $\tilde{z}$ such that it has a^n as a suffix. If a^n is a suffix of v, and hence also a prefix of v, then we are already done. The remaining case is where a^n is as a factor only inside v.

For this case we take the proper palindromic closure of $w^{(++)} = zv\tilde{z}$. If $x = \mathrm{lpps}(w^{(++)})$ would be longer than $v\tilde{z}$ then it would induce another occurrence of v inside of $v\tilde{z}$ (not as a suffix) and hence also inside of $zv = w$. This is impossible because $v = \mathrm{lpps}(w)$. So x cannot be longer than $v\tilde{z}$, which gives us two subcases: 1) $v\tilde{z} = v'x$, where v' is a prefix of $v = v'v''$, and 2) $\tilde{z} = z'x$.

Case 1. Now the word $zv\tilde{z}\tilde{v'}$ is rich and if a^n is a factor of v' then we can cut the word $zv\tilde{z}\tilde{v'}$ inside $\tilde{v'}$ such that a^n is a suffix. If a^n is not a factor of v' then it has to be a factor of v'', which is impossible. This comes from the fact that $v'v''\tilde{z}$ would be an overlap of two palindromes v and x, where v'' is the common part which contains a^n but the other parts v' and $\tilde{z}$ would not contain a^n.

Case 2. Now the word $zvz'x\tilde{z'}v$ is rich and because v contains a^n as a factor we can cut the word $zvz'x\tilde{z'}v$ inside the latter v such that a^n is a suffix.

Clearly, in each case we constructed the rich extension wu such that $|u| < 2|w|$. □

Example 2.2. Let us apply the idea from the previous theorem to the following rich word $w = 11011010101010110011000111000011100$. The word can be extended richly with only letter 0 and the largest powers of both letters 0 and 1 are inside the lpps 00111000011100. We take the proper palindromic closure $w^{(++)} = w011001101010101011011 = wz'x$, where the new lpps is $x = 11011$. Now we get 111 as a suffix: $(w^{(++)})^{(++)} = wz'x\tilde{z'}00111\cdots$ and we can extend the word richly with both 0 and 1.

Remark 2.3. The original word in the previous example has length 35 and the new word, up to the point where we can extend it in two ways, has length 77. We can increase the ratio $77/35 = 2.2$ ultimately close to 3 by making the block 010101010 longer. So the bound from the previous theorem can be reached if we use the idea of the proof on how to extend the word richly. However, the word $w0$ can already be extended richly in two ways. This implies that the bound $|u| < 2|w|$ from Theorem 2.1 can be improved extensively.

Remark 2.4. We can construct words for which the number of consecutive unique rich extensions grows arbitrarily large, i.e. in every step we extend the word we can choose only from one letter. For example, the rich word 0101101110111101111001 has to be extended four times with the letter 1 before we can extend it with both letters 0 and 1. For n unique extensions the general version of the word is $(\prod_{k=1}^{n} 01^k)001$. Notice that the length of the word is $\frac{(n+1)n}{2} + 3$, which means it grows rapidly.

We also have words for which the consecutive unique rich extensions are not made with the same letter. For example, the word 1010010011000110010 has to be extended first with the letter 0 and then with the letter 1.

The next proposition will be used to prove Propositions 2.6 and 2.7. It gives us a necessary condition whether two rich words can appear in a same rich word.

Proposition 2.5. *Two rich words u and v cannot be factors of the same rich word if there are words $u' \in \mathrm{F}(u)$ and $v' \in \mathrm{F}(v)$ such that $\mathrm{lps}(u') = \mathrm{lps}(v')$, $\mathrm{lpp}(u') = \mathrm{lpp}(v')$ and $u' \neq v'$.*

Proof. From Theorem 6 in [9] we understand immediately that if a word would contain such words as u and v as factors, and hence u' and v', then it would not be rich. □

Proposition 2.6. *Not every ternary rich word can be eventually extended richly in 3 ways.*

Proof. The word $w = 0020102202$ is rich. Let us prove that it cannot be eventually extended richly in 3 ways. First, we give some forbidden factors that can never appear in any rich extension of w, by using Proposition 2.5.

The factors 12, 21, 001 and 0202 are forbidden because w has factors 102, 201, 00201 and 020102202, respectively. Suppose some rich extension of w has factor 00. Then we could take the first occurrence of 00 and get that the rich extension has factor $\tilde{w} = 2022010200$. This is because the complete return to 00 has to be a palindrome. Thus, we would have factors 22010200 and 10200, which means that the factors 2200 and 100 are also forbidden.

Now suppose the contrary: there exists a rich extension wu such that $wu0$, $wu1$ and $wu2$ are rich. The last letter of wu has to be 0 because otherwise we would have factors 12 or 21. Now we have three cases depending on which is the second last letter of wu.

1) Suppose $wu = x00$. This would give a forbidden factor 001 in $wu1 = x001$.

2) Suppose $wu = x10$. This would give a forbidden factor 100 in $wu0 = x100$.

3) Suppose $wu = x20$. If the third last letter is 0 then we would get a factor 0202 in $wu2$. If the third last letter is 1 then we would get a factor 12 already in wu. If the third last letter is 2 then we would get a factor 2200 in $wu0$. These are all forbidden factors.

In each case we get a contradiction and the proof is complete. □

Proposition 2.7. *For every $n \geq 3$ there exists an n-ary rich word which cannot be eventually extended richly in n ways.*

Proof. The case $n = 3$ follows from the previous proposition. For $n \geq 4$ we take the word $w = 123 \cdots (n-1)n$.

Similar to the previous proof, factors $13, 31, 24, 42, \ldots, (n-2)n$ and $n(n-2)$ can never appear on any rich extension of w. Suppose to the contrary that wu is such that we can extend it richly in n ways, i.e. with all the letters $1, 2, \ldots, n$. Now, for every last letter of wu we would always get one of the forbidden factors listed above. □

Remark 2.8. For every $n \geq 1$ there also exists an n-ary rich word which can be extended richly in n ways. The word 5010201030102010401020103010201 0 is an example for $n = 6$. You can construct a generalized word of this by starting from a one-letter word. Subsequently, you repeatedly introduce a new letter, adding it between every letter and also to the end and the beginning of the word. After you have $n - 1$ letters you just add the last new letter to the beginning.

Remark 2.9. From the proof of the previous proposition we also see that if a word has factors of the form $0^k 1^l 2^m 3^n$, where $k, l, m, n \geq 1$, then we can never extend the word richly with all the letters 0, 1, 2 and 3.

Proposition 2.10. *Every non-unary rich word w can be extended to an infinite aperiodic rich word.*

Proof. We can construct such a word for every rich w by repeating the procedure of Theorem 2.1 infinitely many times. We first choose the largest power a^n of some letter a in w. Then we extend the word such that this largest power is a suffix. Then we richly extend the resulting word with the letter a and we get a new palindrome a^{n+1}. We further extend that word with the letter that is before the factor a^{n+1}. When we repeat this procedure to the resulting word we always get larger powers of this letter. This kind of word is clearly aperiodic and rich. □

Proposition 2.11. *Every rich word w can be extended to an infinite periodic rich word.*

Proof. From Proposition 1.9 we get that $w^{(++)} = uv$ is a rich palindrome, where v denotes the lpps of $w^{(++)}$. We prove by induction that $w^{(++)^n} = u^n v$, where $w^{(++)^n}$ means taking the proper palindromic closure n times in a row. It holds for $n = 1$. Suppose it holds for $n = k$.

Now we have to prove that $w^{(++)^{k+1}} = u^{k+1}v$. When we use the assumption that the claim is true for $n = k$, we get $w^{(++)^{k+1}} = (u^k v)^{(++)} = (v(\tilde{u})^k)^{(++)} = (\tilde{u}v(\tilde{u})^{k-1})^{(++)}$. Because $v(\tilde{u})^{k-1} = u^{k-1}v$ is a palindrome, we get that $\mathrm{lpps}(\tilde{u}v(\tilde{u})^{k-1}) = v(\tilde{u})^{k-1}$. Otherwise $v(\tilde{u})^{k-1}$ would occur somewhere else than at the end or the beginning of the word $\tilde{u}v(\tilde{u})^{k-1}$ and hence v would not be the lpps of $w^{(++)}$. This all means that $w^{(++)^{k+1}} = (\tilde{u}v(\tilde{u})^{k-1})^{(++)} = \tilde{u}v(\tilde{u})^{k-1}\tilde{u} = u^{k+1}v$.

We get that u^∞ is an infinite periodic rich word which has w as a prefix. □

Finite rich words can always be extended richly with some letter by Proposition 1.7. The next proposition shows that the same also holds for infinite rich words. This means that left and right infinite rich words can be extended richly to two-way infinite words.

Proposition 2.12. *For every right infinite rich word w there exists a letter $a \in \mathrm{Alph}(w)$ such that aw is rich.*

Proof. We suppose to the contrary, that for every $a \in \mathrm{Alph}(w)$ the word aw is not rich. From the definition of a rich infinite word we get that for every $a \in \mathrm{Alph}(w)$ there exists a word u_a such that u_a is a non-rich prefix of aw. Suppose that u_b is longest of them. Now we would get that $b^{-1}u_b$ is a rich word and cannot be extended richly to the left with any letter. This is a contradiction because of Proposition 1.7. □

The next proposition shows that taking the palindromic closure of the reverse of a rich word, i.e. the left palindromic closure, preserves the number of rich extensions that the original word had.

Proposition 2.13. *If w is rich and can be extended richly in n ways, then so can be $\tilde{w}^{(+)}$.*

Proof. Let w be a rich word and $\tilde{w}^{(+)} = uw = \tilde{w}\tilde{u} = uv\tilde{u}$, where $v = \mathrm{lpp}(w)$. If $u = \epsilon$ then w is a palindrome and the claim is true, so suppose $u \neq \epsilon$. We suppose that wa is rich and wb is non-rich, for letters $a, b \in \mathrm{Alph}(w)$. Now we only need to prove that 1) $\tilde{w}^{(+)}a$ is rich and 2) $\tilde{w}^{(+)}b$ is non-rich.

1) Suppose that p is the lps of wa, which means it is unioccurrent in wa. We will prove that p is also unioccurrent lps of $\tilde{w}^{(+)}a = uwa = uv\tilde{u}a$. Clearly p cannot occur inside w nor $\tilde{w}$, otherwise it would not be unioccurrent in wa. So if p would not be unioccurrent in $\tilde{w}^{(+)}a$ then it has to contain v.

Suppose that $p = xvy$, where $xy \neq \epsilon$. If $p = wa = v\tilde{u}a$, then another occurrence of p in $\tilde{w}^{(+)}a = uwa = \tilde{w}\tilde{u}a$ would imply another occurrence of v in w ($u \neq \epsilon$). If $p \neq wa = v\tilde{u}a$, then the occurrence of p in the end of wa would again directly imply another occurrence of v in w ($xy \neq \epsilon$). Suppose that $xy = \epsilon$, i.e. $p = v$. This would directly imply that $wa = v\tilde{u}a = p\tilde{u}a$ has two occurrences of p. So in every case we get a contradiction.

2) The lps of $\tilde{w}^{(+)}b$ clearly cannot be wb, otherwise wb would be a palindrome and hence rich. The lps of $\tilde{w}^{(+)}b$ cannot be strictly longer than $wb = v\tilde{u}b$ because then v would occur at least twice in w, which is impossible because $v = \mathrm{lpp}(w)$. If the lps of $\tilde{w}^{(+)}b = uwb$ is shorter than wb then it is not unioccurrent, because wb was non-rich. This means $\tilde{w}^{(+)}b$ is not rich. □

At the end of this section we prove that every factor of any Sturmian word can always be extended richly in two ways. However, let us first define Sturmian words. In the following we suppose that all words are binary.

A word w is *balanced* if for every two factors $x, y \in \mathrm{F}(w)$ of the same length and for every letter $a \in \mathrm{Alph}(w)$ the number $||x|_a - |y|_a|$ is at most 1. If a word is not balanced then it is *unbalanced*. An infinite word is *Sturmian* if it is balanced and aperiodic. A finite word is *Sturmian* if it is a factor of an infinite Sturmian word. From [10] (Proposition 2) we get that all Sturmian words are actually rich. This was proved for trapezoidal words, of which Sturmian words are a subset.

Proposition 2.14. *(See [4, Proposition 2.1.17].) A finite word is Sturmian if and only if it is balanced.*

Proposition 2.15. *(See [4, Proposition 2.1.3].) A word u is unbalanced if and only if there exists a palindrome v such that $0v0$ and $1v1$ are factors of u.*

Proposition 2.16. *Every finite Sturmian word can always be extended richly in two ways.*

Proof. Suppose u is a finite Sturmian word and $\mathrm{Alph}(u) = \{0, 1\}$. From Proposition 2.14 we get that u is balanced. We only have to prove that both $u1$ and $u0$ are rich.

Suppose to the contrary that $u1$ is not rich (the case $u0$ is identical). Now $u1$ is unbalanced because otherwise it would be Sturmian and hence rich. Proposition 2.15 tells us that there is a palindrome v such that $0v0$ and $1v1$ are factors of $u1$. Because u was balanced, the factor $1v1$ has to be a suffix of $u1$ and it cannot occur anywhere else in the word. This means that $1v1$ is a new palindrome and $u1$ is rich, which is a contradiction. □

3. The infinite defect

Rich words were defined such that they contain the maximum number of possible palindromes. We can define other words with respect to how many palindromes they lack compared to rich words, i.e. the defect of a word. This concept has been studied in various papers from different angles, for example in [13,6,7,2] and [3]. In this section we define a new concept, the infinite defect.

The *defect* of a finite word w is defined by $\mathrm{D}(w) = |w| + 1 - |\mathrm{Pal}(w)|$. The *defect* of a (right, left or two-way) infinite word w is defined by $\mathrm{D}(w) = \sup\{\mathrm{D}(u) \mid u \text{ is a factor of } w\}$. If the supremum does not exist then the defect is defined to be ∞. Clearly, finite and infinite rich words are exactly those words with defect equal to 0.

We can also study how much the defect must grow if a word has to be extended to an infinite word. The *infinite defect* of a finite word w is $\mathrm{D}_\infty(w) = \min\{\mathrm{D}(z) \mid z \text{ is an infinite word which has factor } w\}$, where we suppose $\mathrm{Alph}(z) \subseteq \mathrm{Alph}(w)$. We need the next theorem to guarantee that the min-function in the definition is always defined.

Theorem 3.1. *The infinite defect $\mathrm{D}_\infty(w)$ of a finite word w is finite.*

Proof. Let $u = w^{(+)}$. We will prove that u^∞ has finite defect. More precisely, we will prove $\mathrm{D}(u^\infty) = \mathrm{D}_\infty(u^2)$ with induction. Thus, the claim is that $\mathrm{D}_\infty(u^n) = \mathrm{D}_\infty(u^2)$ holds for all $n \geq 2$. If $n = 2$ then the claim is trivial. Yet, suppose it holds for $n = k$.

If v is any non-empty prefix of u, then clearly $\tilde{v}u^{k-1}v$ is a palindromic suffix of $u^k v$. The word $\tilde{v}u^{k-1}v$ is longer than half of the word $u^k v$, so if $\tilde{v}u^{k-1}v$ is not unioccurrent in $u^k v$ then it must overlap with itself. We can take the longest such overlap and see that it is the unioccurrent lps of $u^k v$.

We get that for every non-empty prefix v of u the word $u^k v$ has a unioccurrent lps. The word u^{k+1} has therefore the same defect as u^k because we get a new palindrome in every step when we extend u^k into u^{k+1}. This completes the induction.

We now get our claim because clearly $\mathrm{D}_\infty(w) \leq \mathrm{D}_\infty(u^2)$, where $\mathrm{D}_\infty(u^2)$ is finite. □

Remark 3.2. The definition of the infinite defect is useful because there exist words for which $\mathrm{D}(w) \neq \mathrm{D}_\infty(w)$, for example $w = 110100110111011001011$: clearly $\mathrm{D}(w) = 2$ but for every $w0, w1, 0w$ and $1w$ the defect is equal to 3. This means that no matter how we extend the word we always create new defects.

Both the defect and the infinite defect of a rich word are equal to 0. The defect and the infinite defect of a finite word can also be the same for non-rich words. For example $\mathrm{D}(00101100) = \mathrm{D}_\infty(00101100) = 1$.

Remark 3.3. We could also define right and left infinite defects of w separately such that the right (left) infinite defect means the lowest defect of a right (resp. left) infinite word that contains w. This is also reasonable since there are words for which they are different. For example the word $w = 101100111010111011$ has defect equal to 1. We can extend w to the right with an infinite word 1^∞ and get a word which also has defect equal to 1. But the words $0w$ and $1w$ already have defects equal to 2. So to get an infinite word that has the lowest defect we sometimes have to extend it to the left and sometimes to the right. Sometimes it does not matter.

We could also study how a right or left infinite word can be extended into a two-way infinite word. The *infinite defect* of a right or left infinite word w could be defined by $\mathrm{D}_\infty(w) = \inf\{\mathrm{D}(z) \mid z \text{ is a two-way infinite word that contains } w\}$.

If a word is rich then we know that the infinite defect is always zero. But how can we determine the infinite defect for other words? We know that it is always finite. An algorithm to solve the problem might not exist, but we can at least find some upper bounds for it. Clearly the normal defect is always a lower bound, which is sometimes achieved.

Proposition 3.4. *For a finite word w we have inequalities $\mathrm{D}_\infty(w) \leq \mathrm{D}(w^{(+)}w^{(+)})$ and $\mathrm{D}_\infty(w) \leq \mathrm{D}(\tilde{w}^{(+)}\tilde{w}^{(+)})$.*

Proof. This comes directly from the proof of Theorem 3.1. □

Remark 3.5. For some words this bound reaches the infinite defect. The word $w = 00101100$ is such: both w and $w^{(+)}w^{(+)} = 00101100110100001011001101 00$ have defect equal to 1.

On the other hand, the word $w = 110010010110010$ has defect equal to 4, but the words $w^{(+)} = 1100100101100100\cdots$ and $\tilde{w}^{(+)} = 0100110100100110\cdots$ have defects equal to 5. We can do better: $\mathrm{D}(1^\infty w) = 4$, which means that $\mathrm{D}_\infty(w) = 4$. Therefore, the bound does not reach the infinite defect.

Note that the words $w^{(+)}w^{(+)}$ and $\tilde{w}^{(+)}\tilde{w}^{(+)}$ can actually have different defects. Such a word is, for example, $w = 0010110001010$.

Proposition 3.6. *Let w be a finite word and let n be the length of the longest rich suffix or prefix of w. Then $\mathrm{D}_\infty(w) \leq |w| - n$.*

Proof. We can take the longest rich suffix (prefix) u of $w = vu$ and extend it to be a right (resp. left) infinite rich word. When we add the leftover v to the beginning (resp. to the end) we clearly get at most $|v|$ defects. □

Remark 3.7. This bound also sometimes reaches the infinite defect. For example, the word $w = 00101100101$ has defect 3 and so does the word $00101100101(0)^\infty$, where we have extended the longest rich suffix 01100101. We see that $|w| - n = 11 - 8 = 3$.

Sometimes the bound does not reach the infinite defect. For example, the word $w = 110110011101001101100110110111 0011011$ has defect 16. The longest rich suffix and prefix are $v = 01110011011$ and $\tilde{v}$, respectively. If we want to extend those to be right and left infinite rich words, respectively, we have to extend them both first with 0. That creates a new defect. After that, both extensions 0 and 1 also create new defects. So, we have $\mathrm{D}_\infty(w00)$, $\mathrm{D}_\infty(w01)$, $\mathrm{D}_\infty(00w)$, $\mathrm{D}_\infty(10w) \geq 18$. But $\mathrm{D}_\infty(w) = 17$ because in the word $w1^\infty$ the first letter 1 creates a new defect (which we cannot avoid) but after that we always get a new palindrome 1^k, where $k \geq 4$.

Proposition 3.8. *Let w be a finite word and let n be the length of the largest power of any letter in w. Then $\mathrm{D}_\infty(w) \leq \mathrm{D}(w) + n$. If w ends or begins with a^k then we can choose $n = n_a - k$, where a^{n_a} is the largest power of letter a in w.*

Proof. Suppose a^n is the largest power of letter a in w. If we now extend the word w with a^∞ then we create a new palindrome a^m, where $m > n$, in every step after we have extended w with at least a^n. So we create at most n new defects. Clearly, if w ends or begins with a^k then we create at most $n - k$ new defects. □

Remark 3.9. This bound also sometimes reaches the infinite defect; for example, if we look at the previous remark where the defect was 16. Now, the largest power of letter 1 is 1^3 and the word w ends with 1^2. So we get that $\mathrm{D}_\infty(w) \leq \mathrm{D}(w) + 3 - 2 = 17$.

For the word $w = 101001111000111101001$, which has defect 4, this bound does not reach the infinite defect. The largest powers of letters in w are 0^3 and 1^4. If we now extend the word with 0^∞ or 1^∞ to the left or right we get at least one defect more in each case. But we can see that $\mathrm{D}(w(01)^\infty) = 4$, which means $\mathrm{D}_\infty(w) = 4$.

Note that the previous three propositions also give us a method to construct the infinite word which contains a given word w. Note also that the constructions in Propositions 3.6 and 3.8 may sometimes give a smaller defect than what the bound always guarantees.

Proposition 3.10. *Each finite word can be extended to both periodic and aperiodic infinite words with finite defects.*

Proof. Let w be a finite word. Using Theorem 3.1 we can construct a periodic infinite word $(w^{(+)})^\infty$. Using Propositions 3.6 and 2.10 we can construct an aperiodic infinite word: we take the longest rich suffix of w and extend it to an aperiodic rich word. Clearly these both have finite defect. □

Next, let us look at how the defect and the infinite defect can differ. First, we define some functions for an integer n:

$$\mathrm{D}(n) = \max\{\mathrm{D}(w) \mid w \text{ is a word of length at most } n\},$$

$$\mathrm{D}_\infty(n) = \max\{\mathrm{D}_\infty(w) \mid w \text{ is a word of length at most } n\},$$

$$\mathrm{D}_{\mathrm{dif}}(n) = \max\{\mathrm{D}_\infty(w) - \mathrm{D}(w) \mid w \text{ is a word of length at most } n\}.$$

We clearly see that $\mathrm{D}(n)$ and $\mathrm{D}_\infty(n)$ are unbounded growing functions. For the function $\mathrm{D}_{\mathrm{dif}}(n)$ we can prove the following inequality.

Proposition 3.11. $\mathrm{D}_\infty(n) - \mathrm{D}(n) \leq \mathrm{D}_{\mathrm{dif}}(n)$.

Proof. Let us choose w to be a word for which $\mathrm{D}_\infty(w) = \mathrm{D}_\infty(n)$. Now clearly $\mathrm{D}(w) \leq \mathrm{D}(n)$. From this we get that $\mathrm{D}_\infty(n) - \mathrm{D}(n) \leq \mathrm{D}_\infty(w) - \mathrm{D}(w) \leq \mathrm{D}_{\mathrm{dif}}(n)$. □

Proposition 3.12. *The function $\mathrm{D}_{\mathrm{dif}}(n)$ is a unbounded growing function.*

Proof. It is trivially growing, so we only need to prove the unboundedness.

Suppose to the contrary, that there exists $k > 1$ such that $\forall n : \mathrm{D}_{\mathrm{dif}}(n) < k$. We will consider the word $u = 001^{k+1}01w101^{k+1}00$, where w is a word which does not have 1^{k+1} as a factor but still has every palindrome of length at most $2k + 2$ as a factor that is possible under this restriction. Now we need to prove that no matter how we extend w to be an infinite word, the defect will always grow by at least k.

We cannot get a lower defect by extending the word at the both ends, so we suppose that we only extend it to the right. Every time we put a letter at the end of u we cannot get a new palindromic factor that would contain 1^{k+1}. This comes from the fact that w does not contain 1^{k+1} and the factor 1^{k+1} is preceded by 10 and followed by 00. This means that after we have extended the word u with k letters we have no new palindromes because every palindrome of length $2k + 2$ or shorter is contained in w. □

4. The number of rich words

All binary words of length 7 or shorter are rich. The shortest non-rich binary words are of length 8 and there are four of them: 00101100, 00110100, 11010011, 11001011. Since not every word is rich it is natural to study how many of them exist.

Let us mark with $r_k(n)$ the number of k-ary rich words of length n. The next proposition states that there is an exponentially decreasing upper bound for $r_2(n)/2^n$ (i.e. the ratio of rich binary words from all binary words).

Proposition 4.1. $r_2(n)/2^n \le (63/64)^{\lfloor n/8 \rfloor}$.

Proof. First we give a recursive upper bound for $r_2(n)$. We start with the exact initial values: $r_2(k) = 2^k$ for $0 \le k \le 7$. Because there are four words of length 8 that are not rich and every binary rich word can be extended richly in at most two ways, we get a recursive inequality $r_2(n) \le 2^8 r_2(n-8) - 4r_2(n-8) = 252 r_2(n-8)$ for $n \ge 8$. This recursive inequality is easy to solve: $r_2(n) \le 252^{\lfloor n/8 \rfloor} 2^{n-8\lfloor n/8 \rfloor}$.

Now we have $r_2(n)/2^n \le 252^{\lfloor n/8 \rfloor} 2^{n-8\lfloor n/8 \rfloor}/2^n = 252^{\lfloor n/8 \rfloor} 2^{-8\lfloor n/8 \rfloor} = (63/64)^{\lfloor n/8 \rfloor}$. □

Using the idea from the previous proof, we can trivially enhance the upper bound. We just need to note that there are 16 non-rich words of length 9, 44 non-rich words of length 10, 108 non-rich words of length 11 and 266 non-rich words of length 12, such that all these non-rich words do not contain the shorter ones as a suffix (so they really do create completely new non-rich words).

Corollary 4.2. *For $n \ge 12$ we have $r_2(n) \le 2r_2(n-1) - 4r_2(n-8) - 16r_2(n-9) - 44r_2(n-10) - 108r_2(n-11) - 266r_2(n-12)$, where we have the exact values of $r_2(k)$ for $0 \le k \le 11$.*

We can of course do the same for every size of the alphabet if we change a few numbers in the above proof. Suppose $k \ge 3$. All k-ary words of length 3 or shorter are rich but all the words of form 0120 are non-rich and there are $k(k-1)(k-2)$ of them.

Corollary 4.3. $r_k(n)/k^n \le (1 - k(k-1)(k-2)/k^4)^{\lfloor n/4 \rfloor}$.

For the lower bound of $r_k(n)$ we can use Theorem 2.1.

Proposition 4.4. *For $n \ge 1$ we have $r_k(n) \ge r_k(n-1) + r_k(\lfloor n/3 \rfloor)$, where $r_k(0) = 1$.*

Proof. Every rich word can be extended richly in at least one way. From Theorem 2.1 we get that every rich word w of length $\lfloor n/3 \rfloor$ can be extended richly in at least two ways after it has been extended with a proper word u such that $|wu| < n$. These facts clearly give us our recursive formula when we notice that $r_k(0) = 1$. □

Remark 4.5. We do not know whether the function that comes from the recursive formula of the lower bound is exponentially growing or not. Note that we can improve the lower bound if we can improve Theorem 2.1. Note also that Remark 2.4 gives some limitations for improving it.

For $0 \le n \le 25$ the exact numbers of $r_2(n)$ are:

n	0	1	2	3	4	5	6	7	8	9	10	11
$r_2(n)$	1	2	4	8	16	32	64	128	252	488	932	1756

12	13	14	15	16	17	18	19
3246	5916	10618	18800	32846	56704	96702	163184

20	21	22	23	24	25
272460	450586	738274	1199376	1932338	3089518

In Fig. 1 there are both the bounds and the 26 first exact numbers of rich binary words. The picture would imply that the upper bound cannot be enhanced very much but the lower bound can be. This means that there is much to improve in Theorem 2.1.

5. Two-dimensional rich words

Let A be a finite alphabet. We define a *two-dimensional word* to be an infinite rectangular grid $\mathbb{Z}^2$ where every pair $(i, j) \in \mathbb{Z}^2$ gets a value from $A \cup \{\epsilon\}$. We denote the letter (or the empty word) from the pair of indices $(i, j) \in \mathbb{Z}^2$ of a

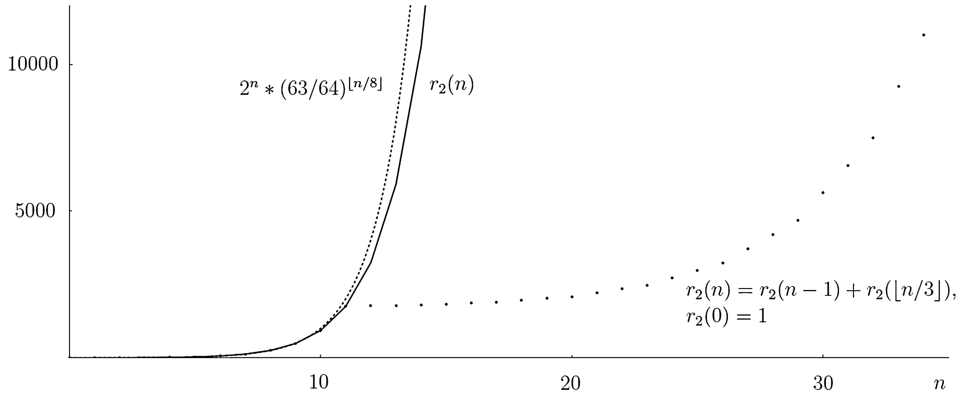

Fig. 1. The exact number of rich binary words $r_2(n)$ and the bounds.

```
                 01011001
  11010          00010110
01  01           01100101
11100            01011001
   11            01101001
                 10110010
                 01011001
                 00101110
```

Fig. 2. Two rich two-dimensional words.

two-dimensional word w by $w(i, j)$. Two-dimensional words have been studied in various papers, for example in [5,12,17] and [16], but the idea has not been applied to rich words. Here we define rich two-dimensional words and introduce some notions about them.

A two-dimensional word w is *rich* if $\forall i, j \in \mathbb{Z}$: the words $w(i, j)w(i, j+1)\cdots w(i, j+n)$ and $w(i, j)w(i+1, j)\cdots w(i+n, j)$ are rich, where $n \geq 0$ and $w(i, j+k), w(i+k, j) \neq \epsilon$ for every k $(0 \leq k \leq n)$.

We say that a two-dimensional word w can be extended to a *rich plane* if for every $(i, j) \in \mathbb{Z}^2$ for which $w(i, j) = \epsilon$ there exists a letter $a \in \text{Alph}(w)$ such that if we set $w(i, j) = a$ then the new two-dimensional word is rich.

Example 5.1. The two-dimensional words in Fig. 2 are both rich. The first one can trivially be extended to a rich plane by just adding the letter 1 for every empty spot, but the latter cannot. This comes from the fact that we would be forced to extend the (8×8)-square to the right so that the vertical word becomes 00101100, which is non-rich.

Let us suppose we have a binary alphabet $\{0, 1\}$. We saw that rich (8×8)-squares cannot always be extended to rich planes. The next proposition states that every (6×6)-square can always be extended to a rich plane. The (6×6)-square does not need to be full with letters because every binary word of length 6 or shorter is rich.

Proposition 5.2. *If w is a two-dimensional binary word such that every $w(i, j) \neq \epsilon$ is inside a (6×6)-square, then w can be extended to a rich plane.*

Proof. The proof is constructive. We suppose that the corners of the (6×6)-square of w are $(1, 1, 1, 1)$, $(1, 1, 1, 0)$, $(0, 0, 1, 1)$ or $(0, 1, 0, 1)$, where the order of the corners is left lower, left upper, right upper and right lower. Clearly all other possibilities are isomorphic in terms of the orientation of the plane and/or swapping the letters. If the (6×6)-square is not full, we can extend it to be full in any way we want.

Now, we can extend all the four possibilities in a way that is described in Fig. 3. The big letters 0 and 1 mean that the whole part of the plane is filled with that letter. All the horizontal and vertical words inside the (6×6)-square can be extended to be infinite rich words. The words w_1, w_2, w_3 and w_4 are any words that satisfy this.

Now we can see that the whole plane is rich by using the fact that the words $0^\infty u 1^\infty$, $0^\infty u 01^\infty$ and $0^\infty 1 u 1^\infty$ are always rich if $|u| \leq 4$. $\square$

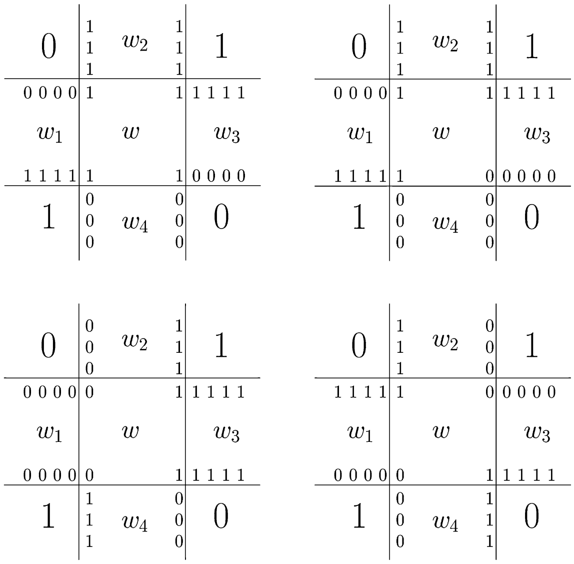

Fig. 3. The rich extensions of all 6×6-squares.

Remark 5.3. A finite word w is *strongly* rich if w^∞ is rich. If a two-dimensional word can be extended to a rectangle such that every horizontal and vertical word is strongly rich, then it can be extended to a rich plane: we just replicate the rectangle and fill the whole plane so that they are side by side. For more about strongly rich words, see [13] and [18].

The problem of whether a given two-dimensional word w can be extended to a rich plane is of course semi-decidable, i.e., there exists an algorithm that gives "yes" if the word cannot be extended. It just tries every possible way to fill all the empty indices $(i, j) \in \mathbb{Z}^2$ of w in some order. If at some point there is no possible choice, the algorithm halts and returns "yes". However, the problem is probably not decidable.

6. Open problems

Here we list some open problems from the previous sections.

Open problem 6.1. Let w be a rich word. How long is the shortest u such that wu can always be extended in at least two ways?

Open problem 6.2. Is the condition in Proposition 2.5 sufficient for two rich words u and v to be factors of the same rich word?

Open problem 6.3. Does there exist an algorithm to determine the infinite defect of a given finite word w?

Open problem 6.4. Is the function $r_k(n)$, i.e. the number of rich words, exponentially increasing?

Open problem 6.5. Can every (7×7)-square be extended to a rich plane?

Open problem 6.6. Is the following problem decidable: "Given a two-dimensional word w, can it be extended to a rich plane"?

Acknowledgements

I would like to thank my supervisors Luca Zamboni and Tero Harju for several fruitful discussions and for introducing me to rich words. The remarks and suggestions of the anonymous referees were also very helpful and substantially improved the paper.

References

[1] P. Ambrož, C. Frougny, Z. Masáková, E. Pelantová, Palindromic complexity of infinite words associated with simple Parry numbers, Ann. Inst. Fourier (Grenoble) 56 (2006) 2131–2160.

[2] L. Balková, E. Pelantová, Š. Starosta, Infinite words with finite defect, Adv. in Appl. Math. 47 (2011) 562–574.
[3] L. Balková, E. Pelantová, Š. Starosta, Proof of the Brlek–Reutenauer conjecture, Theoret. Comput. Sci. 475 (2013) 120–125.
[4] J. Berstel, P. Séébold, Sturmian words, in: M. Lothaire (Ed.), Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, UK, 2002.
[5] V. Berthe, L. Vuillon, Tilings and rotations: a two-dimensional generalization of Sturmian sequences, Discrete Math. 223 (2000) 27–53.
[6] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer, On the palindromic complexity of infinite words, Internat. J. Found. Comput. Sci. 15 (2004) 293–306.
[7] S. Brlek, C. Reutenauer, Complexity and palindromic defect of infinite words, Theoret. Comput. Sci. 412 (2011) 493–497.
[8] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni, A connection between palindromic and factor complexity using return words, Adv. in Appl. Math. 42 (2009) 60–74.
[9] M. Bucci, A. De Luca, A. Glen, L.Q. Zamboni, A new characteristic property of rich words, Theoret. Comput. Sci. 410 (2009) 2860–2863.
[10] A. de Luca, A. Glen, L.Q. Zamboni, Rich, Sturmian, and trapezoidal words, Theoret. Comput. Sci. 407 (2008) 569–573.
[11] X. Droubay, J. Justin, G. Pirillo, Episturmian words and some constructions of de Luca and Rauzy, Theoret. Comput. Sci. 255 (2001) 539–553.
[12] C. Epifanio, M. Koskas, F. Mignosi, On a conjecture on bidimensional words, Theoret. Comput. Sci. 299 (1–3) (2003) 123–150.
[13] A. Glen, J. Justin, S. Widmer, L.Q. Zamboni, Palindromic richness, European J. Combin. 30 (2009) 510–531.
[14] M. Lothaire, Combinatorics on Words, Encyclopedia of Mathematics and Its Applications, vol. 17, Addison-Wesley, Reading, MA, 1983.
[15] M. Lothaire, Algebraic Combinatorics on Words, Encyclopedia of Mathematics and Its Applications, vol. 90, Cambridge University Press, UK, 2002.
[16] S.A. Puzynina, S.V. Avgustinovich, On periodicity of two-dimensional words, Theoret. Comput. Sci. 391 (2008) 178–187.
[17] A. Quas, L.Q. Zamboni, Periodicity and local complexity, Theoret. Comput. Sci. 319 (1–3) (2004) 229–240.
[18] A. Restivo, G. Rosone, Burrows–Wheeler transform and palindromic richness, Theoret. Comput. Sci. 410 (2009) 3018–3026.

Publication II

Tero Harju, Jetro Vesti, Luca Q. Zamboni, *On a question of Hof, Knill and Simon on palindromic substitutive systems*, Monatshefte für Mathematik, Vol. 179, pp. 379-388, 2016, DOI 10.1007/s00605-015-0828-2.

Publication III

Jetro Vesti, *Rich square-free words*, preprint, submitted to Theoretical Computer Science 03/2016, available in arXiv.

Rich square-free words

Jetro Vesti

University of Turku, Department of Mathematics and Statistics, 20014 Turku, Finland

Abstract

A word w is *rich* if it has $|w|+1$ many distinct palindromic factors, including the empty word. A word is *square-free* if it does not have a factor uu, where u is a non-empty word.

Pelantová and Starosta (Discrete Math. 313 (2013)) proved that every infinite rich word contains a square. We will give another proof for that result. Pelantová and Starosta denoted by $r(n)$ the length of a longest rich square-free word on an alphabet of size n. The exact value of $r(n)$ was left as an open question. We will give an upper and a lower bound for $r(n)$. The lower bound is conjectured to be exact but it is not explicit.

We will also generalize the notion of repetition threshold for a limited class of infinite words. The repetition thresholds for episturmian and rich words are left as an open question.

Keywords: Combinatorics on words, Palindromes, Rich words, Square-free words, Repetition threshold.
2000 MSC: 68R15

1. Introduction

In recent years, rich words and palindromes have been studied extensively in combinatorics on words. A word is a *palindrome* if it is equal to its reversal. In [DJP], the authors proved that every word w has at most $|w|+1$ many distinct palindromic factors, including the empty word. The class of words which achieve this limit was introduced in [BHNR] with the term *full* words. When the authors of [GJWZ] studied these words thoroughly they called them *rich* (in palindromes). Since then, rich words have been studied in various papers, for example in [AFMP], [BDGZ1], [BDGZ2], [DGZ], [RR] and [V].

The *defect* of a finite word w, denoted $D(w)$, is defined as $D(w) = |w| + 1 - |\mathrm{Pal}(w)|$, where $\mathrm{Pal}(w)$ is the set of palindromic factors in w. The *defect* of an infinite word w is defined as $D(w) = \sup\{D(u) |\ u \text{ is a factor of } w\}$. In other words, the defect is a measure of

Email address: `jejove@utu.fi` (Jetro Vesti)

how many palindromes the word lacks. Rich words are exactly those whose defect is equal to 0.

The authors of [PS] proved, in Theorem 4 of the article, that every recurrent word with finite Θ-defect contains infinitely many overlapping factors. An *overlapping* word is a word of form uuv, where v is a non-empty prefix of u. A word is a Θ-*palindrome* if it is a fixed point of an involutory antimorphism Θ. The reversal mapping R is an involutory antimorphism, which means that if $\Theta = R$ then Θ-defect is equal to the defect. This means Theorem 4 in [PS] holds also for normal defect and normal palindromes. In this article we will restrict ourselves to the case where Θ is the reversal mapping.

Since every rich word has a zero defect and every overlapping factor uuv has a square uu, a corollary of Theorem 4 in [PS] is that every recurrent rich word contains a square. This was noted in [PS] as Remark 6, where the word *recurrent* was replaced with *infinite*. This can be done, since every infinite rich word x has a recurrent point y in the shift orbit closure of x (see e.g. Section 4 of [Q]). We know y has a square, which means x has a square. We will give another proof of the result in Remark 6 of [PS] (Corollary 2.9).

In Remark 6 of [PS] there was also noted that since every rich square-free word is finite, we can look for a longest one. The length of a longest such word, on an alphabet of size n, was denoted by $r(n)$. An explicit formula for $r(n)$ was left as an open question.

In Section 2.1 we will construct recursively a sequence of rich square-free words, the lengths of which give us a lower bound for $r(n)$. We will also make a conjecture that $r(n)$ can be achieved using these words. In Section 2.2 we will prove an upper bound for $r(n)$.

1.1. Repetition threshold

Square-free words are a special case of unavoidable repetitions of words, which has been a central topic in combinatorics on words since Thue (see [T1] and [T2]). The *repetition threshold*, on an alphabet of size n, is the smallest number r such that there exists an infinite word which avoids greater than r-powers. This number is denoted by $RT(n)$ and it was first studied in [D], where Dejean gave her famous conjecture. This conjecture has now been proven, in many parts and by several authors (see [R] and [CR]).

The repetition threshold can be studied also for a limited class of infinite words. In [MP], it was proven that the infinite Fibonacci word does not contain a power with exponent greater than $2+\varphi$, where φ is the golden ratio $\frac{\sqrt{5}+1}{2}$, but every smaller fractional power is contained. In [CD], the authors proved that among *Sturmian* words, the Fibonacci word is optimal with respect to this property. Sturmian words are equal to *episturmian* words when $n = 2$ (see [DJP]). We define the *episturmian repetition threshold*, on an alphabet of size n, to be the smallest number r such that there exists an episturmian word which avoids greater than r-powers, and denote this number by $ERT(n)$. We get $ERT(2) = 2 + \varphi$. From [GJ], we get the n-bonacci word is episturmian and it has critical exponent $2 + 1/(\varphi_n - 1)$, where φ_n is the generalized golden ratio. This means $ERT(n) \leq 2 + 1/(\varphi_n - 1)$. Notice, from [HPS] we get φ_n converges to 2.

We define the *rich repetition threshold*, on an alphabet of size n, to be the smallest number r such that there exists an infinite rich word which avoids greater than r-powers, and denote this number by $RRT(n)$. From [PS] we get $RRT(n) \geq 2$. Since episturmian words are rich (see [DJP]), we also know $RRT(n) \leq 2 + 1/(\varphi_n - 1)$ and $ERT(n) \geq 2$. This means $2 \leq RRT(n), ERT(n) \leq 2 + 1/(\varphi_n - 1)$. The exact values of $ERT(n)$ and $RRT(n)$ are left as an open problem.

Open problem 1.1. *Determine the repetition threshold for episturmian words and for rich words, on an alphabet of size n.*

1.2. Preliminaries

An *alphabet* A is a non-empty finite set of symbols, called *letters*. A *word* is a finite sequence of letters from A. The *empty* word ϵ is the empty sequence. The set A^* of all finite words over A is a *free monoid* under the operation of concatenation. The set $\mathrm{Alph}(w)$ is the set of all letters that occur in w. If $|\mathrm{Alph}(w)| = n$ then we say that w is *n-ary*.

An *infinite word* is a sequence indexed by $\mathbb{N}$ with values in A. We denote the set of all infinite words over A by A^ω and define $A^\infty = A^* \cup A^\omega$.

The *length* of a word $w = a_1 a_2 \ldots a_n$, with each $a_i \in A$, is denoted by $|w| = n$. The empty word ϵ is the unique word of length 0. By $|w|_a$ we denote the number of occurrences of a letter a in w.

A word x is a *factor* of a word $w \in A^\infty$, denoted $x \in w$, if $w = uxv$ for some $u \in A^*$,$v \in A^\infty$. If x is not a factor of w, we denote $x \notin w$. If $u = \epsilon$ (resp. $v = \epsilon$) then we say that x is a *prefix* (resp. *suffix*) of w. If $w = uv \in A^*$ is a word, we use the notation $u^{-1}w = v$ or $wv^{-1} = u$ to mean the removal of a prefix or a suffix of w. We say that a prefix or a suffix of w is *proper* if it is not the whole of w.

A factor x of a word w is said to be *unioccurrent* in w if x has exactly one occurrence in w. Two occurrences of factor x are said to be *consecutive* if there is no occurrence of x between them. A factor of w having exactly two occurrences of a non-empty factor u, one as a prefix and the other as a suffix, is called a *complete return* to u in w.

The *reversal* of $w = a_1 a_2 \ldots a_n$ is defined as $\widetilde{w} = a_n \ldots a_2 a_1$. A word w is called a *palindrome* if $w = \widetilde{w}$. The empty word ϵ is assumed to be a palindrome.

Other basic definitions and notation in combinatorics on words can be found in [Lot1] and [Lot2].

Proposition 1.2. *([DJP], Prop. 2) A word w has at most $|w| + 1$ distinct palindromic factors, including the empty word.*

Definition 1.3. *([GJWZ], Def. 2.2 and 2.9) A word w is* rich *if it has exactly $|w|+1$ distinct palindromic factors, including the empty word. An infinite word is* rich *if all of its factors are rich.*

Proposition 1.4. *([GJWZ], Thm. 2.14) A finite or infinite word w is rich if and only if all complete returns to any palindromic factor in w are themselves palindromes.*

Let $w = vu$ be a word and u its longest palindromic suffix. The *palindromic closure* of w is defined as $w^{(+)} = vu\widetilde{v}$. If u is the longest *proper* palindromic suffix of w, called lpps, we define the *proper palindromic closure* of w the same way as $w^{(++)} = vu\tilde{v}$. We refer to the longest proper palindromic prefix of w as lppp and define the *proper palindromic prefix closure* of w as ${}^{(++)}w = \widetilde{\widetilde{w}^{(++)}}$.

Proposition 1.5. *([GJWZ], Prop. 2.6) Palindromic closure preserves richness.*

Proposition 1.6. *([GJWZ], Prop. 2.8) Proper palindromic (prefix) closure preserves richness.*

2. The length of a longest rich square-free word

A word of form uu, where $u \neq \epsilon$, is called a *square* and a word w which does not have a square as a factor is called *square-free*. For example 1212 is a square and 01210 is square-free.

In [PS], Theorem 4 and Remark 6, it was proved that every infinite rich word contains a square. This means that every rich square-free word is of finite length. The length of a longest such word, on an alphabet of size n, is denoted with $r(n)$. An explicit formula for $r(n)$ was left as an open problem in [PS].

The first seven exact values of $r(n)$ are $r(1) = 1, r(2) = 3, r(3) = 7, r(4) = 15, r(5) = 33, r(6) = 67$ and $r(7) = 145$. These can be found from https://oeis.org/A269560. The longest rich square-free word on a given alphabet is not unique. Here are all the longest non-isomorphic ones, up to permutating the letters and taking the reversal, for $n = 1, \ldots, 7$:

$w_{1,1} = 1$
$w_{2,1} = 121$
$w_{3,1} = 2131213$
$w_{3,2} = 1213121$
$w_{4,1} = 131214121312141$
$w_{4,2} = 123121412131214$
$w_{4,3} = 213121343121312$
$w_{4,4} = 121312141213121$
$w_{5,1} = 421242131213531213124213121353135$

$w_{5,2} = 1312421312135312131242131213531 35$

$w_{6,1} = 151312131513121412131214161412131214121315131214121312141614121416 1$

$w_{6,2} = 121412131513121412131214161412131214121315131214121312141614121416 1$

$w_{6,3} = 421242131213531213124213121353135653135312131242131213531213124212 4$

$w_{6,4} = 131242131213531213124213121353135653135312131242131213531213124212 4$

$w_{6,5} = 531353121312421312135312131242131613124213121353121312421312135313 5$

$w_{6,6} = 131242131213531213124213121353135653135312131242131213531213124213 1$

$w_{7,1} = 2421312135312131242131613124213121353121312421312135313575313531213 12$
$42131213531213124213161312421312135312131242131213531357531353121353135753 57$

$w_{7,2} = 2421312135312131242131613124213121353121312421312135313575313531213 12$
$42131213531213124213161312421312135312131242131213531357531353121353135753 13$

$w_{7,3} = 2421312135312131242124642124213121353121312421312135313575313531213 12$
$42131213531213124212464212421312135312131242131213531357531353121353135753 57$

$w_{7,4} = 2421312135312131242124642124213121353121312421312135313575313531213 12$
$42131213531213124212464212421312135312131242131213531357531353121353135753 13$

We can see that

$$w_{2,1} = w_{1,1}2w_{1,1},\ w_{3,2} = w_{2,1}3w_{2,1},\ w_{4,3} = w_{3,1}4\widetilde{w_{3,1}},\ w_{4,4} = w_{3,2}4w_{3,2},$$

$$w_{6,3} = w_{5,1}6\widetilde{w_{5,1}},\ w_{6,4} = w_{5,2}6\widetilde{w_{5,1}},\ w_{6,5} = \widetilde{w_{5,2}}6w_{5,2} \text{ and } w_{6,6} = w_{5,2}6\widetilde{w_{5,2}}.$$

Generally, we can construct rich square-free words by using a basic recursion

$$b_n = ba\widetilde{b},$$

where b is a longest rich square-free word over an $(n-1)$-ary alphabet A and $a \notin A$ is a new letter. It is very easy to see that b_n is rich and square-free. This gives us a recursive lower bound for $r(n)$: $r(n) \geq 2r(n-1)+1$, for all $n \geq 2$. We will repeatedly use this equality later in Section 2.2, when we prove an upper bound for $r(n)$. The closed-form solution for the recursion $r(1) = 1, r(n) \geq 2r(n-1)+1$ is $r(n) \geq 2^n - 1$.

The case $n = 5$ reveals that the basic recursion $b_n = ba\widetilde{b}$ is not always optimal, since neither $w_{5,1}$ nor $w_{5,2}$ is of that form: $|w_{5,1}| = r(5) = 33 > 31 = 2 \cdot r(4) + 1$.

We can also see that

$$w_{3,1} = 2w_{1,1}3w_{1,1}2w_{1,1}3,\ w_{4,1} = 13w_{2,1}4w_{2,1}3w_{2,1}41,\ w_{4,2} = 213w_{2,1}4w_{2,1}3w_{2,1}4,$$

$$w_{5,1} = 42124w_{3,1}5\widetilde{w_{3,1}}4w_{3,1}53135,\ w_{5,2} = 13124w_{3,1}5\widetilde{w_{3,1}}4w_{3,1}53135,$$

$$w_{6,1} = 1513121315w_{4,1}6\widetilde{w_{4,1}}5w_{4,1}6141214161,\ w_{6,2} = 1214121315w_{4,1}6\widetilde{w_{4,1}}5w_{4,1}6141214161,$$

$$w_{7,1} = u_{1,2}6w_{5,2}7\widetilde{w_{5,2}}6w_{5,2}7v_{1,3},\ w_{7,2} = u_{1,2}6w_{5,2}7\widetilde{w_{5,2}}6w_{5,2}7v_{2,4},$$

$$w_{7,3} = u_{3,4}6w_{5,1}7\widetilde{w_{5,1}}6w_{5,1}7v_{1,3},\ \ w_{7,4} = u_{3,4}6w_{5,1}7\widetilde{w_{5,1}}6w_{5,1}7v_{2,4},$$

$$\text{where } u_{1,2} = 24213121353121312421 31, u_{3,4} = 2421312135312131242124,$$

$$v_{1,3} = 5313531213531357535 7 \text{ and } v_{2,4} = 53135312135313575313.$$

This gives us a hint as to how to get, in some cases, a better recursion than the basic recursion. We will define this recursion explicitly in the next subsection.

2.1. A lower bound

In this subsection, we will prove another lower bound for $r(n)$. We will use an alphabet $\{A_0, A_1, A_2, A_3, B_3, A_4, B_4, A_5, B_5, \ldots\}$. The following construction of rich square-free words w_n is recursive. The first six words are

$$w_1 = A_1, w_2 = A_0A_2A_0, w_3 = v_3A_3w_1B_3w_1A_3w_1B_3u_3, w_4 = v_4A_4w_2B_4w_2A_4w_2B_4u_4,$$

$$w_5 = v_5A_5w_3B_5w_3A_5w_3B_5u_5, w_6 = v_6A_6w_4B_6w_4A_6w_4B_6u_6,$$

where $v_3, u_3 = \epsilon$, $v_4, u_4 = A_0$, $v_5 = A_5A_3A_1A_3$, $u_5 = B_3A_1A_3A_1$, $v_6 = A_0A_6A_0A_4A_0A_2A_0A_4A_0$ and $u_6 = A_0B_4A_0A_2A_0A_4A_0A_2A_0$. Notice that w_6 is isomorphic ($\cong$) to $w_{6,2}$, $w_5 \cong w_{5,2}$, $w_4 \cong w_{4,1}$ and $w_3 \cong w_{3,1}$. For $n \geq 7$, we define

$$w_n = v_nA_nw_{n-2}B_n\widetilde{w_{n-2}}A_nw_{n-2}B_nu_n,$$

where $v_n = (P_nc_n)^{-1}\widetilde{v_{n-4}}A_{n-2}\widetilde{v_{n-2}}A_nv_{n-2}A_{n-2}v_{n-4}A_{n-4}w_{n-6}B_{n-4}\widetilde{w_{n-6}}A_{n-4}\widetilde{v_{n-4}}A_{n-2}\widetilde{v_{n-2}}$ and $u_n = \widetilde{u_{n-2}}B_{n-2}\widetilde{w_{n-4}}A_{n-2}w_{n-4}B_{n-2}\widetilde{u_{n-4}}B_{n-4}\widetilde{w_{n-6}}(d_n\widetilde{P_n})^{-1}$, where P_n is the largest common prefix of w_{n-6} and $\widetilde{v_{n-4}}$, c_n is the first letter of $(P_n)^{-1}\widetilde{v_{n-4}}A_{n-2}$ and d_n is the first letter of $(P_n)^{-1}w_{n-6}B_{n-4}$.

We can see that $\mathrm{Alph}(w_{2k}) = \{A_0, A_2, A_4, B_4, A_6, B_6, \ldots, A_{2k}, B_{2k}\}$ and $\mathrm{Alph}(w_{2k+1}) = \{A_1, A_3, B_3, A_5, B_5, \ldots, A_{2k+1}, B_{2k+1}\}$. This means we really have $|\mathrm{Alph}(w_n)| = n$. We also have $c_n \neq d_n$, since $A_{n-2} \notin w_{n-6}$ and $B_{n-4} \notin \widetilde{v_{n-4}}$.

Before we prove that w_n is rich and square-free, we will define some notation in order to make the proof easier to read. We mark that $E_n = A_nw_{n-2}B_n\widetilde{w_{n-2}}A_nw_{n-2}B_n$, $F_n = (P_nc_n)^{-1}\widetilde{v_{n-4}}A_{n-2}\widetilde{v_{n-2}}$, $G_n = \widetilde{w_{n-6}}A_{n-4}\widetilde{v_{n-4}}A_{n-2}\widetilde{v_{n-2}}$ and $H_n = \widetilde{P_n}A_{n-4}w_{n-6}B_{n-4}G_n$. Now $w_n = v_nE_nu_n$, $v_n = F_nA_n\widetilde{G_n}B_{n-4}G_n$ and $w_{n-2} = \widetilde{H_n}d_n\widetilde{u_n}$. We can also see that H_n is a suffix of v_n and F_n is a suffix of G_n.

Proposition 2.1. *The word w_n is square-free for all $n \geq 1$.*

Proof. We prove the claim by induction. It is easy to check that w_n is square-free when $1 \leq n \leq 6$. Suppose w_n is square-free for all $n < k$, where $k \geq 7$. Now we need to prove that w_k is square-free.

The word $A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k w_{k-2} B_k u_k$ is square-free because w_{k-2} is square-free, $A_k, B_k \notin w_{k-2}$ and $\widetilde{u_k}$ is a proper suffix of w_{k-2}. The words G_k and F_k are suffixes of $\widetilde{w_{k-2}}$ and $A_k, B_{k-4} \notin G_k, F_k$, which means that $v_k = F_k A_k \widetilde{G_k} B_{k-4} G_k$ is square-free.

Now, the only way $w_k = F_k A_k \widetilde{G_k} B_{k-4} G_k A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k w_{k-2} B_k u_k$ can have a square is if the square is equal to either 1) $x A_k w_{k-2} B_k y x A_k w_{k-2} B_k y$, where x is a suffix of both v_k and $\widetilde{w_{k-2}}$, and y is a prefix of both u_k and $\widetilde{w_{k-2}}$, or 2) $x A_k y x A_k y$, where x is a suffix of both F_k and $\widetilde{G_k} B_{k-4} G_k$, and y is a prefix of both $\widetilde{w_{k-2}}$ and $\widetilde{G_k} B_{k-4} G_k$.

1) Case $x A_k w_{k-2} B_k y x A_k w_{k-2} B_k y$. Now $yx = \widetilde{w_{k-2}} = u_k d_k H_k$. Because y is a prefix of u_k and x is suffix of v_k, we have $d_k H_k$ is a suffix of v_k. We also know that $c_k H_k$ is always a suffix of v_k. This is a contradiction since $c_k \neq d_k$.

2) Case $x A_k y x A_k y$. Now y is a prefix of $\widetilde{w_{k-2}}$, which means that x has to have a suffix $P_k^{-1} \widetilde{v_{k-4}} A_{k-2} \widetilde{v_{k-2}}$. This is a contradiction, since x is also a suffix of $(P_k c_k)^{-1} \widetilde{v_{k-4}} A_{k-2} \widetilde{v_{k-2}}$. □

Proposition 2.2. *The word w_n is rich for all $n \geq 1$.*

Proof. We prove the claim by induction. It is easy to check that w_n is rich when $1 \leq n \leq 6$. Suppose w_n is rich for all $n < k$, where $k \geq 7$. Now we need to prove that w_k is rich.

Since w_{k-2} is rich and $A_k, B_k \notin w_{k-2}$, we get $A_k w_{k-2} B_k$ is rich. Proposition 1.5 gives now that $A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k$ is rich. The lpps of $A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k$ is A_k, which means $A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k$ is rich by Proposition 1.6. The word u_k is a prefix of $\widetilde{w_{k-2}}$, so the factor $A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k w_{k-2} B_k u_k$ is also rich.

The lppp of $A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k w_{k-2} B_k u_k$ is $A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k$, which means that also the proper palindromic prefix closure $\widetilde{u_k} B_k \widetilde{w_{k-2}} A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k w_{k-2} B_k u_k$ is rich. The word H_k is a suffix of $\widetilde{w_{k-2}}$, which means $H_k A_k w_{k-2} B_k \widetilde{w_{k-2}} A_k w_{k-2} B_k u_k = H_k E_k u_k$ is also rich.

The word $c_k H_k E_k u_k$ has a palindromic prefix $PP = c_k \widetilde{P_k} A_{k-4} w_{k-6} B_{k-4} \widetilde{w_{k-6}} A_{k-4} P_k c_k$. The following paragraph proves that it is unioccurrent in $c_k H_k E_k u_k$.

The letter B_{k-4} occurs only once in $c_k H_k$, in the middle of our palindromic prefix PP. This occurrence of B_{k-4} is preceded by $c_k \widetilde{P_k} A_{k-4} w_{k-6}$ and succeeded by $\widetilde{w_{k-6}} A_{k-4} P_k c_k$. The last occurrence of B_{k-4} in $E_k u_k$ is succeeded by $\widetilde{w_{k-6}} (d_k \widetilde{P_k})^{-1}$ and nothing more. Since the word $\widetilde{w_{k-6}} (d_k \widetilde{P_k})^{-1}$ is clearly a proper prefix of $\widetilde{w_{k-6}} A_{k-4} P_k c_k$, this last occurrence of B_{k-4} in $E_k u_k$ cannot occur in a factor PP. All other occurrences of B_{k-4} in $E_k u_k$ are preceded by $B_{k-4} \widetilde{w_{k-6}} A_{k-4} w_{k-6}$ or succeeded by $\widetilde{w_{k-6}} A_{k-4} w_{k-6} B_{k-4}$. The word $B_{k-4} \widetilde{w_{k-6}} A_{k-4} w_{k-6}$ has a suffix $d_k \widetilde{P_k} A_{k-4} w_{k-6}$, which means that it cannot have a suffix $c_k \widetilde{P_k} A_{k-4} w_{k-6}$ because $c_k \neq d_k$. Hence no B_{k-4} in $c_k H_k E_k u_k$ can occur in a factor PP, except the first one.

Since PP is unioccurrent palindromic prefix in $c_kH_kE_ku_k$, we get $c_kH_kE_ku_k$ is rich and PP is the lppp of $c_kH_kE_ku_k$. Now, all we need to do is to take the proper palindromic prefix closure of $c_kH_kE_ku_k$, which is rich by Proposition 1.6. It has a suffix w_k, which concludes the proof:

$$^{(++)}(c_kH_kE_ku_k) = \widetilde{u_k}B_k\widetilde{w_{k-2}}A_kw_{k-2}B_k\widetilde{w_{k-2}}A_k\widetilde{G_k}B_{k-4}G_kE_ku_k$$

$$\overset{*}{=} XF_kA_k\widetilde{G_k}B_{k-4}G_kE_ku_k = Xv_kE_ku_k = Xw_k \ (^*F_k \text{ is a suffix of } \widetilde{w_{k-2}}).$$

□

Now we know that w_n is rich and square-free. Hence $r(n) \geq |w_n|$ for all $n \geq 1$. We can compute $|w_7| = 145$, $|w_8| = 291$, $|w_9| = 629$ and $|w_{10}| = 1255$. Notice that $w_7 = w_{7,4}$, which means our lower bound is exact when $n = 7$. The cases $r(8)$ and $r(9)$ are too large to compute the exact value. However, by creating a partial tree of rich square-free words for $n = 8$ and 9, by leaving some branches out of it, the longest words we could find were of length 291 and 629, respectively. These are exactly the lengths of $|w_8|$ and $|w_9|$. Notice that $|w_8| = 291 = 2 \cdot 145 + 1 = 2|w_7| + 1$, which means the basic recursion b_n is as good as our recursion w_n when $n = 8$. Notice also that $|w_9| = 629 > 583 = 2 \cdot 291 + 1 = 2|w_8| + 1$ and $|w_{10}| = 1255 < 1259 = 2 \cdot 629 + 1 = 2|w_9| + 1$, which mean w_n is better than b_n when $n = 9$ and b_n is better than w_n when $n = 10$.

The previous paragraph suggests that it is reasonable to make the following conjecture.

Conjecture 2.3. *$r(n) = max\{|w_n|, 2 \cdot |w_{n-1}| + 1\}$ for all $n \geq 1$.*

The recursion for the length of w_n might be too complex to be solved in a closed-form, but we want to get at least an estimate for it. Let us first estimate the length of v_n, which will be used in Proposition 2.5.

Lemma 2.4. *$|v_n| \geq 3|v_{n-2}| + 2|w_{n-6}| + 2|v_{n-4}| + 6$, for $n \geq 7$.*

Proof.

$$|v_n| = |(P_nc_n)^{-1}\widetilde{v_{n-4}}A_{n-2}\widetilde{v_{n-2}}A_nv_{n-2}A_{n-2}v_{n-4}A_{n-4}w_{n-6}B_{n-4}\widetilde{w_{n-6}}A_{n-4}\widetilde{v_{n-4}}A_{n-2}\widetilde{v_{n-2}}|$$

$$\geq |\widetilde{v_{n-2}}A_nv_{n-2}A_{n-2}v_{n-4}A_{n-4}w_{n-6}B_{n-4}\widetilde{w_{n-6}}A_{n-4}\widetilde{v_{n-4}}A_{n-2}\widetilde{v_{n-2}}|$$

$$\geq 3|v_{n-2}| + 2|w_{n-6}| + 2|v_{n-4}| + 6,$$

where $|(P_nc_n)^{-1}\widetilde{v_{n-4}}A_{n-2}| \geq 0$, since c_n is a letter and P_n is a prefix of $\widetilde{v_{n-4}}$. □

Proposition 2.5. *$r(n) \geq |w_n| > 2.008^n$ for $n \geq 5$.*

Proof. From our recursion of w_n, we get for $n \geq 11$:

$$|w_n| = |v_n A_n w_{n-2} B_n \widetilde{w_{n-2}} A_n w_{n-2} B_n u_n| = 3|w_{n-2}| + |v_n| + |u_n| + 4$$

$$= 3|w_{n-2}| + |(P_n c_n)^{-1} \widetilde{v_{n-4}} A_{n-2} \widetilde{v_{n-2}} A_n v_{n-2} A_{n-2} v_{n-4} A_{n-4} w_{n-6} B_{n-4} \widetilde{w_{n-6}} A_{n-4} \widetilde{v_{n-4}} A_{n-2} \widetilde{v_{n-2}}|$$

$$+ |\widetilde{u_{n-2}} B_{n-2} \widetilde{w_{n-4}} A_{n-2} w_{n-4} B_{n-2} \widetilde{u_{n-4}} B_{n-4} \widetilde{w_{n-6}} (d_n \widetilde{P_n})^{-1}| + 4$$

$$= 3|w_{n-2}| + |(P_n c_n)^{-1} \widetilde{v_{n-4}} A_{n-2} \widetilde{v_{n-2}} A_n v_{n-2} A_{n-2} v_{n-4}| - |d_n \widetilde{P_n}| + 4$$

$$+ |\widetilde{u_{n-2}} B_{n-2} \widetilde{w_{n-4}} A_{n-2} w_{n-4} B_{n-2} \widetilde{u_{n-4}} B_{n-4} \widetilde{w_{n-6}}| + |A_{n-4} w_{n-6} B_{n-4} \widetilde{w_{n-6}} A_{n-4} \widetilde{v_{n-4}} A_{n-2} \widetilde{v_{n-2}}|$$

$$= 4|w_{n-2}| + |(P_n c_n)^{-1} \widetilde{v_{n-4}} A_{n-2} \widetilde{v_{n-2}} A_n v_{n-2} A_{n-2} v_{n-4}| - |d_n \widetilde{P_n}| + 4$$

$$= 4|w_{n-2}| + 2(|\widetilde{v_{n-4}}| - |P_n|) + 2|v_{n-2}| + |A_{n-2} A_n A_{n-2}| - |d_n| - |c_n| + 4$$

$$\geq 4|w_{n-2}| + 2|v_{n-2}| + 5 \geq 4|w_{n-2}| + 2(3|v_{n-4}| + 2|w_{n-8}| + 2|v_{n-6}| + 6) + 5$$

$$\geq 4|w_{n-2}| + 2(3(3|v_{n-6}| + 2|w_{n-10}| + 2|v_{n-8}| + 6) + 2|w_{n-8}| + 2|v_{n-6}| + 6) + 5$$

$$> 4|w_{n-2}| + 4|w_{n-8}| + 12|w_{n-10}|.$$

From our recursion of w_n we also know that $|w_{10}| = 1255 > 1164 = 4|w_8|$, $|w_9| = 629 > 580 = 4|w_7|$, $|w_8| = 291 > 268 = 4|w_6|$ and $|w_7| = 145 > 132 = 4|w_5|$.

Now, for $n \geq 15$ we have

$$|w_n| > 4|w_{n-2}| + 4|w_{n-8}| + 12|w_{n-10}| > 4(4|w_{n-4}| + 4|w_{n-10}|) + 4|w_{n-8}| + 12|w_{n-10}|$$

$$= 16|w_{n-4}| + 4|w_{n-8}| + 28|w_{n-10}| > 16 \cdot 4 \cdot 4 \cdot 4|w_{n-10}| + 4 \cdot 4|w_{n-10}| + 28|w_{n-10}|$$

$$= 1068|w_{n-10}| > 2.008^{10}|w_{n-10}|.$$

We can also easily check that $|w_n| > 2.008^n$ for all $5 \leq n \leq 14$. This means we have our result

$$|w_n| > 2.008^n \text{ for } n \geq 5.$$

□

From the basic recursion b_n alone, we get $r(n) \geq 2^n - 1$. Our new recursion gives a slightly better bound $r(n) > 2.008^n$.

Remark 2.6. *Sébastien Labbé pointed out, through private communication, that this lower bound can be improved to* 2.0634^n *by solving our recursions in a closed form without estimating them that roughly.*

2.2. An upper bound

In this subsection, we will prove an upper bound for $r(n)$. First, we will prove two useful lemmas. To this end, let us mention that every square-free palindrome has to be of odd length, because palindromes of even length create a square of two letters to the middle, for example 12011021 has a square 11 in the middle.

Lemma 2.7. *The middle letter of a rich square-free palindrome is unioccurrent.*

Proof. Since all square-free palindromes are of odd length, there always exists the middle letter. Then, suppose the contrary: $zb\widetilde{z}$ is rich and square-free and the letter b has another occurrence inside z. We can take the other occurrence of b to be consecutive to the b in the middle and then we have $zb\widetilde{z} = z_1bz_2bz_2b\widetilde{z_1}$, where z_2 is a palindrome because of Proposition 1.4. We reach a contradiction because bz_2bz_2 is a square. □

Lemma 2.8. *Suppose* $w = u_1a_1u_2a_1 \cdots a_1u_{k-1}a_1u_k \in \{a_1, a_2, \ldots, a_n\}^*$ *is rich and square-free, where* $n, k \geq 3$ *(possibly* $u_k = \epsilon$*),* $Alph(u_1) = \{a_2, \ldots, a_n\}$ *and* $\forall i : a_1 \notin Alph(u_i)$*.*

$$For\ 2 \leq i \leq k-1:\ Alph(u_{i+1}) \subseteq Alph(u_i) \setminus \{a_i\},\ where\ u_i = v_ia_i\widetilde{v_i}.$$

Proof. Since $\forall i : a_1 \notin \mathrm{Alph}(u_i)$, we get from Proposition 1.4 that $u_2, \ldots, u_{k-1}$ are palindromes, and because w is square-free, they are of odd length and non-empty. By permutating the letters, we can suppose for $2 \leq i \leq k-1$: a_i is the middle letter of $u_i = v_ia_i\widetilde{v_i}$, where $a_i \notin \mathrm{Alph}(v_i)$ by Lemma 2.7.

We will prove the claim by induction on i.

1) The base case $i = 2$. Since $a_2 \in \mathrm{Alph}(u_1) = \{a_2, \ldots, a_n\}$, we get from Proposition 1.4 that $u_1 = v_1a_2\widetilde{v_2}$. If $a_2 \in \mathrm{Alph}(u_3)$ then, by Proposition 1.4, we have $u_3 = v_2a_2v_3'$, which creates a square $(a_2\widetilde{v_2}a_1v_2)^2$ in $u_1a_1u_2a_1u_3 = v_1a_2\widetilde{v_2}a_1v_2a_2\widetilde{v_2}a_1v_2a_2v_3'$. This means $a_2 \notin \mathrm{Alph}(u_3)$.

Suppose then that $b \in \mathrm{Alph}(u_3) \setminus \mathrm{Alph}(u_2)$, which implies $b \in \mathrm{Alph}(v_1)$. The word between the first occurrence of b in u_3 and the last occurrence of b in v_1 is a palindrome by Proposition 1.4: $u_1a_1u_2a_1u_3 = t_1bt_2a_2\widetilde{v_2}a_1v_2a_2\widetilde{v_2}a_1v_2a_2\widetilde{t_2}bt_3$, where $v_1 = t_1bt_2$ and $u_3 = v_2a_2\widetilde{t_2}bt_3$. We reach a contradiction since we have a square $(a_2\widetilde{v_2}a_1v_2)^2$. This means that $\mathrm{Alph}(u_3) \subseteq \mathrm{Alph}(u_2) \setminus \{a_2\}$.

2) The induction hypothesis. We can now suppose $k \geq 4$, since the base case proves our claim if $k = 3$. Suppose then that for every j, where $2 \leq j \leq i < k-1$, we have: $\mathrm{Alph}(u_{j+1}) \subseteq \mathrm{Alph}(u_j) \setminus \{a_j\}$.

3) The induction step. Now we need to prove that $\mathrm{Alph}(u_{i+2}) \subseteq \mathrm{Alph}(u_{i+1}) \setminus \{a_{i+1}\}$. From the induction hypothesis we get $a_{i+1} \in \mathrm{Alph}(u_{i+1}) \subseteq \mathrm{Alph}(u_i) \setminus \{a_i\}$, which means $u_i = v_{i+1}a_{i+1}xa_i\widetilde{x}a_{i+1}\widetilde{v_{i+1}}$ by Proposition 1.4. If $a_{i+1} \in \mathrm{Alph}(u_{i+2})$ then, by Proposition 1.4,

we have $u_{i+2} = v_{i+1}a_{i+1}y$, which creates a square $(a_{i+1}\widetilde{v_{i+1}}a_1v_{i+1})^2$ inside $u_ia_1u_{i+1}a_1u_{i+2} = v_{i+1}a_{i+1}xa_i\widetilde{x}a_{i+1}\widetilde{v_{i+1}}a_1v_{i+1}a_{i+1}\widetilde{v_{i+1}}a_1v_{i+1}a_{i+1}y$. This means $a_{i+1} \notin \text{Alph}(u_{i+2})$

Suppose then that $c \in \text{Alph}(u_{i+2}) \setminus \text{Alph}(u_{i+1})$, which implies $c \in \text{Alph}(u_1a_1 \dots a_1u_i)$. Without loss of generality, we can assume that c is the letter from $\text{Alph}(u_{i+2}) \setminus \text{Alph}(u_{i+1})$ that has the rightmost occurrence in $u_1a_1 \dots a_1u_i$. The word between the leftmost occurrence of c in $u_{i+2} = zcz'$ and the rightmost occurrence of c in $u_1a_1 \dots a_1u_i$ has to be a palindrome by Proposition 1.4. We divide this into two cases.

- Suppose $c \notin \text{Alph}(u_i)$. Now $c\widetilde{z}a_1u_{i+1}Pu_{i+1}a_1zc$ is a palindrome, where $\text{Alph}(P) \subseteq \text{Alph}(a_1u_{i+1})$ because of the way we chose c. Now the middle letter of the palindrome $a_1u_{i+1}Pu_{i+1}a_1$ belongs to P and therefore has other occurrences inside it, in a_1u_{i+1} and in $u_{i+1}a_1$. This is a contradiction by Lemma 2.7.

- Suppose $c \in \text{Alph}(u_i)$. Now $c\widetilde{z}a_1v_{i+1}a_{i+1}\widetilde{v_{i+1}}a_1zc$ is a palindrome, where a_{i+1} is its middle letter and $c\widetilde{z}$ is a suffix of u_i. If $a_{i+1} \in \text{Alph}(z)$ then it is not unioccurrent in the palindrome $\widetilde{z}a_1v_{i+1}a_{i+1}\widetilde{v_{i+1}}a_1z$ and we reach a contradiction by Lemma 2.7. Since $a_{i+1} \in \text{Alph}(u_i)$ by the induction hypothesis, we can take the rightmost occurrence of it in u_i and get $a_{i+1}v_i'c\widetilde{z}a_1v_{i+1}a_{i+1}$ is a palindrome, where $v_i'c\widetilde{z} = \widetilde{v_{i+1}}$. We reach a contradiction since this would mean $c \in \text{Alph}(v_{i+1}) \subset \text{Alph}(u_{i+1})$.

Both cases yield a contradiction, which means $\text{Alph}(u_{i+2}) \subseteq \text{Alph}(u_{i+1}) \setminus \{a_{i+1}\}$. □

The following corollary gives another proof for the result mentioned in Remark 6 of [PS].

Corollary 2.9. *All rich square-free words are finite.*

Proof. We prove this by induction. Suppose w is rich and square-free word for which $|\text{Alph}(w)| = n \geq 4$. Suppose that all rich square-free words on an alphabet of size $n-1$ or smaller are finite. Cases $n = 1, 2, 3$ are trivial.

Suppose that a_1 is the letter of w for which $w = u_1a_1w'$, where $\text{Alph}(u_1) = \text{Alph}(w)\setminus\{a_1\}$. We partition w such that $w = u_1a_1u_2a_1u_3a_1u_4a_1\dots$, where $a_1 \notin \text{Alph}(u_i)$ for all i. From Lemma 2.8 we now get $|\text{Alph}(u_i)| > |\text{Alph}(u_{i+1})|$ for all $i \geq 2$. This means there are finitely many words u_i, at most n, and they are all over an alphabet of size $n-1$ or smaller, which concludes the proof. □

The proof of the above corollary gives us a way to get an upper bound for $r(n)$: $r(n) \leq r(n-1) + 1 + \sum_{i=1}^{n-1}(r(n-i)+1)$. This bound can be easily improved if we examine the word also from the right side, i.e. we suppose that a_1 is the letter of w for which $w = w'a_1u_1$, where $\text{Alph}(u_1) = \text{Alph}(w) \setminus \{a_1\}$. This notice makes it reasonable to make the following definition.

Definition 2.10. *Let $w = uav$ be a word, where a is a letter. If $Alph(u) = Alph(w) \setminus \{a\}$ then the leftmost occurrence of the letter a in w is called the* left special letter *of w. If*

Alph(v) = *Alph*$(w) \setminus \{a\}$ *then the rightmost occurrence of the letter* a *in* w *is called the* right special letter *of* w.

In Subsection 2.1, where we constructed the words w_n for our lower bound, the rightmost occurrence of A_n is always the right special letter of w_n and the leftmost occurrence of B_n is always the left special letter of w_n, for $n \geq 3$. In Lemma 2.8 and Corollary 2.9, the first occurrence of letter a_1 is the left special letter of w.

Before we go to our upper bound for $r(n)$, we will state a helpful lemma.

Lemma 2.11. *Suppose* $w_n = xB_nyA_nz$ *is a rich square-free* n*-ary word, where* $n \geq 3$ *and the letters* A_n *and* B_n *are the right and left special letters of* w_n*, respectively. Now* *Alph*(y) = *Alph*$(w_n) \setminus \{A_n, B_n\}$ *and* $A_n \neq B_n$.

Proof. First we prove that $A_n, B_n \notin y$. Suppose to the contrary that $B_n \in y$ (the case $A_n \in y$ is symmetric). We can take the leftmost occurrence of B_n in y and get $w_n = xB_ny_1c\widetilde{y_1}B_ny_2A_nz$, where $B_n \notin y_1c\widetilde{y_1}$ and c is a letter. Since A_n is the right special letter of w_n, we have $c \in z$. Since B_n is the left special letter of w_n, we get from Lemma 2.8 that $c \notin y_2A_nz$, i.e. $c \notin z$. This is a contradiction.

Now we prove that $A_n \neq B_n$. Suppose to the contrary that $A_n = B_n$. Now, since $A_n, B_n \notin y$, we get from Proposition 1.4 that y is a palindrome. From Lemma 2.8 we get the middle letter of y cannot be in x nor in z. This is a contradiction, since x and z has to contain all the letters except A_n.

Now we prove that if $a \in \mathrm{Alph}(w_n) \setminus \{A_n, B_n\}$ then $a \in y$. Suppose to the contrary that $a \in \mathrm{Alph}(w_n) \setminus (\{A_n, B_n\} \cup \mathrm{Alph}(y))$. Since A_n and B_n are the right and left special letters, we have $a \in x, z$. If we take the leftmost occurrence of a in z and the rightmost occurrence of a in x, then we get from Proposition 1.4 that $w = x'auB_nyA_nvaz'$, where auB_nyA_nva is a palindrome, $x = x'au$ and $z = vaz'$. The middle letter of the palindrome auB_nyA_nva cannot be inside u nor v, since it would mean $B_n \in u$ or $A_n \in v$, which is impossible since A_n and B_n are special letters. The middle letter of auB_nyA_nva cannot be inside y neither, since that would mean $B_n \in yA_n$ or $A_n \in B_ny$, which we proved above to be impossible. The only possibility is that the middle letter of auB_nyA_nva is either A_n or B_n. Since these cases are symmetric, we can suppose B_n is the middle letter. This means $w = x'a\widetilde{v}A_n\widetilde{y}B_nyA_nvaz'$. Since B_n is the left special letter of w, we have $B_n \in z = vaz'$ and $B_n \notin v$. This means $B_n \in z'$. If we take the leftmost occurrence of B_n in z', we get $w = x'a\widetilde{v}A_n\widetilde{y}B_nyA_nvav'B_nz''$, where $B_nyA_nvav'B_n$ is a palindrome which has A_n as the middle letter. This means $\widetilde{y} = vav'$ and hence $a \in y$, which is a contradiction. □

There are only three cases in which the right and left special letters can appear inside a word, with respect to each other. If w_n is a rich square-free n-ary word which has A_n and B_n

as the right and left special letters, respectively, then one of the following cases must hold (the visible occurrences of A_n and B_n in w_n are the special letters):

1) $w_n = xB_nyA_nz$. Now $A_n \neq B_n$ by Lemma 2.11.

2) $w_n = xA_nyB_nz$. Now $A_n \neq B_n$ by the definition of special letters.

3) $w_n = xA_nz = xB_nz$. Now $A_n = B_n$.

Proposition 2.12. *Suppose w_n is a rich square-free n-ary word, where $n \geq 3$.*

1) If $w_n = xB_nyA_nz$, where the letters A_n and B_n are the right and left special letters of w_n, respectively, then $|w_n| \leq 2r(n-1) + r(n-2) + 2$.

2) If $w_n = xA_nyB_nz$, where the letters A_n and B_n are the right and left special letters of w_n, respectively, then $|w_n| \leq r(n-1) + r(n-2) + r(n-3) + 2 \leq 2r(n-1)$ and $|x|, |z| \leq r(n-2) + r(n-3) + 1$, where $r(n-3) = 0$ if $n = 3$.

3) If $w_n = xA_nz = xB_nz$, where the letter $A_n = B_n$ is both the right and left special letter of w_n, then $|w_n| \leq 2r(n-1) + 1$.

Proof. Let $A = \text{Alph}(w_n)$.

1) By the definition of special letters, we have $\text{Alph}(x) = A \setminus \{B_n\}$ and $\text{Alph}(z) = A \setminus \{A_n\}$. These mean $|x|, |z| \leq r(n-1)$. From Lemma 2.11 we get $\text{Alph}(y) = A \setminus \{A_n, B_n\}$, which means $|y| \leq r(n-2)$, since $A_n \neq B_n$. Now

$$|w_n| = |x| + |B_n| + |y| + |A_n| + |z| \leq r(n-1) + 1 + r(n-2) + 1 + r(n-1) = 2r(n-1) + r(n-2) + 2.$$

2) If $A_n \notin x$, then $|x| \leq r(n-2)$. If $A_n \in x$ then we can take the rightmost occurrence of it in x and get $xA_n = x_2A_nx_1c\widetilde{x_1}A_n$, where $A_n \notin x_1c\widetilde{x_1}$ and by Lemma 2.8 $c \notin x_2A_nx_1$. Now $\text{Alph}(x_2A_nx_1) = A \setminus \{c, B_n\}$ and $\text{Alph}(\widetilde{x_1}) = A \setminus \{c, A_n, B_n\}$, where $c \neq B_n$ since B_n is the left special letter of w_n. This means $|x| = |x_2A_nx_1| + |c| + |\widetilde{x_1}| \leq r(n-2) + r(n-3) + 1$, where $r(n-3) = 0$ if $n = 3$. The same holds for z.

We have $\text{Alph}(yB_nz) = A \setminus \{A_n\}$, which means $|yB_nz| \leq r(n-1)$. Now

$$|w_n| = |x| + |A_n| + |yB_nz| \leq [r(n-2) + r(n-3) + 1] + 1 + r(n-1) = r(n-1) + r(n-2) + r(n-3) + 2.$$

From the basic recursion we know that $r(n) \geq 2r(n-1) + 1$. This means that $r(n-1) + r(n-2) + r(n-3) + 2 \leq r(n-1) + r(n-2) + 2r(n-3) + 2 \leq r(n-1) + 2r(n-2) + 1 \leq 2r(n-1)$, which we needed to prove.

3) By the definition of special letters, we have $\text{Alph}(x) = \text{Alph}(z) = A \setminus \{A_n\}$, which means $|x|, |z| \leq r(n-1)$. Now

$$|w_n| = |x| + |A_n| + |z| \leq r(n-1) + 1 + r(n-1) = 2r(n-1) + 1.$$

□

Corollary 2.13. $r(n) \leq 2r(n-1) + r(n-2) + 2$, *for* $n \geq 3$.

Proof. The claim follows from Proposition 2.12, since the proposition covered all the three different possible cases for w_n. $\square$

We do not solve the recursion $r(n) \leq 2r(n-1) + r(n-2) + 2, r(2) = 3, r(1) = 1$, in a closed-form, but we will estimate it. We use the inequality $r(n) \geq 2r(n-1) + 1$ from the basic recursion, and the fact that $r(4) = 15 > 13$. For $n \geq 8$ we have

$$r(n) \leq 2r(n-1)+r(n-2)+2 \leq 2(2r(n-2)+r(n-3)+2)+r(n-2)+2 = 5r(n-2)+2r(n-3)+6$$

$$\leq 5(2r(n-3) + r(n-4) + 2) + 2r(n-3) + 6 = 12r(n-3) + 5r(n-4) + 16$$

$$< 12r(n-3) + 5r(n-4) + 16 + (r(n-4) - 13) = 12r(n-3) + 6r(n-4) + 3 \leq 15r(n-3)$$

$$< 2.47^3 r(n-3) < 2.47^n,$$

where the last inequality comes from the fact that $r(n) < 2.47^n$ for $1 \leq n \leq 7$. Together with the lower bound, we now have $2.008^n < r(n) < 2.47^n$ for $n \geq 5$.

This upper bound can still be improved. Cases 2 and 3 from Proposition 2.12 already give better or equal upper bounds than the basic recursion, i.e. $r(n) \leq 2r(n-1) + 1$. This means we need to look closer only for the case 1.

Proposition 2.14. $r(n) \leq 5r(n-2) + 4$, *for* $n \geq 7$.

Proof. Suppose $w_n = xB_nyA_nz$ is a rich square-free n-ary word, where $n \geq 7$ and the letters A_n and B_n are the right and left special letters of w_n, respectively. This means $A_n \neq B_n$. If w_n is not of this form, then we already know from Proposition 2.12 that $|w_n| \leq 2r(n-1)+1$, which means we can use the upper bound of Corollary 2.13 and get

$$|w_n| \leq 2(2r(n-2) + r(n-3) + 2) + 1 = 4r(n-2) + 2r(n-3) + 5 \leq 5r(n-2) + 4,$$

where the last inequality comes from the basic recursion $r(n) \geq 2r(n-1) + 1$. From now on, we will use the basic recursion without mentioning it.

By the definition of special letters, we have $A_n \in x$ and $B_n \in z$. From Lemma 2.11 we know that $A_n, B_n \notin y$. Since $A_n \neq B_n$, we can take the rightmost occurrence of A_n in x and the leftmost occurrence of B_n in z and hence by Proposition 1.7 we have $w_n = x_1A_n\widetilde{y}B_nyA_n\widetilde{y}B_nz_1$.

We divide this proof into three different cases depending on whether $A_n \in x_1$ or $A_n \notin x_1$ and whether $B_n \in z_1$ or $B_n \notin z_1$.

Case 1) $A_n \notin x_1, B_n \notin z_1$.

Now we have $A_n, B_n \notin x_1, z_1, y$. This means $|x_1|, |z_1|, |y| \le r(n-2)$. Together we get

$$|w_n| = |x_1 A_n \widetilde{y} B_n y A_n \widetilde{y} B_n z_1| \le 5r(n-2) + 4.$$

Case 2) $A_n \in x_1, B_n \notin z_1$ (the case $A_n \notin x_1, B_n \in z_1$ is symmetric).

If we take the rightmost occurrence of A_n in x_1 we get, by Proposition 1.4, Lemma 2.7 and Lemma 2.8, that $w_n = x_2 A_n \widetilde{x_B} B x_B A_n \widetilde{y} B_n y A_n \widetilde{y} B_n z_1$, where B $(\neq A_n, B_n)$ is a letter, $A_n, B \notin x_B$, $B \notin x_2$ and $x_1 = x_2 A_n \widetilde{x_B} B x_B$. Since B_n is a left special letter of w_n, we have $B_n \notin x_2 A_n \widetilde{x_B}$ and $B_n \notin x_B$. We also have $A_n, B_n \notin y, z_1$. Together we have $|y|, |z_1|, |x_2 A_n \widetilde{x_B}| \le r(n-2)$ and $|x_B| \le r(n-3)$.

Let us set the left special letter of $\widetilde{y}$ to be B_{n-2}. Now we divide this into two cases depending on whether $B \neq B_{n-2}$ or $B = B_{n-2}$.

Case 2.1) $B \neq B_{n-2}$.

Since B_{n-2} is the left special letter of $\widetilde{y}$, we must have $B_{n-2} \notin x_B$. Otherwise we would have, by Proposition 1.4, that $B \in x_B$, which is impossible by Lemma 2.7. From Lemma 2.8 we now get $B_{n-2} \notin x_2$. Earlier, we already noted that $B_n, B \notin x_2 A_n \widetilde{x_B}$ and $A_n, B_n, B \notin x_B$. Together we now get $|x_2 A_n \widetilde{x_B}| \le r(n-3)$ and $|x_B| \le r(n-4)$, and therefore

$$|w_n| = |x_2 A_n \widetilde{x_B}| + |B| + |x_B| + |A_n \widetilde{y} B_n y A_n \widetilde{y} B_n| + |z_1|$$

$$\le r(n-3) + 1 + r(n-4) + [3r(n-2) + 4] + r(n-2) = 4r(n-2) + r(n-3) + r(n-4) + 5$$
$$< 4r(n-2) + r(n-3) + r(n-4) + 5 + r(n-4) \le 5r(n-2) + 3,$$

where we added the extra $r(n-4)$ after the second inequality only to make the use of the basic recursion simpler.

Case 2.2) $B = B_{n-2}$.

If we can prove that $|z_1| \le r(n-3)$, then we get

$$|w_n| = |x_2 A_n \widetilde{x_B}| + |B| + |x_B| + |A_n \widetilde{y} B_n y A_n \widetilde{y} B_n| + |z_1|$$

$$\le r(n-2) + 1 + r(n-3) + [3r(n-2) + 4] + r(n-3) = 4r(n-2) + 2r(n-3) + 5 \le 5r(n-2) + 4.$$

So we need to prove there exists some letter, different from A_n and B_n, such that it does not belong to z_1. We divide this into three cases depending on the form of $\widetilde{y}$.

Case 2.2.1) $\widetilde{y} = y_1 A_{n-2} y_3 B_{n-2} y_2$, where the letters A_{n-2} and B_{n-2} are the right and left special letters of $\widetilde{y}$, respectively.

Because $B = B_{n-2}$, we have $\widetilde{x_B} = y_1 A_{n-2} y_3$, by Proposition 1.4 and Lemma 2.7. Now $A_{n-2} \notin z_1$, since otherwise we could take the leftmost occurrence of A_{n-2} in z_1 and get a square in w_n:

$$\widetilde{y_1} A_n \widetilde{y} B_n \widetilde{y_2} B_{n-2} \widetilde{y_3} A_{n-2} \widetilde{y_1} A_n \widetilde{y} B_n \widetilde{y_2} B_{n-2} \widetilde{y_3} A_{n-2},$$

where the rightmost $\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}$ is a prefix of z_1 and the leftmost $\widetilde{y_1}$ is a suffix x_1.

Case 2.2.2) $\widetilde{y} = y_1 B_{n-2} y_2$, where B_{n-2} is also the right special letter of $\widetilde{y}$.

Because $B = B_{n-2}$, we have $\widetilde{x_B} = y_1$. Now $B_{n-2} \notin z_1$, since otherwise, similar to Case 2.2.1, we could take the leftmost occurrence of B_{n-2} in z_1 and get a square in w_n:

$$\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}.$$

Case 2.2.3) $\widetilde{y} = y_1 A_{n-2} y_3 B_{n-2} \widetilde{y_3} A_{n-2} y_3 B_{n-2} y_2$, where the rightmost A_{n-2} and the leftmost B_{n-2} are the right and left special letters of $\widetilde{y}$, respectively.

Because $B = B_{n-2}$, we have $\widetilde{x_B} = y_1 A_{n-2} y_3$. Again $A_{n-2} \notin z_1$, since otherwise, similar to Case 2.2.1, we could take the leftmost occurrence of A_{n-2} in z_1 and get a square in w_n:

$$y_3B_{n-2}\widetilde{y_3}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}y_3B_{n-2}\widetilde{y_3}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}.$$

Case 3) $A_n \in x_1, B_n \in z_1$.

If we take the rightmost occurrence of A_n in x_1 and the leftmost occurrence of B_n in z_1, we get $w_n = x_2 A_n \widetilde{x_B} B x_B A_n \widetilde{y} B_n y A_n \widetilde{y} B_n z_A A \widetilde{z_A} B_n z_2$, where A, B ($\neq A_n, B_n$) are letters and $x_1 = x_2 A_n \widetilde{x_B} B x_B$, $z_1 = z_A A \widetilde{z_A} B_n z_2$. Similar to Case 2, we have $|y|, |x_1|, |z_1|, |x_2 A_n \widetilde{x_B}|$, $|\widetilde{z_A} B_n z_2| \leq r(n-2)$ and $|x_B|, |z_A| \leq r(n-3)$.

We divide this case now into three cases depending on the form of $\widetilde{y}$.

Case 3.1) $\widetilde{y} = y_1 B_{n-2} y_2$, where B_{n-2} is both the right and left special letter of $\widetilde{y}$.

If $A = B = B_{n-2}$ then $x_B = \widetilde{y_1}$ and $z_A = \widetilde{y_2}$. This would create a square in w_n:

$$B_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}.$$

Now we divide this into two possible cases: $A, B \neq B_{n-2}$ and $A = B_{n-2}$,$B \neq B_{n-2}$.

Case 3.1.1) $A, B \neq B_{n-2}$.

Similar to Case 2.1, we get $B_n, B_{n-2}, B \notin x_2 A_n \widetilde{x_B}$ and $A_n, B_n, B_{n-2}, B \notin x_B$. In the same way, we get $A_n, A, B_{n-2} \notin \widetilde{z_A} B_n z_2$ and $A_n, A, B_n, B_{n-2} \notin z_A$. Together we have

$$|w_n| = |x_2A_n\widetilde{x_B}| + |B| + |x_B| + |A_n\widetilde{y}B_nyA_n\widetilde{y}B_n| + |z_A| + |A| + |\widetilde{z_A}B_nz_2|$$

$$\leq r(n-3)+1+r(n-4)+[3r(n-2)+4]+r(n-4)+1+r(n-3) = 3r(n-2)+2r(n-3)+2r(n-4)+6$$

$$< 3r(n-2) + 2r(n-3) + 2r(n-4) + 6 + 2r(n-4) \leq 5r(n-2) + 2.$$

Case 3.1.2) $A = B_{n-2}$ and $B \neq B_{n-2}$ (the case $A \neq B_{n-2}$ and $B = B_{n-2}$ is symmetric).

Now $z_A = \widetilde{y_2}$. Let us set $y_1 = u_1 B_{n-4} u_2$ and $y_2 = v_1 A_{n-4} v_2$, where B_{n-4} and A_{n-4} are the left special letters of y_1 and y_2, respectively.

We prove $B \neq B_{n-4}$. Suppose to the contrary that $B = B_{n-4}$. Since B_{n-2} is the right and left special letter of $\widetilde{y}$, we have $A_{n-4} \in y_1$. If we take the rightmost occurrence of A_{n-4}

in y_1 then we get from Proposition 1.4 that $A_{n-4}\widetilde{v_1}$ is a suffix of y_1 and hence A_{n-4} is the right special letter of y_1. There are now three different cases how A_{n-4} and B_{n-4} can appear inside y_1 with respect to each other. These all yield a square and hence a contradiction:

- If $y_1 = u_1' A_{n-4} u_3 B_{n-4} \widetilde{u_3} A_{n-4} u_3 B_{n-4} u_2'$, where $u_1 = u_1' A_{n-4} u_3$, $u_2 = \widetilde{u_3} A_{n-4} u_3 B_{n-4} u_2'$ and $v_1 = \widetilde{u_2'} B_{n-4} \widetilde{u_3}$, then we have a square in w_n:

$$A_{n-4}u_3B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1'}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{u_2'}B_{n-4}\widetilde{u_3}A_{n-4}u_3B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1'}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{u_2'}B_{n-4}\widetilde{u_3}.$$

- If $y_1 = u_1 B_{n-4} u_2 = u_1 A_{n-4} \widetilde{v_1}$ (i.e. $A_{n-4} = B_{n-4}$), then $u_2 = \widetilde{v_1}$ and we have a square in w_n:

$$B_{n-4}\widetilde{u_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{u_2}B_{n-4}\widetilde{u_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{u_2}.$$

- If $y_1 = u_1' A_{n-4} u_3 B_{n-4} u_2$, where $u_1 = u_1' A_{n-4} u_3$ and $v_1 = \widetilde{u_2} B_{n-4} \widetilde{u_3}$, then we have a square in w_n:

$$B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1'}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{u_2}B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1'}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{u_2}.$$

This means $B \neq B_{n-4}$. Similar to Case 2.1 we now get $B_n, B_{n-2}, B_{n-4}, B \notin x_2 A_n \widetilde{x_B}$ and $A_n, B_n, B_{n-2}, B_{n-4}, B \notin x_B$. Together we have

$$\begin{aligned} |w_n| &= |x_2A_n\widetilde{x_B}| + |B| + |x_B| + |A_n\widetilde{y}B_nyA_n\widetilde{y}B_n| + |z_A| + |B_{n-2}| + |\widetilde{z_A}B_nz_2| \\ &\leq r(n-4) + 1 + r(n-5) + [3r(n-2)+4] + r(n-3) + 1 + r(n-2) \\ &= 4r(n-2) + r(n-3) + r(n-4) + r(n-5) + 6 \\ &< 4r(n-2) + r(n-3) + r(n-4) + r(n-5) + 6 + r(n-5) \leq 5r(n-2) + 3. \end{aligned}$$

Case 3.2) $\widetilde{y} = y_1 A_{n-2} y_3 B_{n-2} y_2$, where the letters A_{n-2} and B_{n-2} are the right special letter and the left special letter of $\widetilde{y}$, respectively.

If $A = A_{n-2}$, $B = B_{n-2}$ then we would have a square in w_n:

$$A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}.$$

This means we can divide this case, similar to Case 3.1, into two different cases: $A = A_{n-2}$, $B \neq B_{n-2}$ and $A \neq A_{n-2}$, $B \neq B_{n-2}$.

Case 3.2.1) $A \neq A_{n-2}$, $B \neq B_{n-2}$.

Similar to Case 2.1, we get $B_n, B_{n-2}, B \notin x_2 A_n \widetilde{x_B}$ and $A_n, B_n, B_{n-2}, B \notin x_B$. In the same way, we get $A_n, A_{n-2}, A \notin \widetilde{z_A} B_n z_2$ and $A_n, A_{n-2}, A, B_n \notin z_A$. Together we have

$$|w_n| = |x_2A_n\widetilde{x_B}| + |B| + |x_B| + |A_n\widetilde{y}B_nyA_n\widetilde{y}B_n| + |z_A| + |A| + |\widetilde{z_A}B_nz_2|$$

$\leq r(n-3)+1+r(n-4)+[3r(n-2)+4]+r(n-4)+1+r(n-3) = 3r(n-2)+2r(n-3)+2r(n-4)+6$
$$< 3r(n-2) + 2r(n-3) + 2r(n-4) + 6 + 2r(n-4) \leq 5r(n-2) + 2.$$

Case 3.2.2) $A = A_{n-2}$, $B \neq B_{n-2}$ (the case $A \neq A_{n-2}$,$B = B_{n-2}$ is symmetric).

Now $z_A = \widetilde{y_2} B_{n-2} \widetilde{y_3}$. We divide this case into two cases: $A_{n-2} \notin y_1$ and $A_{n-2} \in y_1$.

Case 3.2.2.1) $A_{n-2} \notin y_1$.

We must have $A_{n-2} \notin x_1$. Otherwise we could take the rightmost occurrence of A_{n-2} in x_1 and get a square in w_n:

$$A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}.$$

Similar to Case 2.1, we have $B_n, B_{n-2} \notin x_1$. Since B_n and B_{n-2} are the left special letters of w_n and $\widetilde{y}$, respectively, we have $B_n, B_{n-2} \notin y_1$. Together with the previous paragraph we get $A_{n-2}, B_n, B_{n-2} \notin x_1 A_n y_1$. Since A_{n-2} is the right special letter of $\widetilde{y}$ we have $A_n, B_n, A_{n-2} \notin y_3 B_{n-2} y_2$. This all means $|x_1 A_n y_1| \leq r(n-3)$ and $|y_3 B_{n-2} y_2| \leq r(n-3)$. Together we have

$$|w_n| = |x_1 A_n y_1| + |A_{n-2}| + |y_3 B_{n-2} y_2| + |B_n y A_n \widetilde{y} B_n| + |z_A| + |A_{n-2}| + |\widetilde{z_A} B_n z_2|$$

$$\leq r(n-3) + 1 + r(n-3) + [2r(n-2) + 3] + r(n-3) + 1 + r(n-2)$$

$$= 3r(n-2) + 3r(n-3) + 5 < 3r(n-2) + 3r(n-3) + 5 + r(n-3) \leq 5r(n-2) + 3.$$

Case 3.2.2.2) $A_{n-2} \in y_1$.

If we take the rightmost occurrence of A_{n-2} in y_1, we get $\widetilde{y} = y_1' A_{n-2} y_4 B_y \widetilde{y_4} A_{n-2} y_3 B_{n-2} y_2$, where B_y is a letter, $y_1 = y_1' A_{n-2} y_4 B_y \widetilde{y_4}$ and $A_{n-2} \notin y_4 B_y \widetilde{y_4}$. Let us set $y_3 = u_1 B_{n-4} u_2$, where B_{n-4} is the left special letter of y_3. We will prove $B_{n-4} \notin x_1$.

Suppose $B_{n-4} \notin y_1$. Now $B_{n-4} \notin x_1$, since otherwise we could take the rightmost occurrence of B_{n-4} in x_1 and get a square in w_n:

$$A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}.$$

Suppose $B_{n-4} \in y_1$. Because of Lemma 2.7, we must have $B_y = B_{n-4}$ and $y_4 = \widetilde{u}_1$. Also now $B_{n-4} \notin x_1$, since otherwise we would have a square in w_n:

$$B_{n-4}\widetilde{u}_1 A_{n-2}\widetilde{y'}_1 A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}\widetilde{u}_1 B_{n-4}\widetilde{u}_1 A_{n-2}\widetilde{y'}_1 A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}\widetilde{u}_1.$$

This means we have $B_{n-4} \notin x_1$.

If $B_y = B_{n-4}$ then we get from Lemma 2.8 that $B_{n-4} \notin y_1' A_{n-2} y_4$. If $B_y \neq B_{n-4}$ then, since B_{n-4} is the left special letter of y_3, we also get from Lemma 2.7 and 2.8 that $B_{n-4} \notin y_1' A_{n-2} y_4$. These mean $B_{n-4} \notin x_1 A_n y_1' A_{n-2} y_4$.

From Lemma 2.7 we get $B_y \notin \widetilde{y_4}$, which means $A_n, A_{n-2}, B_n, B_{n-2}, B_y \notin \widetilde{y_4}$. Since A_{n-2} is the right special letter of $\widetilde{y}$, we have $A_n, A_{n-2}, B_n \notin y_3 B_{n-2} y_2$. Together we have

$$|w_n| = |x_1 A_n y_1' A_{n-2} y_4| + |B_y| + |\widetilde{y_4}| + |A_{n-2}| + |y_3 B_{n-2} y_2| + |B_n y A_n \widetilde{y} B_n| + |z_A| + |A_{n-2}| + |\widetilde{z_A} B_n z_2|$$

$$\leq r(n-3) + 1 + r(n-5) + 1 + r(n-3) + [2r(n-2) + 3] + r(n-3) + 1 + r(n-2)$$

$$= 3r(n-2) + 3r(n-3) + r(n-5) + 6 < 3r(n-2) + 3r(n-3) + r(n-5) + 6 + 3r(n-5) \leq 5r(n-2) + 1.$$

Case 3.3) $\widetilde{y} = y_1 A_{n-2} y_3 B_{n-2} \widetilde{y_3} A_{n-2} y_3 B_{n-2} y_2$, where the rightmost A_{n-2} and the leftmost B_{n-2} are the right and left special letters of $\widetilde{y}$, respectively.

If $A = A_{n-2}$, $B = B_{n-2}$ then we would have a square in w_n:

$$B_{n-2} \widetilde{y_3} A_{n-2} \widetilde{y_1} A_n \widetilde{y} B_n \widetilde{y_2} B_{n-2} \widetilde{y_3} A_{n-2} y_3 B_{n-2} \widetilde{y_3} A_{n-2} \widetilde{y_1} A_n \widetilde{y} B_n \widetilde{y_2} B_{n-2} \widetilde{y_3} A_{n-2} y_3.$$

This means we can divide this case, similar to Case 3.1, into two different cases: $A = A_{n-2}$, $B \neq B_{n-2}$ and $A \neq A_{n-2}$, $B \neq B_{n-2}$.

Case 3.3.1) $A \neq A_{n-2}$ and $B \neq B_{n-2}$.

Similar to Case 2.1, we get $B_n, B_{n-2}, B \notin x_2 A_n \widetilde{x_B}$ and $A_n, B_n, B_{n-2}, B \notin x_B$. In the same way, we get $A_n, A_{n-2}, A \notin \widetilde{z_A} B_n z_2$ and $A_n, A_{n-2}, A, B_n \notin z_A$. Together we have

$$|w_n| = |x_2 A_n \widetilde{x_B}| + |B| + |x_B| + |A_n \widetilde{y} B_n y A_n \widetilde{y} B_n| + |z_A| + |A| + |\widetilde{z_A} B_n z_2|$$

$$\leq r(n-3) + 1 + r(n-4) + [3r(n-2) + 4] + r(n-4) + 1 + r(n-3) = 3r(n-2) + 2r(n-3) + 2r(n-4) + 6$$

$$< 3r(n-2) + 2r(n-3) + 2r(n-4) + 6 + 2r(n-4) \leq 5r(n-2) + 2.$$

Case 3.3.2) $A = A_{n-2}$, $B \neq B_{n-2}$ (the case $A \neq A_{n-2}$, $B = B_{n-2}$ is symmetric).

Let A_{n-4} be the right special letter of y_3. We will divide this into two cases: $A_{n-4} \notin y_2$ and $A_{n-4} \in y_2$.

Case 3.3.2.1) $A_{n-4} \notin y_2$.

If $A_{n-4} \in z_2$ then we could take the leftmost occurrence of it in z_2, which would create a square in w_n:

$$\widetilde{y_3} A_{n-2} y_3 B_{n-2} y_2 B_n \widetilde{y_2} B_{n-2} \widetilde{y_3} A_{n-2} y_3 B_{n-2} y_2 B_n \widetilde{y_2} B_{n-2}.$$

This means $A_{n-4} \notin z_2$. Let us now mark $y_3 = u_1 A_{n-4} u_2$, where the letter A_{n-4} is the right special letter. We get $A_{n-4} \notin u_2 B_{n-2} y_2 B_n z_2$. Similar to Case 2.1, we also have $A_n, A_{n-2} \notin u_2 B_{n-2} y_2 B_n z_2$. From Proposition 2.12 we get $|u_1| \leq r(n-5) + r(n-6) + 1$. Similar to Case 2.1, we get $B_n, B_{n-2}, B \notin x_2 A_n \widetilde{x_B}$ and $A_n, B_n, B_{n-2}, B \notin x_B$. Together we have

$$|w_n| = |x_2 A_n \widetilde{x_B}| + |B| + |x_B| + |A_n \widetilde{y} B_n y A_n \widetilde{y} B_n| + |\widetilde{y_2} B_{n-2} \widetilde{u_2}| + |A_{n-4} \widetilde{u_1} A_{n-2} u_1 A_{n-4}| + |u_2 B_{n-2} y_2 B_n z_2|$$

$$\leq r(n-3)+1+r(n-4)+[3r(n-2)+4]+r(n-4)+[2(r(n-5)+r(n-6)+1)+3]+r(n-3)$$
$$= 3r(n-2)+2r(n-3)+2r(n-4)+2r(n-5)+2r(n-6)+10$$
$$< 3r(n-2)+2r(n-3)+2r(n-4)+2r(n-5)+2r(n-6)+10+2r(n-6) \leq 5r(n-2)+2.$$

Case 3.3.2.2) $A_{n-4} \in y_2$.

We will divide this case into three cases depending on the form of y_3.

Case 3.3.2.2.1) $y_3 = u_1 A_{n-4} u_3 B_{n-4} u_2$, where A_{n-4} and B_{n-4} are the right and left special letters of y_3, respectively.

Since $A_{n-4} \in y_2$, we have $y_2 = \widetilde{u_2} B_{n-4} \widetilde{u_3} A_{n-4} y_2'$, where the A_{n-4} is the leftmost occurrence of A_{n-4} in y_2. If $B_{n-4} \in y_1$ then $y_1 = y_1' B_{n-4} \widetilde{u_3} A_{n-4} \widetilde{u_1}$, where the B_{n-4} is the rightmost occurrence of B_{n-4} in y_1. This would create a square in $\widetilde{y}$:

$$B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1}A_{n-2}y_3B_{n-2}\widetilde{u_2}B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1}A_{n-2}y_3B_{n-2}\widetilde{u_2}.$$

So $B_{n-4} \notin y_1$. Now, if $B = B_{n-4}$ then $x_B = \widetilde{u_3} A_{n-4} \widetilde{u_1} A_{n-2} \widetilde{y_1}$ by Lemma 2.7, since $B_{n-4} \notin y_1$. This would create a square in w_n:

$$B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}y_3B_{n-2}\widetilde{u_2}$$
$$B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}y_3B_{n-2}\widetilde{u_2}.$$

So $B \neq B_{n-4}$. This means that, in similar way as in Case 2.1, we get $B_n, B_{n-2}, B_{n-4}, B \notin x_2 A_n \widetilde{x_B}$ and $A_n, B_n, B_{n-2}, B_{n-4}, B \notin x_B$. Together we have

$$|w_n| = |x_2A_n\widetilde{x_B}| + |B| + |x_B| + |A_n\widetilde{y}B_nyA_n\widetilde{y}B_n| + |z_A| + |A_{n-2}| + |\widetilde{z_A}B_nz_2|$$
$$\leq r(n-4)+1+r(n-5)+[3r(n-2)+4]+r(n-3)+1+r(n-2)$$
$$= 4r(n-2)+r(n-3)+r(n-4)+r(n-5)+5$$
$$< 4r(n-2)+r(n-3)+r(n-4)+r(n-5)+5+r(n-5) \leq 5r(n-2)+2.$$

Case 3.3.2.2.2) $y_3 = u_1 B_{n-4} u_2$, where B_{n-4} is both the right and left special letter.

This case is very similar to the previous, Case 3.3.2.2.1.

Now B_{n-4} is both the right and left special letter, which means $A_{n-4} = B_{n-4}$. Since this case is a subcase of Case 3.3.2.2, we have $A_{n-4} = B_{n-4} \in y_2$, which means $y_2 = \widetilde{u_2} B_{n-4} y_2'$. If $B_{n-4} \in y_1$ then $y_1 = y_1' B_{n-4} \widetilde{u_1}$ and we would have a square in $\widetilde{y}$:

$$B_{n-4}\widetilde{u_1}A_{n-2}y_3B_{n-2}\widetilde{u_2}B_{n-4}\widetilde{u_1}A_{n-2}y_3B_{n-2}\widetilde{u_2}.$$

So $B_{n-4} \notin y_1$. If $B = B_{n-4}$ then $x_B = \widetilde{u_1} A_{n-2} \widetilde{y_1}$. This would create a square in w_n:

$$B_{n-4}\widetilde{u_1}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}y_3B_{n-2}\widetilde{u_2}$$

$$B_{n-4}\widetilde{u_1}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}y_3B_{n-2}\widetilde{u_2}.$$

So $B \neq B_{n-4}$. This means that, in similar way as in Case 2.1, we get $B_n, B_{n-2}, B_{n-4}, B \notin x_2A_n\widetilde{x_B}$ and $A_n, B_n, B_{n-2}, B_{n-4}, B \notin x_B$. Again, we have

$$|w_n| = |x_2A_n\widetilde{x_B}| + |B| + |x_B| + |A_n\widetilde{y}B_nyA_n\widetilde{y}B_n| + |z_A| + |A_{n-2}| + |\widetilde{z_A}B_nz_2|$$

$$\leq r(n-4) + 1 + r(n-5) + [3r(n-2)+4] + r(n-3) + 1 + r(n-2)$$
$$= 4r(n-2) + r(n-3) + r(n-4) + r(n-5) + 5$$
$$< 4r(n-2) + r(n-3) + r(n-4) + r(n-5) + 5 + r(n-5) \leq 5r(n-2) + 2.$$

Case 3.3.2.2.3) $y_3 = u_1A_{n-4}u_3B_{n-4}\widetilde{u_3}A_{n-4}u_3B_{n-4}u_2$, where the rightmost A_{n-4} and the leftmost B_{n-4} are the right and left special letters of y_3, respectively.

We divide this case into two subcases: $B_{n-4} \notin y_1$ and $B_{n-4} \in y_1$.

Case 3.3.2.2.3.1) $B_{n-4} \notin y_1$.

Now $B \neq B_{n-4}$, since otherwise we would have a square in w_n:

$$A_{n-4}u_3B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}y_3B_{n-2}\widetilde{u_2}B_{n-4}\widetilde{u_3}$$

$$A_{n-4}u_3B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1}A_{n-2}\widetilde{y_1}A_n\widetilde{y}B_n\widetilde{y_2}B_{n-2}\widetilde{y_3}A_{n-2}y_3B_{n-2}\widetilde{u_2}B_{n-4}\widetilde{u_3}.$$

Similar to Case 2.1, we get $B_n, B_{n-2}, B_{n-4}, B \notin x_2A_n\widetilde{x_B}$ and $A_n, B_n, B_{n-2}, B_{n-4}, B \notin x_B$. Again, we have

$$|w_n| = |x_2A_n\widetilde{x_B}| + |B| + |x_B| + |A_n\widetilde{y}B_nyA_n\widetilde{y}B_n| + |z_A| + |A_{n-2}| + |\widetilde{z_A}B_nz_2|$$

$$\leq r(n-4) + 1 + r(n-5) + [3r(n-2)+4] + r(n-3) + 1 + r(n-2)$$
$$= 4r(n-2) + r(n-3) + r(n-4) + r(n-5) + 5$$
$$< 4r(n-2) + r(n-3) + r(n-4) + r(n-5) + 5 + r(n-5) \leq 5r(n-2) + 2.$$

Case 3.3.2.2.3.2) $B_{n-4} \in y_1$.

Now $y_1 = y_1'B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1}$, where the B_{n-4} is the rightmost occurrence of B_{n-4} in y_1, and $y_2 = \widetilde{u_2}B_{n-4}\widetilde{u_3}A_{n-4}y_2'$, where the A_{n-4} is the leftmost occurrence of A_{n-4} in y_2. Remember that we really have $A_{n-4} \in y_2$, since this is a subcase of Case 3.3.2.2.

If $A_{n-2} \in y_1$ then we can take the rightmost occurrence of A_{n-2} in y_1' and get $y_1 = y_1''A_{n-2}u_1A_{n-4}u_3B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1}$, which creates a square in $\widetilde{y}$:

$$A_{n-4}u_3B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1}A_{n-2}y_3B_{n-2}\widetilde{u_2}B_{n-4}\widetilde{u_3}A_{n-4}u_3B_{n-4}\widetilde{u_3}A_{n-4}\widetilde{u_1}A_{n-2}y_3B_{n-2}\widetilde{u_2}B_{n-4}\widetilde{u_3}.$$

This means $A_{n-2} \notin y_1$.

Now we divide this case into two subcases: $B \neq A_{n-2}$ and $B = A_{n-2}$.

Case 3.3.2.2.3.2.1) $B \neq A_{n-2}$.

Now, in similar way as in Case 2.1, we again get $A_{n-2}, B_n, B_{n-2}, B \notin x_2 A_n \widetilde{x_B}$ and $A_n, A_{n-2}, B_n, B_{n-2}, B \notin x_B$. Again, we have

$$|w_n| = |x_2 A_n \widetilde{x_B}| + |B| + |x_B| + |A_n \widetilde{y} B_n y A_n \widetilde{y} B_n| + |z_A| + |A_{n-2}| + |\widetilde{z_A} B_n z_2|$$

$$\leq r(n-4) + 1 + r(n-5) + [3r(n-2) + 4] + r(n-3) + 1 + r(n-2)$$

$$= 4r(n-2) + r(n-3) + r(n-4) + r(n-5) + 5$$

$$< 4r(n-2) + r(n-3) + r(n-4) + r(n-5) + 5 + r(n-5) \leq 5r(n-2) + 2.$$

Case 3.3.2.2.3.2.2) $B = A_{n-2}$.

Now we have $x_B = \widetilde{y_1} = u_1 A_{n-4} u_3 B_{n-4} \widetilde{y_1'}$. We will first show that $A_{n-4} \notin y_1', x_2$ and $B_{n-4} \notin y_2', z_2$.

If $A_{n-4} \in y_1'$ then we have $y_1 = y_1'' A_{n-4} u_3 B_{n-4} \widetilde{u_3} A_{n-4} \widetilde{u_1}$. This creates a square in w_n:

$$u_3 B_{n-4} \widetilde{u_3} A_{n-4} \widetilde{u_1} A_{n-2} \widetilde{y_1} A_n \widetilde{y} B_n \widetilde{y_2} B_{n-2} \widetilde{y_3} A_{n-2} y_3 B_{n-2} \widetilde{u_2} B_{n-4} \widetilde{u_3} A_{n-4}$$

$$u_3 B_{n-4} \widetilde{u_3} A_{n-4} \widetilde{u_1} A_{n-2} \widetilde{y_1} A_n \widetilde{y} B_n \widetilde{y_2} B_{n-2} \widetilde{y_3} A_{n-2} y_3 B_{n-2} \widetilde{u_2} B_{n-4} \widetilde{u_3} A_{n-4}.$$

So $A_{n-4} \notin y_1'$. If $B_{n-4} \in y_2'$ then we have $y_2 = \widetilde{u_2} B_{n-4} \widetilde{u_3} A_{n-4} u_3 B_{n-4} y_2''$. Also this creates a square in w_n:

$$B_{n-4} \widetilde{u_3} A_{n-4} \widetilde{u_1} A_{n-2} \widetilde{y_1} A_n \widetilde{y} B_n \widetilde{y_2} B_{n-2} \widetilde{y_3} A_{n-2} y_3 B_{n-2} \widetilde{u_2} B_{n-4} \widetilde{u_3} A_{n-4} u_3$$

$$B_{n-4} \widetilde{u_3} A_{n-4} \widetilde{u_1} A_{n-2} \widetilde{y_1} A_n \widetilde{y} B_n \widetilde{y_2} B_{n-2} \widetilde{y_3} A_{n-2} y_3 B_{n-2} \widetilde{u_2} B_{n-4} \widetilde{u_3} A_{n-4} u_3.$$

So $B_{n-4} \notin y_2'$. If $A_{n-4} \in x_2$ then we could take the rightmost occurrence of A_{n-4} in x_2 and get a square in w_n:

$$A_{n-4} u_3 B_{n-4} \widetilde{y_1'} A_n y_1' B_{n-4} \widetilde{u_3} A_{n-4} \widetilde{u_1} A_{n-2} u_1 A_{n-4} u_3 B_{n-4} \widetilde{y_1'} A_n y_1' B_{n-4} \widetilde{u_3} A_{n-4} \widetilde{u_1} A_{n-2} u_1.$$

So $A_{n-4} \notin x_2$. If $B_{n-4} \in z_2$ then we could take the leftmost occurrence of B_{n-4} in z_2 and get a square in w_n:

$$u_2 B_{n-2} \widetilde{y_3} A_{n-2} y_3 B_{n-2} \widetilde{u_2} B_{n-4} \widetilde{u_3} A_{n-4} y_2' B_n \widetilde{y_2'} A_{n-4} u_3 B_{n-4}$$

$$u_2 B_{n-2} \widetilde{y_3} A_{n-2} y_3 B_{n-2} \widetilde{u_2} B_{n-4} \widetilde{u_3} A_{n-4} y_2' B_n \widetilde{y_2'} A_{n-4} u_3 B_{n-4}.$$

So $B_{n-4} \notin z_2$. Now we know that $A_{n-4} \notin x_2 A_n y_1' B_{n-4} \widetilde{u_3}$ and $B_{n-4} \notin \widetilde{u_3} A_{n-4} y_2' B_n z_2$.

Similar to Case 2.1, we get $A_{n-2}, B_n, B_{n-2} \notin x_2 A_n y_1' B_{n-4} \widetilde{u_3}$ and $A_n, A_{n-2}, B_n, B_{n-2} \notin u_3 B_{n-4} \widetilde{y_1'}$ and $A_n, A_{n-2} \notin \widetilde{u_3} A_{n-4} y_2' B_n z_2$. From Proposition 2.12 we get $|u_1|, |u_2| \leq r(n-6)+r(n-7)+1$, where $r(n-7)=0$ if $n=7$. Since $A_n, B_n \notin y$ and A_{n-2} is the right special letter of $\widetilde{y}$, we trivially have $A_n, A_{n-2}, B_n \notin \widetilde{y_2} B_{n-2} \widetilde{y_3}$. From Lemma 2.11 we also get easily that $A_n, A_{n-2}, B_n, B_{n-2} \notin y_3$. Together we finally have

$$|w_n| = |x_2 A_n y_1' B_{n-4} \widetilde{u_3}| + |A_{n-4} \widetilde{u_1} A_{n-2} u_1 A_{n-4}| + |u_3 B_{n-4} \widetilde{y_1'}| + |A_n \widetilde{y} B_n y A_n \widetilde{y} B_n|$$

$$+|\widetilde{y_2} B_{n-2} \widetilde{y_3}| + |A_{n-2}| + |y_3| + |B_{n-2}| + |\widetilde{u_2}| + |B_{n-4}| + |\widetilde{u_3} A_{n-4} y_2' B_n z_2|$$

$$\leq r(n-4)+[2r(n-6)+2r(n-7)+5]+r(n-5)+[3r(n-2)+4]+r(n-3)+1+r(n-4)+1+$$
$$[r(n-6)+r(n-7)+1]+1+r(n-3) = 3r(n-2)+2r(n-3)+2r(n-4)+r(n-5)+3r(n-6)+3r(n-7)+13$$
$$< 3r(n-2)+2r(n-3)+2r(n-4)+r(n-5)+3r(n-6)+3r(n-7)+13+r(n-6)+r(n-7) \leq 5r(n-2)+2.$$

□

As we can see, improving our upper bound was very exhausting. If we would like to achieve Conjecture 2.3, we would need to use a slightly different approach.

Let us still estimate our upper bound in a closed form. Suppose first $n \geq 7$ is even:

$$r(n) \leq 5r(n-2)+4 \leq 5(5r(n-4)+4)+4 \leq \ldots \leq 5^{(n-6)/2} r(6) + 4(5^{(n-8)/2} + \ldots + 5 + 1)$$
$$< 5^{(n-6)/2} \cdot (5^3 - 58) + (5^{(n-8)/2+1} + \ldots + 5) = 5^{n/2} - 58 \cdot 5^{(n-6)/2} + (5^{(n-8)/2+1} + \ldots + 5) < 5^{n/2} < 2.237^n.$$

Suppose now that $n \geq 7$ is odd:

$$r(n) \leq 5r(n-2)+4 \leq 5(5r(n-4)+4)+4 \leq \ldots \leq 5^{(n-5)/2} r(5) + 4(5^{(n-7)/2} + \ldots + 5 + 1)$$
$$< 5^{(n-5)/2} \cdot (5^{2.5} - 22) + (5^{(n-7)/2+1} + \ldots + 5) = 5^{n/2} - 22 \cdot 5^{(n-5)/2} + (5^{(n-7)/2+1} + \ldots + 5) < 5^{n/2} < 2.237^n.$$

Together with the lower bound, we finally get $2.008^n < r(n) < 2.237^n$, for $n \geq 5$.

Remark 2.15. *Sébastien Labbé pointed out, through private communication, that our upper bound 2.237^n can be improved to 2.21432^n by solving our recursions in a closed form without estimating them that roughly.*

Acknowledgements

I want to thank Š. Starosta for making me aware of this problem, L. Zamboni and T. Harju for insightful comments, and J. Peltomäki for calculating the exact value of $r(7)$ and adding all the values of $r(n)$ up to $n = 7$ to the OEIS database https://oeis.org/A269560. I also want to thank my PhD thesis pre-examiners Amy Glen and Sébastien Labbé for their comments which greatly improved the quality of this paper.

References

[AFMP] P. Ambrož, C. Frougny, Z. Masáková, E. Pelantová: *Palindromic complexity of infinite words associated with simple Parry numbers*, Ann. Inst. Fourier (Grenoble) 56 (2006), pp. 2131-2160.

[BDGZ1] M. Bucci, A. De Luca, A. Glen, L. Q. Zamboni: *A connection between palindromic and factor complexity using return words*, Advances in Applied Mathematics 42 (2009), pp. 60-74.

[BDGZ2] M. Bucci, A. De Luca, A. Glen, L. Q. Zamboni: *A new characteristic property of rich words*, Theoret. Comput. Sci. 410 (2009), pp. 2860-2863.

[BHNR] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer: *On the palindromic complexity of infinite words*, Internat. J. Found. Comput. 15 (2004), pp. 293-306.

[CD] A. Carpi, A. de Luca: *Special factors, periodicity, and an application to Sturmian words*, Acta Informatica 36 (2000), pp. 983-1006.

[CR] J. D. Currie, N. Rampersad: *A proof of Dejeans conjecture*, Math. Comput. 80(274) (2011), pp. 1063-1070.

[D] F. Dejean: *Sur un théorème de Thue*, J. Combin. Theory Ser. A 13 (1972), pp. 90-99.

[DGZ] A. de Luca, A. Glen, L. Q. Zamboni: *Rich, Sturmian, and trapezoidal words*, Theoret. Comput. Sci. 407 (2008), pp. 569-573.

[DJP] X. Droubay, J. Justin, G. Pirillo: *Episturmian words and some constructions of de Luca and Rauzy*, Theoret. Comput. Sci. 255 (2001), pp. 539-553.

[GJ] A. Glen, J. Justin: *Episturmian words: A survey*, Theor. Inform. Appl. 43(3) (2009), pp. 403-442.

[GJWZ] A. Glen, J. Justin, S. Widmer, L. Q. Zamboni: *Palindromic richness*, European Journal of Combinatorics 30 (2009), pp. 510-531.

[HPS] K. Hare, H. Prodinger, J. Shallit: *Three Series for the Generalized Golden Mean*, Fibonacci Quart. 52 (2014), pp. 307-313.

[Lot1] M. Lothaire: *Combinatorics on Words,* in: Encyclopedia of Mathematics and its Applications, vol 17, Addison-Wesley, Reading, MA, 1983.

[Lot2] M. Lothaire: *Algebraic Combinatorics on Words,* in: Encyclopedia of Mathematics and its Applications, vol 90, Cambridge University Press, UK, 2002.

[MP] F. Mignosi, G. Pirillo: *Repetitions in the Fibonacci infinite word*, Informatique théorique et applications, tome 26, n. 3 (1992), pp. 199-204.

[PS] E. Pelantová, Š. Starosta: *Languages invariant under more symmetries: overlapping factors versus palindromic richness*, Discrete Math. 313 (2013), pp. 2432-2445.

[Q] M. Queffélec: *Substitution Dynamical Systems - Spectral Analysis,* Lecture Notes in Mathematics. 2nd ed. 2010. Springer-Verlag, Berlin.

[R] M. Rao: *Last cases of Dejeans conjecture*, Theoret. Comput. Sci. 412(27) (2011), pp. 3010-3018.

[RR] A. Restivo, G. Rosone: *Burrows-Wheeler transform and palindromic richness*, Theoret. Comput. Sci. 410 (2009), pp. 3018-3026.

[T1] A. Thue: *Uber unendliche Zeichenvihen*, Kra. Vidensk. Selsk. Skrifter. I Mat.-Nat.Kl., Christiana, Nr. 7, 1906.

[T2] A. Thue: *Uber die gegenseitige Lage gleicher Teile gewisser Zeichenreihen*, Kra. Vidensk. Selsk. Skrifter. I Mat.-Nat.Kl., Christiana, Nr. 12, 1912.

[V] J. Vesti: *Extensions of rich words*, Theoret. Comput. Sci. 548 (2014), pp. 14-24.

Publication IV

Jetro Vesti, *The most unbalanced words* $0^{q-p}1^p$ *and majorization*, Discrete Mathematics, Algorithms and Applications, Vol. 7, No. 3, 1550028, 22 pages, 2015, DOI 10.1142/S1793830915500287. Revised version of the preprint.

The webpage of the journal:
http://www.worldscientific.com/worldscinet/dmaa

The most unbalanced words $0^{q-p}1^p$ and majorization

Jetro Vesti

University of Turku, Department of Mathematics and Statistics, 20014 Turku, Finland

---

Abstract

A finite word $w \in \{0,1\}^*$ is (cyclically) *balanced* if for every equal-length factors u and v of every cyclic shift of w we have $||u|_1 - |v|_1| \leq 1$. This new notion of balanced words was defined in [3].

In [2], the authors considered finite balanced words and majorization. One of the main results was that the base-2 orbit of the balanced word is the least element in the set of orbits with respect to partial sum. It was also proved that the product of the elements in the base-2 orbit of a word is maximized precisely when the word is balanced.

It turns out that the words $0^{q-p}1^p$ have similar extremal properties, opposite to the balanced words, which makes it meaningful to call these words *the most unbalanced words*. This article contains the analogues of the results mentioned above. We will prove that the orbit of the word $u = 0^{q-p}1^p$, where p and q are integers with $1 \leq p < q$, is the greatest element in the set of orbits with respect to partial sum and that it has the smallest product. We will also prove that u is the greatest element in the set of orbits with respect to partial product.

Keywords: Combinatorics on words, Balanced word, Majorization.
2000 MSC: 68R15

---

1. Introduction

Sturmian words were first studied by Morse and Hedlund in [6], since then being one of the core interests in combinatorics on words, and *finite balanced words* were first introduced and studied in [3]. These words are closely linked because every Sturmian word is *balanced*, as an infinite word, and every finite balanced word is a factor of some Sturmian word. Every factor of a Sturmian word is not however necessarily a finite balanced word. Notice that our

Email address: `jejove@utu.fi` (Jetro Vesti)

definition of balanced words in this article is not the usual one (see Chapter 2 of [5]). For a well-known survey on Sturmian words one should look [1].

Besides balanced words, this paper mostly concerns *the most unbalanced words* $0^{q-p}1^p$. In [2] the authors proved many new properties for balanced words in terms of majorization. Majorization is a common notion in many branches of mathematics and has many applications, for example in probability, statistics and graph theory. We will notice that also the words 0^p1^{q-p} have many extremal properties in terms of majorization, opposite to the balanced words. This makes it meaningful to call these words the most unbalanced words.

In Section 2 we will give an analogue of Theorem 2.3 from [2], which says that the base-2 orbit of the balanced word is the least element in the set of orbits with respect to partial sum. In this article we will prove that the base-2 orbit of the most unbalanced word is the greatest element. Hence, this result places every other word, with the same number of ones and zeros, between these two extremal words.

In Section 3 we will give an analogue of Theorem 1.2 from [2], which says that the product of the elements in the base-2 orbit of a word is maximized precisely when the word is balanced. In this article we will prove that the product is minimized when the word is the most unbalanced word. This is done in the case where the number of zeros is greater than the number of ones.

In Section 4 we will prove a similar result for partial product as we proved for partial sum in Section 2. The result that the base-2 orbit of the balanced word is the least element also with respect to partial product has not been proved, but it seems very likely to be true. In any case, we will prove that the base-2 orbit of the most unbalanced word is the greatest element also with respect to partial product. For this, we will use the result from Section 3, which means that the result is proved in the case where the number of zeros is greater than the number of ones.

1.1. Definitions and notation

We denote by A an *alphabet*, i.e. a non-empty finite set of symbols called *letters*. A *word* w over A is a finite sequence $w = w_1w_2 \dots w_n$, where $\forall i : w_i \in A$. The *empty word* ϵ is the empty sequence. The set A^* of all words over A is a free monoid under the operation of concatenation with identity element ϵ and set of generators A. The free semigroup $A^+ = A^* \setminus \{\epsilon\}$ is the set of non-empty words over A.

The *length* of a word $w = w_1w_2 \dots w_n \in A^n$ is denoted by $|w| = n$. The empty word is the unique word of length 0. By $|w|_a$, where $a \in A$, we denote the number of occurrences of the letter a in w. A word x is a *factor* of a word $w \in A^*$ if $w = uxv$, for some $u, v \in A^*$. If $u = \epsilon$ ($v = \epsilon$) then we say that x is a *prefix* (resp. *suffix*) of w. The set $\mathrm{F}(w)$ is the set of all factors of w and the set $\mathrm{Alph}(w)$ is the set of all letters that occur in w.

Other basic definitions and notation in combinatorics on words can be found in [4] and [5].

1.2. Preliminaries

In this paper we will restrict ourselves to a binary alphabet, i.e. $A = \{0, 1\}$, where we define that $0 < 1$. The *lexicographic order* on words $u = u_1 \dots u_n$ and $v = v_1 \dots v_n$ in A^n is defined by: $u < v$ if there exists $j \in \{1, \dots, n\}$ such that $u_k = v_k$ for all $k = 1, \dots, j-1$ and $u_j < v_j$. We denote $u \leq v$ if either $u < v$ or $u = v$. The *cyclic shift* $\sigma : A^n \to A^n$ is defined by $\sigma(w_1 \dots w_n) = w_2 \dots w_n w_1$. The *orbit* $\mathcal{O}(w)$ of a word $w \in A^n$ is the vector

$$\mathcal{O}(w) = (\mathcal{O}_1(w), \dots, \mathcal{O}_n(w)),$$

where the words $\mathcal{O}_i(w)$ are the iterated cyclic shifts $w, \sigma(w), \dots, \sigma^{n-1}(w)$ arranged in lexicographic order from the smallest to the largest. We will set $(w)_2 = \sum_{i=1}^{n} w_i 2^{n-i}$ for a word $w = w_1 w_2 \dots w_n$ and define the *base-2 orbit* of w by

$$\mathcal{I}(w) = (\mathcal{I}_1(w), \dots, \mathcal{I}_n(w)) = ((\mathcal{O}_1(w))_2, \dots, (\mathcal{O}_n(w))_2).$$

A finite word $w \in \{0, 1\}^*$ is (cyclically) *balanced* if for every equal-length factors u and v of every cyclic shift of w we have $||u|_1 - |v|_1| \leq 1$. If a word is not balanced then it is *unbalanced*. Notice that for example the word 001010 is not balanced even though it is a factor of a Sturmian word and hence a factor of an *infinite* balanced word.

Let p and q be coprime integers such that $1 \leq p < q$. $\mathcal{W}_{p,q}$ will denote the set of binary words $w \in A^q$ such that $|w|_1 = p$ and $|w|_0 = q - p$. From [1] we know that there are q balanced words in $\mathcal{W}_{p,q}$ and they are all in the same orbit. We define $\mathbb{W}_{p,q}$ to be the set of all orbits in $\mathcal{W}_{p,q}$ and get that there is a unique balanced orbit in each $\mathbb{W}_{p,q}$. Because the orbit depends only on one of its components, we will use the lexicographically smallest component $\mathcal{O}_1(w)$ to represent the orbit. This smallest component is a Lyndon word. For example the orbit $(00101, 01001, 01010, 10010, 10100)$ will be represented notationally by 00101.

Example 1.1. *If* $(p, q) = (2, 5)$ *then the set of all orbits is* $\mathbb{W}_{2,5} = \{00011, 00101\}$*, where* $00011 = (00011, 00110, 01100, 10001, 11000)$*,* $00101 = (00101, 01001, 01010, 10010, 10100)$*. The base-2 orbits are* $\mathcal{I}(00011) = (3, 6, 12, 17, 24)$ *and* $\mathcal{I}(00101) = (5, 9, 10, 18, 20)$*.*

For $w, w' \in \mathbb{W}_{p,q}$ the base-2 orbit $\mathcal{I}(w)$ of w is said to *majorize* (notice that we use the same definition as in [2]) the base-2 orbit $\mathcal{I}(w')$ of w', denoted $w' \prec w$, if

$$\sum_{k=1}^{i} \mathcal{I}_k(w') \geq \sum_{k=1}^{i} \mathcal{I}_k(w) \quad \text{for } 1 \leq i \leq q.$$

This majorization defines a partial order on the set $\mathbb{W}_{p,q}$. We can easily determine that $\sum_{k=1}^{q} \mathcal{I}_k(w') = \sum_{k=1}^{q} \mathcal{I}_k(w) = (2^q - 1)p$, which was stated already in [3] after Definition 2.1. We denote the partial sums of the orbit of w by $\mathcal{S}_i(w) = \sum_{k=1}^{i} \mathcal{I}_k(w)$.

Similarly, for $w, w' \in \mathbb{W}_{p,q}$ the base-2 orbit of w is said to *majorize with respect to product* the base-2 orbit of w', denoted $w' \prec_p w$, if

$$\prod_{k=1}^{i} \mathcal{I}_k(w') \geq \prod_{k=1}^{i} \mathcal{I}_k(w) \quad \text{for } 1 \leq i \leq q.$$

The majorization with respect to product also defines a partial order on the set $\mathbb{W}_{p,q}$. We denote the partial products of the orbit of w by $\mathcal{P}_i(w) = \prod_{k=1}^{i} \mathcal{I}_k(w)$.

Let us now present Jenkinson's theorems from [2].

Theorem 1.2. *([2], Thm. 2.3) For any coprime integers $1 \leq p < q$, the unique balanced orbit $b \in \mathbb{W}_{p,q}$ is the least element in $(\mathbb{W}_{p,q}, \prec)$. In other words, for any $w \in \mathbb{W}_{p,q}$,*

$$\mathcal{S}_i(b) \geq \mathcal{S}_i(w) \quad \text{for all } 1 \leq i \leq q.$$

Theorem 1.3. *([2], Thm. 1.2) Suppose $1 \leq p < q$ are coprime integers. For $w \in \mathbb{W}_{p,q}$ the product $P(w) = \prod_{i=1}^{q} \mathcal{I}_i(w)$ is maximized precisely when w is balanced.*

The next conjecture, stating Theorem 1.2 for partial products, is very likely to be true. One might be able to prove it using similar ideas that Jenkinson used for partial sums. In any case, we will prove an analogue of it for the most unbalanced words.

Conjecture 1.4. *For any coprime integers $1 \leq p < q$, the unique balanced orbit $b \in \mathbb{W}_{p,q}$ is the least element in $(\mathbb{W}_{p,q}, \prec_p)$. In other words, for any $w \in \mathbb{W}_{p,q}$,*

$$\mathcal{P}_i(b) \geq \mathcal{P}_i(w) \quad \text{for all } 1 \leq i \leq q.$$

2. Partial sum

In this section we will prove an analogue of Theorem 1.2 for the most unbalanced words. Let p, q be integers such that $1 \leq p < q$. We will not need the condition that p and q are coprime. The word $u = 0^{q-p}1^p$ is called *the most unbalanced word* in $\mathcal{W}_{p,q}$ and the orbit of u is called *the most unbalanced orbit* in $\mathbb{W}_{p,q}$. Notice that if $p = 1$ or $q - p = 1$ then u is also balanced.

Example 2.1. *If $(p, q) = (3, 8)$ then the set of all orbits is $\mathbb{W}_{3,8} = \{00000111, 00001011, 00001101, 00010011, 00010101, 00011001, 00100101\}$. These are all Lyndon words. The base-2 orbits and the partial sums of those orbits are listed in Table 1. From these partial sums we can see the partial ordering of the set $\mathbb{W}_{3,8}$, which is drawn in Figure 1.*

00000111		00001011		00001101		00010011		00010101		00011001		00100101	
$\mathcal{I}_i$	$\mathcal{S}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$	$\mathcal{I}_i$	$\mathcal{S}_i$
7	7	11	11	13	13	19	19	21	21	25	25	37	37
14	21	22	33	26	39	38	57	42	63	35	60	41	78
28	49	44	77	52	91	49	106	69	132	50	110	73	151
56	105	88	165	67	158	76	182	81	213	70	180	74	225
112	217	97	262	104	262	98	280	84	297	100	280	82	307
131	348	133	395	134	396	137	417	138	435	140	420	146	453
193	541	176	571	161	557	152	569	162	597	145	565	148	601
224	765	194	765	208	765	196	765	168	765	200	765	164	765

Table 1: The base-2 orbits and the partial sums in $\mathbb{W}_{3,8}$.

From now on, to make the notation easier, we will suppose that the base-2 expansion $(a_1a_2\dots a_n)_2 = \sum_{i=1}^n a_i 2^{n-i}$ can contain numbers also different from 0 or 1, i.e. $a_i \in \mathbb{N}$. If there is a power a_i^k inside a base-2 expansion $(a_1a_2\dots a_n)_2$ then we suppose it means that the number a_i appears k times in a row. If we have a number with two or more digits then we put parentheses around it. For example $(0012^3013(14))_2 = (001222013(14))_2 = 1\cdot 2^7 + 2\cdot 2^6 + 2\cdot 2^5 + 2\cdot 2^4 + 1\cdot 2^2 + 3\cdot 2^1 + 14\cdot 2^0 = 376$.

We will start with a lemma that states some simple formulas on base-2 numbers, which we will need in the proof of Theorem 2.3. Notice that for example $(003000)_2 = (000600)_2$, $(0040)_2 = (0200)_2$, $(010000000)_2 = (001111112)_2$ and $(001111111)_2 < (010000000)_2$. The next lemma uses these kind of facts.

Lemma 2.2. *1)* $(040^{q-2})_2 = (0160^{q-3})_2 = (01280^{q-4})_2 = \dots = (0123\dots(q-3)(q-2)(2q))_2$.

2) $(0123\dots(q-3)(q-2)(2q))_2 > (0123\dots(p-2)(p-1)p^{q-2p+1}(p-1)(p-2)\dots 321)_2$, *if* $p \le q-p$.

3) $(0123\dots(q-3)(q-2)(2q))_2 > (0123\dots(q-p-1)(q-p)^{2p-q+1}(q-p-1)\dots 321)_2$, *if* $p > q-p$.

4) $(040^{q-2})_2 = (0320^{q-3})_2 = (03120^{q-4})_2 = (02320^{q-4})_2$.

5) $(02311210^{q-7})_2 = (02312010^{q-7})_2 = (02320010^{q-7})_2 > (02320^{q-4})_2$.

6) $(0220^{q-3})_2 = (0140^{q-3})_2 = (012^{q-3}4)_2 > (012^{q-3}1)_2$.

7) $(02310^{q-4})_2 = (01270^{q-4})_2 = (01234^{q-5}8)_2 > (01234^{q-7}321)_2$.

8) $(022232110^{q-8})_2 = (022240110^{q-8})_2 > (022240^{q-5})_2 = (022400^{q-5})_2 = (02320^{q-4})_2$.

9) $(02222110^{q-7})_2 = (02302110^{q-7})_2 = (02310110^{q-7})_2 > (02310^{q-4})_2 > (01234^{q-7}321)_2$.

10) $(0221210^{q-6})_2 > (022120^{q-5})_2 = (0150^{q-3})_2 = (0123^{q-4}6)_2 > (0123^{q-5}21)_2$.

11) $(021120^{q-5})_2 = (013120^{q-5})_2 = (014000^{q-5})_2 = (012^{q-3}4)_2 > (012^{q-3}1)_2$.

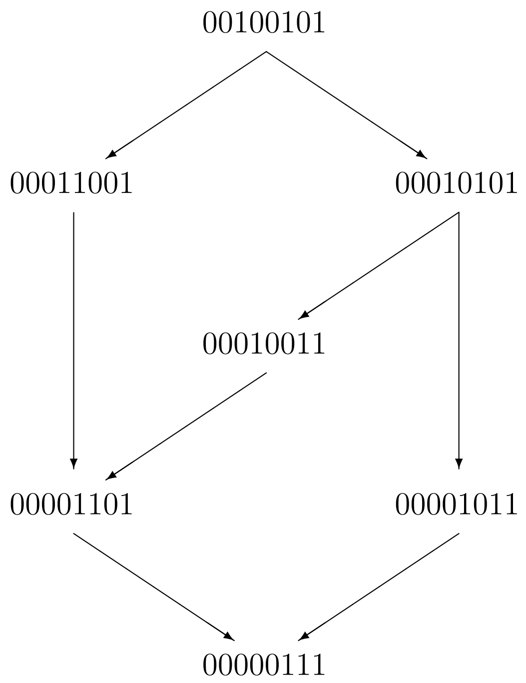

Figure 1: The partially ordered set $(\mathbb{W}_{3,8}, \prec)$. If p and q grow large, the poset $(\mathbb{W}_{p,q}, \prec)$ grows very complex and it is hard to yield any other general results except the two extremal elements.

Theorem 2.3. *For any integers $1 \le p < q$, the most unbalanced orbit $u = 0^{q-p}1^p \in \mathbb{W}_{p,q}$ is the greatest element in $(\mathbb{W}_{p,q}, \prec)$. In other words, for any $w \in \mathbb{W}_{p,q}$,*

$$\mathcal{S}_i(u) \le \mathcal{S}_i(w) \quad \text{for all } 1 \le i \le q.$$

Proof. We set $w = 0^{r_1}1^{s_1}0^{r_2}1^{s_2}\dots 0^{r_n}1^{s_n}$, where $\sum_{i=1}^n s_i = p$, $\sum_{i=1}^n r_i = q-p$, $n \ge 2$ and $\forall i : r_i, s_i > 0$. The orbits of w and u are marked with $(w_1, \dots, w_q)$ and $(u_1, \dots, u_q)$.

We get Table 2 by writing the orbits of u and w in (lexicographic) order. There are p ones, and since $n \ge 2$, the words w_{q-p+1} and w_{q-p+2} start with 10 (the rest of the word is marked with w_i'). For the same reasons the words w_{q-p} and w_{q-p-1} start with 01.

For words from w_1 to w_{q-p-2} we see that w_i cannot be smaller than a word which we get by increasing the number of zeros in front of the word by one, starting from w_{q-p-1}. This is because the number of zeros in front of the word cannot increase by more than one, when moving one word upwards, and we clearly get a larger word if it does not increase. Similarly, for words from w_{q-p+3} to w_q we see that w_i cannot be larger than a word which we get by increasing the number of ones in front of the word by one, starting from w_{q-p+2}. We will suppose that all these words w_i start as described.

Now we see that $(u_i)_2 < (w_i)_2$ for $1 \le i \le q-p-1, i = q-p+1$, because we

i	u_i	$>/<$	w_i
1	$0^{q-p}1^p$	$<$	$0^{q-p-1}1w'_1$
2	$0^{q-p-1}1^p0$	$<$	$0^{q-p-2}1w'_2$
$\dots$	$\dots$	$\dots$	$\dots$
$q-p-3$	00001^p0^{q-p-4}	$<$	$0001w'_{q-p-3}$
$q-p-2$	0001^p0^{q-p-3}	$<$	$001w'_{q-p-2}$
$q-p-1$	001^p0^{q-p-2}	$<$	$01w'_{q-p-1}$
$q-p$	01^p0^{q-p-1}	$>$	$01w'_{q-p}$
$q-p+1$	$10^{q-p}1^{p-1}$	$<$	$10w'_{q-p+1}$
$q-p+2$	$110^{q-p}1^{p-2}$	$>$	$10w'_{q-p+2}$
$q-p+3$	$1110^{q-p}1^{p-3}$	$>$	$110w'_{q-p+3}$
$\dots$	$\dots$	$\dots$	$\dots$
$q-1$	$1^{p-1}0^{q-p}1$	$>$	$1^{p-2}0w'_{q-1}$
q	1^p0^{q-p}	$>$	$1^{p-1}0w'_q$

Table 2: Orbits of u and w from Theorem 2.3.

estimated the words from w_1 to w_{q-p-2} to be the smallest possible. Similarly $(u_i)_2 > (w_i)_2$ for $q-p+2 \leq i \leq q, i = q-p$, because we estimated the words from w_{q-p+3} to w_q to be the largest possible.

Now we see that $\mathcal{S}_i(u) \leq \mathcal{S}_i(w)$ for $1 \leq i \leq q-p-1$. If we suppose that $\mathcal{S}_{q-p}(u) \leq \mathcal{S}_{q-p}(w)$ then we clearly see that also $\mathcal{S}_{q-p+1}(u) \leq \mathcal{S}_{q-p+1}(w)$, since $(u_{q-p+1})_2 < (w_{q-p+1})_2$. We already deduced in the preliminaries that $\mathcal{S}_q(u) = \mathcal{S}_q(w) = (2^q-1)p$. Because $(u_i)_2 > (w_i)_2$ for $q-p+2 \leq i \leq q$, we have $\mathcal{S}_i(u) \leq \mathcal{S}_i(w)$ for $q-p+2 \leq i \leq q$.

The only thing left to prove is our assumption $\mathcal{S}_{q-p}(u) \leq \mathcal{S}_{q-p}(w)$ in the previous paragraph. Direct calculation gives:

$$\mathcal{S}_{q-p}(u) = \sum_{k=1}^{q-p} \mathcal{I}_k(u) = (0123\dots(p-2)(p-1)p^{q-2p+1}(p-1)(p-2)\dots321)_2 \text{ if } p \leq q-p,$$

$$\mathcal{S}_{q-p}(u) = \sum_{k=1}^{q-p} \mathcal{I}_k(u) = (0123\dots(q-p-1)(q-p)^{2p-q+1}(q-p-1)\dots321)_2 \text{ if } p > q-p.$$

Now we divide the proof into three cases: 1) $n \geq 4$, 2) $n = 3$ and 3) $n = 2$. See Tables 3 and 4 for the prefixes of w_i in each case.

1) Because $n \geq 4$ the words $w_{q-p}, w_{q-p-1}, w_{q-p-2}$ and w_{q-p-3} start with 01. It is enough to take only these four words for the partial sum $\mathcal{S}_{q-p}(w)$ and even suppose that the remaining

i	Prefixes of w_i							
	1	2.1	2.2	2.3	2.4	3.1	3.2	3.3
$q-p-4$		001						
$q-p-3$	01	001	001			0011		
$q-p-2$	01	01	01	0101	010101	0011	0011011	
$q-p-1$	01	01	0101	0101	010101	011	011	011
$q-p$	01	01	0101	011	010101	011	011011	011

Table 3: Prefixes of w_i from cases 1-3.3 of Theorem 2.3.

parts of these four words are zeros. Hence we have

$$\mathcal{S}_{q-p}(w) \geq \sum_{k=q-p-3}^{q-p} \mathcal{I}_k(w) \geq (040^{q-2})_2 = (0123\ldots(q-2)(2q))_2 > \mathcal{S}_{q-p}(u),$$

where the equality comes from Lemma 2.2, 1 and the last inequality from Lemma 2.2, 2&3.

2) The case $n = 3$ is similar to the previous one. We divide it into four subcases depending on the values of r_i and s_i. Because $n = 3$ the words w_{q-p}, w_{q-p-1} and w_{q-p-2} start with 01 in all cases. From now on, we will use Lemma 2.2 without stating it explicitly.

2.1) $\exists i, j$ $(i \neq j)$: $r_i, r_j \geq 2$. This means there are at least two blocks of zeros of length at least 2, which means that the words w_{q-p-3} and w_{q-p-4} starts with 001. Now it is enough to take only five words to the partial sum $\mathcal{S}_{q-p}(w)$ and suppose that the remaining parts are zeros. We get

$$\mathcal{S}_{q-p}(w) \geq \sum_{k=q-p-4}^{q-p} \mathcal{I}_k(w) \geq (0320^{q-3})_2 > \mathcal{S}_{q-p}(u).$$

2.2) $\exists! i : r_i \geq 2$. Because we have two blocks of zeros of length 1, we get the words w_{q-p}, w_{q-p-1} start with $011, 011$ or $011, 0101$ or $0101, 0101$ (depending on the values of s_i). We can estimate the partial sum $\mathcal{S}_{q-p}(w)$ downwards so we suppose they start with $0101, 0101$. Since $r_i \geq 2$, the word w_{q-p-3} starts with 001. We get

$$\mathcal{S}_{q-p}(w) \geq \sum_{k=q-p-3}^{q-p} \mathcal{I}_k(w) \geq (03120^{q-4})_2 > \mathcal{S}_{q-p}(u).$$

2.3) $r_1, r_2, r_3 = 1$, $\exists i : s_i \geq 2$. Notice that $q-p = r_1+r_2+r_3 = 3$. Because there is at least one block of ones of length at least 2 and $r_1, r_2, r_3 = 1$, we get the words $w_{q-p}, w_{q-p-1}, w_{q-p-2}$

start with $011, 011, 011$ or $011, 011, 0101$ or $011, 0101, 0101$. From these, we again choose the smallest ones $011, 0101, 0101$ in order to estimate $\mathcal{S}_{q-p}(w)$ downwards. We get

$$\mathcal{S}_{q-p}(w) = \sum_{1}^{3} \mathcal{I}_k(w) \geq (03120^{q-4})_2 > \mathcal{S}_{q-p}(u).$$

2.4) $\forall i : r_i, s_i = 1$, i.e. $u = 000111$ and $w = 010101$. Trivially we get

$$\mathcal{S}_{q-p}(w) = (030303)_2 > (012321)_2 = \mathcal{S}_{q-p}(u).$$

3) The case $n = 2$ is similarly divided into several subcases depending on the values of r_i and s_i. Because $n = 2$ the words w_{q-p} and w_{q-p-1} start with 01 in all cases.

3.1) $s_1, s_2 \geq 2$ and $r_1, r_2 \geq 2$. Because $s_1, s_2 \geq 2$ the words w_{q-p} and w_{q-p-1} start with 011, and because $r_1, r_2 \geq 2$ the words w_{q-p-2} and w_{q-p-3} start with 0011. We get

$$\mathcal{S}_{q-p}(w) \geq \sum_{k=q-p-3}^{q-p} \mathcal{I}_k(w) \geq (02420^{q-4})_2 > (02320^{q-4})_2 > \mathcal{S}_{q-p}(u).$$

3.2) $s_1, s_2 \geq 2$ and $r_1 = 1, r_2 \geq 2$. Because $r_1 = 1$ and $s_1, s_2 \geq 2$ the word w_{q-p} starts with 0111 or 011011 and because $r_2 \geq 2$ the word w_{q-p-2} starts with 00111 or 0011011. From these we choose the smaller ones 011011 and 0011011. The word w_{q-p-1} starts with 011, since $s_1, s_2 \geq 2$. We get

$$\mathcal{S}_{q-p}(w) \geq \sum_{k=q-p-2}^{q-p} \mathcal{I}_k(w) \geq (02311210^{q-7})_2 > (02320^{q-4})_2 > \mathcal{S}_{q-p}(u).$$

3.3) $s_1, s_2 \geq 2$ and $r_1, r_2 = 1$. Notice that $q - p = r_1 + r_2 = 2$. Because $s_1, s_2 \geq 2$ the words w_1 and w_2 start with 011. We get

$$\mathcal{S}_{q-p}(w) = \sum_{k=1}^{2} \mathcal{I}_k(w) \geq (0220^{q-3})_2 > (012^{q-3}1)_2 = \mathcal{S}_{q-p}(u).$$

3.4.1) $s_1 = 1, s_2 \geq 2$ and $r_1 \geq 3, r_2 \geq 2$ or $r_1 \geq 2, r_2 \geq 3$. We can easily see that the words $w_{q-p}, w_{q-p-1}, w_{q-p-2}, w_{q-p-3}, w_{q-p-4}$ start with $01, 01, 001, 001, 0001$. Because $s_2 \geq 2$ we additionally get that from these words w_{q-p} and w_{q-p-2} start with 011 and 0011. Together we have

$$\mathcal{S}_{q-p}(w) \geq \sum_{k=q-p-4}^{q-p} \mathcal{I}_k(w) \geq (02320^{q-4})_2 > \mathcal{S}_{q-p}(u).$$

i	Prefixes of w_i								
	3.4.1-2	3.5.1-2	3.5.3	3.6.1-3	3.7	3.8	3.9	3.10	3.11
$q-p-4$	0001	00001101		00001011					
$q-p-3$	001	0001101		0001011		001	0001		
$q-p-2$	0011	001101	001101	001011		001	001	00101	
$q-p-1$	01	01	01001	01011	01011	01	01	01001	0101
$q-p$	011	01101	01101	011	01101	01	0101	01010	0101

Table 4: Prefixes of w_i from cases 3.4-3.11 of Theorem 2.3.

3.4.2) $s_1 = 1, s_2 \geq 2$ and $r_1, r_2 = 2$. Notice that $q - p = r_1 + r_2 = 4$. This is similar to the previous case except we do not have the word w_{q-p-4}. We get

$$\mathcal{S}_{q-p}(w) = \sum_{k=1}^{4} \mathcal{I}_k(w) \geq (02310^{q-4})_2 > (01234^{q-7}321)_2 = \mathcal{S}_{q-p}(u).$$

3.5.1) $s_1 = 1, s_2 \geq 2$ and $r_1 = 1, r_2 \geq 4$. Because $r_1 = 1$ and $s_2 \geq 2$ the word w_{q-p} starts with 0111 or 01101. We choose the smaller one 01101. Since $r_2 \geq 4$, the same applies to the words w_{q-p-2}, w_{q-p-3} and w_{q-p-4}, which are estimated to start with $001101, 0001101$ and 00001101. We get

$$\mathcal{S}_{q-p}(w) \geq \sum_{k=q-p-4}^{q-p} \mathcal{I}_k(w) \geq (022232110^{q-8})_2 > (02320^{q-4})_2 > \mathcal{S}_{q-p}(u).$$

3.5.2) $s_1 = 1, s_2 \geq 2$ and $r_1 = 1, r_2 = 3$. Notice that $q - p = r_1 + r_2 = 4$. This is similar to the previous case except we do not have the word w_{q-p-4}. We get

$$\mathcal{S}_{q-p}(w) = \sum_{k=1}^{4} \mathcal{I}_k(w) \geq (02222110^{q-7})_2 > (01234^{q-7}321)_2 = \mathcal{S}_{q-p}(u).$$

3.5.3) $s_1 = 1, s_2 \geq 2$ and $r_1 = 1, r_2 = 2$. Notice that $q - p = r_1 + r_2 = 3$. This is similar to the previous case except we do not have the word w_{q-p-3} and we know that the word w_{q-p-1} starts with 01001. We get

$$\mathcal{S}_{q-p}(w) = \sum_{k=1}^{3} \mathcal{I}_k(w) \geq (0221210^{q-6})_2 > (0123^{q-5}21)_2 = \mathcal{S}_{q-p}(u).$$

3.6.1) $s_1 = 1, s_2 \geq 2$ and $r_1 \geq 4, r_2 = 1$. Because $s_2 \geq 2$ the word w_{q-p} starts with 011. Because $s_1, r_2 = 1$ and $r_1 \geq 4$ we know that the words $w_{q-p-1}, w_{q-p-2}, w_{q-p-3}$ and w_{q-p-4} start with $01011, 001011, 0001011$ and 00001011. We get

$$\mathcal{S}_{q-p}(w) \geq \sum_{k=q-p-4}^{q-p} \mathcal{I}_k(w) \geq (022232210^{q-8})_2 > (02320^{q-4})_2 > \mathcal{S}_{q-p}(u).$$

3.6.2) $s_1 = 1, s_2 \geq 2$ and $r_1 = 3, r_2 = 1$. Notice that $q - p = r_1 + r_2 = 4$. This is similar to the previous case except we do not have the word w_{q-p-4}. We get

$$\mathcal{S}_{q-p}(w) = \sum_{k=1}^{4} \mathcal{I}_k(w) \geq (02222210^{q-7})_2 > (01234^{q-7}321)_2 = \mathcal{S}_{q-p}(u).$$

3.6.3) $s_1 = 1, s_2 \geq 2$ and $r_1 = 2, r_2 = 1$. Notice that $q - p = r_1 + r_2 = 3$. This is similar to the previous case except we do not have the word w_{q-p-3}. We get

$$\mathcal{S}_{q-p}(w) = \sum_{k=1}^{3} \mathcal{I}_k(w) \geq (0221210^{q-6})_2 > (0123^{q-5}21)_2 = \mathcal{S}_{q-p}(u).$$

3.7) $s_1 = 1, s_2 \geq 2$ and $r_1, r_2 = 1$. Notice that $q - p = r_1 + r_2 = 2$. Because $s_2 \geq 2$ and $r_1 = 1$ the word w_2 starts with 0111 or 01101 from which we choose the smaller one 01101. The word w_1 starts with 01011. We get

$$\mathcal{S}_{q-p}(w) = \sum_{k=1}^{2} \mathcal{I}_k(w) \geq (021120^{q-5})_2 > (012^{q-3}1)_2 = \mathcal{S}_{q-p}(u).$$

3.8) $s_1, s_2 = 1$ and $r_1, r_2 \geq 2$. Notice that $p = s_1 + s_2 = 2$. Because $r_1, r_2 \geq 2$ the words w_{q-p-2} and w_{q-p-3} start with 001. We get

$$\mathcal{S}_{q-p}(w) \geq \sum_{k=q-p-3}^{q-p} \mathcal{I}_k(w) \geq (0220^{q-3})_2 > (012^{q-3}1)_2 = \mathcal{S}_{q-p}(u).$$

3.9) $s_1, s_2 = 1$ and $r_1 = 1, r_2 \geq 3$. Because $s_1, r_2 = 1$ the word w_{q-p} starts with 0101 and because $r_2 \geq 3$ the words w_{q-p-1}, w_{q-p-2} and w_{q-p-3} start with $01, 001$ and 0001. We get

$$\mathcal{S}_{q-p}(w) \geq \sum_{k=q-p-3}^{q-p} \mathcal{I}_k(w) \geq (02120^{q-4})_2 = (0220^{q-3})_2 > (012^{q-3}1)_2 = \mathcal{S}_{q-p}(u).$$

3.10) $s_1, s_2 = 1$ and $r_1 = 1, r_2 = 2$. Now $u = 00011$ and $w = 01001$. We get

$$\mathcal{S}_{q-p}(w) = (02112)_2 > (01221)_2 = \mathcal{S}_{q-p}(u).$$

3.11) $s_1, s_2 = 1$ and $r_1, r_2 = 1$. Now $u = 0011$ and $w = 0101$. We get

$$\mathcal{S}_{q-p}(w) = (0202)_2 > (0121)_2 = \mathcal{S}_{q-p}(u).$$

□

Remark 2.4. *In the previous proof, we can also use induction on the length of the word u. Instead of dividing the proof into three cases and handling them all separately, we can suppose that the theorem holds for all shorter words and use the fact that $\mathcal{S}_{q-p}(u) = \mathcal{S}_{q-p}(0^{-1}u) \leq \mathcal{S}_{q-p}(0^{-1}w) = \mathcal{S}_{q-p}(w)$. This reduces significantly the length of the proof, but we also lose some combinatorial insight.*

3. Product

In this section we prove an analogue of Theorem 1.3 for the most unbalanced words. We will not need the condition that q and p are coprime. The following two lemmas state some simple inequalities on base-2 numbers that we need in the proof of Theorem 3.3 to follow. We will suppose that the base-2 expansion $(a_1a_2\dots a_n)_2 = \sum_{i=1}^{n} a_i 2^{n-i}$ can contain also rational numbers, i.e. $a_i \in \mathbb{Q}$. For example $(\frac{1}{2}00\frac{3}{2}0)_2 = (010\frac{3}{2}0)_2 = (01011)_2$ and $\frac{2}{3}\cdot(0110)_2 = (0\frac{2}{3}\frac{2}{3}0)_2 = (0\frac{6}{3}00)_2 = (1000)_2$.

Lemma 3.1. *The following inequalities hold for any $w \in \{0,1\}^*$ and $a, b \geq 0$ $(a+b \geq 1)$ such that the words on both sides are equally long and have equally many zeros and ones.*

1) $2\cdot(0^{b_1}1^a0^{b_2})_2 < (0^{b_1-2}1w)_2$, *where $b_1 \geq 3$ and $b_2 \geq 0$.*

2) $4\cdot(000001^a0^b)_2 < (001w)_2$.

3) $\frac{21}{8}\cdot(00001^a0^b)_2 < (0010101w)_2$.

4) $(0^{b_1}1^a0^{b_2})_2 < (0^{b_1-1}1w)_2$, *where $b_1 \geq 2$ and $b_2 \geq 0$.*

5) $\frac{1}{2}\cdot(01^a0^b)_2 < (01w)_2$.

6) $(10^b1^a)_2 \leq (1w)_2$.

7) $\frac{2}{3}\cdot(110^b1^a)_2 < (1w)_2$.

8) $\frac{4}{7}\cdot(1110^b1^a)_2 < (1w)_2$.

9) $\frac{1}{2}\cdot(1^{a_1}0^b1^{a_2})_2 < (1w)_2$, *where $a_1 \geq 1$ and $a_2 \geq 0$.*

Proof. 1) $2\cdot(0^{b_1}1^a0^{b_2})_2 = (0^{b_1-1}1^a0^{b_2+1})_2 < (0^{b_1-2}10^{a+b_2+1})_2 < (0^{b_1-2}1w)_2$.

2) $4\cdot(000001^a0^b)_2 = 2\cdot(00001^a0^{b+1})_2 = (0001^a0^{b+2})_2 < (0010^{a+b+2})_2 < (001w)_2$.

3) $\frac{21}{8}\cdot(00001^a0^b)_2 < \frac{21}{8}\cdot(00010^{a+b})_2 = (000\frac{21}{8}0^{a+b})_2 = (00101010^{a+b-3})_2 < (0010101w)_2$

4) $(0^{b_1}1^a0^{b_2})_2 < (0^{b_1-1}10^{a+b_2})_2 < (0^{b_1-1}1w)_2$.
5) $\frac{1}{2} \cdot (01^a0^b)_2 < \frac{1}{2} \cdot (10^{a+b})_2 = (010^{a+b-1})_2 < (01w)_2$.
6) Trivial.
7) $\frac{2}{3} \cdot (110^b1^a)_2 = (\frac{2}{3}\frac{2}{3}0^b(\frac{2}{3})^a)_2 = (100^b(\frac{2}{3})^a)_2 < (1w)_2$.
8) $\frac{4}{7} \cdot (1110^b1^a)_2 = (\frac{4}{7}\frac{4}{7}\frac{4}{7}0^b(\frac{4}{7})^a)_2 = (\frac{4}{7}\frac{6}{7}00^b(\frac{4}{7})^a)_2 = (1000^b(\frac{4}{7})^a)_2 < (1w)_2$.
9) $\frac{1}{2} \cdot (1^{a_1}0^b1^{a_2})_2 < \frac{1}{2} \cdot (1^{a_1+b+a_2})_2 < \frac{1}{2} \cdot (20^{a_1+b+a_2-1})_2 = (10^{a_1+b+a_2-1})_2 < (1w)_2$. □

Lemma 3.2. *The following inequalities hold for any $w \in \{0,1\}^*$ and $a, b \geq 0$ $(a+b \geq 1)$ such that the words on both sides are equally long and have equally many zeros and ones.*
1) $\frac{3}{2} \cdot (0^{b_1+1}1^a0^{b_2})_2 < (0^{b_1}11w)_2$, *where* $b_1 \geq 1$ *and* $b_2 \geq 0$.
2) $\frac{11}{8} \cdot (0^{b_1+1}1^a0^{b_2})_2 < (0^{b_1}1011w)_2$, *where* $b_1 \geq 1$ *and* $b_2 \geq 0$.
3) $\frac{8}{3} \cdot (0000110^b)_2 < (001w)_2$.
4) $\frac{13}{8} \cdot (0001^a0^b)_2 < (001101w)_2$.
5) $\frac{5}{3} \cdot (000110^b)_2 < (00101w)_2$.
6) $\frac{3}{4} \cdot (01^a0^b)_2 < (011w)_2$.
7) $\frac{2}{3} \cdot (0^{b_1}110^{b_2}1^a)_2 < (0^{b_1}1w)_2$, *where* $a, b_1, b_2 \geq 0$.
8) $\frac{3}{4} \cdot (1^{a_1}0^b1^{a_2})_2 < (11w)_2$, *where* $a_1, b \geq 1$ *and* $a_2 \geq 0$.
9) $\frac{5}{6} \cdot (0^{b_1}110^{b_2}1^a)_2 < (0^{b_1}101w)_2$, *where* $a, b_1 \geq 0$ *and* $b_2 \geq 1$.

Proof. 1) $\frac{3}{2} \cdot (0^{b_1+1}1^a0^{b_2})_2 < \frac{3}{2} \cdot (0^{b_1}10^{a+b_2})_2 = (0^{b_1}\frac{3}{2}0^{a+b_2})_2 = (0^{b_1}110^{a+b_2-1})_2 < (0^{b_1}11w)_2$.
2) $\frac{11}{8} \cdot (0^{b_1+1}1^a0^{b_2})_2 < \frac{11}{8} \cdot (0^{b_1}10^{a+b_2})_2 = (0^{b_1}\frac{11}{8}0^{a+b_2})_2 = (0^{b_1}10110^{a+b_2-3})_2 < (0^{b_1}1011w)_2$.
3) $\frac{8}{3} \cdot (0000110^b)_2 = (0000\frac{8}{3}\frac{8}{3}0^b)_2 = (0000\frac{12}{3}00^b)_2 = (0010000^b)_2 < (001w)_2$.
4) $\frac{13}{8} \cdot (0001^a0^b)_2 < \frac{13}{8} \cdot (0010^{a+b})_2 = (00\frac{13}{8}0^{a+b})_2 = (0011010^{a+b-3})_2 < (001101w)_2$.
5) $\frac{5}{3} \cdot (000110^b)_2 = (000\frac{5}{3}\frac{5}{3}0^b)_2 = (000\frac{6}{3}10^b)_2 = (001010^b)_2 < (00101w)_2$.
6) $\frac{3}{4} \cdot (01^a0^b)_2 < \frac{3}{4} \cdot (10^{a+b})_2 = (\frac{3}{4}0^{a+b})_2 = (0\frac{6}{4}0^{a+b-1})_2 = (0110^{a+b-2})_2 < (011w)_2$.
7) $\frac{2}{3} \cdot (0^{b_1}110^{b_2}1^a)_2 = (0^{b_1}\frac{2}{3}\frac{2}{3}0^{b_2}\frac{2}{3}^a)_2 = (0^{b_1}100^{b_2}\frac{2}{3}^a)_2 < (0^{b_1}1w)_2$
8) $\frac{3}{4} \cdot (1^{a_1}0^b1^{a_2})_2 < \frac{3}{4} \cdot (10^{a_1+b}1^{a_2})_2 = (\frac{3}{4}0^{a_1+b}\frac{3}{4}^{a_2})_2 = (0110^{a_1+b-2}\frac{3}{4}^{a_2})_2 < (11w)_2$ (notice that the length of the base-2 expansion changes after the first and last inequality).
9) $\frac{5}{6} \cdot (0^{b_1}110^{b_2}1^a)_2 = (0^{b_1}\frac{5}{6}\frac{5}{6}0^{b_2}\frac{5}{6}^a)_2 = (0^{b_1}1\frac{3}{6}0^{b_2}\frac{5}{6}^a)_2 = (0^{b_1}1010^{b_2-1}\frac{5}{6}^a)_2 < (0^{b_1}101w)_2$. □

The idea of the proof of Theorem 3.3 to follow is to multiply the base-2 expansions of the words in the orbit $(u_1, \ldots, u_q)$ of the most unbalanced word $u = 0^{q-p}1^p$ with some number so that the base-2 expansion of the corresponding word in the orbit $(w_1, \ldots, w_q)$ of any other word $w \in \mathbb{W}_{p,q}$ is larger. If the product of all the multipliers is at least one then we get the product of u is smaller than the product of w. Table 5 gives the multipliers for each word in case 1 of the proof of Theorem 3.3. If we multiply the base-2 expansion of u_i with Multiplier(i) we get a smaller number than the base-2 expansion of w_i. We find that the product of the multipliers really is at least one: in case 1.1 $\prod_{i=1}^{q}$ Multiplier(i) =

i	u_i	w_i		Multiplier(i)	
		Case 1.1	Case 1.2	Case 1.1	Case 1.2
1	$0^{q-p}1^p$	$0^n 1 w'_1 \quad (n \le q-p-2)$		2	
2	$0^{q-p-1}1^p 0$	$0^n 1 w'_2 \quad (n \le q-p-3)$		2	
...	...	...		...	
$q-p-5$	$0000001^p 0^{q-p-6}$	$0^n 1 w'_{q-p-5} \quad (n \le 4)$		2	
$q-p-4$	$000001^p 0^{q-p-5}$	$001w'_{q-p-4}$	$0001w'_{q-p-4}$	4	2
$q-p-3$	$00001^p 0^{q-p-4}$	$001w'_{q-p-3}$	$0010101w'_{q-p-3}$	2	$\frac{21}{8}$
$q-p-2$	$0001^p 0^{q-p-3}$	$01w'_{q-p-2}$		2	
$q-p-1$	$001^p 0^{q-p-2}$	$01w'_{q-p-1}$		1	
$q-p$	$01^p 0^{q-p-1}$	$01w'_{q-p}$		$\frac{1}{2}$	
$q-p+1$	$10^{q-p}1^{p-1}$	$1w'_{q-p+1}$		1	
$q-p+2$	$110^{q-p}1^{p-2}$	$1w'_{q-p+2}$		$\frac{2}{3}$	
$q-p+3$	$1110^{q-p}1^{p-3}$	$1w'_{q-p+3}$		$\frac{4}{7}$	
$q-p+4$	$11110^{q-p}1^{p-4}$	$1w'_{q-p+4}$		$\frac{1}{2}$	
...	...	...		...	
$q-1$	$1^{p-1}0^{q-p}1$	$1w'_{q-1}$		$\frac{1}{2}$	
q	$1^p 0^{q-p}$	$1w'_q$		$\frac{1}{2}$	

Table 5: Case 1 in Theorem 3.3.

$2^{q-p-5} \cdot 4 \cdot 2 \cdot 2 \cdot 1 \cdot \frac{1}{2} \cdot 1 \cdot \frac{2}{3} \cdot \frac{4}{7} \cdot \frac{1}{2}^{p-3} = \frac{32}{21} \cdot 2^{(q-p)-p-1}$ and in case 1.2 $\prod_{i=1}^{q}$ Multiplier(i) = $2^{q-p-4} \cdot \frac{21}{8} \cdot 2 \cdot 1 \cdot \frac{1}{2} \cdot 1 \cdot \frac{2}{3} \cdot \frac{4}{7} \cdot \frac{1}{2}^{p-3} = 2^{(q-p)-p-1}$, where $(q-p)-p-1 \ge 0$ because we will suppose that there are more zeros than ones, i.e. $p < q-p$. The theorem is probably true even without the assumption that there are more zeros than ones, but it would be more difficult to prove.

Theorem 3.3. *Suppose* $1 \le p < q$ *are integers such that* $p < q-p$. *For* $w \in \mathbb{W}_{p,q}$ *the product* $P(w) = \prod_{i=1}^{q} \mathcal{I}_i(w)$ *is minimized precisely when* $w = 0^{q-p}1^p$.

Proof. We set $u = 0^{q-p}1^p$ and $w = 0^{r_1}1^{s_1}0^{r_2}1^{s_2}\dots 0^{r_n}1^{s_n}$, where $\sum_{i=1}^{n} s_i = p$, $\sum_{i=1}^{n} r_i = q-p$, $n \ge 2$ and $\forall i : r_i, s_i > 0$. The orbits of w and u are marked with $(w_1, \dots, w_q)$ and $(u_1, \dots, u_q)$. Our goal is to prove that $P(u) < P(w)$. We divide the proof into two cases: 1) $n \ge 3$ and 2) $n = 2$.

1) We divide this case into two subcases: 1.1) $\exists i, j\ (i \ne j)$: $r_i, r_j \ge 2$ and 1.2) $\exists! i : r_i \ge 2$. Notice that at least one r_i has to be at least two because otherwise there would not be more zeros than ones.

We get Table 5 by writing the orbits of u and w in (lexicographic) order. There are p ones so the words from w_q to w_{q-p} start with the letter 1 (the rest of the word is marked with w_i'). Because $n \geq 3$ the next three words from w_{q-p-1} to w_{q-p-3} start with 01. In case 1.1 there are at least two blocks of zeros of length at least two, which means that the next two words w_{q-p-3} and w_{q-p-4} can start with $01, 01$ or $01, 001$ or $001, 001$. We suppose that the words start with $001, 001$ because that makes the product $P(w)$ smallest. In case 1.2 there is only one block of zeros which is of length at least 2, which means that the word w_{q-p-3} starts with $01, 0011, 001011$ or 0010101. Similar to the case 1.1, we suppose that it starts with 0010101 because that makes the product $P(w)$ smallest.

The number of zeros in front of the word cannot increase by more than one, when moving one word upwards. Since $(10^a)_2 > (01^a)_2$, we get the smallest possible w_i for the rest by doing exactly that.

From Lemma 3.1 we now get directly the following inequalities:

$$\begin{aligned}
2 \cdot (u_{q-p-i})_2 &< (w_{q-p-i})_2 \text{ for every } 2 \leq i \leq q-p-1 \\
4 \cdot (u_{q-p-4})_2 &< (w_{q-p-4})_2 \text{ (case 1.1)} \\
21/8 \cdot (u_{q-p-3})_2 &< (w_{q-p-3})_2 \text{ (case 1.2)} \\
(u_{q-p-1})_2 &< (w_{q-p-1})_2 \\
1/2 \cdot (u_{q-p})_2 &< (w_{q-p})_2 \\
(u_{q-p+1})_2 &< (w_{q-p+1})_2 \\
2/3 \cdot (u_{q-p+2})_2 &< (w_{q-p+2})_2 \\
4/7 \cdot (u_{q-p+3})_2 &< (w_{q-p+3})_2 \\
1/2 \cdot (u_{q-p+i})_2 &< (w_{q-p+i})_2 \text{ for every } 4 \leq i \leq p.
\end{aligned}$$

We already determined that the products of the multipliers are at least one: $2^{q-p-5} \cdot 4 \cdot 2 \cdot 2 \cdot 1 \cdot \frac{1}{2} \cdot 1 \cdot \frac{2}{3} \cdot \frac{4}{7} \cdot \frac{1}{2}^{p-3} = \frac{32}{21} \cdot 2^{(q-p)-p-1} > 1$ and $2^{q-p-4} \cdot \frac{21}{8} \cdot 2 \cdot 1 \cdot \frac{1}{2} \cdot 1 \cdot \frac{2}{3} \cdot \frac{4}{7} \cdot \frac{1}{2}^{p-3} = 2^{(q-p)-p-1} \geq 1$, where $(q-p)-p-1 \geq 0$ because $p < q-p$. From these facts we get our claim:

1.1) $P(u) = \prod_{i=1}^{q} (u_i)_2 < \prod_{i=1}^{q-p-5} [2(u_i)_2] \cdot 4(u_{q-p-3})_2 \cdot 2(u_{q-p-2})_2 (u_{q-p-1})_2 \cdot 1/2(u_{q-p})_2 (u_{q-p+1})_2 \cdot 2/3(u_{q-p+2})_2 \cdot 4/7(u_{q-p+3})_2 \cdot \prod_{i=q-p+4}^{q} 1/2(u_i)_2 < \prod_{i=1}^{q} (w_i)_2 = P(w)$.

1.2) $P(u) = \prod_{i=1}^{q} (u_i)_2 \leq \prod_{i=1}^{q-p-4} [2(u_i)_2] \cdot 21/8(u_{q-p-3})_2 \cdot 2(u_{q-p-2})_2 (u_{q-p-1})_2 \cdot 1/2(u_{q-p})_2 (u_{q-p+1})_2 \cdot 2/3(u_{q-p+2})_2 \cdot 4/7(u_{q-p+3})_2 \cdot \prod_{i=q-p+4}^{q} 1/2(u_i)_2 < \prod_{i=1}^{q} (w_i)_2 = P(w)$.

2) This case is similar to the previous one. We divide it into five subcases depending on the values of r_1, r_2, s_1 and s_2. Notice that case $r_1, r_2 = 1$ is impossible because then we would have $2 \leq s_1 + s_2 = p < q - p = r_1 + r_2 = 2$.

2.1) $r_1, r_2 \geq 2$ and $(s_1, s_2) \neq (1, 1)$

2.2) $r_1, r_2 \geq 2$ and $s_1, s_2 = 1$

2.3) $r_1 = 1, r_2 \geq 2$ and $s_1 \geq 1, s_2 \geq 2$

i	u_i	Prefixes of w_i					Multiplier(i)				
		2.1	2.2	2.3	2.4	2.5	2.1	2.2	2.3	2.4	2.5
1	$0^{q-p}1^p$	0^n1	0^n1	0^n1101	0^n1011	0^n101	2	$\frac{8}{3}$	$\frac{13}{8}$	$\frac{11}{8}$	$\frac{5}{3}$
...	...	...	...	...	...	...	...	...	...	...	...
$q-p-3$	00001^p0^{q-p-4}	001	001	...	...	...	2	$\frac{8}{3}$	...	...	...
$q-p-2$	0001^p0^{q-p-3}	0011	001	001101	...	00101	$\frac{3}{2}$	1	$\frac{13}{8}$	...	$\frac{5}{3}$
$q-p-1$	001^p0^{q-p-2}	01	01	01	01011	01	1	1	1	$\frac{11}{8}$	1
$q-p$	01^p0^{q-p-1}	011	01	011	011	0101	$\frac{3}{4}$	$\frac{2}{3}$	$\frac{3}{4}$	$\frac{3}{4}$	$\frac{5}{6}$
$q-p+1$	$10^{q-p}1^{p-1}$	10	10	10	10	10	1	1	1	1	1
$q-p+2$	$110^{q-p}1^{p-2}$	10	10	101	101	101	$\frac{2}{3}$	$\frac{2}{3}$	$\frac{5}{6}$	$\frac{5}{6}$	$\frac{5}{6}$
$q-p+3$	$1110^{q-p}1^{p-3}$	11		11	11		$\frac{3}{4}$		$\frac{3}{4}$	$\frac{3}{4}$	
...	...	...		...	...		...		...	...	
q	1^p0^{q-p}	11		11	11		$\frac{3}{4}$		$\frac{3}{4}$	$\frac{3}{4}$	

Table 6: Case 2 in Theorem 3.3.

2.4) $r_1 = 1, r_2 \geq 2$ and $s_1 \geq 2, s_2 \geq 1$

2.5) $r_1 = 1, r_2 \geq 2$ and $s_1, s_2 = 1$.

We get Table 6 by using the same kind of reasoning as in case 1 (the suffixes w'_i of w_i have been left out to save space):

There are p ones and $n = 2$ so the words from w_q to w_{q-p+3} start with 11 and the words w_{q-p+2} and w_{q-p+1} start with 10. In cases 2.3, 2.4 and 2.5 we have $r_1 = 1$ so we additionally know that the word w_{q-p+2} starts with 101.

There are $q-p$ zeros and $n = 2$ so the words w_{q-p} and w_{q-p-1} start with 01. In addition, in cases 2.1, 2.3 and 2.4 we have s_1 or $s_2 \geq 2$, which means the word w_{q-p} starts with 011. In case 2.5 we have $s_1, s_2 = 1$, which means the word w_{q-p} starts with 0101. In addition, in case 2.4 we have $r_1 = 1$ and $s_1 \geq 2$, which means the word w_{q-p-1} starts with either 011 or 01011, from which we choose the smaller one 01011.

In cases 2.1 and 2.3 the word w_{q-p-2} starts with 0011 because $r_1, s_1 \geq 2$ or $r_2, s_2 \geq 2$. In addition, in case 2.3 we have $r_1 = 1$, which means it starts with 00111 or 001101, from which we choose the smaller one 001101. In cases 2.2 and 2.5 the word w_{q-p-2} starts with 001 because r_1 or $r_2 \geq 2$. In addition, in case 2.5 we have $r_1, s_1, s_2 = 1$, which means the word w_{q-p-2} starts with 00101. In cases 2.1 and 2.2 the word w_{q-p-3} starts with 001 because $r_1, r_2 \geq 2$.

We get the smallest possible w_i for the rest of the words by increasing the number of zeros in front of the word by one, until $i = 1$.

From Lemmas 3.1 and 3.2 we now get the following inequalities:

2.1)

$$
\begin{aligned}
2 \cdot (u_{q-p-i})_2 &< (w_{q-p-i})_2 \text{ for every } 3 \leq i \leq q-p-1 \text{ (Lemma 3.1, 1)}\\
3/2 \cdot (u_{q-p-2})_2 &< (w_{q-p-2})_2 \text{ (Lemma 3.2, 1)}\\
(u_{q-p-1})_2 &< (w_{q-p-1})_2 \text{ (Lemma 3.1, 4)}\\
3/4 \cdot (u_{q-p})_2 &< (w_{q-p})_2 \text{ (Lemma 3.2, 6)}\\
(u_{q-p+1})_2 &< (w_{q-p+1})_2 \text{ (Lemma 3.1, 6)}\\
2/3 \cdot (u_{q-p+2})_2 &< (w_{q-p+2})_2 \text{ (Lemma 3.2, 7)}\\
3/4 \cdot (u_{q-p+i})_2 &< (w_{q-p+i})_2 \text{ for every } 3 \leq i \leq p \text{ (Lemma 3.2, 8).}
\end{aligned}
$$

2.2)

$$
\begin{aligned}
8/3 \cdot (u_{q-p-i})_2 &< (w_{q-p-i})_2 \text{ for every } 3 \leq i \leq q-p-1 \text{ (Lemma 3.2, 3)}\\
(u_{q-p-2})_2 &< (w_{q-p-2})_2 \text{ (Lemma 3.1, 4)}\\
(u_{q-p-1})_2 &< (w_{q-p-1})_2 \text{ (Lemma 3.1, 4)}\\
2/3 \cdot (u_{q-p})_2 &< (w_{q-p})_2 \text{ (Lemma 3.2, 7)}\\
(u_{q-p+1})_2 &< (w_{q-p+1})_2 \text{ (Lemma 3.1, 6)}\\
2/3 \cdot (u_{q-p+2})_2 &< (w_{q-p+2})_2 \text{ (Lemma 3.2, 7).}
\end{aligned}
$$

2.3)

$$
\begin{aligned}
13/8 \cdot (u_{q-p-i})_2 &< (w_{q-p-i})_2 \text{ for every } 2 \leq i \leq q-p-1 \text{ (Lemma 3.2, 4)}\\
(u_{q-p-1})_2 &< (w_{q-p-1})_2 \text{ (Lemma 3.1, 4)}\\
3/4 \cdot (u_{q-p})_2 &< (w_{q-p})_2 \text{ (Lemma 3.2, 6)}\\
(u_{q-p+1})_2 &< (w_{q-p+1})_2 \text{ (Lemma 3.1, 6)}\\
5/6 \cdot (u_{q-p+2})_2 &< (w_{q-p+2})_2 \text{ (Lemma 3.2, 9)}\\
3/4 \cdot (u_{q-p+i})_2 &< (w_{q-p+i})_2 \text{ for every } 3 \leq i \leq p \text{ (Lemma 3.2, 8).}
\end{aligned}
$$

2.4)

$$
\begin{aligned}
11/8 \cdot (u_{q-p-i})_2 &< (w_{q-p-i})_2 \text{ for every } 1 \leq i \leq q-p-1 \text{ (Lemma 3.2, 2)}\\
3/4 \cdot (u_{q-p})_2 &< (w_{q-p})_2 \text{ (Lemma 3.2, 6)}\\
(u_{q-p+1})_2 &< (w_{q-p+1})_2 \text{ (Lemma 3.1, 6)}\\
5/6 \cdot (u_{q-p+2})_2 &< (w_{q-p+2})_2 \text{ (Lemma 3.2, 9)}\\
3/4 \cdot (u_{q-p+i})_2 &< (w_{q-p+i})_2 \text{ for every } 3 \leq i \leq p \text{ (Lemma 3.2, 8).}
\end{aligned}
$$

2.5)

$$
\begin{aligned}
5/3 \cdot (u_{q-p-i})_2 &< (w_{q-p-i})_2 \text{ for every } 2 \leq i \leq q-p-1 \text{ (Lemma 3.2, 5)}\\
(u_{q-p-1})_2 &< (w_{q-p-1})_2 \text{ (Lemma 3.1, 4)}\\
5/6 \cdot (u_{q-p})_2 &< (w_{q-p})_2 \text{ (Lemma 3.2, 9)}\\
(u_{q-p+1})_2 &< (w_{q-p+1})_2 \text{ (Lemma 3.1, 6)}\\
5/6 \cdot (u_{q-p+2})_2 &< (w_{q-p+2})_2 \text{ (Lemma 3.2, 9).}
\end{aligned}
$$

All that remains to be done is to calculate the products of the multipliers are at least one:

2.1) $2^{q-p-3} \cdot \frac{3}{2} \cdot \frac{3}{4} \cdot \frac{2}{3} \cdot \frac{3}{4}^{p-3} = \frac{9}{8} \cdot \frac{3}{2}^{p-4} 2^{q-2p} > 1$

2.2) $\frac{8}{3}^{q-p-3} \cdot \frac{2}{3} \cdot \frac{2}{3} = \frac{32}{27} \cdot \frac{8}{3}^{q-p-4} > 1$

2.3) $\frac{13}{8}^{q-p-2} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdot \frac{3}{4}^{p-3} = \frac{195}{192} \cdot \frac{39}{32}^{p-3} \frac{13}{8}^{q-2p} > 1$

2.4) $\frac{11}{8}^{q-p-1} \cdot \frac{3}{4} \cdot \frac{5}{6} \cdot \frac{3}{4}^{p-3} = \frac{605}{512} \cdot \frac{33}{32}^{p-3} \frac{11}{8}^{q-2p} > 1$

2.5) $\frac{5}{3}^{q-p-2} \cdot \frac{5}{6} \cdot \frac{5}{6} = \frac{125}{108} \cdot \frac{5}{3}^{q-p-3} > 1.$ □

We know that the lexicographically smallest orbit, the most unbalanced orbit u, gives the smallest product and that the lexicographically largest orbit, the balanced orbit b, gives the largest product. This does not apply generally to all the words between these two extremal words, i.e. a word may have a smaller product than a word which has smaller lexicoraphical order.

In [3] it was observed a permutation between the lexicographic ordering of an orbit $w \in \mathbb{W}_{p,q}$ and the *dynamic* ordering $w, \sigma(w), \ldots, \sigma^{q-1}(w)$ of that same orbit. They called it the *lexidynamic permutation* for the word w. We can also examine a permutation from the lexicographical order of an orbit in the whole $\mathbb{W}_{p,q}$ to the productional order of that orbit. This means that the *lexiproductional* permutation for the word $w \in \mathbb{W}_{p,q}$ always maps $1 \mapsto 1$ and $q \mapsto q$, where $|w| = q$. Here is the permutation for $\mathbb{W}_{4,9}$ which is plotted in Figure 2 (the product of the latter word is in parentheses):

$$000001111 \mapsto 000001111 \ (17057310054912000000)$$

$$000010111 \mapsto 000010111 \ (69309861547173120000)$$

$$000011011 \mapsto 000011101 \ (103115999585285683200)$$

$$000011101 \mapsto 000011011 \ (106107230996504524800)$$

$$000100111 \mapsto 000100111 \ (184709385608811148800)$$

$$000101011 \mapsto 000111001 \ (225726106934040832512)$$

$$000101101 \mapsto 000101101 \ (287935726164372000000)$$

$$000110011 \mapsto 000110011 \ (288046371229598615040)$$

$$000110101 \mapsto 000101011 \ (294762710705942322432)$$

$$000111001 \mapsto 000110101 \ (359572755909315080448)$$

$$001001011 \mapsto 001001011 \ (450633542546718000000)$$

$$001001101 \mapsto 001001101 \ (480928605792476688000)$$

$$001010011 \mapsto 001010011 \ (524261153928446022528)$$
$$001010101 \mapsto 001010101 \ (678501146123915400000)$$

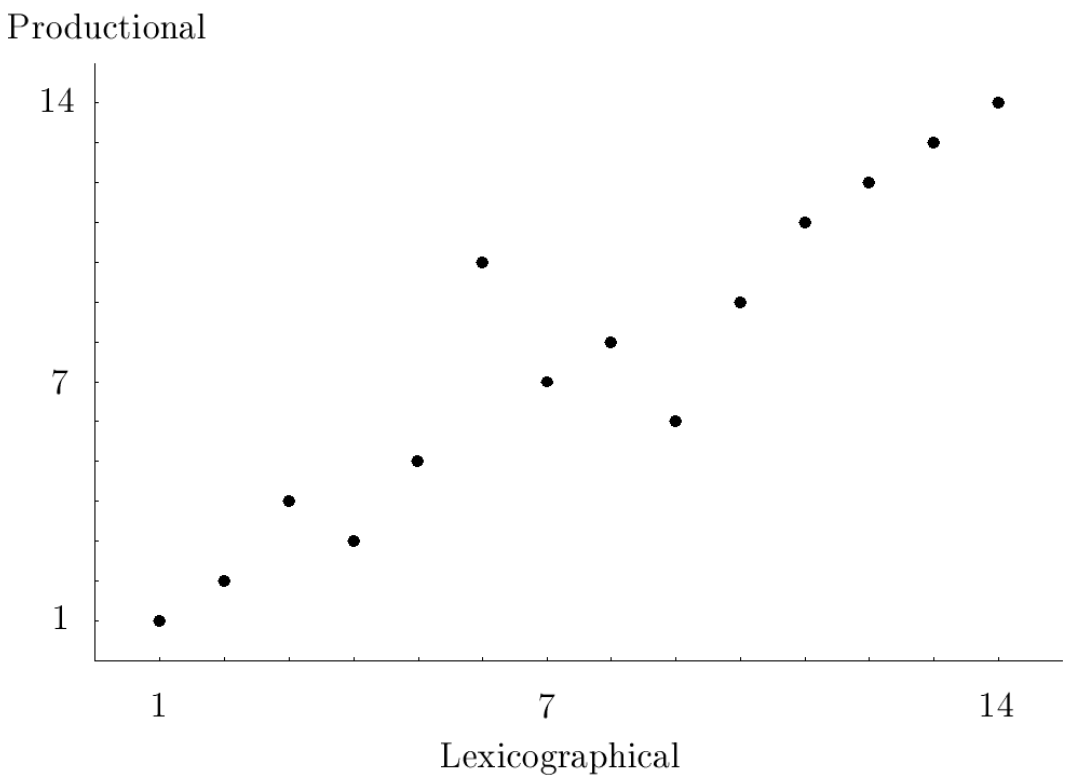

Figure 2: The permutation between lexicographical and productional orders in $\mathbb{W}_{4,9}$.

We can define the balancedness of a word in $\mathbb{W}_{p,q}$ by the productional ordering, i.e. a word is more balanced than words with smaller product. We can see that the words 000011101 and 000111001 have larger productional order than lexicographical order and that the words 000011011, 000101011 and 000110101 have smaller productional order than lexicographical order. For the rest of the words these orders are the same. We can therefore define that the words 000011101 and 000111001 are *over balanced* and that the words 000011011, 000101011 and 000110101 are *under balanced*. The rest of the words are *equally balanced*.

4. Partial product

In this section we prove an analogue of Conjecture 1.4 for the most unbalanced words.

Example 4.1. *If $(p, q) = (3, 8)$ then the set of all orbits is $\mathbb{W}_{3,8} = \{00000111, 00001011, 00001101, 00010011, 00010101, 00011001, 00100101\}$. The base-2 orbits and the approximated partial products of those orbits are listed in Table 7. From these partial products we can see the partial ordering of the set $\mathbb{W}_{3,8}$ with respect to product, which is drawn in Figure 3. Notice that it is different from the Figure 1.*

00000111		00001011		00001101		00010011		00010101		00011001		00100101		$\cdot 10^{x_i}$
$\mathcal{I}_i$	$\mathcal{P}_i$	$\mathcal{I}_i$	$\mathcal{P}_i$	$\mathcal{I}_i$	$\mathcal{P}_i$	$\mathcal{I}_i$	$\mathcal{P}_i$	$\mathcal{I}_i$	$\mathcal{P}_i$	$\mathcal{I}_i$	$\mathcal{P}_i$	$\mathcal{I}_i$	$\mathcal{P}_i$	x_i
7	7	11	11	13	13	19	19	21	21	25	25	37	37	0
14	0.98	22	2.4	26	3.4	38	7.2	42	8.8	35	8.7	41	15	2
28	0.27	44	1.1	52	1.8	49	3.5	69	6.1	50	4.4	73	11	4
56	0.15	88	0.94	67	1.2	76	2.7	81	4.9	70	3.1	74	8.2	6
112	0.17	97	0.91	104	1.2	98	2.6	84	4.1	100	3.1	82	6.7	8
131	0.23	133	1.2	134	1.6	137	3.6	138	5.7	140	4.3	146	9.8	10
193	0.44	176	2.1	161	2.6	152	5.5	162	9.3	145	6.2	148	15	12
224	0.97	194	4.1	208	5.5	196	11	168	16	200	12	164	24	14

Table 7: The base-2 orbits and the (approximated) partial products in $\mathbb{W}_{3,8}$. The numbers $\mathcal{P}_i$ are to be multiplied by 10^{x_i}.

Theorem 4.2. *For any integers $1 \le p < q - p$, the most unbalanced orbit $u = 0^{q-p}1^p \in \mathbb{W}_{p,q}$ is the greatest element in $(\mathbb{W}_{p,q}, \prec_p)$. In other words, for any $w \in \mathbb{W}_{p,q}$,*

$$\mathcal{P}_i(u) \le \mathcal{P}_i(w) \quad \text{for all } 1 \le i \le q.$$

Proof. We set $w = 0^{r_1}1^{s_1}0^{r_2}1^{s_2}\dots 0^{r_n}1^{s_n}$, where $\sum_{i=1}^{n} s_i = p$, $\sum_{i=1}^{n} r_i = q - p$, $n \ge 2$ and $\forall i : r_i, s_i > 0$. The orbits of w and u are marked with $(w_1, \dots, w_q)$ and $(u_1, \dots, u_q)$.

We use Table 2 from the proof of Theorem 2.3 and the same kind of deduction. We get $\mathcal{P}_i(u) \le \mathcal{P}_i(w)$ for $1 \le i \le q - p - 1$. If we suppose that $\mathcal{P}_{q-p}(u) \le \mathcal{P}_{q-p}(w)$ then we get $\mathcal{P}_{q-p+1}(u) \le \mathcal{P}_{q-p+1}(w)$, since $(u_{q-p+1})_2 < (w_{q-p+1})_2$. From Theorem 3.3 we directly get $\mathcal{P}_q(u) < \mathcal{P}_q(w)$. Because $(u_i)_2 > (w_i)_2$ for $q - p + 2 \le i \le q$, we get $\mathcal{P}_i(u) \le \mathcal{P}_i(w)$ for $q - p + 2 \le i \le q$.

Again, the only thing we need to prove now is our assumption $\mathcal{P}_{q-p}(u) \le \mathcal{P}_{q-p}(w)$ in the previous paragraph. We divide the proof into two cases: 1) $n \ge 3$ and 2) $n = 2$. See Table 8 for the prefixes and multipliers of w_i in each case.

1) Because $n \ge 3$ the words w_{q-p}, w_{q-p-1} and w_{q-p-2} start with 01. From Lemma 3.1, 1 and 5, we now directly get $1/2 \cdot (u_{q-p})_2 < (w_{q-p})_2$ and $2 \cdot (u_{q-p-2})_2 < (w_{q-p-2})_2$. This gives

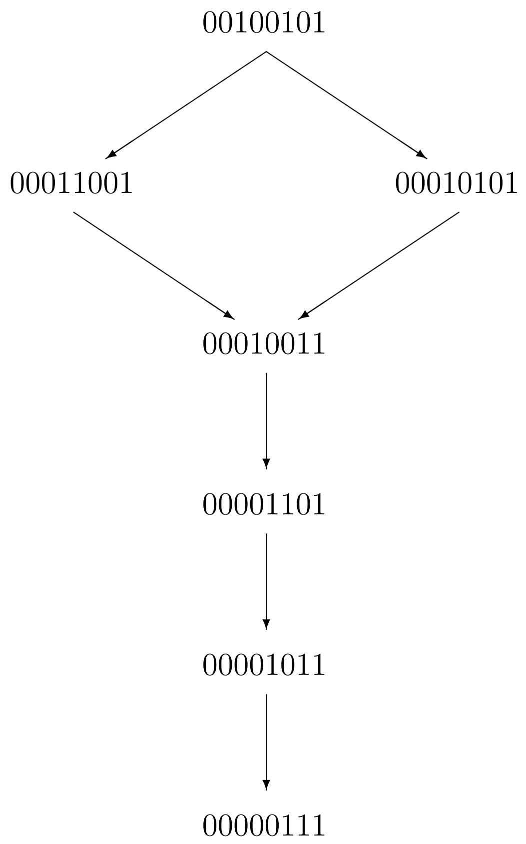

Figure 3: The partially ordered set $(\mathbb{W}_{3,8}, \prec_p)$. Similar to the partial sum, if p and q grow large it is hard to yield any other general results from the poset $(\mathbb{W}_{p,q}, \prec_p)$ except the two extremal elements.

our claim:

$$\mathcal{P}_{q-p}(u) = \prod_{i=1}^{q-p} (u_i)_2 = \prod_{i=1}^{q-p-3} [(u_i)_2] \cdot 2(u_{q-p-2})_2 \cdot (u_{q-p-1})_2 \cdot 1/2(u_{q-p})_2 < \prod_{i=1}^{q-p} (w_i)_2 = \mathcal{P}_{q-p}(w).$$

2) We divide this case into four subcases depending on the values of r_i and s_i. Notice that case $r_1, r_2 = 1$ is impossible because then we would have $2 \leq s_1 + s_2 = p < q-p = r_1 + r_2 = 2$.

2.1) $r_1, r_2 \geq 2$. Now the words w_{q-p} and w_{q-p-1} start with 01 and the words w_{q-p-2} and w_{q-p-3} start with 001. From Lemma 3.1, 1 and 5, we again get $1/2 \cdot (u_{q-p})_2 < (w_{q-p})_2$ and $2 \cdot (u_{q-p-3})_2 < (w_{q-p-3})_2$. This gives our claim:

$$\mathcal{P}_{q-p}(u) = \prod_{i=1}^{q-p-4} [(u_i)_2] \cdot 2(u_{q-p-3})_2 \cdot (u_{q-p-2})_2 \cdot (u_{q-p-1})_2 \cdot 1/2(u_{q-p})_2 < \prod_{i=1}^{q-p} (w_i)_2 = \mathcal{P}_{q-p}(w).$$

i	Prefixes of w_i					Multiplier(i)				
	1	2.1	2.2	2.3	2.4	1	2.1	2.2	2.3	2.4
$q-p-3$		001					2			
$q-p-2$	01	001	001101	001011	001	2		13/8	11/8	4/3
$q-p-1$	01	01	01	01011	01				11/8	4/3
$q-p$	01	01	011	011	01	1/2	1/2	3/4	3/4	2/3

Table 8: Prefixes and multipliers of w_i from the proof of Theorem 4.2.

2.2) $r_1 = 1, r_2 \geq 2$ and $s_1 \geq 1, s_2 \geq 2$. This is identical to the case 2.3 in the proof of Theorem 3.3, from which we get w_{q-p} starts with 011,w_{q-p-1} starts with 01 and w_{q-p-2} starts with 001101. From Lemma 3.2, 4 and 6, we similarly get $3/4 \cdot (u_{q-p})_2 < (w_{q-p})_2$ and $13/8 \cdot (u_{q-p-2})_2 < (w_{q-p-2})_2$. This gives our claim (notice that $13/8 \cdot 3/4 > 1$):

$$\mathcal{P}_{q-p}(u) < \prod_{i=1}^{q-p-3} [(u_i)_2] \cdot 13/8(u_{q-p-2})_2 \cdot (u_{q-p-1})_2 \cdot 3/4(u_{q-p})_2 < \prod_{i=1}^{q-p} (w_i)_2 = \mathcal{P}_{q-p}(w).$$

2.3) $r_1 = 1, r_2 \geq 2$ and $s_1 \geq 2, s_2 \geq 1$. This is similar to the case 2.4 in the proof of Theorem 3.3, from which we get w_{q-p} starts with 011, w_{q-p-1} starts with 01011 and w_{q-p-2} starts with 001011. From Lemma 3.2, 2 and 6, we similarly get $3/4 \cdot (u_{q-p})_2 < (w_{q-p})_2$, $11/8 \cdot (u_{q-p-1})_2 < (w_{q-p-1})_2$ and $11/8 \cdot (u_{q-p-2})_2 < (w_{q-p-2})_2$. This gives our claim (notice that $11/8 \cdot 11/8 \cdot 3/4 > 1$):

$$\mathcal{P}_{q-p}(u) < \prod_{i=1}^{q-p-3} [(u_i)_2] \cdot 11/8(u_{q-p-2})_2 \cdot 11/8(u_{q-p-1})_2 \cdot 3/4(u_{q-p})_2 < \prod_{i=1}^{q-p} (w_i)_2 = \mathcal{P}_{q-p}(w).$$

2.4) $r_1 = 1, r_2 \geq 2$ and $s_1, s_2 = 1$. Because $r_2 \geq 2$ the words w_{q-p}, w_{q-p-1} and w_{q-p-2} start with 01, 01 and 001. Because $q-p = s_1 + s_2 = 2$ we have $u_{q-p} = 0110^{q-3}$, $u_{q-p-1} = 00110^{q-4}$ and $u_{q-p-2} = 000110^{q-5}$. Now we get $2/3 \cdot (u_{q-p})_2 = (0\frac{2}{3}\frac{2}{3}0^{q-3})_2 = (010^{q-2})_2 < (w_{q-p})_2$, $4/3 \cdot (u_{q-p-1})_2 = (00\frac{4}{3}\frac{4}{3}0^{q-4})_2 = (00\frac{6}{3}0^{q-3})_2 = (010^{q-2})_2 < (w_{q-p-1})_2$ and $4/3 \cdot (u_{q-p-2})_2 = (000\frac{4}{3}\frac{4}{3}0^{q-5})_2 = (000\frac{6}{3}0^{q-4})_2 = (0010^{q-3})_2 < (w_{q-p-2})_2$. This gives our claim (notice that $4/3 \cdot 4/3 \cdot 2/3 > 1$):

$$\mathcal{P}_{q-p}(u) < \prod_{i=1}^{q-p-3} [(u_i)_2] \cdot 4/3(u_{q-p-2})_2 \cdot 4/3(u_{q-p-1})_2 \cdot 2/3(u_{q-p})_2 < \prod_{i=1}^{q-p} (w_i)_2 = \mathcal{P}_{q-p}(w).$$

□

Acknowledgements

I would like to thank my supervisor Luca Zamboni for introducing me to this problem. I also want to thank my PhD thesis pre-examiners Amy Glen and Sébastien Labbé for their comments which greatly improved the quality of this paper.

References

[1] J. Berstel, P. Séébold: *Sturmian words,* in: M. Lothaire (Ed.), Algebraic Combinatorics on Words, Cambridge University Press, Cambridge, UK, 2002.

[2] O. Jenkinson: *Balanced words and majorization*, Discrete Math., Alg. and Appl. 1(4) (2009) 463-484.

[3] O. Jenkinson, L. Q. Zamboni: *Characterisations of balanced words via orderings*, Theoret. Comput. Sci. 310 (1-3) (2004) 247-271.

[4] M. Lothaire: *Combinatorics on Words,* in: Encyclopedia of Mathematics and its Applications, vol 17, Addison-Wesley, Reading, MA, 1983.

[5] M. Lothaire: *Algebraic Combinatorics on Words,* in: Encyclopedia of Mathematics and its Applications, vol 90, Cambridge University Press, UK, 2002.

[6] M. Morse, G. A. Hedlund: *Symbolic dynamics, II; Sturmian trajectories*, Amer. J. Math. 62 (1940) 1-42.

Turku Centre for Computer Science
TUCS Dissertations

1. **Marjo Lipponen**, On Primitive Solutions of the Post Correspondence Problem
2. **Timo Käkölä**, Dual Information Systems in Hyperknowledge Organizations
3. **Ville Leppänen**, Studies on the Realization of PRAM
4. **Cunsheng Ding**, Cryptographic Counter Generators
5. **Sami Viitanen**, Some New Global Optimization Algorithms
6. **Tapio Salakoski**, Representative Classification of Protein Structures
7. **Thomas Långbacka**, An Interactive Environment Supporting the Development of Formally Correct Programs
8. **Thomas Finne**, A Decision Support System for Improving Information Security
9. **Valeria Mihalache**, Cooperation, Communication, Control. Investigations on Grammar Systems.
10. **Marina Waldén**, Formal Reasoning About Distributed Algorithms
11. **Tero Laihonen**, Estimates on the Covering Radius When the Dual Distance is Known
12. **Lucian Ilie**, Decision Problems on Orders of Words
13. **Jukkapekka Hekanaho**, An Evolutionary Approach to Concept Learning
14. **Jouni Järvinen**, Knowledge Representation and Rough Sets
15. **Tomi Pasanen**, In-Place Algorithms for Sorting Problems
16. **Mika Johnsson**, Operational and Tactical Level Optimization in Printed Circuit Board Assembly
17. **Mats Aspnäs**, Multiprocessor Architecture and Programming: The Hathi-2 System
18. **Anna Mikhajlova**, Ensuring Correctness of Object and Component Systems
19. **Vesa Torvinen**, Construction and Evaluation of the Labour Game Method
20. **Jorma Boberg**, Cluster Analysis. A Mathematical Approach with Applications to Protein Structures
21. **Leonid Mikhajlov**, Software Reuse Mechanisms and Techniques: Safety Versus Flexibility
22. **Timo Kaukoranta**, Iterative and Hierarchical Methods for Codebook Generation in Vector Quantization
23. **Gábor Magyar**, On Solution Approaches for Some Industrially Motivated Combinatorial Optimization Problems
24. **Linas Laibinis**, Mechanised Formal Reasoning About Modular Programs
25. **Shuhua Liu**, Improving Executive Support in Strategic Scanning with Software Agent Systems
26. **Jaakko Järvi**, New Techniques in Generic Programming – C++ is more Intentional than Intended
27. **Jan-Christian Lehtinen**, Reproducing Kernel Splines in the Analysis of Medical Data
28. **Martin Büchi**, Safe Language Mechanisms for Modularization and Concurrency
29. **Elena Troubitsyna**, Stepwise Development of Dependable Systems
30. **Janne Näppi**, Computer-Assisted Diagnosis of Breast Calcifications
31. **Jianming Liang**, Dynamic Chest Images Analysis
32. **Tiberiu Seceleanu**, Systematic Design of Synchronous Digital Circuits
33. **Tero Aittokallio**, Characterization and Modelling of the Cardiorespiratory System in Sleep-Disordered Breathing
34. **Ivan Porres**, Modeling and Analyzing Software Behavior in UML
35. **Mauno Rönkkö**, Stepwise Development of Hybrid Systems
36. **Jouni Smed**, Production Planning in Printed Circuit Board Assembly
37. **Vesa Halava**, The Post Correspondence Problem for Market Morphisms
38. **Ion Petre**, Commutation Problems on Sets of Words and Formal Power Series
39. **Vladimir Kvassov**, Information Technology and the Productivity of Managerial Work
40. **Frank Tétard**, Managers, Fragmentation of Working Time, and Information Systems

41. **Jan Manuch**, Defect Theorems and Infinite Words
42. **Kalle Ranto**, Z_4-Goethals Codes, Decoding and Designs
43. **Arto Lepistö**, On Relations Between Local and Global Periodicity
44. **Mika Hirvensalo**, Studies on Boolean Functions Related to Quantum Computing
45. **Pentti Virtanen**, Measuring and Improving Component-Based Software Development
46. **Adekunle Okunoye**, Knowledge Management and Global Diversity – A Framework to Support Organisations in Developing Countries
47. **Antonina Kloptchenko**, Text Mining Based on the Prototype Matching Method
48. **Juha Kivijärvi**, Optimization Methods for Clustering
49. **Rimvydas Rukšėnas**, Formal Development of Concurrent Components
50. **Dirk Nowotka**, Periodicity and Unbordered Factors of Words
51. **Attila Gyenesei**, Discovering Frequent Fuzzy Patterns in Relations of Quantitative Attributes
52. **Petteri Kaitovaara**, Packaging of IT Services – Conceptual and Empirical Studies
53. **Petri Rosendahl**, Niho Type Cross-Correlation Functions and Related Equations
54. **Péter Majlender**, A Normative Approach to Possibility Theory and Soft Decision Support
55. **Seppo Virtanen**, A Framework for Rapid Design and Evaluation of Protocol Processors
56. **Tomas Eklund**, The Self-Organizing Map in Financial Benchmarking
57. **Mikael Collan**, Giga-Investments: Modelling the Valuation of Very Large Industrial Real Investments
58. **Dag Björklund**, A Kernel Language for Unified Code Synthesis
59. **Shengnan Han**, Understanding User Adoption of Mobile Technology: Focusing on Physicians in Finland
60. **Irina Georgescu**, Rational Choice and Revealed Preference: A Fuzzy Approach
61. **Ping Yan**, Limit Cycles for Generalized Liénard-Type and Lotka-Volterra Systems
62. **Joonas Lehtinen**, Coding of Wavelet-Transformed Images
63. **Tommi Meskanen,** On the NTRU Cryptosystem
64. **Saeed Salehi**, Varieties of Tree Languages
65. **Jukka Arvo**, Efficient Algorithms for Hardware-Accelerated Shadow Computation
66. **Mika Hirvikorpi**, On the Tactical Level Production Planning in Flexible Manufacturing Systems
67. **Adrian Costea**, Computational Intelligence Methods for Quantitative Data Mining
68. **Cristina Seceleanu**, A Methodology for Constructing Correct Reactive Systems
69. **Luigia Petre**, Modeling with Action Systems
70. **Lu Yan**, Systematic Design of Ubiquitous Systems
71. **Mehran Gomari**, On the Generalization Ability of Bayesian Neural Networks
72. **Ville Harkke**, Knowledge Freedom for Medical Professionals – An Evaluation Study of a Mobile Information System for Physicians in Finland
73. **Marius Cosmin Codrea**, Pattern Analysis of Chlorophyll Fluorescence Signals
74. **Aiying Rong**, Cogeneration Planning Under the Deregulated Power Market and Emissions Trading Scheme
75. **Chihab BenMoussa**, Supporting the Sales Force through Mobile Information and Communication Technologies: Focusing on the Pharmaceutical Sales Force
76. **Jussi Salmi**, Improving Data Analysis in Proteomics
77. **Orieta Celiku**, Mechanized Reasoning for Dually-Nondeterministic and Probabilistic Programs
78. **Kaj-Mikael Björk**, Supply Chain Efficiency with Some Forest Industry Improvements
79. **Viorel Preoteasa**, Program Variables – The Core of Mechanical Reasoning about Imperative Programs
80. **Jonne Poikonen**, Absolute Value Extraction and Order Statistic Filtering for a Mixed-Mode Array Image Processor
81. **Luka Milovanov**, Agile Software Development in an Academic Environment
82. **Francisco Augusto Alcaraz Garcia**, Real Options, Default Risk and Soft Applications
83. **Kai K. Kimppa**, Problems with the Justification of Intellectual Property Rights in Relation to Software and Other Digitally Distributable Media
84. **Dragoş Truşcan**, Model Driven Development of Programmable Architectures
85. **Eugen Czeizler**, The Inverse Neighborhood Problem and Applications of Welch Sets in Automata Theory

86. **Sanna Ranto**, Identifying and Locating-Dominating Codes in Binary Hamming Spaces
87. **Tuomas Hakkarainen**, On the Computation of the Class Numbers of Real Abelian Fields
88. **Elena Czeizler**, Intricacies of Word Equations
89. **Marcus Alanen**, A Metamodeling Framework for Software Engineering
90. **Filip Ginter**, Towards Information Extraction in the Biomedical Domain: Methods and Resources
91. **Jarkko Paavola**, Signature Ensembles and Receiver Structures for Oversaturated Synchronous DS-CDMA Systems
92. **Arho Virkki**, The Human Respiratory System: Modelling, Analysis and Control
93. **Olli Luoma**, Efficient Methods for Storing and Querying XML Data with Relational Databases
94. **Dubravka Ilić**, Formal Reasoning about Dependability in Model-Driven Development
95. **Kim Solin**, Abstract Algebra of Program Refinement
96. **Tomi Westerlund**, Time Aware Modelling and Analysis of Systems-on-Chip
97. **Kalle Saari**, On the Frequency and Periodicity of Infinite Words
98. **Tomi Kärki**, Similarity Relations on Words: Relational Codes and Periods
99. **Markus M. Mäkelä**, Essays on Software Product Development: A Strategic Management Viewpoint
100. **Roope Vehkalahti**, Class Field Theoretic Methods in the Design of Lattice Signal Constellations
101. **Anne-Maria Ernvall-Hytönen**, On Short Exponential Sums Involving Fourier Coefficients of Holomorphic Cusp Forms
102. **Chang Li**, Parallelism and Complexity in Gene Assembly
103. **Tapio Pahikkala**, New Kernel Functions and Learning Methods for Text and Data Mining
104. **Denis Shestakov**, Search Interfaces on the Web: Querying and Characterizing
105. **Sampo Pyysalo**, A Dependency Parsing Approach to Biomedical Text Mining
106. **Anna Sell**, Mobile Digital Calendars in Knowledge Work
107. **Dorina Marghescu**, Evaluating Multidimensional Visualization Techniques in Data Mining Tasks
108. **Tero Säntti**, A Co-Processor Approach for Efficient Java Execution in Embedded Systems
109. **Kari Salonen**, Setup Optimization in High-Mix Surface Mount PCB Assembly
110. **Pontus Boström**, Formal Design and Verification of Systems Using Domain-Specific Languages
111. **Camilla J. Hollanti**, Order-Theoretic Mehtods for Space-Time Coding: Symmetric and Asymmetric Designs
112. **Heidi Himmanen**, On Transmission System Design for Wireless Broadcasting
113. **Sébastien Lafond**, Simulation of Embedded Systems for Energy Consumption Estimation
114. **Evgeni Tsivtsivadze**, Learning Preferences with Kernel-Based Methods
115. **Petri Salmela**, On Commutation and Conjugacy of Rational Languages and the Fixed Point Method
116. **Siamak Taati**, Conservation Laws in Cellular Automata
117. **Vladimir Rogojin**, Gene Assembly in Stichotrichous Ciliates: Elementary Operations, Parallelism and Computation
118. **Alexey Dudkov**, Chip and Signature Interleaving in DS CDMA Systems
119. **Janne Savela**, Role of Selected Spectral Attributes in the Perception of Synthetic Vowels
120. **Kristian Nybom**, Low-Density Parity-Check Codes for Wireless Datacast Networks
121. **Johanna Tuominen**, Formal Power Analysis of Systems-on-Chip
122. **Teijo Lehtonen**, On Fault Tolerance Methods for Networks-on-Chip
123. **Eeva Suvitie**, On Inner Products Involving Holomorphic Cusp Forms and Maass Forms
124. **Linda Mannila**, Teaching Mathematics and Programming – New Approaches with Empirical Evaluation
125. **Hanna Suominen**, Machine Learning and Clinical Text: Supporting Health Information Flow
126. **Tuomo Saarni**, Segmental Durations of Speech
127. **Johannes Eriksson**, Tool-Supported Invariant-Based Programming

128. **Tero Jokela**, Design and Analysis of Forward Error Control Coding and Signaling for Guaranteeing QoS in Wireless Broadcast Systems
129. **Ville Lukkarila**, On Undecidable Dynamical Properties of Reversible One-Dimensional Cellular Automata
130. **Qaisar Ahmad Malik**, Combining Model-Based Testing and Stepwise Formal Development
131. **Mikko-Jussi Laakso**, Promoting Programming Learning: Engagement, Automatic Assessment with Immediate Feedback in Visualizations
132. **Riikka Vuokko**, A Practice Perspective on Organizational Implementation of Information Technology
133. **Jeanette Heidenberg**, Towards Increased Productivity and Quality in Software Development Using Agile, Lean and Collaborative Approaches
134. **Yong Liu**, Solving the Puzzle of Mobile Learning Adoption
135. **Stina Ojala**, Towards an Integrative Information Society: Studies on Individuality in Speech and Sign
136. **Matteo Brunelli**, Some Advances in Mathematical Models for Preference Relations
137. **Ville Junnila**, On Identifying and Locating-Dominating Codes
138. **Andrzej Mizera**, Methods for Construction and Analysis of Computational Models in Systems Biology. Applications to the Modelling of the Heat Shock Response and the Self-Assembly of Intermediate Filaments.
139. **Csaba Ráduly-Baka**, Algorithmic Solutions for Combinatorial Problems in Resource Management of Manufacturing Environments
140. **Jari Kyngäs**, Solving Challenging Real-World Scheduling Problems
141. **Arho Suominen**, Notes on Emerging Technologies
142. **József Mezei**, A Quantitative View on Fuzzy Numbers
143. **Marta Olszewska**, On the Impact of Rigorous Approaches on the Quality of Development
144. **Antti Airola**, Kernel-Based Ranking: Methods for Learning and Performace Estimation
145. **Aleksi Saarela**, Word Equations and Related Topics: Independence, Decidability and Characterizations
146. **Lasse Bergroth**, Kahden merkkijonon pisimmän yhteisen alijonon ongelma ja sen ratkaiseminen
147. **Thomas Canhao Xu**, Hardware/Software Co-Design for Multicore Architectures
148. **Tuomas Mäkilä**, Software Development Process Modeling – Developers Perspective to Contemporary Modeling Techniques
149. **Shahrokh Nikou**, Opening the Black-Box of IT Artifacts: Looking into Mobile Service Characteristics and Individual Perception
150. **Alessandro Buoni**, Fraud Detection in the Banking Sector: A Multi-Agent Approach
151. **Mats Neovius**, Trustworthy Context Dependency in Ubiquitous Systems
152. **Fredrik Degerlund**, Scheduling of Guarded Command Based Models
153. **Amir-Mohammad Rahmani-Sane**, Exploration and Design of Power-Efficient Networked Many-Core Systems
154. **Ville Rantala**, On Dynamic Monitoring Methods for Networks-on-Chip
155. **Mikko Pelto**, On Identifying and Locating-Dominating Codes in the Infinite King Grid
156. **Anton Tarasyuk**, Formal Development and Quantitative Verification of Dependable Systems
157. **Muhammad Mohsin Saleemi**, Towards Combining Interactive Mobile TV and Smart Spaces: Architectures, Tools and Application Development
158. **Tommi J. M. Lehtinen**, Numbers and Languages
159. **Peter Sarlin**, Mapping Financial Stability
160. **Alexander Wei Yin**, On Energy Efficient Computing Platforms
161. **Mikołaj Olszewski**, Scaling Up Stepwise Feature Introduction to Construction of Large Software Systems
162. **Maryam Kamali**, Reusable Formal Architectures for Networked Systems
163. **Zhiyuan Yao**, Visual Customer Segmentation and Behavior Analysis – A SOM-Based Approach
164. **Timo Jolivet**, Combinatorics of Pisot Substitutions
165. **Rajeev Kumar Kanth**, Analysis and Life Cycle Assessment of Printed Antennas for Sustainable Wireless Systems
166. **Khalid Latif**, Design Space Exploration for MPSoC Architectures

167. **Bo Yang**, Towards Optimal Application Mapping for Energy-Efficient Many-Core Platforms
168. **Ali Hanzala Khan**, Consistency of UML Based Designs Using Ontology Reasoners
169. **Sonja Leskinen**, m-Equine: IS Support for the Horse Industry
170. **Fareed Ahmed Jokhio**, Video Transcoding in a Distributed Cloud Computing Environment
171. **Moazzam Fareed Niazi**, A Model-Based Development and Verification Framework for Distributed System-on-Chip Architecture
172. **Mari Huova**, Combinatorics on Words: New Aspects on Avoidability, Defect Effect, Equations and Palindromes
173. **Ville Timonen**, Scalable Algorithms for Height Field Illumination
174. **Henri Korvela**, Virtual Communities – A Virtual Treasure Trove for End-User Developers
175. **Kameswar Rao Vaddina**, Thermal-Aware Networked Many-Core Systems
176. **Janne Lahtiranta**, New and Emerging Challenges of the ICT-Mediated Health and Well-Being Services
177. **Irum Rauf**, Design and Validation of Stateful Composite RESTful Web Services
178. **Jari Björne**, Biomedical Event Extraction with Machine Learning
179. **Katri Haverinen**, Natural Language Processing Resources for Finnish: Corpus Development in the General and Clinical Domains
180. **Ville Salo**, Subshifts with Simple Cellular Automata
181. **Johan Ersfolk**, Scheduling Dynamic Dataflow Graphs
182. **Hongyan Liu**, On Advancing Business Intelligence in the Electricity Retail Market
183. **Adnan Ashraf**, Cost-Efficient Virtual Machine Management: Provisioning, Admission Control, and Consolidation
184. **Muhammad Nazrul Islam**, Design and Evaluation of Web Interface Signs to Improve Web Usability: A Semiotic Framework
185. **Johannes Tuikkala**, Algorithmic Techniques in Gene Expression Processing: From Imputation to Visualization
186. **Natalia Díaz Rodríguez**, Semantic and Fuzzy Modelling for Human Behaviour Recognition in Smart Spaces. A Case Study on Ambient Assisted Living
187. **Mikko Pänkäälä**, Potential and Challenges of Analog Reconfigurable Computation in Modern and Future CMOS
188. **Sami Hyrynsalmi**, Letters from the War of Ecosystems – An Analysis of Independent Software Vendors in Mobile Application Marketplaces
189. **Seppo Pulkkinen**, Efficient Optimization Algorithms for Nonlinear Data Analysis
190. **Sami Pyöttiälä**, Optimization and Measuring Techniques for Collect-and-Place Machines in Printed Circuit Board Industry
191. **Syed Mohammad Asad Hassan Jafri**, Virtual Runtime Application Partitions for Resource Management in Massively Parallel Architectures
192. **Toni Ernvall**, On Distributed Storage Codes
193. **Yuliya Prokhorova**, Rigorous Development of Safety-Critical Systems
194. **Olli Lahdenoja**, Local Binary Patterns in Focal-Plane Processing – Analysis and Applications
195. **Annika H. Holmbom**, Visual Analytics for Behavioral and Niche Market Segmentation
196. **Sergey Ostroumov**, Agent-Based Management System for Many-Core Platforms: Rigorous Design and Efficient Implementation
197. **Espen Suenson**, How Computer Programmers Work – Understanding Software Development in Practise
198. **Tuomas Poikela**, Readout Architectures for Hybrid Pixel Detector Readout Chips
199. **Bogdan Iancu**, Quantitative Refinement of Reaction-Based Biomodels
200. **Ilkka Törmä**, Structural and Computational Existence Results for Multidimensional Subshifts
201. **Sebastian Okser**, Scalable Feature Selection Applications for Genome-Wide Association Studies of Complex Diseases
202. **Fredrik Abbors**, Model-Based Testing of Software Systems: Functionality and Performance
203. **Inna Pereverzeva**, Formal Development of Resilient Distributed Systems
204. **Mikhail Barash**, Defining Contexts in Context-Free Grammars
205. **Sepinoud Azimi**, Computational Models for and from Biology: Simple Gene Assembly and Reaction Systems
206. **Petter Sandvik**, Formal Modelling for Digital Media Distribution

207. **Jongyun Moon**, Hydrogen Sensor Application of Anodic Titanium Oxide Nanostructures
208. **Simon Holmbacka**, Energy Aware Software for Many-Core Systems
209. **Charalampos Zinoviadis**, Hierarchy and Expansiveness in Two-Dimensional Subshifts of Finite Type
210. **Mika Murtojärvi**, Efficient Algorithms for Coastal Geographic Problems
211. **Sami Mäkelä**, Cohesion Metrics for Improving Software Quality
212. **Eyal Eshet**, Examining Human-Centered Design Practice in the Mobile Apps Era
213. **Jetro Vesti**, Rich Words and Balanced Words

Turku Centre *for* Computer Science

http://www.tucs.fi
tucs@abo.fi

University of Turku
Faculty of Mathematics and Natural Sciences
- Department of Information Technology
- Department of Mathematics and Statistics

Turku School of Economics
- Institute of Information Systems Science

Åbo Akademi University
Faculty of Science and Engineering
- Computer Engineering
- Computer Science

Faculty of Social Sciences, Business and Economics
- Information Systems

ISBN 978-952-12-3404-0
ISSN 1239-1883

Jetro Vesti

Rich Words and Balanced Words

Jetro Vesti

Rich Words and Balanced Words

Jetro Vesti

Rich Words and Balanced Words